# Weighted Likelihood Estimation under Two-phase Sampling

Takumi Saegusa

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Jon A. Wellner, Chair

Norman Breslow

Peter B. Gilbert

Kwun C. Chan

Program Authorized to Offer Degree:
Biostatistics

University of Washington

## Abstract

Weighted Likelihood Estimation under Two-phase Sampling

Takumi Saegusa

Chair of the Supervisory Committee:
Professor Jon A. Wellner
Statistics

Two-phase sampling is a sampling technique for cost reduction and improved efficiency of estimation, adopted in many epidemiological studies. In this dissertation, we study weighted likelihood estimation, a standard estimation method in this study design. Though sampling without replacement at the second phase induces dependence among observations, independence is often assumed in practice for theoretical convenience, leading to overestimating the asymptotic variance. The main contribution of this dissertation is to develop asymptotic theory for weighted likelihood estimation taking account of the dependence of observations due to the sampling scheme, for both cases where the nuisance parameter is estimable at a regular($\sqrt{n}$-rate) and non-regular rates. To this end, we develop a set of empirical process tools including a Glivenko-Cantelli theorem, a theorem for rates of convergence of $M$-estimators, and a Donsker theorem for the inverse probability weighted empirical processes under two-phase sampling and sampling without replacement at the second phase. For variance estimation, we propose two different bootstrap procedures. The first method is to estimate the phase I and II variances separately which allows us to evaluate how much information we lose by two-phase designs. The second method, which accounts for the phase I and II variances at the same time, provides valid variance estimates even under model misspecification. We also develop the method, within-stratum centered calibration, to improve efficiency over generally inefficient weighted likelihood estimators and study its theoretical properties.

# TABLE OF CONTENTS

Page

## LIST OF TABLES

## ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my advisor, Jon A. Wellner, for his mentorship over the past years. His knowledge, inspiration, and insight have strongly influenced me as a researcher. This dissertation would not have been possible without his encouragement and patience.

I am grateful to my committee members, Norman E. Breslow, KC Gary Chan, and Peter B. Gilbert for helpful discussions and suggestion of various research directions. I also thank Lisa E. Manhart for serving as the graduate school representative. My thanks also go to the faculty, staff, and students in my department for providing a stimulating and supportive research environment.

Finally, I wish to show my deep gratitude to my parents, Yasuhiro and Kiyoko Saegusa, for their unceasing encouragement and support.

# DEDICATION

to my parents, Yasuhiro and Kiyoko

Chapter 1

# INTRODUCTION

In this dissertation, we consider weighted likelihood estimation under two-phase sampling. In the following, we describe a sampling scheme of our interest in contrast to survey sampling and stratified Bernoulli sampling. We discuss advantages to the use of weighted likelihood estimation particularly in the context of two-phase sampling. Along the way, we introduce statistical problems studied in this dissertation, and discuss differences between our results and previous work.

## *1.1 Two-phase Sampling*

Two-phase sampling is a sampling technique that aims at cost reduction and improved efficiency of estimation. At the first phase, a large sample is drawn from a population, and information on variables that are easier to measure is collected. These phase I variables may play an important role in statistical analysis such as exposure in a regression model, or they may simply be auxiliary variables that are correlated with unavailable variables at the first phase but are not of interest in themselves. Based on the all variables observed at the first phase, the sample space is stratified using only the phase I data. At the second phase, a subsample is drawn without replacement from each stratum, and subsequently phase II variables that are costly or difficult to obtain are measured. Strata formation is intended either to oversample subjects with important phase I variables, or to effectively sample subjects with phase II variables correlated with phase I variables, or both. This way, two-phase sampling achieves effective access to important variables with less cost and, as a result, enhances efficiency of estimation.

Two-phase sampling was originally introduced in survey sampling by [38] for estimation of the "finite population mean" of some variable. Since then, this sampling method together with the Horvitz-Thompson estimator [20] as a standard estimator has been widely adopted in survey sampling. Much later, two-phase designs were introduced to biostatistical applications where, in contrast to survey sampling, the population is infinite and the parameter in the statistical model is of interest. Notable examples of two-phase designs include a case-cohort study [42, 55] for the Cox proportional hazards model with right-censored data where the censoring indicator is used as a stratification, a stratified case-control study [61] where additional stratification on rare exposure was considered in the setting of the case control design and a stratified case-cohort study [4] which extended the case cohort study of [42] by stratifying on covariates as well as a censoring indicator. More recently the broad applicability and importance of two phase designs has been emphasized by [5, 6].

A motivating example for this thesis is the RV 144 case control study [19]. This study is based on the RV 144 randomized clinical trial (Thai trial, [44]) that established vaccine efficacy against HIV-1 infection. The primary objective of the RV 144 case control study is to find immune correlates of protection against the HIV-1 infection. Due to limited resources to measure immune responses, a stratified subsample was chosen without replacement from the original study based on the infection status, risk behavior, the number of vaccinations and gender. The outcome of interest is time to infection that unlike the case-cohort study, was subject to interval censoring. We assume the Cox proportional hazards model and consider weighted likelihood estimation. Our primary goal is to estimate the regression parameter for potential immune correlates and test significance of their effects on protection against HIV-1.

As illustrated in this study, there are several statistical issues arising from two-phase designs. The first and main statistical problem throughout this dissertation is dependence among observations induced by the "without replacement' sampling scheme. Because most available tools for asymptotics assume independence, establishing consistency, a rate of convergence and an asymptotic normality is not an easy task. The second issue is efficiency of estimation. In the setting of the above example, there is not a known efficient estimator even if independence is assumed for convenience, while weighted likelihood estimation available in many applications including this example is known to be inefficient. Thus, improving efficiency of the weighted likelihood estimator (WLE) is of practical interest. The third issue is variance estimation. Because the asymptotic variance contains unknown functions in this example, there is no simple consistent estimator such as a sample mean of a squared influence function even when complete data are observed at the first phase.

We describe these statistical problems and relevant previous work in the next section. Before that, we discuss differences between the sampling schemes of interest here and other related study designs.

### *1.1.1 Contrast to Asymptotics in Survey Sampling*

Survey sampling has established asymptotic results for dependent observations from stratified sampling without replacement and even more complicated designs (see [23] and references therein). However, even when a model is postulated, the main focus of survey statisticians is often a “finite population parameter”, which is usually defined as the solution of some estimating equation in a finite population [48], not a “super-population parameter” which determines a scientific phenomenon for an infinite population. Asymptotic theory in survey sampling thus usually treats a sequence of finite populations with increasing sample sizes based on conditions regarding a growing sequence of finite populations [23]. Because biostatisticians are more interested in super-population parameters, the asymptotic theory developed in this dissertation is relevant in the usual biostatistical settings.

One notable exception in the survey sampling literature is [48] (see also references therein and [28], [12]). In [48], the authors define the product of the model space and the design space as a probability space, and decompose their normalized estimator into the contributions from a super-population and a sampling design. Two distinct sets of conditions for the model space and design space are used to guarantee the asymptotic normality of each contribution respectively. The former and the latter conditions are familiar to biostatisticians and survey statisticians, respectively, but not vice versa. See [47] for two sets of conditions in an application to the Cox model under cluster sampling.

Our approach, which relies on the framework and some of the results developed in [8], shares a similar idea with [48] in view of decomposition. In [8], the inverse probability weighted empirical process is decomposed into the usual empirical process (phase I contribution) and the weighted sum of finite sampling empirical processes (phase II contribution).

(Compare (10) of [8] with the decomposition (A.8) of [48].) Then conditional on the phase I data, the results for exchangeably weighted bootstrap empirical processes [41], which covers our sampling scheme, is applied to show the weak convergence of the phase II contribution.

Despite this similarity, our framework is different from [48] in the following important points. First of all, we do not need additional conditions for the phase II contribution unlike design conditions imposed in [48] because the same conditions for the phase I contribution suffice for the exchangeably weighted bootstrap empirical process theory to apply in our setting. Second, our method is more general since our decomposition is at the process level, not the level of random variables. Third, our formulae for the asymptotic variances have more natural interpretations than the formulae in the framework of [48] that consist of two incongruent parts, one that depends on the model conditions and the other that depends on the design conditions. For these reasons, our approach should be distinguished from those in survey sampling.

#### *1.1.2 Contrast to Stratified Bernoulli Sampling*

Stratified Bernoulli sampling is the sampling scheme where each subject is independently sampled at the second phase with probability which depends only on the stratum where an observation belongs. Since the phase II sample size is random, this sampling scheme is usually not adopted in practice where cost of sampling is determined in advance. Rather, it is assumed in statistical analysis for convenience when the actual sampling scheme is stratified sampling without replacement. The justification behind this approximation is that the sampling fraction (i.e., the number of observation sampled at the second phase over the number of observation in a certain stratum) in a certain stratum under sampling without replacement is plausibly assumed to converge to a sampling probability in the stratum under Bernoulli sampling (see [8] for more detailed comparison) so that it is hoped that difference between both sampling schemes would be negligible in the limit. As a consequence, well established asymptotic results (e.g. [58]) can be applied to carry out statistical analysis under the independence assumption.

Despite the hope in practice, these two sampling schemes yield different statistical results. [8] showed that the asymptotic variance of the WLE under sampling without replacement is always smaller than that under Bernoulli sampling. This implies that assuming Bernoulli sampling for convenience leads to conservative results so that power is lower than expected. Another difference comes from methods of improving efficiency of the WLE. All proposed methods (estimated weights [46], calibration [15], modified calibration [10]) have been, so far, justified only under Bernoulli sampling, but their performances are unknown under sampling without replacement.

## ***1.2 Weighted Likelihood Estimation***

In weighted likelihood estimation, WLE's are obtained from maximizing the weighted log likelihood or from solving weighted likelihood equations. Weights are attached to the individual observations each of which contributes to the likelihood. In this thesis, these weights are inverses of (estimated or calibrated) sampling probabilities at the second phase. Thus, our weighted likelihood is an application of the Horvitz-Thompson estimator [20] to the likelihood.

### *1.2.1 Advantages to the WLE's*

The weighted likelihood estimation methods we study in this dissertation have several advantages, especially in the context of two-phase studies. Major arguments for using WLE's are (a) their wide availability in many models, and (b) ease of computation. For example, it may be true that one is willing to assume independence for obtaining efficient estimators because the WLE is generally inefficient under the independence assumption. However, even when this assumption holds, there are not many statistical models where efficient estimators are known (see [46], [45], and [7] for some exceptions). Moreover, efficient estimators, if known, may require sophisticated numerical techniques (e.g., solving integral equations, [36]) or restrictive assumptions that are not imposed when complete data is available (e.g. a parametric covariate distribution [11], discrete covariates [36]). In contrast, the maximum likelihood estimator with complete data is often available in many applications, and the corresponding WLE is obtained simply via a weighted likelihood version of the same likelihood equations. Furthermore, theory for the WLE only requires almost identical conditions to those for the MLE with complete data (see Theorems 3.2.1 and 3.3.1 below).

Another advantage of the WLE and their variants involving estimated weights or calibration is robustness to model misspecification in the following sense: When the underlying model is misspecified, the WLE's and their relatives continue to estimate the same parameters as would be estimated under model miss-specification with complete data. (For example, see [25], [16], [52, 53], [62], and [26]. See also [54] for careful further considerations of this issue.)

#### *1.2.2 Methods to Improve Efficiency*

To improve the efficiency of the WLE, several methods have been advocated. The common idea behind these methods is to use the phase I information, which is largely ignored (for observations not sampled at the second phase), by adjusting weights in the weighted likelihood. [46] proposed estimating weights by binary regression in a missing data problem even though the true weights are known. [15] developed the method of calibration by which adjusted inverse probability weighted mean of some phase I variables are equated to the known phase I sample means. [10] proposed a variant of calibration in a missing response problem.

Efficiency gains, if any, from these methods are achieved through (approximate) projection of the influence function of the plain WLE under Bernoulli sampling (in $L_2(P)$) onto the orthocomplement of the space spanned by the influence function of the estimator of the parameter in adjusting weights. Thus, efficiency gains are larger when the influence function of the plain WLE and the phase I variables used for adjusting weights are highly correlated. An important implication from this projection argument is that efficiency improvements are guaranteed under stratified Bernoulli sampling but not under stratified sampling without replacement, though these methods work well in simulations. In fact, a desirable projection under sampling without replacement should be in the $L_2^0(P)$ sense, i.e., a projected function has mean zero. This difference by sampling schemes is due to the difference between having the conditional expectations and the conditional variance given a stratum membership in the asymptotic variances for Bernoulli sampling and sampling without replacement, respectively (compare (21) and (22) of [8]).

### *1.2.3 Variance Estimation*

The asymptotic variance of the WLE is a sum of the inverse of the efficient information for the complete data model and a linear combination of the conditional variances of the efficient influence function for a complete data model given a stratum membership (see (21) of [8]). Thus, if this efficient influence function is a known function, the Horvitz-Thompson estimators of the efficient information and the conditional variances will consistently estimate the asymptotic variance of the WLE with the aid of the Glivenko-Cantelli theorem for the inverse probability weighted empirical processes, discussed in Chapter 6.

In general, the efficient influence function contains unknown functions, especially in a general semiparametric model, so that the simple method discussed above cannot be applied easily. There are several methods for variance estimation of the MLE in the complete data model where the asymptotic variance of the MLE of a finite dimensional parameter is the inverse of the efficient information. A discretized version of the efficient information at observed data points can be computed for estimation of the efficient information [39, 22, 33]. Another method for the complete data model is (numerical) differentiation of the profile log likelihood by [35] where the nuisance parameter is profiled out. Although obvious modifications of these methods by inverse probability weighting are available for two-phase sampling data, they only estimate the phase I variance.

Bootstrap is a more general method for complete data. For example, exchangeably weighted bootstraps can be used to estimate the asymptotic variance of $Z$-estimators including the MLE under appropriate regularity conditions regardless of whether or not the asymptotic variance involves unknown functions [60]. For two-phase sampling, several bootstrap procedures have been proposed for estimation in a finite population in survey sampling. Bootstrap procedures in survey sampling are often targeted at variance estimation of a rather simple finite population parameter (e.g. finite population mean). For instance, [43] modifies the values of the variables in the bootstrap sample to obtain the unbiased estimate of the variance of simple estimators. Rescaled bootstrap methods such as [43] fail to have distributional consistency in general. Also, it is not obvious how to extend this method to estimation of a relatively complicated estimator such $Z$-estimators in a general semiparametric model. Another example in survey sampling is the method of [18, 3] that creates an artificial population by copying the original sample several times proportional to the inverses of the sampling probabilities. Although this method has distributional consistency for weighted sample averages under our sampling scheme, the extension to the process level, which is usually required for studying a general semiparametric model, has not yet been justified. More importantly, all bootstrap methods in survey sampling aim at yielding phase II variances only. There is no known bootstrap methods to yield phase I and II variances at the same time to the best of our knowledge.

### *1.2.4 Previous Results*

The WLE in a general semiparametric model is already well-studied in cases where nuisance parameters are estimable at regular rates. [8] derived the asymptotic distribution of the WLE under stratified Bernoulli sampling and stratified sampling without replacement. [9] studied the WLE with estimated weights under stratified Bernoulli sampling. [5, 6] obtained in a heuristic way the asymptotic distributions of the the WLE's with estimated weights and the calibrated WLE under stratified sampling without replacement. One of the difficulties in the derivations in [5, 6] involves the lack of a proof of asymptotic equicontinuity of certain stochastic processes under sampling without replacement. A similar difficulty is also recognized by [28] in the context of complex surveys. Direct application of empirical process theory does not help due to lack of independence among observations. Another difficulty, which is also seen in other papers, concerns (lack of) proofs of consistency of estimators under sampling without replacement . When a nuisance parameter is not estimable at a regular rate, no general consistency, rate of convergence, or asymptotic normality results are known in the framework of two-phase designs to the best of our knowledge.

Estimated weights and calibration are less studied under sampling without replacement. Efficiency improvements by estimated weights and a specific type of calibration (within-stratum calibration, see Chapter 3) has previously been proved only under Bernoulli sampling [9, 5, 6] for the plain WLE in a general semiparametric model. The modified calibration of [10] has not yet been studied under both sampling schemes for a general semiparametric model. [15] showed the asymptotic equivalence of the calibrated Horvitz-Thompson estimators of a certain variable among different calibrations in the complex survey including sampling without replacement under the standard conditions of [23] to survey sampling, but the efficiency gain over the plain Horvitz-Thompson estimator was not discussed. Although simulation studies often show efficiency improvements by these methods under sampling without replacement, their theoretical behaviors have not been well-studied with rigor. A possible reason for this would be the lack of theoretical tools to study the inverse probability weighted empirical processes under sampling without replacement.

Research on variance estimation for the WLE's of a finite dimensional parameter has been restricted to the case where the asymptotic variances are represented as sums of expectations of known functions. The asymptotic variances, in this case, are easily estimated by Horvitz-Thompson estimators. If the asymptotic variances involve unknown functions, variance estimation has been studied only for the MLE in the complete data model (e.g.,[35]).

Several bootstrap methods are proposed for estimation of parameters in the finite population in survey sampling as briefly discussed above. For bootstrap methods in survey sampling other than [43, 18, 3], see, for example, [2] and references therein. In this dissertation, we revisit the method of [18, 3] because distributional consistency is our primary interest. Another issue related to the result in this dissertation is how to carry out calibration when bootstrapping. Rescaled bootstrap methods cannot accommodate calibration at all, while other methods often fail to provide theoretical justifications of calibrated bootstrap. For instance, an often proposed method of calibrating to the same known population means has not yet been rigorously studied.

## *1.3 Goals and Outline of the Thesis*

The main goal of this thesis is to develop asymptotic theory for weighted likelihood estimation methods in the setting of general semiparametric models taking account of the dependence of observations due to the "without replacement" sampling scheme. Our aim is to provide theoretical results both when the infinite-dimensional nuisance parameter is estimable at a regular rate ($\sqrt{n}$) and when the infinite-dimensional nuisance parameter is estimable at a non-regular rate (slower than $\sqrt{n}$). Specifically, we prove two $Z$-theorems giving weak sufficient conditions for asymptotic distributions of the WLE's in general semi-parametric models. The first theorem, which covers the case where the nuisance parameter is estimable at a regular ($\sqrt{n}$) rate, provides a rigorous justification of the results of [5, 6] under weaker conditions. The second theorem covers the case when only estimators with slower than $\sqrt{n}$-convergence rates are available for the infinite-dimensional nuisance parameter. In addition to the plain WLE, we include the WLE's with estimated weights and (variants of) calibration in the formulations of both theorems. The conditions of our theorems are formulated in terms of complete data, not two-phase sampling data, and, moreover, they are almost identical to those for the MLE with complete data. Thus, most of them have been already established in many applications. For the conditions requiring verification, tools from empirical process theory will be applied.

To achieve the main objective, we develop a set of empirical process tools including a Glivenko-Cantelli theorem, a theorem for rates of convergence of $Z$-estimators, and a Donsker theorem for the inverse probability weighted empirical processes under two-phase sampling and sampling without replacement at the second phase. Some results such as Glivenko-Cantelli theorem (Theorem 6.1.1) and Donsker Theorem (Theorem 6.3.1) are of interest in their own right. These results, accounting for dependence of observations due to the "without replacement" sampling design, are used to prove our $Z$-theorems in place of the usual empirical process theory. More importantly, they are useful in verifying the conditions of $Z$-theorems in applications. For instance, our Theorem 6.2.1 easily establishes rates of convergence under our "without replacement" sampling scheme. Also, consistency can be verified with the aid of the Glivenko-Cantelli theorem. We illustrate application of the general results with examples in Chapter 5.

To improve efficiency, we introduce a new method called within-stratum centered calibration. We prove that this method yields improved asymptotic efficiency over the plain WLE under our sampling scheme. We compare this method with estimated weights and other types of calibration under stratified Bernoulli sampling and stratified sampling without replacement. We note that our method of centered calibration is the only method guaranteed to gain efficiency under both sampling schemes while other methods are warranted only for stratified Bernoulli sampling.

For variance estimation, we propose two different procedures based on bootstrap techniques. The first method is to estimate the phase I and II variances, separately, which allows us to evaluate how much information we lose by two-phase sampling. The second method, which accounts for the phase I and II variances at the same time, provides valid estimates even under model misspecification.

The rest of this thesis is organized as follows. In Chapter 2, we introduce our sampling scheme and estimation procedures in a general semiparametric model. The WLE and methods involving adjusted weights intended to improve on the efficiency of the WLE are discussed. Two $Z$-theorems are presented in Chapter 3 to derive asymptotic distributions of the WLE's of the finite dimensional parameter. All estimators are compared under Bernoulli sampling and sampling without replacement with different methods of adjusting the weights. Two bootstrap-based procedures for variance estimation are introduced in Chapter 4. We apply our $Z$-theorems to the Cox model, both with right censoring and interval censoring, in Chapter 5. The WLE of the cumulative baseline hazard function has regular rate of convergence in the first example, while it has (non-regular) cube-root rate in the second example. In addition, simulation studies and data analysis are presented. Chapter 6 consists of general results for IPW empirical processes while Chapter 7 presents general results for bootstrap IPW empirical processes.

# Chapter 2

# SAMPLING, MODELS, AND ESTIMATORS

## *2.1 Sampling*

We now introduce our sampling scheme. Most of the following notation is based on [8]. Let $W = (X, U) \in \mathcal{W} = \mathcal{X} \times \mathcal{U}$ be the complete data with distribution $\tilde{P}_0$ where $X$ is the vector of the variables of interest with distribution $P_0$ and $U$ is a vector of auxiliary variables. At the first phase, only a coarsening $\tilde{X} = \tilde{X}(X)$ of $X$ and the auxiliary variables $U$ are available for all $N$ subjects. The phase I data $V = (\tilde{X}, U) \in \mathcal{V} = \tilde{\mathcal{X}} \times \mathcal{U}$ are used to form the $J$ sampling strata $\mathcal{V}_j$ with $\sum_{j=1}^{J} \mathcal{V}_j = \mathcal{V}$, the $j$th of which consists of $N_j$ subjects for $j = 1, \ldots, J$. After stratified sampling, $X$ is fully observed for $n_j$ subjects in the $j$th stratum at the second phase. The observed data is $(V, X\xi, \xi)$ where $\xi$ is the indicator of being sampled at the second phase. We use a doubly subscripted notation by which $V_{j,i}$, for example, denotes $V$ for the $i$th subject in stratum $j$. We denote the stratum probability for the $j$th stratum by $\nu_j \equiv \tilde{P}_0(V \in \mathcal{V}_j)$, and the conditional expectation given membership in the $j$th stratum by $P_{0|j}(\cdot) \equiv \tilde{P}_0(\cdot | V \in \mathcal{V}_j)$.

At the second phase, samples of size $n_j \leq N_j$ are drawn at random without replacement from each of the $J$ strata. The sampling probability is $P(\xi = 1 | V_i) = \pi_0(V_i) = n_j / N_j$ for $V_i \in \mathcal{V}_j$. These sampling probabilities are assumed to be strictly positive; that is, there is a strictly positive constant $\sigma > 0$ such that $0 < \sigma \leq \pi_0(v) \leq 1 \quad \text{for} \quad v \in \mathcal{V}$. We assume that $n_j / N_j \to p_j > 0$ for $j = 1, \ldots, J$ as $N \to \infty$. Although dependence is induced among the observations $(V_i, \xi_i X_i, \xi_i)$ by the sampling indicators, the vector of sampling indicators $(\xi_{j1}, \ldots, \xi_{jN_j})$ within strata, $j = 1, \ldots, J$, are exchangeable for $j = 1, \ldots, J$, and the $J$ random vectors $(\xi_{j1}, \ldots, \xi_{jN_j})$ are independent.

One of the most important tools in empirical process theory is the empirical measure. However, the empirical measure is not directly applicable to estimation under two-phase sampling because some observations are not observed at the second phase. Instead, we define the inverse probability weighted (IPW) empirical measure by

$$\mathbb{P}_N^\pi = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_0(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{j=1}^J \sum_{i=1}^{N_j} \frac{\xi_{j,i}}{n_j/N_j}\delta_{X_{j,i}},$$

where $\delta_{X_i}$ denotes a Dirac measure placing unit mass on $X_i$. The identity in the last display is justified by the arguments in Appendix A of [8]. We also define the IPW empirical process by $\mathbb{G}_N^\pi = \sqrt{N}(\mathbb{P}_N^\pi - P_0)$ and the phase II empirical process for the $j$th stratum by

$$\mathbb{G}_{j,N_j}^\xi \equiv \sqrt{N_j}\left(\mathbb{P}_{j,N_j}^\xi - \frac{n_j}{N_j}\mathbb{P}_{j,N_j}\right), \quad j = 1, \ldots, J,$$

where, for $j \in \{1, \ldots, J\}$, $\mathbb{P}_{j,N_j}^\xi \equiv N_j^{-1}\sum_{i=1}^{N_j} \xi_{j,i}\delta_{X_{j,i}}$ is the phase II empirical measure for the $j$th stratum, and $\mathbb{P}_{j,N_j} \equiv N_j^{-1}\sum_{i=1}^{N_j} \delta_{X_{j,i}}$ is the empirical measure for all the data in the $j$th stratum; note the latter empirical measure is not observed. Then, following [8], page 207, we decompose $\mathbb{G}_N^\pi$ as follows:

$$\mathbb{G}_N^\pi \quad = \quad \mathbb{G}_N + \sum_{j=1}^J \sqrt{\frac{N_j}{N}}\left(\frac{N_j}{n_j}\right)\mathbb{G}_{j,N_j}^\xi. \tag{2.1}$$

where $\mathbb{P}_N = N^{-1}\sum_{j=1}^J N_j\mathbb{P}_{j,N_j}$ and $\mathbb{G}_N = \sqrt{N}(\mathbb{P}_N - P_0)$. Notice that the phase II empirical processes $\mathbb{G}_{j,N_j}^\xi$ correspond to "exchangeably weighted bootstrap" versions of the stratum-wise complete data empirical processes $\mathbb{G}_{j,N_j} \equiv \sqrt{N_j}(\mathbb{P}_{j,N_j} - P_{0|j})$ where $P_{0|j}$ is the conditional distribution of $X$ given membership in the $j$th stratum and $\mathbb{P}_{j,N_j}$ is as defined above. This observation allows application of the "exchangeably weighted bootstrap" theory of [41].

### 2.2 Improving Efficiency by Adjusting Weights

Efficiency of estimators based on IPW empirical processes can be improved by adjusting weights, either by estimated weights [46] or by calibration [15] via use of the phase I information; see also [30]. In addition to these two methods, we discuss two variants of calibration, modified calibration [10], and our proposed method, centered calibration.

Let $Z_i \equiv g(V_i)$ be the auxiliary variables for the $i$th subject for a known transformation $g$. For estimated weights through binary regression, the first $J$ elements of $Z_i$ are the membership indicators for the strata, $I_{\mathcal{V}_j}(V_i), j = 1, \ldots, J$. Furthermore, observations with

$\pi_0(V) = 1$ are dropped from binary regression, and the original weight 1 is used. For notational simplicity, we write $Z_i$ for either method, and assume that sampling probabilities are strictly less than 1 for all strata.

### *2.2.1 Estimated Weights*

The method of estimated weights adjusts weights through binary regression on the phase I variables. The sampling probability for the $i$th subject is modelled by $p_\alpha(\xi_i|Z_i) = G_e(Z_i^T\alpha)^{\xi_i}(1 - G_e(Z_i^T\alpha))^{1-\xi_i} \equiv \pi_\alpha(V_i)^{\xi_i}\{1 - \pi_\alpha(V_i)\}^{1-\xi_i}$, where $\alpha \in \mathcal{A}_e \subset \mathbb{R}^{J+k}$ is a regression parameter and $G_e : \mathbb{R} \mapsto [0, 1]$ is a known function. If $G_e(x) = e^x/(1 + e^x)$ for instance, then the adjustment simply involves logistic regression. Let $\hat{\alpha}_N$ be the estimator of $\alpha$ that maximizes the composite likelihood

$$\prod_{i=1}^{N} p_\alpha(\xi_i|Z_i) = \prod_{i=1}^{N} G_e(Z_i^T\alpha)^{\xi_i}(1 - G_e(Z_i^T\alpha))^{1-\xi_i}. \tag{2.2}$$

Note that this is not an ordinary likelihood due to the dependence among observations under our sampling scheme while this becomes an ordinary likelihood under Bernoulli sampling. We define the IPW empirical measure with estimated weights by

$$\mathbb{P}_N^{\pi,e} = \frac{1}{N}\sum_{i=1}^{N} \frac{\xi_i}{\pi_{\hat{\alpha}_N}(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{i=1}^{N} \frac{\xi_i}{\pi_0(V_i)}\frac{\pi_0(V_i)}{G_e(Z_i^T\hat{\alpha}_N)}\delta_{X_i},$$

and the IPW empirical process with estimated weights by $\mathbb{G}_N^{\pi,e} = \sqrt{N}(\mathbb{P}_N^{\pi,e} - P_0)$.

### *2.2.2 Calibration*

Calibration adjusts weights so that the inverse probability weighted average from the phase II sample is equated to the phase I average, whereby the phase I information is taken into account for estimation. Consider the problem of choosing the weights $\{w_i\}_{i=1}^N$ subject to the condition

$$\frac{1}{N}\sum_{i=1}^{N} \xi_i w_i Z_i = \frac{1}{N}\sum_{i=1}^{N} Z_i; \tag{2.3}$$

since the $Z_i$'s take values in $\mathbb{R}^k$, this is a system of equations in $\mathbb{R}^k$. In general there are many solutions to this system of equations, and the inverse probability weights $1/\pi_0(V_i)$

will typically not satisfy it. Because weights differing greatly from the inverse probability weights are unlikely to improve on the plain weighted likelihood estimates, calibration involves choosing weights closest to the inverse probability weights in a certain distance measure. Let $D_i(w,d)$ be a distance measure between the weights $w$ and $d$ for the $i$th subject, where for every fixed $d > 0$, $D_i(w,d)$ is nonnegative, continuously differentiable with respect to $w$, and strictly convex in $w$, and $(\partial/\partial w)D_i(w,d)$ is strictly increasing and is zero at $w = d$ (see [15] for various choices of $D_i$). The resulting problem is a convex optimization problem: find positive weights $w_i$ that minimize the average distance $N^{-1}\sum_{i=1}^{N} D_i(w_i, 1/\pi_0(V_i))$ subject to the constraint (2.3). The method of Lagrange multipliers leads to $(\partial/\partial w)D_i(w_i, 1/\pi_0(V_i))+Z_i^T\alpha = 0$ for the subjects with $\xi_i = 1$ where $\alpha$ is a Lagrange multiplier. The invertibility of $(\partial/\partial w)D_i$ leads to the solution $w_i = G_i(Z_i^T\alpha)/\pi_0(V_i)$ for some function $G_i$ where $G_i(0) = 1$ and $\dot{G}_i(0) > 0$. Substitution in (2.3) gives the calibration equation $N^{-1}\sum_{i=1}^{N}\xi_i(G_i(Z_i^T\alpha)/\pi_0(V_i))Z_i = N^{-1}\sum_{i=1}^{N} Z_i$. The solution $\hat{\alpha}$ to the calibration equation determines the calibrated weights $\hat{w}_i = G_i(Z_i^T\hat{\alpha})/\pi_0(V_i)$.

One easy choice, as in [5, 6, 30], is to take the distance measures $D_i$ to be the same for all subjects; i.e., $D_i = D$ and $G_i = G_c$ for $i = 1, \ldots, N$. An alternative subject specific choice of the $D_i$'s, leading to "modified calibration", will be discussed in the next subsection. In both cases we formulate the calibration in terms of the calibration equation rather than the problem of minimizing a distance with the inverse probability weights. These assumptions simplify the condition on $G_i$'s and $D_i$'s. In our formulation with equal $D_i$'s, we find an estimator $\hat{\alpha}_N$ that is the solution for $\alpha \in \mathcal{A}_c \subset \mathbb{R}^k$ of the following calibration equation,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i G_c(V_i;\alpha)}{\pi_0(V_i)}Z_i = \frac{1}{N}\sum_{i=1}^{N} Z_i, \tag{2.4}$$

where $G_c(V;\alpha) \equiv G(g(V)^T\alpha) = G(Z^T\alpha)$, and $G$ is a known function with $G(0) = 1$ and $\dot{G}(0) > 0$. We call $\pi_\alpha(V_i) \equiv \pi_0(V_i)/G_c(V_i;\alpha)$ the calibrated sampling probability for the $i$th subject. We define the calibrated IPW empirical measure by

$$\mathbb{P}_N^{\pi,c} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_{\hat{\alpha}_N}(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_0(V_i)}G\left(Z_i^T\hat{\alpha}_N\right)\delta_{X_i},$$

and the calibrated IPW empirical process by $\mathbb{G}_N^{\pi,c} = \sqrt{N}(\mathbb{P}_N^{\pi,c} - P_0)$.

### *2.2.3 Modified Calibration*

[15] discussed subject-dependent distance measures $D_i$ when $(\partial/\partial w)D_i = \tilde{D}(w/d)/q_i$ where $\tilde{D}(x)$ is a continuous, strictly increasing function on $\mathbb{R}$ with $\tilde{D}(1) = 0$ and $(d/dx)\tilde{D}(1) = 1$, independent of the index $i$, and $q_i > 0$. Solving the convex optimization problem in this case with some choice of $q_i$'s leads to the calibration equation $N^{-1}\sum_{i=1}^N \xi_i(G(q_i Z_i^T\alpha)/\pi_0(V_i))Z_i = N^{-1}\sum_{i=1}^N Z_i$ for the inverse $G = (\tilde{D})^{-1}$ of $\tilde{D}$. Recently the choice $q_i = (1-\pi_0(V_i))/\pi_0(V_i)$ was proposed by [10] in a missing response problem. When $\pi_0(V_i) < 1$, $i = 1, \dots, N$, this choice means that when the sampling probability is larger, the subject contributes more to the average distance. Note that $q_i = 0$ when $\pi_0(V_i) = 1$. Although $q_i$ must be strictly positive in the original formulation of the problem to minimize the distance with the inverse probability weights, $q_i = 0$ is valid in the calibration equation. One implication of this choice is that we do not modify the weights if subjects are always sampled at the second phase. We call the method of choosing weights by solving the calibration equation with $q_i = (1-\pi_0(V_i))/\pi_0(V_i)$ *modified calibration.*

In modified calibration, we find the estimator $\hat{\alpha}_N$ that is the solution for $\alpha \in \mathcal{A}_{mc} \subset \mathbb{R}^k$ of the following calibration equation:

$$\frac{1}{N}\sum_{i=1}^N \frac{\xi_i G_{mc}(V_i;\alpha)}{\pi_0(V_i)} Z_i = \frac{1}{N}\sum_{i=1}^N Z_i, \tag{2.5}$$

where

$$G_{mc}(V;\alpha) \equiv G\left(\frac{1-\pi_0(V)}{\pi_0(V)} g(V)^T\alpha\right) = G\left(\frac{1-\pi_0(V)}{\pi_0(V)} Z^T\alpha\right).$$

Here $G$ is a known function with $G(0) = 1$ and $\dot{G}(0) > 0$. We call $\pi_\alpha(V_i) \equiv \pi_0(V_i)/G_{mc}(V_i;\alpha)$ the calibrated sampling probability with modified calibration for the $i$th subject. We define the IPW empirical measure with modified calibration by

$$\mathbb{P}_N^{\pi,mc} = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_{\hat{\alpha}_N}(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_0(V_i)} G\left(\frac{1-\pi_0(V_i)}{\pi_0(V_i)} Z_i^T\hat{\alpha}_N\right)\delta_{X_i},$$

and the IPW empirical process with modified calibration by $\mathbb{G}_N^{\pi,mc} = \sqrt{N}(\mathbb{P}_N^{\pi,mc} - P_0)$.

### *2.2.4 Centered Calibration*

We propose a new method, centered calibration, that calibrates on centered auxiliary variables with modified calibration. This method in fact improves the plain WLE under our sampling scheme, while retaining the good properties of modified calibration. We discuss advantages of centered calibration and connections to other methods in Section 3.4.3.

In centered calibration, we find the estimator $\hat{\alpha}_N$ that is the solution for $\alpha \in \mathcal{A}_{cc} \subset \mathbb{R}^k$ of the following calibration equation:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i G_{cc}(V_i;\alpha)}{\pi_0(V_i)}(Z_i - \overline{Z}_N) = 0, \tag{2.6}$$

where

$$G_{cc}(V;\alpha) \equiv G\left(\frac{1-\pi_0(V)}{\pi_0(V)}\{Z - \overline{Z}_N\}^T\alpha\right),$$

with $\overline{Z}_N = N^{-1}\sum_{i=1}^{N} Z_i$ suppressed in the definition of $G_{cc}$. Here $G$ is a known function with $G(0) = 1$ and $\dot{G}(0) > 0$. We call $\pi_\alpha(V_i) \equiv \pi_0(V_i)/G_{cc}(V_i;\alpha)$ the calibrated sampling probability with centered calibration for the $i$th subject. We define the IPW empirical measure with centered calibration by

$$\mathbb{P}_N^{\pi,cc} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_{\hat{\alpha}_N}(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_0(V_i)}G_{cc}(V_i;\hat{\alpha}_N)\delta_{X_i},$$

and the IPW empirical process with centered calibration by $\mathbb{G}_N^{\pi,cc} = \sqrt{N}(\mathbb{P}_N^{\pi,cc} - P_0)$.

## 2.3 Models and Estimators

We study the asymptotic distribution of the weighted likelihood estimator of a finite dimensional parameter $\theta$ in a general semiparametric model $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$ where $\Theta \subset \mathbb{R}^p$ and the nuisance parameter space $H$ is a subset of some Banach space $\mathcal{B}$. Let $P_0 = P_{\theta_0,\eta_0}$ denote the true distribution.

The maximum likelihood estimator with complete data is often obtained as a solution of the infinite dimensional likelihood equations. In such models, the WLE under two-phase sampling is obtained by solving the corresponding infinite dimensional inverse probability weighted likelihood equations. Specifically, the WLE $(\hat{\theta}_N, \hat{\eta}_N)$ is a solution of the following weighted likelihood equations

$$\Psi^{\pi}_{N,1}(\theta,\eta) = \mathbb{P}^{\pi}_N \dot{\ell}_{\theta,\eta} = o_{P^*}\left(N^{-1/2}\right),$$
$$\left\|\Psi^{\pi}_{N,2}(\theta,\eta)h\right\|_{\mathcal{H}} = \|\mathbb{P}^{\pi}_N(B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h)\|_{\mathcal{H}} = o_{P^*}\left(N^{-1/2}\right), \tag{2.7}$$

where $\dot{\ell}_{\theta,\eta} \in \mathcal{L}^0_2(P_{\theta,\eta})^p$ is the score function for $\theta$, and the score operator $B_{\theta,\eta} : \mathcal{H} \mapsto \mathcal{L}^0_2(P_{\theta,\eta})$ is the bounded linear operator mapping a direction $h$ in some Hilbert space $\mathcal{H}$ of one-dimensional submodels for $\eta$ along which $\eta \to \eta_0$. The corresponding WLE with estimated weights $(\hat{\theta}_{N,e}, \hat{\eta}_{N,e})$, the calibrated WLE $(\hat{\theta}_{N,c}, \hat{\eta}_{N,c})$, the WLE with modified calibration $(\hat{\theta}_{N,mc}, \hat{\eta}_{N,mc})$, and the WLE with centered calibration $(\hat{\theta}_{N,cc}, \hat{\eta}_{N,cc})$ are obtained by replacing $\mathbb{P}^{\pi}_N$ by $\mathbb{P}^{\pi,e}_N$, $\mathbb{P}^{\pi,c}_N$, $\mathbb{P}^{\pi,mc}_N$ or $\mathbb{P}^{\pi,cc}_N$ in (2.7), respectively. Let $\dot{\ell}_0 = \dot{\ell}_{\theta_0,\eta_0}$ and $B_0 = B_{\theta_0,\eta_0}$.

# Chapter 3

# ASYMPTOTICS FOR THE WLE IN GENERAL SEMIPARAMETRIC MODELS

We consider two cases: in the first case the nuisance parameter $\eta$ is estimable at a regular (i.e., $\sqrt{n}$) rate and, for ease of exposition, $\eta$ is assumed to be a measure. In the second case $\eta$ is only estimable at a non-regular (slower than $\sqrt{n}$) rate. Our theorem (Theorem 3.3.1) concerning the second case nearly covers the former case, but requires slightly more smoothness and a separate proof of the rate of convergence for an estimator of $\eta$. On the other hand, our theorem (Theorem 3.2.1) concerning the former case includes a proof of the (regular) ($\sqrt{n}$) rate of convergence, and hence is of interest by itself.

In the following, we present two $Z$-theorems for weighted likelihood estimation. Then we compare asymptotic variances of WLE's with different adjusting weights under different designs (Bernoulli sampling vs. sampling without replacement). We introduce a new calibration method that guarantees efficiency gains over the plain WLE regardless of the different designs. All proofs are presented at the end of this chapter.

## *3.1 Conditions for Adjusting Weights*

To derive asymptotic distributions of WLE's with estimated weights and (modified and centered) calibration, we need to establish asymptotic results on estimators of $\alpha$ with estimating weights and (modified and centered) calibration. To this end, we assume the following. Throughout this dissertation, we may assume both Conditions 3.1.1 and 3.1.2 at the same time, but it should be understood that the former condition is used exclusively for the estimators regarding estimated weights and the latter condition is imposed only for estimators regarding (modified and centered) calibration. Moreover, it should be understood that Conditions 3.2(a)(i) and 3.2(d)(i) are assumed for the estimators regarding calibration, Conditions 3.2(a)(ii) and 3.2(d)(ii) are imposed for the estimators regarding modified calibration and that Conditions 3.2(a)(iii) and 3.2(d)(iii) are used for the estimators regarding centered calibration, respectively.

**Condition 3.1.1** (Estimated weights). *(a) The estimator $\hat{\alpha}_N$ is a maximizer of the composite likelihood (2.2).*

*(b) $Z \in \mathbb{R}^{J+k}$ is not concentrated on a $(J+k)$-dimensional affine space of $\mathbb{R}^{J+k}$ and has bounded support.*

*(c) $G_e : \mathbb{R} \mapsto [0,1]$ is a twice continuously differentiable, monotone function.*

*(d) $S_0 \equiv P_0 \left( \{\dot{G}_e(Z^T\alpha_0)\}^2 \{\pi_0(V)(1-\pi_0(V))\}^{-1} Z^{\otimes 2} \right)$ is finite and nonsingular, where $\dot{G}_e$ is a derivative of $G_e$.*

*(e) The "true" parameter $\alpha_0 = (\alpha_{0,1}, \ldots, \alpha_{0,J+k})$ is given by $\alpha_{0,j} = G_e^{-1}(p_j)$, for $j = 1, \ldots, J$, and $\alpha_{0,j} = 0$, for $j = J+1, \ldots, J+k$. The parameter $\alpha$ is identifiable, that is, $p_\alpha = p_{\alpha_0}$ almost surely implies $\alpha = \alpha_0$.*

*(f) For a fixed $p_j \in (0,1)$, $n_j$ satisfies $n_j = [N_j p_j]$ for $j = 1, \ldots, J$.*

**Condition 3.1.2** ((modified and centered) Calibration). *(a) (i) The estimator $\hat{\alpha}_N = \hat{\alpha}_N^c$ is a solution of the calibration equation (2.4). (ii) The estimator $\hat{\alpha}_N = \hat{\alpha}_N^{mc}$ is a solution of the calibration equation (2.5). (iii) The estimator $\hat{\alpha}_N = \hat{\alpha}_N^{cc}$ is a solution of the calibration equation (2.6).*

*(b) The distribution of $Z \in \mathbb{R}^k$ is not concentrated at 0 and has bounded support.*

*(c) $G$ is a strictly increasing continuously differentiable function on $\mathbb{R}$ such that $G(0) = 1$ and for all $x$, $-\infty < m_1 \leq G(x) \leq M_1 < \infty$ and $0 < \dot{G}(x) \leq M_2 < \infty$, where $\dot{G}$ is the derivative of $G$.*

*(d) (i) $P_0 Z^{\otimes 2}$ is finite and positive definite. (ii)$P_0[\pi_0(V)^{-1}(1-\pi_0(V))Z^{\otimes 2}]$ is finite and positive definite. (iii)$P_0[\pi_0(V)^{-1}(1-\pi_0(V))(Z-\mu_Z)^{\otimes 2}]$ is finite and positive definite where $\mu_Z = PZ$.*

*(e) The "true" parameter $\alpha_0 = 0$.*

Condition 3.1.1 (f) may seem unnatural at first, but in practice the phase II sample size $n_j$ can be chosen by the investigator so that the sampling probability $p_j$ can be understood to be automatically chosen to satisfy $n_j = [N_j p_j]$. The other parts of Condition 3.1.1 are standard in binary regression, and Condition 3.1.2 is similar to Condition 3.1.1.

Asymptotic properties of $\hat{\alpha}_N$ for all cases (estimated weights and (modified and centered) calibration) are proved in [50].

### 3.2 Regular Rate for a Nuisance Parameter

We assume the following conditions.

**Condition 3.2.1** (Consistency)**.** *The estimator $(\widehat{\theta}_N, \widehat{\eta}_N)$ is consistent for $(\theta_0, \eta_0)$ and solves the weighted likelihood equations (2.7), where $\mathbb{P}_N^{\pi}$ may be replaced by $\mathbb{P}_N^{\pi,e}$, $\mathbb{P}_N^{\pi,c}$, $\mathbb{P}_N^{\pi,mc}$ or $\mathbb{P}_N^{\pi,cc}$ for the estimators with estimated weights, calibration, modified calibration or centered calibration.*

**Condition 3.2.2** (Asymptotic equicontinuity)**.** *Let $\mathcal{F}_1(\delta) = \{\dot{\ell}_{\theta,\eta} : |\theta - \theta_0| + \|\eta - \eta_0\| < \delta\}$ and $\mathcal{F}_2(\delta) = \{B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h : h \in \mathcal{H}, |\theta - \theta_0| + \|\eta - \eta_0\| < \delta\}$. There exists a $\delta_0 > 0$ such that (1) $\mathcal{F}_k(\delta_0), k = 1, 2,$ are $P_0$-Donsker and $\sup_{h\in\mathcal{H}} P_0|f_j - f_{0,j}|^2 \to 0$, as $|\theta - \theta_0| + \|\eta - \eta_0\| \to 0$, for every $f_j \in \mathcal{F}_j(\delta_0), j = 1, 2,$ where $f_{0,1} = \dot{\ell}_{\theta_0,\eta_0}$ and $f_{0,2} = B_0h - P_0B_0h$, (2) $\mathcal{F}_k(\delta_0), k = 1, 2,$ have integrable envelopes.*

**Condition 3.2.3.** *The map $\Psi = (\Psi_1, \Psi_2) : \Theta \times H \mapsto \mathbb{R}^p \times \ell^\infty(\mathcal{H})$ with components*

$$\begin{aligned}\Psi_1(\theta, \eta) &\equiv P_0\Psi_{N,1}(\theta, \eta) = P_0\dot{\ell}_{\theta,\eta},\\ \Psi_2(\theta, \eta)h &\equiv P_0\Psi_{N,2}(\theta, \eta) = P_0B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h, \quad h \in \mathcal{H},\end{aligned}$$

*has a continuously invertible Fréchet derivative map $\dot{\Psi}_0 = (\dot{\Psi}_{11}, \dot{\Psi}_{12}, \dot{\Psi}_{21}, \dot{\Psi}_{22})$ at $(\theta_0, \eta_0)$ given by $\dot{\Psi}_{ij}(\theta_0, \eta_0)h = P_0(\dot{\psi}_{i,j,\theta_0,\eta_0,h})$, $i, j \in \{1, 2\}$ in terms of $L_2(P_0)$ derivatives of $\psi_{1,\theta,\eta,h} = \dot{\ell}_{\theta,\eta}$ and $\psi_{2,\theta,\eta,h} = B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h$; that is,*

$$\begin{aligned}\sup_{h\in\mathcal{H}} \left\{P_0\left(\psi_{i,\theta,\eta_0,h} - \psi_{i,\theta_0,\eta_0,h} - \dot{\psi}_{i1,\theta_0,\eta_0,h}(\theta - \theta_0)\right)^2\right\}^{1/2} &= o(\|\theta - \theta_0\|),\\ \sup_{h\in\mathcal{H}} \left\{P_0\left(\psi_{i,\theta_0,\eta,h} - \psi_{i,\theta_0,\eta_0,h} - \dot{\psi}_{i2,\theta_0,\eta_0,h}(\eta - \eta_0)\right)^2\right\}^{1/2} &= o(\|\eta - \eta_0\|).\end{aligned}$$

*Furthermore, $\dot{\Psi}_0$ admits a partition*

$$(\theta - \theta_0, \eta - \eta) \mapsto \begin{pmatrix} \dot{\Psi}_{11} & \dot{\Psi}_{12} \\ \dot{\Psi}_{21} & \dot{\Psi}_{22} \end{pmatrix} \begin{pmatrix} \theta - \theta_0 \\ \eta - \eta_0 \end{pmatrix},$$

*where*

$$\dot{\Psi}_{11}(\theta-\theta_0) = -P_{\theta_0,\eta_0}\dot{\ell}_{\theta_0,\eta_0}\dot{\ell}^T_{\theta_0,\eta_0}(\theta-\theta_0),$$
$$\dot{\Psi}_{12}(\eta-\eta_0) = -\int B^*_{\theta_0,\eta_0}\dot{\ell}_{\theta_0,\eta_0}d(\eta-\eta_0),$$
$$\dot{\Psi}_{21}(\theta-\theta_0)h = -P_{\theta_0,\eta_0}B_{\theta_0,\eta_0}h\dot{\ell}^T_{\theta_0,\eta_0}(\theta-\theta_0),$$
$$\dot{\Psi}_{22}(\eta-\eta_0)h = -\int B^*_{\theta_0,\eta_0}B_{\theta_0,\eta_0}hd(\eta-\eta_0),$$

*and* $B^*_{\theta_0,\eta_0}B_{\theta_0,\eta_0}$ *is continuously invertible.*

Let $\tilde{I}_0 = P_0[(I-B_0(B_0^*B_0)^{-1}B_0^*)\dot{\ell}_0\dot{\ell}_0^T]$ be the efficient information for $\theta$ and $\tilde{\ell}_0 = \tilde{I}_0^{-1}(I-B_0(B_0^*B_0)^{-1}B_0^*)\dot{\ell}_0$ be the efficient influence function for $\theta$ for the semiparametric model with complete data.

**Theorem 3.2.1.** *Under Conditions 3.1.1-3.2.3,*

$$\begin{array}{lllll}
\sqrt{N}(\hat{\theta}_N-\theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z & \sim & N_p(0,\Sigma), \\
\sqrt{N}(\hat{\theta}_{N,e}-\theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,e}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_e & \sim & N_p(0,\Sigma_e), \\
\sqrt{N}(\hat{\theta}_{N,c}-\theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,c}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_c & \sim & N_p(0,\Sigma_c), \\
\sqrt{N}(\hat{\theta}_{N,mc}-\theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_{mc} & \sim & N_p(0,\Sigma_{mc}), \\
\sqrt{N}(\hat{\theta}_{N,cc}-\theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,cc}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_{cc} & \sim & N_p(0,\Sigma_{cc}),
\end{array}$$

*where*

$$\Sigma \equiv I_0^{-1} + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}(\tilde{\ell}_0), \tag{3.1}$$

$$\Sigma_e \equiv I_0^{-1} + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}((I-Q_e)\tilde{\ell}_0), \tag{3.2}$$

$$\Sigma_c \equiv I_0^{-1} + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}((I-Q_c)\tilde{\ell}_0), \tag{3.3}$$

$$\Sigma_{mc} \equiv I_0^{-1} + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}((I-Q_{mc})\tilde{\ell}_0), \tag{3.4}$$

$$\Sigma_{cc} \equiv I_0^{-1} + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}((I-Q_{cc})\tilde{\ell}_0), \tag{3.5}$$

*and (recall Conditions 3.1.1 and 3.1.2)*

$$
\begin{aligned}
Q_e f &\equiv P_0[\pi_0^{-1}(V) f \dot{G}_e(Z^T\alpha_0) Z^T] S_0^{-1} (1-\pi_0(V))^{-1} \dot{G}_e(Z^T\alpha_0) Z, \\
Q_c f &\equiv P_0[f Z^T]\{P_0 Z^{\otimes 2}\}^{-1} Z, \\
Q_{mc} f &\equiv P_0[(\pi_0^{-1}(V)-1) f Z^T]\{P_0[(\pi_0^{-1}(V)-1) Z^{\otimes 2}]\}^{-1} Z, \\
Q_{cc} f &\equiv P_0[(\pi_0^{-1}(V)-1) f (Z-\mu_Z)^T]\{P_0[(\pi_0^{-1}(V)-1)(Z-\mu_Z)^{\otimes 2}]\}^{-1}(Z-\mu_Z).
\end{aligned}
$$

**Remark 3.2.1.** *Our conditions in Theorem 3.2.1 are the same as those in [9] except the integrability condition. Our Condition 3.2.2 (2) requires existence of integrable envelopes for class of scores while the condition $(A1^*)$ in [9] requires square integrable envelopes. Note that this integrability condition is required only for the WLE with estimated weights and (modified and centered) calibration, as in [8].*

### 3.3 Non-regular Rate for a Nuisance Parameter

Set

$$B_{\theta,\eta}\left[\underline{h}\right] = (B_{\theta,\eta}h_1, \ldots, B_{\theta,\eta}h_p)^T$$

for $\underline{h} = (h_1, \ldots, h_p)^T$ where $h_k \in H$ for each $k = 1, \ldots, p$. We assume the following conditions.

**Condition 3.3.1** (Consistency and rate of convergence)**.** *An estimator $(\widehat{\theta}_N, \widehat{\eta}_N)$ of $(\theta_0, \eta_0)$ satisfies $|\widehat{\theta}_N - \theta_0| = o_P(1)$, and $\|\widehat{\eta}_N - \eta_0\| = O_P(N^{-\beta})$ for some $\beta > 0$.*

**Condition 3.3.2** (Positive information)**.** *There is an $\underline{h}^* = (h_1^*, \ldots, h_p^*)$, where $h_k^* \in H$ for $k = 1, \ldots, p$, such that*

$$P_0\left\{\left(\dot{\ell}_0 - B_0[\underline{h}^*]\right) B_0 h\right\} = 0$$

*for all $h \in H$. Furthermore, the efficient information $I_0 \equiv P_0\left(\dot{\ell}_0 - B_0[\underline{h}^*]\right)^{\otimes 2}$ for $\theta$ for the semiparametric model with complete data is finite and nonsingular. Denote the efficient influence function for the semiparametric model with complete data by $\tilde{\ell}_0 \equiv I_0^{-1}(\dot{\ell}_0 - B_0[\underline{h}^*])$.*

**Condition 3.3.3** (Asymptotic equicontinuity)**.** *(1) For any $\delta_N \downarrow 0$ and $C > 0$,*

$$
\begin{aligned}
\sup_{|\theta-\theta_0|\le\delta_N, \|\eta-\eta_0\|\le CN^{-\beta}} \left|\mathbb{G}_N(\dot{\ell}_{\theta,\eta} - \dot{\ell}_0)\right| &= o_P(1), \\
\sup_{|\theta-\theta_0|\le\delta_N, \|\eta-\eta_0\|\le CN^{-\beta}} |\mathbb{G}_N(B_{\theta,\eta} - B_0)[\underline{h}^*]| &= o_P(1).
\end{aligned}
$$

*(2) There exists a $\delta > 0$ such that the classes $\left\{\dot{\ell}_{\theta,\eta} : |\theta-\theta_0| + \|\eta-\eta_0\| \le \delta\right\}$ and $\{B_{\theta,\eta}[\underline{h}^*] : |\theta-\theta_0| + \|\eta-\eta_0\| \le \delta\}$ are $P_0$-Glivenko-Cantelli and have integrable envelopes. Moreover, $\dot{\ell}_{\theta,\eta}$ and $B_{\theta,\eta}[\underline{h}^*]$ are continuous with respect to $(\theta,\eta)$ either pointwise or in $L_1(P_0)$.*

**Condition 3.3.4** (Smoothness of the model)**.** *For some $\alpha > 1$ satisfying $\alpha\beta > 1/2$ and for $(\theta,\eta)$ in the neighborhood $\{(\theta,\eta) : |\theta-\theta_0| \le \delta_N, \|\eta-\eta_0\| \le CN^{-\beta}\}$,*

$$\begin{aligned}
&\left|P_0\left\{\dot{\ell}_{\theta,\eta} - \dot{\ell}_0 + \dot{\ell}_0(\dot{\ell}_0^T(\theta-\theta_0) + B_0[\eta-\eta_0])\right\}\right| \\
&\quad = o\left(|\theta-\theta_0|\right) + O\left(\|\eta-\eta_0\|^\alpha\right), \\
&\left|P_0\left\{(B_{\theta,\eta} - B_0)[\underline{h}^*] + B_0[\underline{h}^*](\dot{\ell}_0^T(\theta-\theta_0) + B_0[\eta-\eta_0])\right\}\right| \\
&\quad = o\left(|\theta-\theta_0|\right) + O\left(\|\eta-\eta_0\|^\alpha\right).
\end{aligned}$$

In the previous section, we required that the WLE solves the weighted likelihood equations (2.7) for all $h \in \mathcal{H}$. Here, we only assume that the WLE $(\hat{\theta}_N, \hat{\eta}_N)$ satisfies the weighted likelihood equations

$$\begin{aligned}
\Psi^\pi_{N,1}(\theta,\eta,\alpha) &= \mathbb{P}^\pi_N \dot{\ell}_{\theta,\eta} = o_{P^*}\left(N^{-1/2}\right), \\
\Psi^\pi_{N,2}(\theta,\eta,\alpha)\left[\underline{h}^*\right] &= \mathbb{P}^\pi_N B_{\theta,\eta}[\underline{h}^*] = o_{P^*}\left(N^{-1/2}\right).
\end{aligned} \tag{3.6}$$

The corresponding WLE with estimated weights, $(\hat{\theta}_{N,e}, \hat{\eta}_{N,e})$, the calibrated WLE $(\hat{\theta}_{N,c}, \hat{\eta}_{N,c})$, the WLE $(\hat{\theta}_{N,mc}, \hat{\eta}_{N,mc})$ with modified calibration and the WLE $(\hat{\theta}_{N,cc}, \hat{\eta}_{N,cc})$ with centered calibration satisfy (3.6) with $\mathbb{P}^\pi_N$ replaced by $\mathbb{P}^{\pi,e}_N$, $\mathbb{P}^{\pi,c}_N$, $\mathbb{P}^{\pi,mc}_N$ or $\mathbb{P}^{\pi,cc}_N$, respectively.

**Theorem 3.3.1.** *Suppose that the WLE is a solution of (3.6) where $\mathbb{P}^\pi_N$ may be replaced by $\mathbb{P}^{\pi,e}_N$ $\mathbb{P}^{\pi,c}_N$, $\mathbb{P}^{\pi,mc}_N$ or $\mathbb{P}^{\pi,cc}_N$ for the estimators with estimated weights, calibration, modified calibration and centered calibration. Under Conditions 3.1.1, 3.1.2 and 3.3.1-3.3.4,*

$$\begin{array}{rclcccl}
\sqrt{N}(\hat{\theta}_N - \theta_0) &=& \sqrt{N}\mathbb{P}^\pi_N \tilde{\ell}_0 + o_{P^*}(1) &\rightsquigarrow& Z &\sim& N_p(0,\Sigma), \\
\sqrt{N}(\hat{\theta}_{N,e} - \theta_0) &=& \sqrt{N}\mathbb{P}^{\pi,e}_N \tilde{\ell}_0 + o_{P^*}(1) &\rightsquigarrow& Z_e &\sim& N_p(0,\Sigma_e), \\
\sqrt{N}(\hat{\theta}_{N,c} - \theta_0) &=& \sqrt{N}\mathbb{P}^{\pi,c}_N \tilde{\ell}_0 + o_{P^*}(1) &\rightsquigarrow& Z_c &\sim& N_p(0,\Sigma_c), \\
\sqrt{N}(\hat{\theta}_{N,mc} - \theta_0) &=& \sqrt{N}\mathbb{P}^{\pi,mc}_N \tilde{\ell}_0 + o_{P^*}(1) &\rightsquigarrow& Z_{mc} &\sim& N_p(0,\Sigma_{mc}), \\
\sqrt{N}(\hat{\theta}_{N,cc} - \theta_0) &=& \sqrt{N}\mathbb{P}^{\pi,cc}_N \tilde{\ell}_0 + o_{P^*}(1) &\rightsquigarrow& Z_{cc} &\sim& N_p(0,\Sigma_{cc}),
\end{array}$$

*where $\Sigma$, $\Sigma_e$, $\Sigma_c$, $\Sigma_{mc}$ and $\Sigma_{cc}$ are as defined in (3.1) - (3.5) of Theorem 3.2.1, but now $I_0$ and $\tilde{\ell}_0$ are defined in Condition 3.3.2, and $Q_e$, $Q_c$, $Q_{mc}$ and $Q_{cc}$ are defined in Theorem 3.2.1.*

**Remark 3.3.1.** *Our conditions are identical to those of the $Z$-theorem of [21] except Condition 3.3.3 (2). This additional condition is not stringent. First, the Glivenko-Cantelli condition is usually assumed to prove consistency of estimators before deriving asymptotic distributions. Second, a stronger $L_2(P_0)$-continuity condition is standard as is seen in Condition 3.2.2 (See also 25.8 of [56] for a nice discussion of regularity conditions for efficient score equations with complete data). Note that the $L_1(P_0)$-continuity condition is only required for the WLE's with estimated weights and (modified and centered) calibration. Another way to understand the relative weakness of the Condition 3.3.3 (2) is to compare it with standard conditions for bootstrapping $Z$-estimators because the IPW empirical process is closely related to the exchangeably weighted bootstrap empirical process. See, for example, conditions A.4 and A.5 in [60]. The differentiability condition A.4, which implies continuity, corresponds to our $L_1(P_0)$-continuity condition. However, we do not impose a condition similar to the weak $L_2(P_0)$ condition A.5. In fact, our Lemma 6.3.2 in Chapter 6 with the Glivenko-Cantelli condition can be used to relax condition A.5 of [60].*

## 3.4 Comparisons of Methods

We compare asymptotic variances of five WLE's in view of improvement by adjusting weights and change of design. To make these comparisons clearly, we first need to give a clear statement of the result corresponding to Theorem 3.2.1 for stratified Bernoulli sampling.

### 3.4.1 Stratified Bernoulli Sampling

We present asymptotic normality of the WLE's, $\hat{\theta}_N^{Bern}$, $\hat{\theta}_{N,e}^{Bern}$, $\hat{\theta}_{N,c}^{Bern}$, $\hat{\theta}_{N,mc}^{Bern}$, $\hat{\theta}_{N,cc}^{Bern}$ under stratified Bernoulli sampling where all subjects are independent with the sampling probability $p_j$ if $V \in \mathcal{V}_j$.

**Theorem 3.4.1.** *Suppose Conditions 3.1.1 (except 3.1.1(f)) and 3.1.2 hold. Let $\xi_i \in \{0,1\}$ be i.i.d. with $E[\xi_1|V] = \pi_0(V) = \sum_{j=1}^{J} p_j I(V \in \mathcal{V}_j)$.*

*(1) Suppose that the WLE is a solution of (3.6) where* $\mathbb{P}_N^{\pi}$ *may be replaced by* $\mathbb{P}_N^{\pi,e}$ $\mathbb{P}_N^{\pi,c}$, $\mathbb{P}_N^{\pi,mc}$ *or* $\mathbb{P}_N^{\pi,cc}$ *for the estimators with estimated weights, calibration, modified calibration and centered calibration. Under the same conditions as in Theorem 3.2.1,*

$$\begin{array}{lclclcl}
\sqrt{N}(\hat{\theta}_N^{Bern} - \theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z^{Bern} & \sim & N_p(0, \Sigma^{Bern}), \\
\sqrt{N}(\hat{\theta}_{N,e}^{Bern} - \theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,e}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_e^{Bern} & \sim & N_p(0, \Sigma_e^{Bern}), \\
\sqrt{N}(\hat{\theta}_{N,c}^{Bern} - \theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,c}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_c^{Bern} & \sim & N_p(0, \Sigma_c^{Bern}), \\
\sqrt{N}(\hat{\theta}_{N,mc}^{Bern} - \theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_{mc}^{Bern} & \sim & N_p(0, \Sigma_{mc}^{Bern}), \\
\sqrt{N}(\hat{\theta}_{N,cc}^{Bern} - \theta_0) & = & \sqrt{N}\mathbb{P}_N^{\pi,cc}\tilde{\ell}_0 + o_{P^*}(1) & \rightsquigarrow & Z_{cc}^{Bern} & \sim & N_p(0, \Sigma_{cc}^{Bern}),
\end{array}$$

*where*

$$\Sigma^{Bern} \equiv I_0^{-1} + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j}(\tilde{\ell}_0)^{\otimes 2}, \tag{3.7}$$

$$\Sigma_e^{Bern} \equiv I_0^{-1} + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j}((I - Q_e)\tilde{\ell}_0)^{\otimes 2}, \tag{3.8}$$

$$\Sigma_c^{Bern} \equiv I_0^{-1} + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j}((I - Q_c)\tilde{\ell}_0)^{\otimes 2}, \tag{3.9}$$

$$\Sigma_{mc}^{Bern} \equiv I_0^{-1} + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j}((I - Q_{mc})\tilde{\ell}_0)^{\otimes 2}, \tag{3.10}$$

$$\Sigma_{cc}^{Bern} \equiv I_0^{-1} + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j}((I - Q_{cc})\tilde{\ell}_0)^{\otimes 2}, \tag{3.11}$$

*where* $Q_e$, $Q_c$, $Q_{mc}$ *and* $Q_{cc}$ *are defined in Theorem 3.2.1.*
*(2) Under the same conditions as in Theorem 3.3.1, the same conclusion in (1) holds with* $I_0$ *and* $\tilde{\ell}_0$ *replaced by those defined in Condition 3.3.2.*

Comparing the variance-covariance matrices in Theorem 3.4.1 to those in Theorems 3.2.1 and 3.3.1, we obtain the following corollary comparing designs. All estimators have smaller variances under sampling without replacement.

**Corollary 3.4.1.**

$$\begin{aligned}
\Sigma &= \Sigma^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \{P_{0|j}\tilde{\ell}_0\}^{\otimes 2}, \\
\Sigma_e &= \Sigma_e^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \{P_{0|j}(I-Q_e)\tilde{\ell}_0\}^{\otimes 2}, \\
\Sigma_c &= \Sigma_c^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \{P_{0|j}(I-Q_c)\tilde{\ell}_0\}^{\otimes 2}, \\
\Sigma_{mc} &= \Sigma_{mc}^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \{P_{0|j}(I-Q_{mc})\tilde{\ell}_0\}^{\otimes 2}, \\
\Sigma_{cc} &= \Sigma_{cc}^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \{P_{0|j}(I-Q_{cc})\tilde{\ell}_0\}^{\otimes 2}.
\end{aligned}$$

Variance formulae (3.8), (3.10) and (3.11) have the following alternative representations which show the efficiency gains over the plain WLE under Bernoulli sampling.

**Corollary 3.4.2.** *Under the same conditions as in Theorem 3.4.1,*

$$\begin{aligned}
\Sigma_e^{Bern} &= \Sigma^{Bern} - \text{Var}\left(\frac{\xi - \pi_0(V)}{\pi_0(V)} Q_e \tilde{\ell}_0\right), \\
\Sigma_{mc}^{Bern} &= \Sigma^{Bern} - \text{Var}\left(\frac{\xi - \pi_0(V)}{\pi_0(V)} Q_{mc} \tilde{\ell}_0\right), \\
\Sigma_{cc}^{Bern} &= \Sigma^{Bern} - \text{Var}\left(\frac{\xi - \pi_0(V)}{\pi_0(V)} Q_{cc} \tilde{\ell}_0\right).
\end{aligned}$$

Thus modified calibration and centered calibration yield improved efficiency over the plain WLE, while (ordinary) calibration does not yield a guaranteed improvement in general. We do not have similar formulas for sampling without replacement except for the special case involving within-stratum calibration described in part (2) of Corollary 3.4.3 below.

### *3.4.2 Within-stratum Adjustment of Weights*

[5] proposed calibration within each stratum to improve the calibrated WLE. Let $Z^{(j)} \equiv I(V \in \mathcal{V}_j) Z^T$ and $\tilde{Z} \equiv (Z^{(1)}, \ldots, Z^{(J)})^T$, and consider calibration on $\tilde{Z}$. The calibration equation (2.4) becomes

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\xi_i G_c(\tilde{Z}_i; \alpha)}{\pi_0(V_i)} Z_i I(V_i \in \mathcal{V}_j) = \frac{1}{N} \sum_{i=1}^{N} Z_i I(V_i \in \mathcal{V}_j), \quad j = 1, \ldots, J,$$

where $\alpha \in \mathbb{R}^{Jk}$. We call this special case *within-stratum calibration*. We define *within-stratum modified and centered calibration* analogously.

We also call the method of adjusting weights *within-stratum estimated weights* when binary regression is done within each stratum. Recall that the first $J$ elements of $Z$ for estimated weights are stratum membership indicators and the rest are other auxiliary variables, say $Z^{[2]}$. Within-stratum estimated weights uses $\tilde{Z} \equiv (Z^{(1)}, \ldots, Z^{(J)})^T$ where $Z^{(j)} \equiv I(V \in \mathcal{V}_j)(Z^{[2]})^T$ with 1 included in $Z^{[2]}$. The "true" parameter $\tilde{\alpha}_0$ has zero for all elements except having $G_e^{-1}(p_j)$ for the element corresponding to $I(V \in \mathcal{V}_j)$, $j = 1, \ldots, J$.

The following corollary summarizes within-stratum adjustment of weights under stratified Bernoulli sampling and sampling without replacement.

**Corollary 3.4.3.** *(1) (Bernoulli) Under the same conditions as in Theorem 3.4.1 with $Z$ replaced by $\tilde{Z}$ and $\alpha_0$ replaced by $\tilde{\alpha}_0$ for within-stratum estimated weights,*

$$\Sigma_e^{Bern} = \Sigma^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j} \left( Q_e^{(j)} \tilde{\ell}_0 \right)^{\otimes 2}, \tag{3.12}$$

$$\Sigma_c^{Bern} = \Sigma_{mc}^{Bern} = \Sigma^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j} \left( Q_c^{(j)} \tilde{\ell}_0 \right)^{\otimes 2}, \tag{3.13}$$

$$\Sigma_{cc}^{Bern} = \Sigma^{Bern} - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} P_{0|j} \left( Q_{cc}^{(j)} \tilde{\ell}_0 \right)^{\otimes 2}, \tag{3.14}$$

*where*

$$\begin{aligned} Q_e^{(j)} f &\equiv P_{0|j} \left[ f \dot{G}_e(\tilde{Z}^T \tilde{\alpha}_0)(Z^{[2]})^T \right] \left\{ P_{0|j} \dot{G}_e^2(\tilde{Z}^T \tilde{\alpha}_0)(Z^{[2]})^{\otimes 2} \right\}^{-1} \\ &\quad \times \dot{G}_e((\tilde{Z}^T \tilde{\alpha}_0) I(V \in \mathcal{V}_j) Z^{[2]}, \\ Q_c^{(j)} f &\equiv P_{0|j}[f Z^T]\{P_{0|j}[Z^{\otimes 2}]\}^{-1} I(V \in \mathcal{V}_j) Z, \\ Q_{cc}^{(j)} f &\equiv P_{0|j}[f(Z - \mu_{Z,j})^T]\{P_{0|j}[(Z - \mu_{Z,j})^{\otimes 2}]\}^{-1} I(V \in \mathcal{V}_j)(Z - \mu_{Z,j}), \end{aligned}$$

*with $\mu_{Z,j} \equiv E[I(V \in \mathcal{V}_j)Z]$ for $j = 1, \ldots, J$.*

*(2) (without replacement) Under the same conditions as in Theorem 3.2.1 or Theorem 3.3.1 with $Z$ replaced by $\tilde{Z}$,*

$$\Sigma_{cc} = \Sigma - \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j} \left( Q_{cc}^{(j)} \tilde{\ell}_0 \right). \tag{3.15}$$

### *3.4.3 Comparisons*

We summarize Corollaries 3.4.1-3.4.3. All estimators have reduced variance under the sampling without replacement design in comparison to Bernoulli sampling. Every method of adjusting weights improves efficiency over the plain WLE in a certain design and with a certain range of adjustment of weights (within-stratum or "across-strata" adjustment). However, particularly notable among all methods is centered calibration. While other methods gain efficiency only under stratified Bernoulli sampling, centered calibration improves efficiency over the plain WLE under both sampling schemes. There is no known method of "across-strata" adjustment that is guaranteed to gain efficiency over the plain WLE under stratified sampling without replacement.

There are close connections among all methods. When the auxiliary variables have mean zero, then centered and modified calibration are essentially the same. Within-stratum calibration and within-stratum modified calibration give the same asymptotic variance. For $Z$ and $\alpha_0$ defined for estimated weights and $\tilde{Z}$ and $\tilde{\alpha}_0$ defined for within-stratum estimated weights, modified calibration based on $(1 - \pi_0(V))^{-1}\dot{G}_e(Z^T\alpha_0)Z$ and within-stratum calibration based on $\dot{G}_e(Z^T\alpha_0)$ perform in the same way as the estimated weights and within-stratum estimated weights, respectively. Because of these connections among methods, there is no single method superior to others in each scenario. Performance depends on choice and transformation of auxiliary variables, the true distribution $P_0$, and the design. For our sampling scheme, within-stratum centered calibration is the only guaranteed method to gain efficiency while other methods may perform even worse than the plain WLE.

## ***3.5 Proofs***

Asymptotic linearity and the limiting distributions of $\hat{\alpha}_N$ in binary regression and (modified and centered) calibration are given by the following proposition. The proof requires a Glivenko-Cantelli theorem for $\mathbb{P}_N^{\pi}$ whose proof is independent of Proposition 3.5.1.

**Proposition 3.5.1.** *Under Condition 3.1.1, $\hat{\alpha}_N$ is consistent for $\alpha_0$, and*

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N - \alpha_0) \\
&= S_0^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{\dot{G}_e(Z_i^T\alpha_0)Z_i}{\pi_0(V_i)(1-\pi_0(V_i))}(\xi_i - \pi_0(V_i)) + o_P^*(1) \\
&\rightsquigarrow S_0^{-1}\sum_{j=1}^{J}\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot{G}_e(Z^T\alpha_0)Z,
\end{aligned}$$

*where $\mathbb{G}_j$ are independent $P_{0|j}$-Brownian bridge processes.*

*Under Condition 3.1.2, $\hat{\alpha}_N^c$, $\hat{\alpha}_N^{mc}$ and $\hat{\alpha}_N^{cc}$ are consistent, and*

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^c - \alpha_0) \\
&= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\dot{G}(0)^{-1}\left\{P_0Z^{\otimes 2}\right\}^{-1}Z_i\left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0Z^{\otimes 2}\right\}^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_jZ,
\end{aligned}$$

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^{mc} - \alpha_0) \\
&= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\dot{G}(0)^{-1}\left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\right\}^{-1}Z_i\left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\right\}^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_jZ,
\end{aligned}$$

*and*

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^{cc} - \alpha_0) \\
&= -\dot{G}(0)^{-1}\left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}(Z-\mu_Z)^{\otimes 2}\right\}^{-1} \\
&\quad\times\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(Z_i-\mu_Z)\left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}(Z-\mu_Z)^{\otimes 2}\right\}^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(Z-\mu_Z),
\end{aligned}$$

*where the $P_{0|j}$-Brownian bridge processes, $\mathbb{G}_j$, are independent.*

*Proof.* We first consider estimated weights. Define $M_N(\alpha) \equiv \mathbb{P}_N m_\alpha$ and $M(\alpha) = P_{\alpha_0}m_\alpha$ where $m_\alpha(Z,\xi) = \log\left(\{p_\alpha(\xi|Z) + p_{\alpha_0}(\xi|Z)\}/2\right)$. We again apply Theorem 5.7 of [56] for a consistency proof. Because $p_\alpha(\xi|Z)$ is a valid marginal density of a single observation $\xi$ given $Z$, the argument of [56], page 66, can be used to verify the second condition of the theorem. We verify the first condition of Theorem 5.7 of [56]. Let $\tilde{G}_e(z;\alpha) \equiv \{G_e(z^T\alpha)+G_e(z^T\alpha_0)\}/2$. Then $m_\alpha(z,\xi) = \xi \log \tilde{G}_e(z;\alpha) + (1-\xi)\log(1-\tilde{G}_e(z;\alpha))$. We rewrite $\mathbb{P}_N m_\alpha$ as

$$\begin{aligned}
\mathbb{P}_N m_\alpha &= \frac{1}{N}\sum_{i=1}^{N} \xi_i \log \tilde{G}_e(Z_i;\alpha) + (1-\xi_i)\log\left(1-\tilde{G}_e(Z_i;\alpha)\right) \\
&= \sum_{j=1}^{J}\left\{\frac{N_j}{N}\frac{n_j}{N_j}\left[\frac{1}{N_j}\sum_{i=1}^{N}\frac{\xi_{j,i}}{n_j/N_j}\log\tilde{G}_e(Z_{j,i};\alpha)\right]\right\} \\
&\quad +\sum_{j=1}^{J}\left\{\frac{N_j}{N}\left(1-\frac{n_j}{N_j}\right)\left[\frac{1}{N_j}\sum_{i=1}^{N}\frac{1-\xi_{j,i}}{1-n_j/N_j}\log\left(1-\tilde{G}_e(Z_{j,i};\alpha)\right)\right]\right\}.
\end{aligned}$$

Thus, if we establish that both $\mathcal{S}_{0,j} \equiv \left\{\log\left(1-\tilde{G}_e(z^T\alpha)\right) : \alpha \in \mathbb{R}^{J+k}, V \in \mathcal{V}_j\right\}$ and $\mathcal{S}_{1,j} \equiv \left\{\log\tilde{G}_e(z^T\alpha) : \alpha \in \mathbb{R}^{J+k}, V \in \mathcal{V}_j\right\}$ are $P_0$-Glivenko-Cantelli for $j = 1,\ldots,J$, it follows from Theorem 6.1.1 applied to sampled subjects and non-sampled subjects in each stratum separately that $\mathbb{P}_N m_\alpha$ converges in probability to

$$\begin{aligned}
P_0 m_\alpha \quad &= \quad \sum_{j=1}^{J}\nu_j p_j P_0\left(\log\tilde{G}_e(Z^T\alpha)\Big| V\in\mathcal{V}_j\right) \\
&\quad\quad +\sum_{j=1}^{J}\nu_j(1-p_j)P_0\left(\log\left(1-\tilde{G}_e(Z^T\alpha)\right)\Big| V\in\mathcal{V}_j\right),
\end{aligned}$$

uniformly in $\alpha$. Note that the method of estimated weights does not estimate the sampling probability for the subjects in a stratum if the sampling probability is 1. Thus, we can assume that $G_e(Z^T\alpha_0) \le \sigma' < 1$. Hence we have $\log(\sigma/2) \le \log\tilde{G}_e(Z^T\alpha) \le 0$ and $\log(\{1-\sigma'\}/2) \le \log\left(1-\tilde{G}_e(Z^T\alpha)\right) \le 0$ for all $j = 1,\ldots,J$ and $\alpha \in \mathbb{R}^{J+k}$. This implies that all sets $\mathcal{S}_{k,j}, k = 0,1$, have integrable envelopes. Now it suffices to show that all sets are VC subgraph classes. Note first that $\{z^T\alpha : \alpha \in \mathbb{R}^{J+k}\}$ is a VC subgraph class by Lemma 2.6.15 of [58]. Note also that $G_e$ and the logarithm are monotone functions. Because a map by a monotone function, addition and multiplication all preserve the property of the VC subgraph class by Lemma 2.6.17 of [58], our claim follows and hence the first condition is verified. Since we have by concavity of the logarithm and the property of $\hat{\alpha}_N$ that

$$\begin{aligned} M_N(\hat{\alpha}_N) &\geq \frac{1}{2}\mathbb{P}_N \log p_{\hat{\alpha}_N}(\xi|V) + \frac{1}{2}\mathbb{P}_N \log p_{\alpha_0}(\xi|V) \\ &\geq \frac{1}{2}\mathbb{P}_N \log p_{\alpha_0}(\xi|V) + \frac{1}{2}\mathbb{P}_N \log p_{\alpha_0}(\xi|V) = M_N(\alpha_0), \end{aligned}$$

consistency follows from Theorem 5.7 of [56].

We apply Theorem 3.3.1 of [58] to show asymptotic normality of $\hat{\alpha}_N$. Define

$$\Phi_{N,e}(\alpha) = \frac{1}{N}\sum_{i=1}^{N} \frac{\dot{G}_e(Z_i^T\alpha)Z_i}{G_e(Z_i^T\alpha)(1-G_e(Z_i^T\alpha))}\left(\xi_i - G_e(Z_i^T\alpha)\right) \equiv \mathbb{P}_N\phi_\alpha(\xi, V),$$

and

$$\Phi_e(\alpha) = P_0\left\{\frac{\dot{G}_e(Z^T\alpha)Z}{G_e(Z^T\alpha)(1-G_e(Z^T\alpha))}\left(\sum_{j=1}^{J} p_j I(V\in\mathcal{V}_j) - G_e(Z^T\alpha)\right)\right\}.$$

Note that $\Phi_{N,e}(\hat{\alpha}_N) = 0$ because $(\partial/\partial\alpha)\mathbb{P}_N \log p_\alpha = \Phi_{N,e}(\alpha)$. Note also that $\Phi_e(\alpha_0) = 0$ since $G_e(Z^T\alpha_0) = p_j$ when $V \in \mathcal{V}_j$. It follows by the decomposition (2.1) of the IPW empirical process (see also [8]) that

$$\begin{aligned} &\sqrt{N}(\Phi_{N,e}(\alpha_0) - \Phi_e(\alpha_0)) = \sqrt{N}\Phi_{N,e}(\alpha_0) \\ &= \sqrt{N}\mathbb{P}_N\frac{\dot{G}_e(Z^T\alpha_0)Z}{G_e(Z^T\alpha_0)(1-G_e(Z^T\alpha_0))}(\xi - G_e(Z^T\alpha_0)) \\ &= \sqrt{N}\mathbb{P}_N^\pi\frac{\pi_0(V)}{G_e(Z^T\alpha_0)}\frac{\dot{G}_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)} - \sqrt{N}\mathbb{P}_N\frac{\dot{G}_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)} \\ &= \sum_{j=1}^{J}\sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\mathbb{G}_j^\xi\frac{\pi_0(V)}{G_e(Z^T\alpha_0)}\frac{\dot{G}_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)} \\ &\quad + \sqrt{N}\mathbb{P}_N\left(\frac{\pi_0(V)}{G_e(Z^T\alpha_0)} - 1\right)\frac{\dot{G}_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)}. \end{aligned}$$

Since $\pi_0(V) = n_j/N_j$ and $G_e(Z^T\alpha_0) = p_j$ when $V \in \mathcal{V}_j$, the first term converges to

$$\sum_{j=1}^{J}\sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\frac{n_j/N_j}{p_j(1-p_j)}\mathbb{G}_j^\xi\dot{G}_e(Z^T\alpha_0)Z \rightsquigarrow \sum_{j=1}^{J}\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot{G}_e(Z^T\alpha_0)Z.$$

The second term can be written as

$$\sum_{j=1}^{J}\sqrt{N_j}\left(\frac{n_j}{N_j} - p_j\right)\sqrt{\frac{N_j}{N}}\frac{1}{p_j(1-p_j)}\frac{1}{N_j}\sum_{i=1}^{N_j}\dot{G}_e(Z_{j,i}^T\alpha_0)Z_{j,i}.$$

Since $n_j = [N_j p_j]$ by assumption, it is easy to see that $-N_j^{-1/2} \leq \sqrt{N_j}(n_j/N_j - p_j) \leq 0$, and hence $\sqrt{N_j}(n_j/N_j - p_j) \to 0$. Since $N_j^{-1} \sum_{i=1}^{N_j} \dot{G}_e(Z_{j,i}^T \alpha_0) Z_{j,i} = O_{P^*}(1)$ by the weak law of large numbers and $\sqrt{N_j/N} \to \sqrt{\nu_j}$, the second term converges to zero in probability. The weak convergence of $\sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha_0)$ follows from Slutsky's theorem.

For the asymptotic equicontinuity of the process, it suffices to consider a compact subset $\mathcal{A}_{e,0} \in \mathbb{R}^{J+k}$ where $\alpha_0$ is its interior point since $\hat{\alpha}_N$ is consistent. Let

$$\phi_{\alpha,1}(v) \equiv \frac{\pi_0(v) z^{\otimes 2}}{G_e(z^T\alpha)\left\{1 - G_e(z^T\alpha)\right\}} \left( \ddot{G}_e(z^T\alpha) - \frac{\left\{\dot{G}_e(z^T\alpha)\right\}^2}{G_e(z^T\alpha)} \right),$$

$$\phi_{\alpha,2}(v) \equiv \frac{z^{\otimes 2}}{1 - G_e(z^T\alpha)} \left( \ddot{G}_e(z^T\alpha) - \frac{\left\{\dot{G}_e(z^T\alpha)\right\}^2}{1 - G_e(z^T\alpha)} \right).$$

Taylor's theorem gives

$$\phi_\alpha(\xi, v) - \phi_{\alpha_0}(\xi, v) = \phi_{\alpha_1^*,1}(v) \frac{\xi}{\pi_0(v)}(\alpha - \alpha_0) + \phi_{\alpha_2^*,2}(v)(\alpha - \alpha_0),$$

where $\alpha_j^*, j = 1, 2$, are some convex combinations of $\alpha$ and $\alpha_0$. Thus,

$$\begin{aligned}
&\sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha) - \sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha_0) \\
&= \sqrt{N}(\mathbb{P}_N - P_0)(\phi_\alpha - \phi_{\alpha_0}) + \sqrt{N} P_0(\phi_\alpha - \phi_{\alpha_0}) \\
&\quad - \sqrt{N}\Phi_e(\alpha) + \sqrt{N}\Phi_e(\alpha_0) \\
&= (\mathbb{P}_N^\pi - P_0)\phi_{\alpha_1^*,1}\sqrt{N}(\alpha - \alpha_0) + (\mathbb{P}_N - P_0)\phi_{\alpha_2^*,2}\sqrt{N}(\alpha - \alpha_0) \\
&\quad + P_0\phi_{\alpha_1^*,1}\left( \xi - \sum_{j=1}^J p_j I(V \in \mathcal{V}_j) \right)\sqrt{N}(\alpha - \alpha_0). \qquad (3.16)
\end{aligned}$$

To show this is $o_{P^*}(1 + \sqrt{N}(\alpha - \alpha_0))$, we first show that the set $\mathcal{T}_k = \{\phi_{\alpha,k} : \alpha \in \mathcal{A}_{e,0}\}, k = 1, 2$, are Glivenko-Cantelli. It is easy to see that $\{z^T\alpha : \alpha \in \mathcal{A}_{e,0}\}$ is Glivenko-Cantelli. Since $G_e \in \mathcal{C}^2$ by assumption, $\phi_{\alpha,k}, k = 1, 2$, are uniformly bounded in $\alpha \in \mathcal{A}_{e,0}$. Thus, the sets $\mathcal{T}_k$, $k = 1, 2$ are both Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). For the third term in (3.16), apply the dominated convergence theorem with $P_0(\xi|V) = \sum_{j=1}^J (n_j/N_j) I(V \in \mathcal{V}_j) \to \sum_{j=1}^J p_j I(V \in \mathcal{V}_j)$.

Since $\dot{\Phi}(\alpha_0) = -S_0$, apply Theorem 3.3.1 of [58] to obtain

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N - \alpha_0) \\
&= S_0^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{\dot{G}_e(Z_i^T\alpha_0)Z_i}{G_e(Z_i^T\alpha_0)(1-G_e(Z_i^T\alpha_0))}\left(\xi_i - G_e(Z_i^T\alpha_0)\right) + o_{P^*}(1) \\
&\rightsquigarrow S_0^{-1}\sum_{j=1}^{J}\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot{G}_e(Z^T\alpha_0)Z.
\end{aligned}$$

This completes the proof.

Next we consider modified calibration with $\hat{\alpha}_N = \hat{\alpha}_N^{mc}$. The cases for (centered) calibration (i.e., $\hat{\alpha}_N = \hat{\alpha}_N^{c}$ and $\hat{\alpha}_N = \hat{\alpha}_N^{cc}$) are similar. Define $\Phi_{N,mc}(\alpha) \equiv \mathbb{P}_N^{\pi}G_{mc}(V;\alpha)Z - \mathbb{P}_N Z$ and $\Phi_{mc}(\alpha) \equiv P_0[(G_{mc}(V;\alpha)-1)Z]$. Note that $\Phi_{N,mc}(\hat{\alpha}_N) = 0$ and $\Psi_{mc}(0) = 0$. We apply Theorem 5.7 of [56] for a consistency proof. For the first condition of the theorem, we have

$$\begin{aligned}
&\sup_{\alpha\in\mathbb{R}^k}\|\Phi_{N,mc}(\alpha) - \Phi_{mc}(\alpha)\| \\
&= \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\xi_i}{\pi_0(V_i)}G_{mc}(V;\alpha) - 1\right)Z_i - P_0\left\{G_{mc}(V;\alpha)-1\right\}Z\right\| \\
&\le \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_0(V_i)}G_{mc}(V_i;\alpha)Z_i - P_0G_{mc}(;\alpha)Z\right\| \\
&\quad + \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}Z_i - P_0Z\right\|,
\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm. Since $\alpha$ is a vector in $\mathbb{R}^k$ and $G$ is monotone, $\{G_{mc}(\cdot;\alpha) : \alpha \in \mathbb{R}^k\}$ is a VC subgraph by Lemmas 2.6.15 and 2.6.18 of [58]. Boundedness of $G$ implies that the set $\left\{G_{mc}(v;\alpha)z : \alpha \in \mathbb{R}^k\right\}$ is $P_0$-Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). Then the first term is $o_{P^*}(1)$ by Theorem 6.1.1. The second term is $o_{P^*}(1)$ by the weak law of large numbers.

The second condition of the theorem is that for any $\epsilon > 0$, $\inf_{|\alpha|>\epsilon}\|\Phi_{mc}(\alpha)\| > 0$. Suppose, to the contrary, that $\inf_{|\alpha|>\epsilon}\|\Phi_{mc}(\alpha)\| = 0$ for some $\epsilon > 0$. Then there exists a sequence $\{\alpha^{(m)}\} \subset \mathbb{R}^k$ with $|\alpha^{(m)}| > \epsilon$ for each $m = 1, 2, \ldots,$ such that

$$\|\Phi_{mc}(\alpha^{(m)})\| \to 0.$$

Let $\Phi_{j,c}(\alpha)$, $j = 1, \ldots, k$, be the $j$th element of $\Phi_{mc}(\alpha)$. Since the norm $\|\cdot\|$ is the Euclidean norm, each element $\Phi_{j,c}(\alpha^{(m)})$ converges to zero. If $\alpha^{(m)}$ converges to $\alpha^{(\infty)}$ with $|\alpha^{(\infty)}| < \infty$, then by the dominated convergence theorem and Taylor's theorem,

$$0 = P_0\left[\left\{G_{mc}(V;\alpha^{(\infty)}) - 1\right\} Z\right] = P_0\left[(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}\alpha^{(\infty)}\right]$$

for some $\alpha^*$ with $|\alpha^*| \le |\alpha^{(\infty)}|$. Because $P_0(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}$ is positive definite by assumption, $\alpha^{(\infty)}$ must be zero, which contradicts the fact that $|\alpha^{(\infty)}| \ge \epsilon$.

We assume that some elements of $\alpha^{(m)}$ diverge. Then, a further subsequence $\alpha^{(m')}$ converges to some $\alpha^{(\infty)}$ whose elements are extended real numbers. Define a unit vector $\beta^{(\infty)} \equiv \lim_{m'\to\infty} \alpha^{(m')}/\|\alpha^{(m')}\|$. Then we have for each $Z$ on the set $\{\pi_0(V) < 1\}$ that

$$\begin{aligned} G_{mc}^{(\infty)}(Z)\, Z^T\beta^{(\infty)} &\equiv \lim_{m'\to\infty} G\left(\frac{1-\pi_0(V)}{\pi_0(V)} Z^T \frac{\alpha^{(m')}}{\|\alpha^{m'}\|}\|\alpha^{m'}\|\right) Z^T\beta^{(\infty)} \\ &= \begin{cases} M_1 Z^T\beta^{(\infty)} & \text{if } Z^T\beta^{(\infty)} > 0, \\ m_1 Z^T\beta^{(\infty)} & \text{if } Z^T\beta^{(\infty)} < 0, \\ 0 & \text{if } Z^T\beta^{(\infty)} = 0. \end{cases} \end{aligned}$$

It follows by the dominated convergence theorem applied to each element of the vector of $\Phi_{mc}(\alpha)$ that

$$\begin{aligned} 0 &= \lim_{m'\to\infty} \Phi_{mc}\left(\alpha^{(m')}\right)^T \beta^{(\infty)} = P_0 \lim_{m'\to\infty}\left\{G_{mc}\left(V;\alpha^{(m')}\right) - 1\right\} Z^T\beta^{(\infty)} \\ &= (M_1 - 1)P_0 I_{\left\{Z^T\beta^{(\infty)}>0,\pi_0(V)<1\right\}} Z^T\beta^{(\infty)} \\ &\quad + (m_1 - 1)P_0 I_{\left\{Z^T\beta^{(\infty)}<0,\pi_0(V)<1\right\}} Z^T\beta^{(\infty)}. \end{aligned}$$

However, this is strictly positive since $m_1 < 1$ and $M_1 > 1$, which is a contradiction. This completes the proof that $\hat{\alpha}_N \to_{P^*} 0$.

We apply Theorem 3.3.1 of [58] to show the asymptotic normality of $\hat{\alpha}_N$. For the asymptotic equicontinuity condition, it follows by Taylor's theorem that

$$\begin{aligned} &\sqrt{N}(\Phi_{N,mc} - \Phi_{mc})(\hat{\alpha}_N) - \sqrt{N}(\Phi_{N,mc} - \Phi_{mc})(\alpha_0) \\ &= \mathbb{G}_N^\pi[G_{mc}(V;\hat{\alpha}_N)Z - G_{mc}(V;\alpha_0)Z] \\ &= (\mathbb{P}_N^\pi - P_0)(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}\sqrt{N}(\hat{\alpha} - \alpha_0) \end{aligned}$$

for some $\alpha^*$ with $|\alpha^* - \alpha_0| \le |\hat{\alpha}_N - \alpha_0|$. This term is $o_P(1 + \sqrt{N}|\hat{\alpha} - \alpha_0)|)$ if $(\mathbb{P}_N^\pi - P_0)(\pi_0(V)^{-1} - 1)Z^{\otimes 2}\dot{G}_{mc}(V;\alpha) \to_{P^*} 0$, uniformly in $\alpha$. Let $\mathcal{A}_{mc,1} \subset \mathbb{R}^k$ be a compact neighborhood of zero. Since $\hat{\alpha}_N$ is consistent, it suffices to show that the set $\{(\pi_0^{-1}(V)-1)Z^{\otimes 2}\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli. Since $|\pi_0^{-1}(V)-1|$ and $Z$ are bounded, the VC subgraph class $\{(\pi_0^{-1}(V)-1)Z\alpha : \alpha \in \mathcal{A}_{mc,1}\}$ (Lemma 2.6.15 of [58]) is $P_0$-Glivenko-Cantelli. Because $\dot{G}$ is continuous and bounded, the set $\{\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). Apply a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) again to conclude $\{(\pi_0^{-1}(V)-1)Z^{\otimes 2}\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli. Hence, asymptotic equicontinuity follows from Theorem 6.1.1. We show the weak convergence of the process $\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\alpha)$ at $\alpha_0 = 0$. Since $G_{mc}(v;\alpha_0) = 1$, it follows from the decomposition (2.1) of the IPW empirical process (see also [8]) that

$$
\begin{aligned}
&\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\alpha_0) = \sqrt{N}\Phi_{N,mc}(0) = \sqrt{N}(\mathbb{P}_N^{\pi}-\mathbb{P}_N)Z \\
&= \sum_{j=1}^{J}\sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\mathbb{G}^{\xi}_{j,N_j}Z \\
&\rightsquigarrow \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_jZ \quad (\text{ by Theorem 6.3.1}).
\end{aligned}
$$

The Fréchet derivative of $\Phi_{mc}(\alpha_0)$ is

$$
\dot{\Phi}_{mc}(\alpha)|_{\alpha=\alpha_0} = \frac{\partial}{\partial\alpha}P_0(G_{mc}(V;\alpha)-1)Z\bigg|_{\alpha=\alpha_0} = \dot{G}(0)P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}.
$$

Thus, by Theorem 3.3.1 of [58] we obtain

$$
\begin{aligned}
\sqrt{N}\hat{\alpha}_N &= -\dot{\Phi}_{mc}(0)\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(0) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}\right\}^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_jZ.
\end{aligned}
$$

□

We give proofs for our main results (Theorems 3.2.1 and 3.3.1). Proofs require results in Chapter 6, which are proved independently below.

*Proof of Theorem 3.2.1.* The asymptotic distributions of $\hat{\theta}_N$ is derived in [8]. Here we derive the asymptotic distribution of $\hat{\theta}_{N,mc}$ that is a solution of the calibrated weighted likelihood equations with modified calibration

$$\Psi^{\pi}_{N,1,mc}(\theta,\eta,\alpha) = \mathbb{P}^{\pi}_{N} G_{mc}(V;\alpha)\dot{\ell}_{\theta,\eta} = 0,$$
$$\Psi^{\pi}_{N,2,mc}(\theta,\eta,\alpha)h = \mathbb{P}^{\pi}_{N} G_{mc}(V;\alpha)(B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h) = 0,$$

for all $h \in \mathcal{H}$ with $\alpha = \hat{\alpha}_N$. Let $\Psi_{mc}(\theta,\eta,\alpha) = (\Psi_{1,mc}(\theta,\eta,\alpha), \Psi_{2,mc}(\theta,\eta,\alpha))$

$$\Psi_{1,mc}(\theta,\eta,\alpha) = P_0 G_{mc}(V;\alpha)\dot{\ell}_{\theta,\eta},$$
$$\Psi_{2,mc}(\theta,\eta,\alpha) = P_0 G_{mc}(V;\alpha)(B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h).$$

The derivative map of $\Psi_{mc}$ with respect to $(\theta,\eta)$ at $(\theta_0,\eta_0,\alpha)$ has components

$$P_0\{G_{mc}(V;\alpha)\dot{\psi}_{ij,\theta_0,\eta_0,h}\}, \quad i,j = 1,2.$$

Our proof proceed by verifying the conditions of Theorem 1 of [9]. The weak convergence of $\sqrt{N}(\Psi_{N,j,mc} - \Psi_{j,mc})(\theta_0,\eta_0,\alpha_0)$ follows from Theorem 6.3.1. The asymptotic equicontinuity conditions

$$\sup_{\theta\in\Theta,\eta\in H}\left\|\sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta,\eta,\hat{\alpha}_N) - \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta,\eta,\alpha_0)\right\|_{\mathcal{H}} = o_{P^*}(1),$$

for $j = 1,2$, follows from Lemma 6.4.3. The other asymptotic equicontinuity condition

$$\left\|\sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\hat{\theta}_{N,mc},\hat{\eta}_{N,mc},\alpha_0) - \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta_0,\eta_0,\alpha_0)\right\|_{\mathcal{H}} = o_{P^*}(1),$$

for $j = 1,2$, follows from Condition 3.2.2 and Lemma 6.3.1. Thus conditions (2) and (3) of [9] are satisfied.

The Fréchet differentiability of the map $(\theta,\eta) \mapsto \Phi_{j,mc}(\theta,\eta,\alpha)$ uniformly over the neighborhood of $\alpha_0$ follows by Condition 3.2.3 and boundedness of $G$;

$$\begin{aligned}
&\left\|\Psi_{mc}(\theta,\eta,\alpha)h - \Psi_{mc}(\theta_0,\eta_0,\alpha)h - \dot{\Psi}_{mc}((\theta,\eta) - (\theta_0,\eta_0))\right\|_{H} \\
&= \sup_{h\in\mathcal{H}}\left|E\left\{G_{mc}(V;\alpha)\left(\psi_{\theta,\eta,h} - \psi_{\theta_0,\eta_0,h} - \dot{\psi}_{\theta_0,\eta_0,h}((\theta,\eta) - (\theta_0,\eta_0))\right)\right\}\right| \\
&\le \left\{EG^2_{mc}(V;\alpha)\right\}^{1/2} \sup_{h\in\mathcal{H}}\left[E\left\{\psi_{\theta,\eta,h} - \psi_{\theta_0,\eta_0,h} - \dot{\psi}_{\theta_0,\eta_0,h}((\theta,\eta) - (\theta_0,\eta_0))\right\}^2\right]^{1/2} \\
&= o_{P^*}\left(\|(\theta,\eta) - (\theta_0,\eta_0)\|\right).
\end{aligned}$$

The Fréchet derivative $\dot{\Psi}_{\alpha,mc}$ of the map $\alpha \mapsto \{\Psi_{mc}(\theta,\eta,\alpha)h : h \in \mathcal{H}\}$ is

$$\frac{\partial}{\partial\alpha}\Psi_{mc}(\theta,\eta,\alpha)h = \frac{\partial}{\partial\alpha}E\left[G_{mc}(V;\alpha)\psi_{\theta,\eta,h}\right] = E\left[\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha)\psi_{\theta,\eta,h}\right].$$

Now proceed in the same way as [9] to obtain

$$\begin{aligned}&\sqrt{N}(\hat{\theta}_{N,mc}-\theta_0)\\&=\sqrt{N}(\hat{\theta}_N-\theta_0)+E\left[\tilde{\ell}_{\theta_0,\eta_0}\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}(0)\right]\sqrt{N}(\hat{\alpha}_N-\alpha_0)+o_{P^*}(1).\end{aligned}$$

Because $\sqrt{N}(\hat{\theta}_N-\theta_0)=\mathbb{G}_N^{\pi}\tilde{\ell}_{\theta_0,\eta_0}+o_{P^*}(1)$ ((16) of [8]), it follows from (6.12) and consistency and asymptotic normality of $\hat{\alpha}_N$ that $\sqrt{N}(\hat{\theta}_{N,mc}-\theta_0) = \mathbb{G}_N^{\pi,mc}\tilde{\ell}_{\theta_0,\eta_0}+o_{P^*}(1)$. Apply Theorem 6.3.1 to complete the proof.

The other three cases are similar. □

*Proof of Theorem 3.3.1.* We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other four cases are similar.

We evaluate the stochastic order of $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0}+\sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}$. Because $\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} = o_{P^*}(N^{-1/2})$ by assumption and $P_0\dot{\ell}_{\theta_0,\eta_0}=0$, we have $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0}+\sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} = -\mathbb{G}_N^{\pi,mc}(\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}-\dot{\ell}_{\theta_0,\eta_0})+o_{P^*}(1)$. Let $\delta_N \downarrow 0$ be arbitrary and define $\mathcal{F}_N \equiv \{\dot{\ell}_{\theta,\eta}-\dot{\ell}_{\theta_0,\eta_0} : |\theta-\theta_0|\le\delta_N, \|\eta-\eta_0\|\le N^{-\beta}\}$. Then $f\in\mathcal{F}_N$ converges to zero either pointwise pointwise or in $L_1(P_0)$ by Condition 3.3.3 as $N\to\infty$. Moreover, it follows from Condition 3.3.3 that $\|\mathbb{G}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$ and that there exists some $N_0$ that $\mathcal{F}_N$ is Glivenko-Cantelli for $N\ge N_0$. Apply Lemma 6.3.2 to obtain $\|\mathbb{G}_N^{\pi,mc}\|_{\mathcal{F}_N}=o_{P^*}(1)$ and conclude $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0}+\sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}=o_{P^*}(1)$. Similarly, $\sqrt{N}\mathbb{P}_N^{\pi,mc}B_{\theta_0,\eta_0}[\underline{h}^*]+\sqrt{N}P_0B_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}[\underline{h}^*]=o_{P^*}(1)$. These stochastic orders and Condition 3.3.4 imply that

$$\begin{aligned}&P_0\left\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\theta}_{N,mc}-\theta_0)+B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc}-\eta_0])\right\}\\&\quad+o\left(|\hat{\theta}_{N,mc}-\theta_0|\right)+O\left(\|\hat{\eta}_{N,mc}-\eta_0\|^{\alpha}\right)+\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0}\\&=P_0\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\theta}_{N,mc}-\theta_0)+B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc}-\eta_0])-\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}+\dot{\ell}_{\theta_0,\eta_0}\}\\&\quad+o\left(|\hat{\theta}_{N,mc}-\theta_0|\right)+O\left(\|\hat{\eta}_{N,mc}-\eta_0\|^{\alpha}\right)+P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}+\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0}\\&=o_{P^*}(N^{-1/2}),\end{aligned}\tag{3.17}$$

and, furthermore, that

$$
\begin{aligned}
&P_0 \left\{ -B_{\theta_0,\eta_0} \left[\underline{h}^*\right] (\dot{\ell}^T_{\theta_0,\eta_0}(\hat{\theta}_{N,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc} - \eta_0]) \right\} \\
&\quad + o\left(|\hat{\theta}_{N,mc} - \theta_0|\right) + O\left(\|\hat{\eta}_{N,mc} - \eta_0\|^\alpha\right) + \mathbb{P}_N^{\pi,mc} B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \\
&= o_{P^*}(N^{-1/2}). \qquad (3.18)
\end{aligned}
$$

By Condition 3.3.1 and $\alpha\beta > 1/2$, $\sqrt{N} O_{P^*}\left(\|\hat{\eta}_N - \eta_0\|^\alpha\right) = o_{P^*}(1)$. So by Condition 3.3.2 and taking the difference of (3.17) and (3.18), we have

$$
\begin{aligned}
&-P_0 \left( \left\{ \dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right\} \dot{\ell}^T_{\theta_0,\eta_0} \right) \left( \hat{\theta}_{N,mc} - \theta_0 \right) + \ o\left( |\hat{\theta}_{N,mc} - \theta_0| \right) \\
&\quad + o_P(N^{-1/2}) - o_P(N^{-1/2}) + \mathbb{P}_N^{\pi,mc} \left( \dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) \\
&= o_P(N^{-1/2}) - o_P(N^{-1/2}),
\end{aligned}
$$

or

$$
-I_0(\hat{\theta}_{N,mc} - \theta_0) = \mathbb{P}_N^{\pi,mc} \left( \dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) + o_{P^*}(N^{-1/2}).
$$

It follows by the invertibility of $I_0$ that

$$
\sqrt{N}\left( \hat{\theta}_{N,mc} - \theta_0 \right) = -\sqrt{N} \mathbb{P}_N^{\pi,mc} I_0^{-1} \left( \dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) + o_P(1).
$$

Now, we recognize that the summand inside $\mathbb{P}_N^{\pi,mc}$ is the efficient influence function for $\theta$ (for complete data) and apply Theorem 6.3.1. □

*Proof of Theorem 3.4.1.* Theorem 3.2.1 for cases for $\hat{\theta}_N^{Bern}$ and $\hat{\theta}_{N,e}^{Bern}$ are proved in [8, 9]. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other four estimators for both theorems are similar.

Under stratified Bernoulli sampling, independence of sampling indicators allows us to proceed in the same as in the proofs of Theorems 3.2.1 and 3.3.1 to conclude $\sqrt{N}(\hat{\theta}_{N,mc}^{Bern} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde{\ell}_0 + o_{P^*}(1)$ and asymptotic linearity of $\hat{\alpha}_N$ in Proposition 3.5.1. In view of (6.12), $\sqrt{N}(\hat{\theta}_{N,mc}^{Bern} - \theta_0) = \sqrt{N}\mathbb{P}_N f + o_{P^*}(1)$ where

$$
f(X, V, \xi) = \frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi - \pi_0(V)}{\pi_0(V)} Q_{mc}\tilde{\ell}_0. \qquad (3.19)
$$

Apply the central limit theorem and compute

$$
\begin{aligned}
&\Sigma_{mc}^{Bern} = \mathrm{Var}(f)\\
&= \mathrm{Var}\left(E\left[\frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi-\pi_0(V)}{\pi_0(V)}Q_{mc}\tilde{\ell}_0\middle| X,V\right]\right)\\
&\quad + E\left[\mathrm{Var}\left(\frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi-\pi_0(V)}{\pi_0(V)}Q_{mc}\tilde{\ell}_0\middle| X,V\right)\right]\\
&= \mathrm{Var}(\tilde{\ell}_0) + E\left[\mathrm{Var}\left(\frac{\xi}{\pi_0(V)}(I-Q_{mc})\tilde{\ell}_0\middle| X,V\right)\right]\\
&= I_0^{-1} + E\left[\frac{1-\pi_0(V)}{\pi_0(V)}\{(I-Q_{mc})\tilde{\ell}_0\}^{\otimes 2}\right]\\
&= I_0^{-1} + \sum_{j=1}^{J}\nu_j\frac{1-p_j}{p_j}P_{0|j}\{(I-Q_{mc})\tilde{\ell}_0\}^{\otimes 2}.
\end{aligned}
$$

□

*Proof of Corollary 3.4.2.* We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other two cases are similar.

Let $Q_{mc}\tilde{\ell}_0 \equiv AZ$ where $A = A_1A_2$ with $A_1 \equiv P_0[(\pi_0^{-1}(V)-1)\tilde{\ell}_0Z^T]$ and $A_2 \equiv \{P_0[(\pi_0^{-1}(V)-1)Z^{\otimes 2}]\}^{-1}$. Recall that $\Sigma^{Bern} = \mathrm{Var}\{(\xi/\pi_0(V))\tilde{\ell}_0\}$. In view of (3.19), it suffices to show that $\mathrm{Cov}\{(\xi/\pi_0(V))\tilde{\ell}_0, (\xi/\pi_0(V)-1)AZ\}$ is equal to $\mathrm{Var}\left((\xi/\pi_0(V)-1)AZ\right)$. This is true since

$$
\begin{aligned}
&\mathrm{Cov}\left\{\frac{\xi}{\pi_0(V)}\tilde{\ell}_0, \frac{\xi-\pi_0(V)}{\pi_0(V)}AZ\right\} = E\left\{\frac{\xi}{\pi_0(V)}\tilde{\ell}_0\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\right\}A^T\\
&= E\left[\tilde{\ell}_0ZE\left\{\frac{\xi}{\pi_0(V)}\frac{\xi-\pi_0(V)}{\pi_0(V)}\middle| X,V\right\}\right]A^T\\
&= E\left[\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0Z\right]A^T = A_1A_2A_1^T,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}AZ\right) = A\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\right)A^T\\
&= AE\left[\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\middle| X,V\right)\right]A^T\\
&\quad + A\mathrm{Var}\left(ZE\left[\frac{\xi-\pi_0(V)}{\pi_0(V)}\middle| X,V\right]\right)A^T\\
&= AE\left[Z^{\otimes 2}\frac{1-\pi_0(V)}{\pi_0(V)}\right]A^T + 0 = A_1A_2A_1^T.
\end{aligned}
$$

□

*Proof of Corollary 3.4.3.* (1). We first consider stratified Bernoulli sampling. The case for $\hat{\theta}_{N,c}$ was proved in [5]. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other two cases, (3.12) and (3.14) corresponding to $\hat{\theta}_{N,e}$ and $\hat{\theta}_{N,cc}$, are similar.

For $\tilde{Z} \equiv (Z^{(1)}, \ldots, Z^{(J)})^T$ with $Z^{(j)} \equiv I(V \in \mathcal{V}_j)Z^T$, we compute $\tilde{A}_1 \equiv P_0[(\pi_0^{-1}(V) - 1)\tilde{\ell}_0\tilde{Z}^T]$ and $\tilde{A}_2 \equiv \{P_0[(\pi_0^{-1}(V) - 1)\tilde{Z}^{\otimes 2}]\}^{-1}$. Note that $Q_{mc}\tilde{\ell}_0 = \tilde{A}_1\tilde{A}_2\tilde{Z}$. The matrix $\tilde{A}_1 = [\tilde{A}_{1,1}, \ldots, \tilde{A}_{1,J}]$ is a partitioned matrix where

$$\tilde{A}_{1,j} \equiv P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0 Z^{(j)}\right) = \nu_j P_{0|j}\left(\frac{1-p_j}{p_j}\tilde{\ell}_0 Z^T\right) \in \mathbb{R}^{p\times k}.$$

and the matrix $\tilde{A}_2$ is the block diagonal matrix the $j$th block of which is

$$\tilde{A}_{2,j} \equiv \left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}[(Z^{(j)})^T]^{\otimes 2}\right\}^{-1} = \left\{\nu_j P_{0|j}\frac{1-p_j}{p_j}Z^{\otimes 2}\right\}^{-1} \in \mathbb{R}^{k\times k}.$$

Thus, the matrix $\tilde{A} \equiv \tilde{A}_1\tilde{A}_2$ is a partitioned matrix $\tilde{A} = [\tilde{A}_1, \ldots, \tilde{A}_J]$ where

$$\tilde{A}_j = \tilde{A}_{1,j}\tilde{A}_{2,j} = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1}.$$

It follows by the definition of the $Z^{(j)}$'s that

$$\begin{aligned} P_{0|j}\left\{(I - Q_{mc})\tilde{\ell}_0\right\}^{\otimes 2} &= P_{0|j}\left\{\tilde{\ell}_0 - \tilde{A}\tilde{Z}\right\}^{\otimes 2} \\ &= P_{0|j}\left\{\tilde{\ell}_0 - \tilde{A}_j Z\right\}^{\otimes 2} = P_{0|j}\left\{(I - Q_c^{(j)})\tilde{\ell}_0\right\}^{\otimes 2}. \end{aligned}$$

Since

$$P_{0|j}\left(\tilde{A}_j Z\right)^{\otimes 2} = \tilde{A}_j P_{0|j} Z^{\otimes 2}\tilde{A}_j^T = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0 Z^T\right)^T,$$

and

$$P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\tilde{A}_j^T = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0 Z^T\right)^T,$$

it follows that

$$P_{0|j}\left\{(I - Q_c^{(j)})\tilde{\ell}_0\right\}^{\otimes 2} = P_{0|j}\tilde{\ell}_0^{\otimes 2} - P_{0|j}\{Q_c^{(j)}\tilde{\ell}_0\}^{\otimes 2}.$$

Substitution of this into (3.10) gives (3.13).

(2). Next, we consider the second part of Corollary 3.4.3 concerning stratified sampling without replacement. For $\tilde{Z} \equiv (Z^{(1)}, \ldots, Z^{(J)})^T$ with $Z^{(j)} \equiv I(V \in \mathcal{V}_j)Z^T$, we compute $\tilde{B}_1 \equiv P_0[(\pi_0^{-1}(V) - 1)\tilde{\ell}_0(\tilde{Z} - \mu_{\tilde{Z}})^T]$ and $\tilde{B}_2 \equiv \{P_0[(\pi_0^{-1}(V) - 1)(\tilde{Z} - \mu_{\tilde{Z}})^{\otimes 2}]\}^{-1}$. Note that $Q_{cc}\tilde{\ell}_0 = \tilde{B}_1\tilde{B}_2\tilde{Z}$ and $\mu_{\tilde{Z}} = (\mu_{Z,1}^T, \ldots, \mu_{Z,J}^T)^T$. The matrix $\tilde{B}_1 = [\tilde{B}_{1,1}, \ldots, \tilde{B}_{1,J}]$ is a partitioned matrix where

$$\tilde{B}_{1,j} \equiv P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0(Z^{(j)} - \mu_{Z,j}^T)\right) = \nu_j P_{0|j}\left(\frac{1-p_j}{p_j}\tilde{\ell}_0(Z - \mu_{Z,j})^T\right).$$

and the matrix $\tilde{B}_2$ is the block diagonal matrix the $j$th block of which is

$$\tilde{B}_{2,j} \equiv \left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}[(Z^{(j)})^T - \mu_{Z,j}]^{\otimes 2}\right\}^{-1} = \left\{\nu_j P_{0|j}\frac{1-p_j}{p_j}(Z - \mu_{Z,j})^{\otimes 2}\right\}^{-1}.$$

Thus, the matrix $\tilde{B} \equiv \tilde{B}_1\tilde{B}_2$ is a partitioned matrix $\tilde{B} = [\tilde{B}_1, \ldots, \tilde{B}_J]$ where

$$\tilde{B}_j = \tilde{B}_{1,j}\tilde{B}_{2,j} = P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{P_{0|j}(Z - \mu_{Z,j})^{\otimes 2}\right\}^{-1}.$$

It follows by the definition of $Z^{(j)}$'s that

$$\begin{aligned}&\mathrm{Var}_{0|j}\left\{(I - Q_{cc})\tilde{\ell}_0\right\} = \mathrm{Var}_{0|j}\left\{\tilde{\ell}_0 - \tilde{B}(\tilde{Z} - \mu_{\tilde{Z}})\right\}\\ &= \mathrm{Var}_{0|j}\left\{\tilde{\ell}_0 - \tilde{B}_j(Z - \mu_{Z,j})\right\} = \mathrm{Var}_{0|j}\left\{(I - Q_{cc}^{(j)})\tilde{\ell}_0\right\}.\end{aligned}$$

Then, since

$$\begin{aligned}&\mathrm{Var}_{0|j}\left(\tilde{B}_j(Z - \mu_{Z,j})\right) = \tilde{B}_j\mathrm{Var}_{0|j}(Z)\tilde{B}_j^T\\ &= P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{\mathrm{Var}_{0|j}(Z)\right\}^{-1}P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)^T,\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Cov}_{0|j}\left(\tilde{\ell}_0, \tilde{B}_j(Z - \mu_{Z,j})\right) = P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\tilde{B}_j^T\\ &= P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{\mathrm{Var}_{0|j}(Z)\right\}^{-1}P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)^T,\end{aligned}$$

it follows that

$$\mathrm{Var}_{0|j}\left\{(I - Q_c^{(j)})\tilde{\ell}_0\right\} = \mathrm{Var}_{0|j}\left(\tilde{\ell}_0\right) - \mathrm{Var}_{0|j}\{Q_c^{(j)}\tilde{\ell}_0\}.$$

Substitution of this last identity into (3.5) gives (3.15). □

# Chapter 4

# VARIANCE ESTIMATION

In the previous chapter, we derived asymptotic distributions of WLE's. A next natural question in order to carry out statistical analysis is variance estimation for a finite-dimensional parameter. However, the presence of an infinite-dimensional nuisance parameter often results in non-explicit expressions for asymptotic variance. This is the case for the Cox proportional hazards model with current status data, which is discussed in the next chapter in detail. Moreover, the phase II variances due to our "without replacement" sampling scheme are not easily estimated by simple modifications of existing methods. Our goal of this chapter is to develop two general methods for variance estimation under two-phase sampling that cover the case where the asymptotic variance involves unknown functions.

Before describing our methods, we illustrate difficulties in this statistical problem in a relatively simple situation. Suppose that the efficient influence function for complete data $\tilde{\ell}_{\theta,\eta}$ is a known function up to a parameter $(\theta, \eta)$, and that a plain WLE $\hat{\theta}_N$ has asymptotic variance

$$\Sigma = \underbrace{I_0^{-1}}_{\text{phase I variance}} + \underbrace{\sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(\tilde{\ell}_0)}_{\text{phase II variances}},$$

as in (3.1). With complete data, the asymptotic variance $I_0^{-1}$ of the MLE $\hat{\theta}$ can be estimated by $\mathbb{P}_N \tilde{\ell}^{\otimes 2}_{\hat{\theta},\hat{\eta}}$ where $\hat{\eta}$ is the MLE of $\eta$ because we expect that the Glivenko-Cantelli theorem implies that

$$\mathbb{P}_N \tilde{\ell}^{\otimes 2}_{\hat{\theta},\hat{\eta}} \to_P P_0 \tilde{\ell}_0^{\otimes 2} = I_0^{-1}.$$

Simple application of this idea to our situation by replacing the empirical measure $\mathbb{P}_N$ by the IPW empirical measure $\mathbb{P}_N^{\pi}$ does not lead to valid variance estimation since our Glivenko-Cantelli theorem (Theorem 6.1.1) is expected to yield

$$\hat{I}_0^{-1} = \mathbb{P}_N^{\pi} \tilde{\ell}^{\otimes 2}_{\hat{\theta}_N,\hat{\eta}_N} \to_P P_0 \tilde{\ell}_0^{\otimes 2} = I_0^{-1} \neq \Sigma,$$

where $\hat{\eta}_N$ is the WLE of $\eta$. The problem here is failure to estimate the phase II variances. In fact, this difficulty is seen in other "naive" applications of available methods for complete data (the method of [35], or a discretized efficient information matrix on observed data) to two-phase sampling in that existing methods with the empirical measure replaced by the IPW empirical measure only estimate the phase I variance. This problem motivates the development of two new bootstrap-based procedures for two-phase sampling described below.

For the sake of completeness, we present a discussion of variance estimation in the above example. In fact, we do not need a new method in this simple example. We estimate $\mathrm{Var}_{0|j}\tilde{\ell}_0$ by $\widehat{\mathrm{Var}}_{0|j}(\tilde{\ell}_0) = \hat{P}_j\tilde{\ell}^{\otimes 2}_{\hat{\theta}_N,\hat{\eta}_N} - \left\{\hat{P}_j\tilde{\ell}_{\hat{\theta}_N,\hat{\eta}_N}\right\}^{\otimes 2}$ where for $j = 1,\ldots,J$, $\hat{P}_j f \equiv \mathbb{P}^{\pi}_N fI(V \in \mathcal{V}_j)$. Also, we estimate $\nu_j$ and $p_j$ by $\hat{\nu}_j = N_j/N$ and $\hat{p}_j = n_j/N_j$, respectively. Then a variance estimate $\hat{\Sigma}$ can be defined by

$$\hat{\Sigma} = \hat{I}_0^{-1} + \sum_{j=1}^{J}\hat{\nu}_j\frac{1-\hat{p}_j}{\hat{p}_j}\widehat{\mathrm{Var}}_{0|j}(\tilde{\ell}_0).$$

The cases for other WLE's are similar with $Q_{\#}, \# \in \{e, c, mc, cc\}$ estimated. For example, $Q_c f$ is estimated by

$$\hat{Q}_c f(V_i, X_i, \xi_i) = \mathbb{P}^{\pi}_N(fZ)\left\{\mathbb{P}^{\pi}_N Z^{\otimes 2}\right\}^{-1}Z,$$

so that we can estimate $\mathrm{Var}_{0|j}\left((I-Q_c)\tilde{\ell}_0\right)$ by

$$\widehat{\mathrm{Var}}_{0|j}\left((I-Q_c)\tilde{\ell}_0\right) = \hat{P}_j\left\{(I-\hat{Q}_c)\tilde{\ell}_{\hat{\theta}_N,\hat{\eta}_N}\right\}^{\otimes 2} - \left\{\hat{P}_j(I-\hat{Q}_c)\tilde{\ell}_{\hat{\theta}_N,\hat{\eta}_N}\right\}^{\otimes 2}.$$

Although variance estimation can be accomplished as above when the asymptotic variance is expressed in terms of a function $\tilde{\ell}_{\theta,\eta}$ known up to $(\theta,\eta)$, our problem of interest is to develop general methods for variance estimation including cases where an explicit formula for the asymptotic variance is unavailable.

In the following, we introduce two general bootstrap-based procedures for estimating asymptotic variances of the WLE's of a finite-dimensional parameter. The first method estimates phase I and II variances separately by an extension of [35] and a survey bootstrap of [18] and [3], respectively. The second method estimates phase I and II variances simultaneously based on the single bootstrap procedure with survey bootstrap as a foundational building block. The first method allows us to compare a design with complete data and a two-phase design while the second method correctly estimates asymptotic variance even under model misspecification. All proofs are presented at the end of this chapter.

## 4.1 Separate Estimation for the Phase I and II Variances

### 4.1.1 Estimation for the Phase I Variance

We estimate the phase I variance by extending the method of [35] to two-phase sampling. This method requires more conditions than imposed in Chapter 3.

A modified version of the conditions imposed in [35] are as follows.

**Condition 4.1.1.** *The weighted likelihood estimator is the maximizer of the weighted likelihood;*

$$(\hat{\theta}_N, \hat{\eta}_N) = \text{argmax}_{\theta\in\Theta,\eta\in H}\mathbb{P}_N^\pi \text{loglik}(\theta,\eta),$$

*where* $\text{loglik}(\theta,\eta)$ *is the log likelihood at* $(\theta,\eta)$*, or equivalently,* $\hat{\theta}_N$ *maximizes the weighted profile likelihood* $\theta \mapsto \mathbb{M}_N^\pi(\theta)$ *given by*

$$\mathbb{M}_N^\pi(\theta) \equiv \sup_{\eta\in H}\mathbb{P}_N^\pi \text{loglik}(\theta,\eta).$$

*The estimator* $\hat{\theta}_N$ *of* $\theta$ *is consistent and asymptotically normal with*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = \tilde{I}_0^{-1}\mathbb{G}_N^\pi \ell_0^* + o_P(1), \tag{4.1}$$

*where* $\ell_0^*$ *is the efficient score and* $\tilde{I}_0 = P_0(\ell_0^*)^{\otimes 2}$ *is the efficient information for* $\theta$ *in the complete data model.*

**Condition 4.1.2.** *Let* $\psi \equiv (\theta,\eta)$*, and, for a fixed* $\theta$*,* $\hat{\eta}_{\theta,N} \equiv \text{argmax}_{\eta\in H}\mathbb{P}_N^\pi \text{loglik}(\theta,\eta)$ *and* $\hat{\psi}_{\theta,N} \equiv (\theta, \hat{\eta}_{\theta,N})$*.*

*(1) For each* $\psi = (\theta,\eta)$ *there exists a map* $t \mapsto \underline{\eta}_t(\psi)$ *from a fixed neighborhood of* $\theta$ *to the parameter set for* $\eta$ *such that the map* $t \mapsto \ell(t,\psi)(x)$ *defined by*

$$\ell(t,\psi)(x) = \text{loglik}(t, \underline{\eta}_t(\psi))(x)$$

*is twice continuously differentiable for all* $x \in \mathcal{X}$ *where the derivatives are denoted by* $\dot{\ell}(t,\psi)(x)$*, and* $\ddot{\ell}(t,\psi)(x)$*, respectively.*

*(2) The p-dimensional submodel with parameters $(t, \underline{\eta}_t(\psi))$ passes through $\psi = (\theta, \eta)$ at $t = \theta$;*

$$\underline{\eta}_\theta(\theta, \eta) = \eta, \quad \textit{every} \quad (\theta, \eta). \tag{4.2}$$

*This submodel is the least favorable in the sense that*

$$\dot{\ell}(\theta_0, \psi_0)(x) = \ell_0^*.$$

*(3) It holds that $\hat{\eta}_{\tilde{\theta}_N, N} \to_P \eta_0$ for every $\tilde{\theta}_N \to_P \theta_0$.*

**Condition 4.1.3.** *For any random sequences $\tilde{\theta}$ and $\overline{\psi}$ such that $\tilde{\theta} \to_P \theta_0$ and $\overline{\psi} \to_P \psi_0$,*

$$\mathbb{G}_N^\pi \dot{\ell}(\tilde{\theta}, \overline{\psi}) = \mathbb{G}_N^\pi \ell_0^* + o_P(1), \tag{4.3}$$

$$\mathbb{P}_N^\pi \ddot{\ell}(\tilde{\theta}, \overline{\psi}) \to_P -\tilde{I}_0, \tag{4.4}$$

$$P_0 \dot{\ell}(\tilde{\theta}, \hat{\psi}_{\tilde{\theta}, N}) = -\tilde{I}_0(\tilde{\theta} - \theta_0) + o_P\left(\|\tilde{\theta} - \theta_0\| + N^{-1/2}\right). \tag{4.5}$$

The weighted likelihood estimator in Condition 4.1.1 is formulated as the $M$-estimator rather than the $Z$-estimator in Chapter 3. This formulation is plausible because the WLE as the $M$-estimator is often obtained from solving the weighted likelihood equations as in Chapter 3. Moreover, the condition

$$(\hat{\theta}_N, \hat{\eta}_N) = \text{argsup}_{\theta \in \Theta, \eta \in H} \mathbb{P}_N^\pi \text{loglik}(\theta, \eta),$$

can be replaced by the conditions

$$\mathbb{P}_N^\pi \text{loglik}(\hat{\theta}_N, \hat{\eta}_N) = \sup_{\theta \in \Theta, \eta \in H} \mathbb{P}_N^\pi \text{loglik}(\theta, \eta) + o_P(N^{-1}),$$

and the weighted likelihood equations (2.7) or (3.6). For ease of the presentation, we do not present the details on this relaxation. The rest of Condition 4.1.1 can be verified either by Theorems 3.2.1 or 3.3.1 after establishing consistency of the estimators. Condition 4.1.2 and the equation (4.5) in Condition 4.1.3 are identical to the conditions imposed on the complete data model in [35]. The equation (4.3) in Condition 4.1.3 can be verified by the result on the asymptotic equicontinuity with appropriate additional conditions such as Theorem 6.3.2, Lemmas 6.3.1 or 6.3.2. The convergence in (4.4) in Condition 4.1.3 can be also verified with the aid of the Glivenko-Cantelli Theorem 6.1.1. See also Lemma 2.2 of [35] for appropriate conditions to verify (4.3) and (4.4).

**Theorem 4.1.1.** *Suppose that Conditions 4.1.1, 4.1.2 and 4.1.3 are satisfied. Then*

$$-2\frac{\mathbb{M}_N^\pi(\hat{\theta}_N + h_N v_N) - \mathbb{M}_N^\pi(\hat{\theta}_N)}{h_N^2} \to_P v_0^T \tilde{I}_0 v_0, \tag{4.6}$$

*for every random sequence* $h_N \to_P 0$ *such that* $(\sqrt{N}h_N)^{-1} = O_P(1)$ *and for every sequence* $v_N \to_P v_0$.

**Remark 4.1.1.** *This theorem is formulated with numerical differentiation of the profile likelihood for the case where the efficient information for the complete data model contains unknown functions. However, this theorem also justifies use of the first or second derivatives of the profile likelihood with respect to the finite-dimensional parameter for estimating the efficient information as seen in the Cox model with right censored data. See a discussion in [35].*

**Remark 4.1.2.** *Under model misspecification, the phase I variance typically takes the form of a "sandwich formula" rather than* $\tilde{I}_0$ *but the method described above will estimate* $\tilde{I}_0$. *In contrast, bootstrap methods we discuss below automatically respect model misspecification.*

### *4.1.2 Estimation for the Phase II Variances*

*Survey Bootstrap*

We introduce the *survey bootstrap* for the IPW empirical processes, which reproduces the phase II variances. This bootstrap procedure is based on the method proposed by [18, 3] for a stratified sample from a finite population. We illustrate the procedure by example. For simplicity, suppose that there is one stratum so that we can suppress indices. The first step is to create a bootstrap equivalent to the phase I population. If $N = 900$ and $n = 300$ so that $N = 3n$, we create an *pre-bootstrap phase I population* or an "artificial population" in the terminology of [18], by copying 300 phase II observations 3 times. This bootstrap population does not contain observations not sampled at the second phase but these observations are not used in weighted likelihood estimation from the first except for adjusting weights. The next step is sample $n$ observations without replacement. If $N$ is not divisible by $n$, say, $N = 1000$ and $n = 350$ ($N = 2n + 300 \equiv kn + r$), we create two pre-bootstrap phase I populations, by copying phase II observations $k = 2$ times and

$k + 1 = 3$ times, respectively. Then we sample $n$ observations from the first pre-bootstrap phase I population with probability $s = (1-r/n)(1-r/(N-1))$ or from the second phase I population with probability $1-s$. For the case of multiple strata, we carry out the procedure described above for each stratum.

More formally, the survey bootstrap is described as follows. For the $j$th stratum, let $(W_1^{(j)}, \ldots, W_{n_j}^{(j)}) \in \mathbb{R}^{n_j}$ be a vector of an exchangeable weights that follows the mixture of the multivariate hypergeometric distribution $MH(n_j k_j, n_j, (k_j, \ldots, k_j))$ with probability $s_j = (1 - r_j/n_j)(1 - r_j/(N_j - 1))$ and the multivariate hypergeometric distribution $MH(n_j(k_j + 1), n_j, (k_j + 1, \ldots, k_j + 1))$ with probability $1 - s_j$ where $N_j = k_j n_j + r_j$, $k_j, r_j \in \mathbb{N}$ with $0 \le r_j < n_j$. See Section 4.3 for notation and basic results on the multivariate hypergeometric distribution. We define $W_{n_j,j,i}$ by $W_{n_j,j,i} = 0$ if $\xi_{j,i} = 0$ and $W_{n_j,j,i} = W_k^{(j)}$ where the observation $(V_{j,i}, \xi_{j,i} X_{j,i}, \xi_{j,i})$ has the $k$th smallest index $i$ among the observations with $\xi_{j,i} = 1$ in the $j$th stratum. We denote $W_{Ni}$ for $W_{n_j,j,i}$ in the same way as other notations when we do not specify the stratum where the observation belongs. Define the *survey bootstrap IPW empirical measure* by

$$\begin{aligned}
\hat{\mathbb{P}}_N^{\pi,S} &= \frac{1}{N}\sum_{i=1}^{N} W_{Ni}\frac{\xi_i}{\pi(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \\
&\equiv \frac{1}{N}\sum_{j=1}^{J} N_j \hat{\mathbb{P}}_{j,n_j}^{\xi,S},
\end{aligned}$$

where the *survey bootstrap phase II empirical measure* for the $j$th stratum is given by

$$\hat{\mathbb{P}}_{j,n_j}^{\xi,S} \equiv \frac{1}{n_j}\sum_{i=1}^{N_j} W_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}}, \quad j = 1, \ldots, J.$$

Also, define the *survey bootstrap IPW empirical process* by

$$\hat{\mathbb{G}}_N^{\pi,S} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S} - \mathbb{P}_N^{\pi}).$$

We decompose the survey IPW empirical process as in the decomposition (2.1) of the IPW empirical process;

$$\begin{aligned}
\hat{\mathbb{G}}_N^{\pi,S} &= \sqrt{N}\left(\hat{\mathbb{P}}_N^{\pi,S} - \mathbb{P}_N^{\pi}\right) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{\xi_{j,i}}{n_j/N_j}(W_{n_j,j,i}-1)\delta_{X_{j,i}} \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{J}N_j\left(\hat{\mathbb{P}}_{j,n_j}^{\xi,S} - \mathbb{P}_{j,n_j}^{\xi}\right) \\
&= \sum_{j=1}^{J}\sqrt{\frac{N_j}{N}}\sqrt{\frac{N_j}{n_j}}\hat{\mathbb{G}}_{j,n_j}^{\xi,S}, \qquad (4.7)
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{P}_{j,n_j}^{\xi} &\equiv \frac{1}{n_j}\sum_{i=1}^{N_j}\xi_{j,i}\delta_{X_{j,i}}, \quad j=1,\ldots,J, \\
\hat{\mathbb{G}}_{j,n_j}^{\xi,S} &\equiv \sqrt{n_j}(\hat{\mathbb{P}}_{j,n_j}^{\xi,S} - \mathbb{P}_{j,n_j}^{\xi}), \quad j=1,\ldots,J.
\end{aligned}$$

*Survey Bootstrap with Adjusted Weights*

We propose two calibration methods for bootstrapping the IPW empirical process. The differences are how to view the original samples to be calibrated leading to different centering of estimators. The first method we propose does not require information on adjusting weights in the original sample in order to calibrate a bootstrap sample. In other words, we calibrate the sample $\{(W_{Ni}(V_i, \xi_i X_i, \xi_i)\}_{i=1}^N$. In the original sample, we calibrate the sample to the known phase I average. The corresponding quantity in the pre-bootstrap phase I population is

$$\frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}k_j\xi_{j,i}Z_{j,i} = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{\xi_{j,i}}{\pi_0(V_{j,i})}Z_{j,i} = \mathbb{P}_N^{\pi}Z,$$

when $N_j = k_j n_j$ for $j = 1, \ldots, J$. This observation leads to calibrating a bootstrap sample to the Horvitz-Thompson estimator of $Z$. Specifically, the estimators of $\alpha$ we propose for calibration, modified calibration and centered calibration in survey bootstrap are the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^{N}W_{Ni}\frac{\xi_i G_c(V_i;\alpha)}{\pi_0(V_i)}Z_i = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_0(V_i)}Z_i, \qquad (4.8)$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_c(V;\hat{\hat{\alpha}}_N^c)Z = \mathbb{P}_N^\pi Z$, the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i G_{mc}(V_i;\alpha)}{\pi_0(V_i)}Z_i = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_0(V_i)}Z_i, \tag{4.9}$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\hat{\hat{\alpha}}_N^{mc})Z = \mathbb{P}_N^\pi Z$, and the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i G_{cc,S}(V_i;\alpha)}{\pi_0(V_i)}\left(Z_i - \mathbb{P}_N^\pi Z\right) = 0, \tag{4.10}$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_{cc}(V;\hat{\hat{\alpha}}_N^{cc})(Z - \mathbb{P}_N^\pi Z) = 0$, where

$$G_{cc,S}(V,\alpha) \equiv G\left(\frac{1-\pi_0(V)}{\pi_0(V)}\left\{Z - \mathbb{P}_N^\pi Z\right\}^T\alpha\right),$$

respectively. As we discussed in Section 3.4.3, certain calibrations with some choice of (transformation of) auxiliary variables lead to the same asymptotic variance of the WLE with estimated weights, and hence we do not propose a method specific to estimating weights. The calibrated survey bootstrap IPW empirical process is defined by $\hat{\mathbb{G}}_N^{\pi,S,c} = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,c} - \mathbb{P}_N^\pi)$ where the calibrated bootstrap IPW empirical measure is defined as

$$\hat{\mathbb{P}}_N^{\pi,S,c} \equiv \frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i G_c(V_i;\hat{\hat{\alpha}}_N^c)}{\pi_0(V_i)}\delta_{X_i} \equiv \frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i}{\pi_{\hat{\hat{\alpha}}_N^c}(V_i)}\delta_{X_i},$$

and $\hat{\hat{\alpha}}_N$ is the solution to (4.8). The calibrated survey bootstrap IPW empirical processes $\hat{\mathbb{G}}_N^{\pi,S,mc}$ and $\hat{\mathbb{G}}_N^{\pi,S,cc}$ with modified calibration and centered calibration and corresponding measures $\hat{\mathbb{P}}_N^{\pi,S,mc}$ and $\hat{\mathbb{P}}_N^{\pi,S,cc}$ are defined analogously. Note that the centering process is $\mathbb{P}_N^\pi$ regardless of the method used to adjust the weights.

The second method we propose uses information on adjusting weights in the original sample. The bootstrap sample to be calibrated in this method is $\{G_\#(V_i;\hat{\alpha}_N^\#)W_{Ni}(V_i,\xi_i X_i,\xi_i)\}_{i=1}^N$, with $\# \in \{c, mc, cc\}$. The estimators of $\alpha$ we propose for calibration, modified calibration and centered calibration in survey bootstrap are the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i G_c(V_i;\alpha)G_c(V_i;\hat{\alpha}_N)}{\pi_0(V_i)}Z_i = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_0(V_i)}G_c(V_i;\hat{\alpha}_N^c)Z_i, \tag{4.11}$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_c(V;\hat{\hat{\alpha}}_N^{dc})G_c(V;\hat{\alpha}_N)Z = \mathbb{P}_N^{\pi,c}Z$, the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^N W_{Ni}\frac{\xi_i G_{mc}(V_i;\alpha)G_{mc}(V_i;\hat{\alpha}_N^{mc})}{\pi_0(V_i)}Z_i = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i G_{mc}(V_i;\hat{\alpha}_N)}{\pi_0(V_i)}Z_i, \tag{4.12}$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\hat{\hat{\alpha}}_N^{dmc}) G_{mc}(V;\hat{\alpha}_N) Z = \mathbb{P}_N^{\pi,mc} Z$, and the solution to the bootstrap calibration equation

$$\frac{1}{N}\sum_{i=1}^{N} W_{Ni} \frac{\xi_i G_{cc,S}(V_i;\alpha) G_{cc}(V_i;\hat{\alpha}_N^{cc})}{\pi_0(V_i)} \left(Z_i - \mathbb{P}_N^{\pi,cc} Z\right) = 0, \tag{4.13}$$

or $\hat{\mathbb{P}}_N^{\pi,S} G_{cc,S}(V;\hat{\hat{\alpha}}_N^{dcc}) G_{cc}(V_i;\hat{\alpha}_N)(Z - \mathbb{P}_N^{\pi,cc} Z) = 0$, where with the abuse of notation

$$G_{cc,S}(V,\alpha) \equiv G\left(\frac{1-\pi_0(V)}{\pi_0(V)} \left\{Z - \mathbb{P}_N^{\pi,cc} Z\right\}^T \alpha\right),$$

respectively. We call these methods of adjusting weights for bootstrap *double (modified and centered) calibrations*. The doubly calibrated survey bootstrap IPW empirical process is defined by $\hat{\mathbb{G}}_N^{\pi,dc} = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,dc} - \mathbb{P}_N^{\pi,c})$ where the calibrated bootstrap IPW empirical measure is defined as

$$\hat{\mathbb{P}}_N^{\pi,S,dc} \equiv \frac{1}{N}\sum_{i=1}^{N} W_{Ni} \frac{\xi_i G_c(V_i;\hat{\hat{\alpha}}_N^{dc}) G_c(V_i;\hat{\alpha}_N^{c})}{\pi_0(V_i)} \delta_{X_i} \equiv \frac{1}{N}\sum_{i=1}^{N} W_{Ni} \frac{\xi_i}{\pi_{\hat{\hat{\alpha}}_N^{dc}}(V_i)} G_c(V_i;\hat{\alpha}_N^{c}) \delta_{X_i},$$

$\hat{\hat{\alpha}}_N^{dc}$ is the solution to (4.11) and $\hat{\alpha}_N^c$ is the solution to the calibration equation (2.4). We suppress $c, dc, mc, dmc, cc, dcc$ for $\hat{\alpha}_N$ and $\hat{\hat{\alpha}}_N$ for ease of presentation. We define the doubly calibrated survey bootstrap IPW empirical processes $\hat{\mathbb{G}}_N^{\pi,S,dmc} = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,dmc} - \mathbb{P}_N^{\pi,mc})$ with modified calibration and $\hat{\mathbb{G}}_N^{\pi,S,dcc} = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,dcc} - \mathbb{P}_N^{\pi,cc})$ with centered calibration where $\hat{\mathbb{P}}_N^{\pi,S,dmc}$ and $\hat{\mathbb{P}}_N^{\pi,S,dcc}$ are defined analogously. Note that centering each bootstrap calibrated IPW empirical process is done by the corresponding IPW empirical process depending on methods of adjusting weights.

We explicitly define the probability space for $(V, \xi X, \xi)$ and $W$. Let $\mathbf{W_j} = \{W_{n_j,j,i} : i = 1,\ldots,n_j, n_j = 1,2,\ldots\}$ be a triangular array defined on the probability space $(\mathcal{Z}_j, \mathcal{E}_j, P_{W_j})$ for $j = 1,\ldots,J$. Define the probability space $(\mathcal{Z}, \mathcal{E}, P_W) = \prod_{j=1}^J (\mathcal{Z}_j, \mathcal{E}_j, P_{W_j})$ for the bootstrap weights. We denote the probability space for $(V_i, \xi_i X_i, \xi_i), i = 1,2,\ldots,$ as $(\mathcal{X}^\infty, \mathcal{B}^\infty, P^\infty)$, and denote

$$(\mathcal{X}^\infty, \mathcal{B}^\infty, P^\infty) \times (\mathcal{Z}, \mathcal{E}, P_W) = (\mathcal{X}^\infty \times \mathcal{Z}, \mathcal{B}^\infty \times \mathcal{E}, Pr),$$

where $Pr \equiv P^\infty \times P_W$. We let $P^*$ and $P_*$ denote the outer and the inner probability, respectively, corresponding to $P^\infty$.

We repeatedly use the definitions and results in [60]. For a real function $\Delta_N$ defined on the joint probability space $(\mathcal{X}^\infty \times \mathcal{Z}, \mathcal{B}^\infty \times \mathcal{E}, Pr)$, we say that $\Delta_N$ is of an order $o_{P_W}(1)$ in

$P^*$-probability if for any $\epsilon > 0$ and $\eta > 0$,

$$P^* \{P_W (|\Delta_N| > \epsilon) > \eta\} \to 0,$$

as $N \to \infty$, and $\Delta_N$ is of an order $O_{P_W}(1)$ in $P^*$-probability if for any $\eta > 0$ and for every $M_N \to \infty$,

$$P^* \{P_W (|\Delta_N| > M_N) > \eta\} \to 0,$$

as $N \to \infty$. The important result is if $\Delta_N$ is of an order $o_{Pr}(1)$ or $O_{Pr}(1)$ then it is of an order $o_{P_W}(1)$ or $O_{P_W}(1)$ in $P^*$-probability, respectively. This result allows us to freely translate stochastic orders in joint probability space to the probability space for bootstrap weights. We summarize several results in the following lemma.

**Lemma 4.1.1.** *Let $\Delta_N$ be a real function defined on the joint probability space $(\mathcal{X}^\infty \times \mathcal{Z}, \mathcal{B}^\infty \times \mathcal{E}, Pr)$.*

*If $\Delta_N = o_{Pr}(1)$, then $\Delta_N = o_{P_W}(1)$ in $P^*$-probability.*

*If $\Delta_N = o_{P_W}(1)$ in $P^*$-probability and $\Delta_N$ is measurable, then $\Delta_N = o_{Pr}(1)$.*

*If $\Delta_N = O_{Pr}(1)$, then $\Delta_N = O_{P_W}(1)$ in $P^*$-probability.*

*If $\Delta_N = O_{P_W}(1)$ in $P^*$-probability and $\Delta_N$ is measurable, then $\Delta_N = O_{Pr}(1)$.*

*Survey Bootstrap Z-theorems*

We require the following conditions in place of Conditions 3.1.2 and 3.2.1, respectively;

**Condition 4.1.4.** *Conditions 3.1.2(b)-(e) hold with either*

*(a-1) (i) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^c$ is a solution of the calibration equation (4.8),*

*(ii) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{mc}$ is a solution of the calibration equation (4.9),*

*(iii) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{cc}$ is a solution of the calibration equation (4.10), or*

*(a-2) (i) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{dc}$ is a solution of the calibration equation (4.11),*

*(ii) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{dmc}$ is a solution of the calibration equation (4.12),*

*(iii) The estimator $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{dcc}$ is a solution of the calibration equation (4.13).*

**Condition 4.1.5** (Consistency). *The bootstrap estimator* $(\hat{\hat{\theta}}_N, \hat{\hat{\eta}}_N)$ *is consistent for* $(\theta_0, \eta_0)$ *in* $P^*$*-probability and solves the bootstrap weighted likelihood equations*

$$\begin{aligned} \hat{\Psi}^{\pi,S}_{N,1}(\theta,\eta) &= \hat{\mathbb{P}}^{\pi,S}_N \dot{\ell}_{\theta,\eta} = o_{P^*_W}\left(N^{-1/2}\right), \\ \left\| \hat{\Psi}^{\pi,S}_{N,2}(\theta,\eta)h \right\|_{\mathcal{H}} &= \left\| \hat{\mathbb{P}}^{\pi,S}_N (B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h) \right\|_{\mathcal{H}} = o_{P^*_W}\left(N^{-1/2}\right), \end{aligned} \tag{4.14}$$

*in* $P^*$*-probability where* $\hat{\mathbb{P}}^{\pi}_N$ *may be replaced by* $\hat{\mathbb{P}}^{\pi,S,c}_N$, $\hat{\mathbb{P}}^{\pi,S,mc}_N$, $\hat{\mathbb{P}}^{\pi,S,cc}_N$ $\hat{\mathbb{P}}^{\pi,S,dc}_N$, $\hat{\mathbb{P}}^{\pi,S,dmc}_N$ *or* $\hat{\mathbb{P}}^{\pi,S,dcc}_N$ *for the estimators with calibration, modified calibration, centered calibration, double calibration, double modified calibration, double centered calibration with the corresponding estimators denoted as* $(\hat{\hat{\theta}}_{N,S,c}, \hat{\hat{\eta}}_{N,S,c})$, $(\hat{\hat{\theta}}_{N,S,mc}, \hat{\hat{\eta}}_{N,S,mc})$, $(\hat{\hat{\theta}}_{N,S,cc}, \hat{\hat{\eta}}_{N,S,cc})$, $(\hat{\hat{\theta}}_{N,S,dc}, \hat{\hat{\eta}}_{N,S,dc})$, $(\hat{\hat{\theta}}_{N,S,dmc}, \hat{\hat{\eta}}_{N,S,dmc})$, *and* $(\hat{\hat{\theta}}_{N,S,dcc}, \hat{\hat{\eta}}_{N,S,dcc})$, *respectively.*

The following is the survey bootstrap version of Theorem 3.2.1. Note that the asymptotic variances only consist of the phase II variances as expected. Note also that double calibrations have right centering of bootstrap WLE's while single calibrations need centering by the plain WLE $\hat{\theta}_N$ to obtain the same phase II asymptotic variances for the WLE's with the original sample.

**Theorem 4.1.2.** *Suppose that Conditions 3.2.2, 3.2.3 and 4.1.5 hold.*

*(1) Under Condition 4.1.4 with 4.1.4(a-1),*

$$\begin{aligned} \sqrt{N}(\hat{\hat{\theta}}_{N,S} - \hat{\theta}_N) &= \hat{\mathbb{G}}^{\pi,S}_N \tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow Z_S &\sim N_p(0, \Sigma_S), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,S,c} - \hat{\theta}_N) &= \hat{\mathbb{G}}^{\pi,S,c}_N \tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow Z_{S,c} &\sim N_p(0, \Sigma_{S,c}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,S,mc} - \hat{\theta}_N) &= \hat{\mathbb{G}}^{\pi,S,mc}_N \tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow Z_{S,mc} &\sim N_p(0, \Sigma_{S,mc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,S,cc} - \hat{\theta}_N) &= \hat{\mathbb{G}}^{\pi,S,cc}_N \tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow Z_{S,cc} &\sim N_p(0, \Sigma_{S,cc}), \end{aligned}$$

*in $P^*$-probability where*

$$\Sigma_S \equiv \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(\tilde{\ell}_0), \tag{4.15}$$

$$\Sigma_{S,c} \equiv \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}((I-Q_c)\tilde{\ell}_0), \tag{4.16}$$

$$\Sigma_{S,mc} \equiv \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}((I-Q_{mc})\tilde{\ell}_0), \tag{4.17}$$

$$\Sigma_{S,cc} \equiv \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}((I-Q_{cc})\tilde{\ell}_0), \tag{4.18}$$

*and $Q_c$, $Q_{mc}$ and $Q_{cc}$ are defined in Theorem 3.2.1.*
*(2) Under Condition 4.1.4 with 4.1.4(a-2),*

$$\begin{array}{lclcccc}
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dc} - \hat{\theta}_{N,c}) & = & \hat{\mathbb{G}}_N^{\pi,S,dc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,c} & \sim & N_p(0,\Sigma_{S,c}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dmc} - \hat{\theta}_{N,mc}) & = & \hat{\mathbb{G}}_N^{\pi,S,dmc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,mc} & \sim & N_p(0,\Sigma_{S,mc}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dcc} - \hat{\theta}_{N,cc}) & = & \hat{\mathbb{G}}_N^{\pi,S,dcc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,cc} & \sim & N_p(0,\Sigma_{S,cc}),
\end{array}$$

*in $P^*$-probability.*

We prove the survey bootstrap version of Theorem 3.3.1. Note again that the asymptotic variances consist of the phase II variances only and that double calibrations have right centering of the bootstrap WLE's. To prove this theorem, we require the following condition.

**Condition 4.1.6** (Consistency and rate of convergence)**.** *(1) The bootstrap estimator $(\hat{\theta}_{N,S,\#}, \hat{\eta}_{N,S,\#})$ satisfies the bootstrap weighted likelihood equations*

$$\begin{aligned}
\Psi^{\pi}_{N,1,S,\#}(\theta,\eta,\alpha) &= \hat{\mathbb{P}}_N^{\pi,S,\#}\dot{\ell}_{\theta,\eta} = o_{P_W^*}\left(N^{-1/2}\right), \\
\Psi^{\pi}_{N,2,S,\#}(\theta,\eta,\alpha)\left[\underline{h}^*\right] &= \hat{\mathbb{P}}_N^{\pi,S,\#}B_{\theta,\eta}[\underline{h}^*] = o_{P_W^*}\left(N^{-1/2}\right),
\end{aligned} \tag{4.19}$$

*in $P^*$-probability where $\#$ is null for the case corresponding to the plain bootstrap WLE or $\# \in \{c, mc, cc, dc, dmc, dcc\}$.*
*(2) The estimator $(\widehat{\theta}_{N,S,\#}, \widehat{\eta}_{N,S,\#})$ of $(\theta_0,\eta_0)$ satisfies $|\widehat{\theta}_{N,S,\#} - \theta_0| = o_{P_W}(1)$, and $\|\widehat{\eta}_{N,S,\#} - \eta_0\| = O_{P_W}(N^{-\beta})$ for some $\beta > 0$ in $P^*$-probability where $\#$ is null for the case corresponding to the plain bootstrap WLE or $\# \in \{c, mc, cc, dc, dmc, dcc\}$.*

**Theorem 4.1.3.** *Under Conditions 3.3.1-3.3.4, 4.1.4 and 4.1.6,*

$$\begin{array}{lllllll}
\sqrt{N}(\hat{\hat{\theta}}_{N,S} - \hat{\theta}_N) & = & \hat{\mathbb{G}}_N^{\pi,S}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_S & \sim & N_p(0, \Sigma_S), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,c} - \hat{\theta}_N) & = & \hat{\mathbb{G}}_N^{\pi,S,c}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,c} & \sim & N_p(0, \Sigma_{S,c}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,mc} - \hat{\theta}_N) & = & \hat{\mathbb{G}}_N^{\pi,S,mc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,mc} & \sim & N_p(0, \Sigma_{S,mc}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,cc} - \hat{\theta}_N) & = & \hat{\mathbb{G}}_N^{\pi,S,cc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,cc} & \sim & N_p(0, \Sigma_{S,cc}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dc} - \hat{\theta}_{N,c}) & = & \hat{\mathbb{G}}_N^{\pi,S,dc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,c} & \sim & N_p(0, \Sigma_{S,c}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dmc} - \hat{\theta}_{N,mc}) & = & \hat{\mathbb{G}}_N^{\pi,S,dmc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,mc} & \sim & N_p(0, \Sigma_{S,mc}), \\
\sqrt{N}(\hat{\hat{\theta}}_{N,S,dcc} - \hat{\theta}_{N,cc}) & = & \hat{\mathbb{G}}_N^{\pi,S,dcc}\tilde{\ell}_0 + o_{P_W^*}(1) & \rightsquigarrow & Z_{S,cc} & \sim & N_p(0, \Sigma_{S,cc})
\end{array}$$

*in $P^*$-probability where $I_0$ and $\tilde{\ell}_0$ are defined in Condition 3.3.2.*

## 4.2 Simultaneous Estimation for the Phase I and II Variances

### 4.2.1 Two-phase Bootstrap

We introduce the *two-phase bootstrap*, which yields both phase I and II variances at the same time. The bootstrap weights we propose consist of two distinct parts, the first part capturing the phase I variance and the second part grasping the phase II variances. The first part, the *phase I bootstrap weights*, are i.i.d. weights, used in the weighted bootstrap [31], that reproduce randomness due to sampling from a population. The second part, the *phase II bootstrap weights*, are exchangeable weights from the survey bootstrap, which is already justified above for the phase II variances. The two-phase bootstrap uses the product of these two types weights as its bootstrap weights to simultaneously yield phase I and II variances.

For each $j = 1, \ldots, J$, let the phase I bootstrap weights $W^{(1)}_{n_j,j,i}$, $i = 1, \ldots, N_j$, for the $j$th stratum be i.i.d. $P^{(1)}_{W_j}$ satisfying $0 \leq W^{(1)}_{n_j,j,i} \leq M < \infty$ for some $M > 0$,

$$EW^{(1)}_{n_j,j,i} = 1, \quad \mathrm{Var}(W^{(1)}_{n_j,j,i}) = p_j/(2 - p_j) \equiv c_j^2.$$

We assume that the phase I bootstrap weights are independent across strata and satisfy $P(W^{(1)}_{n_j,j,i} = 0) = 0$ for $j = 1, \ldots, J$, $i = 1, \ldots, N_j$. The phase II bootstrap weights $W^{(2)}_{n_j,j,i}$ are defined in the same way as survey bootstrap weights, which are independent of the phase I bootstrap weights. Define the two-phase bootstrap weights $W_{n_j,j,i} \equiv W^{(1)}_{n_j,j,i} W^{(2)}_{n_j,j,i}$. We denote $W_{Ni}$, $W^{(1)}_{Ni}$, and $W^{(2)}_{Ni}$ for $W_{n_j,j,i}$, $W^{(1)}_{n_j,j,i}$, and $W^{(2)}_{n_j,j,i}$, respectively in the same way as other notations when we do not specify the stratum where the observation belongs. We also denote the probability measures for the phase I and II bootstrap weights by $P^{(1)}_W \equiv \prod_{j=1}^J P^{(1)}_{W_j}$ and $P^{(2)}_W$ and define $P_W \equiv P^{(1)}_W \times P^{(2)}_W$.

Define the *two-phase bootstrap IPW empirical measure* by

$$\begin{aligned}\hat{\mathbb{P}}^\pi_N &\equiv \frac{1}{N}\sum_{i=1}^N W_{N,i}\frac{\xi_i}{\pi_0(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{j=1}^J \frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}W^{(2)}_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \\ &\equiv \frac{1}{N}\sum_{j=1}^J N_j\hat{\mathbb{P}}^\xi_{j,n_j},\end{aligned}$$

where the bootstrap phase II empirical measure for the $j$th stratum is given by

$$\hat{\mathbb{P}}^\xi_{j,n_j} \equiv \frac{1}{n_j}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}W^{(2)}_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \quad j=1,\ldots,J,$$

and define the *two-phase bootstrap IPW empirical process* by

$$\hat{\mathbb{G}}^\pi_N \equiv \sqrt{N}(\hat{\mathbb{P}}^\pi_N - \mathbb{P}^\pi_N).$$

We decompose the two-phase bootstrap IPW empirical process;

$$\begin{aligned}\hat{\mathbb{G}}^\pi_N &= \sqrt{N}(\hat{\mathbb{P}}^\pi_N - \mathbb{P}^\pi_N) \\ &= \sqrt{N}(\hat{\mathbb{P}}^{\pi,(1)}_N - \mathbb{P}^\pi_N) + \sqrt{N}(\hat{\mathbb{P}}^\pi_N - \hat{\mathbb{P}}^{\pi,(1)}_N) \\ &= \hat{\mathbb{G}}^{\pi,(1)}_N + \hat{\mathbb{G}}^{\pi,(2)}_N,\end{aligned}$$

where $\hat{\mathbb{G}}^{\pi,(1)}_N \equiv \sqrt{N}(\hat{\mathbb{P}}^{\pi,(1)}_N - \mathbb{P}^\pi_N) = \sqrt{N}\mathbb{P}^\pi_N(W^{(1)}_N - 1)\cdot$ is the *phase I bootstrap IPW empirical process*, $\hat{\mathbb{G}}^{\pi,(2)}_N \equiv \sqrt{N}(\hat{\mathbb{P}}^\pi_N - \hat{\mathbb{P}}^{\pi,(1)}_N) = \sqrt{N}\mathbb{P}^\pi_N W^{(1)}_N(W^{(2)}_N - 1)\cdot$ is the *phase II bootstrap IPW empirical process*, and the *phase I bootstrap IPW empirical measure* is given by

$$\begin{aligned}\hat{\mathbb{P}}^{\pi,(1)}_N &\equiv \frac{1}{N}\sum_{i=1}^N W^{(1)}_{Ni}\frac{\xi_i}{\pi_0(V_i)}\delta_{X_i} = \frac{1}{N}\sum_{j=1}^J \frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \\ &\equiv \sum_{j=1}^J \frac{N_j}{N}\hat{\mathbb{P}}^{\xi,(1)}_{j,n_j},\end{aligned}$$

Here the bootstrap phase I empirical measure for the $j$th stratum is given by

$$\hat{\mathbb{P}}^{\xi,(1)}_{j,n_j} \equiv \frac{1}{n_j}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \quad j=1,\ldots,J.$$

As their names suggest, the phase I and II bootstrap IPW empirical processes yield the phase I and II variances, respectively.

We further decompose the phase I bootstrap IPW empirical process as in the decompositions (2.1) and (4.7) of the (survey bootstrap) IPW empirical processes;

$$
\begin{aligned}
\hat{\mathbb{G}}_N^{\pi,(1)} &= \sqrt{N}\left(\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^{\pi}\right) = \sum_{j=1}^{J} \frac{N_j}{\sqrt{N}}\left(\hat{\mathbb{P}}_{j,n_j}^{\xi,(1)} - \mathbb{P}_{j,n_j}^{\xi}\right) \\
&= \sum_{j=1}^{J} \sqrt{\frac{N_j}{N}} \sqrt{\frac{N_j}{n_j}} \hat{\mathbb{G}}_{j,n_j}^{\xi,(1)}, \qquad (4.20)
\end{aligned}
$$

where for each $j = 1, \ldots, J$, $\hat{\mathbb{G}}_{j,n_j}^{\xi,(1)} \equiv \sqrt{n_j}(\hat{\mathbb{P}}_{j,N_j}^{\xi,(1)} - \mathbb{P}_{j,N_j}^{\xi})$ is the phase I bootstrap IPW empirical processes for the $j$th stratum with $\mathbb{P}_{j,n_j}^{\xi} \equiv (N_j/n_j)\mathbb{P}_{j,N_j}^{\xi}$. Also, we decompose the phase II bootstrap IPW empirical process;

$$
\begin{aligned}
\hat{\mathbb{G}}_N^{\pi,(2)} &= \sqrt{N}\left(\hat{\mathbb{P}}_N^{\pi} - \hat{\mathbb{P}}_N^{\pi,(1)}\right) = \sum_{j=1}^{J} \frac{N_j}{\sqrt{N}}\left(\hat{\mathbb{P}}_{j,n_j}^{\xi} - \hat{\mathbb{P}}_{j,n_j}^{\xi,(1)}\right) \\
&= \sum_{j=1}^{J} \sqrt{\frac{N_j}{N}} \sqrt{\frac{N_j}{n_j}} \hat{\mathbb{G}}_{j,n_j}^{\xi,(2)}, \qquad (4.21)
\end{aligned}
$$

where for each $j = 1, \ldots, J$, $\hat{\mathbb{G}}_{j,n_j}^{\xi,(2)} \equiv \sqrt{n_j}(\hat{\mathbb{P}}_{j,n_j}^{\xi} - \hat{\mathbb{P}}_{j,n_j}^{\xi,(1)})$ is the phase II bootstrap IPW empirical process for the $j$th stratum. These decompositions are used for establishing weak convergence of two-phase bootstrap IPW empirical process.

We also introduce several definitions and decompositions for calibrations. Define the two-phase bootstrap (doubly) calibrated bootstrap IPW empirical measures with (modified and centered) calibrations by

$$\begin{aligned}
\hat{\mathbb{P}}_N^{\pi,c} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_c(V;\hat{\hat{\alpha}}_N^c) f = \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_c(V_{j,i};\hat{\hat{\alpha}}_N^c) f(X_{j,i}),\\
\hat{\mathbb{P}}_N^{\pi,dc} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_c(V;\hat{\hat{\alpha}}_N^{dc}) G_c(V;\hat{\alpha}_N^c) f\\
&= \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_c(V_{j,i};\hat{\hat{\alpha}}_N^{dc}) G_c(V_{j,i};\hat{\hat{\alpha}}_N^c) f(X_{j,i}),\\
\hat{\mathbb{P}}_N^{\pi,mc} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_{mc}(V;\hat{\hat{\alpha}}_N^{mc}) f = \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_{mc}(V_{j,i};\hat{\hat{\alpha}}_N^{mc}) f(X_{j,i}),\\
\hat{\mathbb{P}}_N^{\pi,dmc} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_{mc}(V;\hat{\hat{\alpha}}_N^{dmc}) G_c(V;\hat{\alpha}_N^{mc}) f\\
&= \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_{mc}(V_{j,i};\hat{\hat{\alpha}}_N^{dmc}) G_{mc}(V_{j,i};\hat{\alpha}_N^{mc}) f(X_{j,i}),\\
\hat{\mathbb{P}}_N^{\pi,cc} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_{S,cc}(V;\hat{\hat{\alpha}}_N^{cc}) f = \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_{S,cc}(V_{j,i};\hat{\hat{\alpha}}_N^{cc}) f(X_{j,i}),\\
\hat{\mathbb{P}}_N^{\pi,dcc} f &\equiv \hat{\mathbb{P}}_N^{\pi} G_{S,cc}(V;\hat{\hat{\alpha}}_N^{dcc}) G_{cc}(V;\hat{\alpha}_N^{cc}) f\\
&= \frac{1}{N}\sum_{j=1}^{J}\frac{1}{n_j/N_j}\sum_{i=1}^{N_j} W_{n_j,j,i}^{(1)} W_{n_j,j,i}^{(2)} \xi_{j,i} G_{S,cc}(V_{j,i};\hat{\hat{\alpha}}_N^{dcc}) G_{cc}(V_{j,i};\hat{\alpha}_N^{cc}) f(X_{j,i}),
\end{aligned}$$

the two-phase bootstrap calibrated bootstrap IPW empirical processes with (modified and centered) calibrations by

$$\hat{\mathbb{G}}_N^{\pi,\#} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,\#} - \mathbb{P}_N^{\pi}),$$

with $\# \in \{c, mc, cc\}$ and the two-phase bootstrap doubly calibrated bootstrap IPW empirical processes with (modified and centered) calibrations by

$$\hat{\mathbb{G}}_N^{\pi,\#} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,\#} - \mathbb{P}_N^{\pi,\#}),$$

with $\# \in \{dc, dmc, dcc\}$. Here $\hat{\alpha}_N^{\#}$ with $\{c, mc, cc\}$ are obtained from (2.4)-(2.6) and $\hat{\hat{\alpha}}_N^{\#}$ with $\{c, dc, mc, dmc, cc, dcc\}$ are obtained from (4.8)-(4.13).

We decompose the two-phase bootstrap calibrated IPW empirical processes (with modified and centered calibration);

$$\begin{aligned}\hat{\mathbb{G}}_N^{\pi,\#} &= \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,\#} - \mathbb{P}_N^{\pi}) \\ &= \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^{\pi}) + \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,\#} - \hat{\mathbb{P}}_N^{\pi,(1)}) \\ &= \hat{\mathbb{G}}_N^{\pi,(1)} + \hat{\mathbb{G}}_N^{\pi,(2),\#},\end{aligned}$$

where $\# \in \{c, mc, cc\}$, $\hat{\mathbb{G}}_N^{\pi,(1)}$ is the phase I bootstrap IPW empirical process defined above, $\hat{\mathbb{G}}_N^{\pi,(2),\#} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,\#} - \hat{\mathbb{P}}_N^{\pi,(1)})$ is the phase II bootstrap calibrated IPW empirical process (with modified and centered calibration).

We decompose the two-phase bootstrap doubly calibrated IPW empirical processes (with modified and centered calibration);

$$\begin{aligned}\hat{\mathbb{G}}_N^{\pi,d\#} &= \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,d\#} - \mathbb{P}_N^{\pi,\#}) \\ &= \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,(1),d\#} - \mathbb{P}_N^{\pi,\#}) + \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,d\#} - \hat{\mathbb{P}}_N^{\pi,(1),d\#}) \\ &= \hat{\mathbb{G}}_N^{\pi,(1),d\#} + \hat{\mathbb{G}}_N^{\pi,(2),d\#},\end{aligned}$$

where $\# \in \{c, mc, cc\}$, $\hat{\mathbb{G}}_N^{\pi,(1),d\#} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,(1),d\#} - \mathbb{P}_N^{\pi,\#})$ is the phase I bootstrap doubly calibrated IPW empirical process (with modified and centered calibration), $\hat{\mathbb{G}}_N^{\pi,(2),d\#} \equiv \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,d\#} - \hat{\mathbb{P}}_N^{\pi,(1),d\#})$ is the phase II bootstrap doubly calibrated IPW empirical process (with modified and centered calibration), and the phase I bootstrap doubly calibrated IPW empirical measure (with modified and centered calibration) is given by

$$\begin{aligned}\hat{\mathbb{P}}_N^{\pi,(1),dc} &= \frac{1}{N}\sum_{i=1}^N W_{ni}^{(1)} \frac{\xi_i G_c(V_i;\hat{\alpha}_N^c)}{\pi_0(V_i)}\delta_{X_i}, \\ \hat{\mathbb{P}}_N^{\pi,(1),dmc} &= \frac{1}{N}\sum_{i=1}^N W_{ni}^{(1)} \frac{\xi_i) G_{mc}(V_i;\hat{\alpha}_N^c)}{\pi_0(V_i)}\delta_{X_i}, \\ \hat{\mathbb{P}}_N^{\pi,(1),dcc} &= \frac{1}{N}\sum_{i=1}^N W_{ni}^{(1)} \frac{\xi_i G_{cc}(V_i;\hat{\alpha}_N^{cc})}{\pi_0(V_i)}\delta_{X_i},\end{aligned}$$

### *4.2.2 Two-phase Bootstrap Z-theorems*

We prove the two-phase bootstrap version of Theorem 3.2.1. Now, the asymptotic variances of the bootstrap WLE's are the same as in the original sample. The issue of centering seen in Theorems 4.1.2 and 4.1.3 remain in this theorem. We assume the following condition.

**Condition 4.2.1** (Consistency)**.** *The bootstrap estimator* $(\hat{\hat{\theta}}_N, \hat{\hat{\eta}}_N)$ *is consistent for* $(\theta_0, \eta_0)$ *in* $P^*$*-probability and solves the bootstrap weighted likelihood equations*

$$\begin{aligned} \hat{\Psi}^{\pi}_{N,1}(\theta,\eta) &= \hat{\mathbb{P}}^{\pi}_N \dot{\ell}_{\theta,\eta} = o_{P^*_W}\left(N^{-1/2}\right), \\ \left\| \hat{\Psi}^{\pi}_{N,2}(\theta,\eta)h \right\|_{\mathcal{H}} &= \left\| \hat{\mathbb{P}}^{\pi}_N (B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h) \right\|_{\mathcal{H}} = o_{P^*_W}\left(N^{-1/2}\right), \end{aligned} \tag{4.22}$$

*in* $P^*$*-probability where* $\hat{\mathbb{P}}^{\pi}_N$ *may be replaced by* $\hat{\mathbb{P}}^{\pi,c}_N$, $\hat{\mathbb{P}}^{\pi,mc}_N$, $\hat{\mathbb{P}}^{\pi,cc}_N$ $\hat{\mathbb{P}}^{\pi,dc}_N$, $\hat{\mathbb{P}}^{\pi,dmc}_N$ *or* $\hat{\mathbb{P}}^{\pi,dcc}_N$ *for the estimators with calibration, modified calibration, centered calibration, double calibration, double modified calibration, double centered calibration with the corresponding estimators denoted as* $(\hat{\hat{\theta}}_{N,c}, \hat{\hat{\eta}}_{N,c})$, $(\hat{\hat{\theta}}_{N,mc}, \hat{\hat{\eta}}_{N,mc})$, $(\hat{\hat{\theta}}_{N,cc}, \hat{\hat{\eta}}_{N,cc})$, $(\hat{\hat{\theta}}_{N,dc}, \hat{\hat{\eta}}_{N,dc})$, $(\hat{\hat{\theta}}_{N,dmc}, \hat{\hat{\eta}}_{N,dmc})$, *and* $(\hat{\hat{\theta}}_{N,dcc}, \hat{\hat{\eta}}_{N,dcc})$, *respectively.*

**Theorem 4.2.1.** *Suppose that Conditions 3.2.2, 3.2.3 and 4.2.1 hold.*

*(1) Under Condition 4.1.4 with 4.1.4(a-1),*

$$\begin{array}{lllllll} \sqrt{N}(\hat{\hat{\theta}}_N - \hat{\theta}_N) & = & \hat{\mathbb{G}}^{\pi}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z & \sim & N_p(0,\Sigma), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,c} - \hat{\theta}_N) & = & \hat{\mathbb{G}}^{\pi,c}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_c & \sim & N_p(0,\Sigma_c), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,mc} - \hat{\theta}_N) & = & \hat{\mathbb{G}}^{\pi,mc}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_{mc} & \sim & N_p(0,\Sigma_{mc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,cc} - \hat{\theta}_N) & = & \hat{\mathbb{G}}^{\pi,cc}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_{cc} & \sim & N_p(0,\Sigma_{cc}), \end{array}$$

*in* $P^*$*-probability where* $Q_c$, $Q_{mc}$ *and* $Q_{cc}$ *are defined in Theorem 3.2.1.*

*(2) Under Condition 4.1.4 with 4.1.4(a-2),*

$$\begin{array}{lllllll} \sqrt{N}(\hat{\hat{\theta}}_{N,dc} - \hat{\theta}_{N,c}) & = & \hat{\mathbb{G}}^{\pi,dc}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_c & \sim & N_p(0,\Sigma_c), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,dmc} - \hat{\theta}_{N,mc}) & = & \hat{\mathbb{G}}^{\pi,dmc}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_{mc} & \sim & N_p(0,\Sigma_{mc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,dcc} - \hat{\theta}_{N,cc}) & = & \hat{\mathbb{G}}^{\pi,dcc}_N \tilde{\ell}_0 + o_{P^*_W}(1) & \rightsquigarrow & Z_{cc} & \sim & N_p(0,\Sigma_{cc}), \end{array}$$

*in* $P^*$*-probability.*

We prove the two-phase bootstrap version of Theorem 3.3.1. Note again that the asymptotic variances for the bootstrap WLE's are the same as in Theorem 3.3.1 and that the issue of centering is seen in this theorem, too. To this end, we require the following condition.

**Condition 4.2.2** (Consistency and rate of convergence)**.** *(1) The bootstrap estimator* $(\hat{\theta}_{N,\#}, \hat{\eta}_{N,\#})$ *satisfies the bootstrap weighted likelihood equations*

$$\begin{aligned} \Psi^{\pi}_{N,1,\#}(\theta,\eta,\alpha) &= \hat{\mathbb{P}}^{\pi,\#}_{N}\dot{\ell}_{\theta,\eta} = o_{P^*_W}\left(N^{-1/2}\right), \\ \Psi^{\pi}_{N,2,\#}(\theta,\eta,\alpha)\,[\underline{h}^*] &= \hat{\mathbb{P}}^{\pi,\#}_{N}B_{\theta,\eta}[\underline{h}^*] = o_{P^*_W}\left(N^{-1/2}\right), \end{aligned} \tag{4.23}$$

*in* $P^*$*-probability where* $\#$ *is null for the case corresponding to the plain bootstrap WLE or* $\# \in \{c, mc, cc, dc, dmc, dcc\}$*.*
*(2) These estimator* $(\widehat{\theta}_{N,\#}, \widehat{\eta}_{N,\#})$ *of* $(\theta_0, \eta_0)$ *satisfies* $|\widehat{\theta}_{N,\#} - \theta_0| = o_{P_W}(1)$*, and* $\|\widehat{\eta}_{N,\#} - \eta_0\| = O_{P_W}(N^{-\beta})$ *for some* $\beta > 0$ *in* $P^*$*-probability where* $\#$ *is null for the case corresponding to the plain bootstrap WLE or* $\# \in \{c, mc, cc, dc, dmc, dcc\}$*.*

**Theorem 4.2.2.** *Under Conditions 3.3.1-3.3.4, 4.1.4 and 4.2.2,*

$$\begin{array}{lclcclcl} \sqrt{N}(\hat{\hat{\theta}}_N - \hat{\theta}_N) &=& \hat{\mathbb{G}}^{\pi}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z &\sim& N_p(0,\Sigma), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,c} - \hat{\theta}_N) &=& \hat{\mathbb{G}}^{\pi,c}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_c &\sim& N_p(0,\Sigma_c), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,mc} - \hat{\theta}_N) &=& \hat{\mathbb{G}}^{\pi,mc}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_{mc} &\sim& N_p(0,\Sigma_{mc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,cc} - \hat{\theta}_N) &=& \hat{\mathbb{G}}^{\pi,cc}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_{cc} &\sim& N_p(0,\Sigma_{cc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,dc} - \hat{\theta}_{N,c}) &=& \hat{\mathbb{G}}^{\pi,dc}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_c &\sim& N_p(0,\Sigma_c), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,dmc} - \hat{\theta}_{N,mc}) &=& \hat{\mathbb{G}}^{\pi,dmc}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_{mc} &\sim& N_p(0,\Sigma_{mc}), \\ \sqrt{N}(\hat{\hat{\theta}}_{N,dcc} - \hat{\theta}_{N,cc}) &=& \hat{\mathbb{G}}^{\pi,dcc}_N\tilde{\ell}_0 + o_{P^*_W}(1) &\rightsquigarrow& Z_{cc} &\sim& N_p(0,\Sigma_{cc}) \end{array}$$

*in* $P^*$*-probability where* $I_0$ *and* $\tilde{\ell}_0$ *are defined in Condition 3.3.2.*

## *4.3 Proofs*

*Proof of Theorem 4.1.1.* For $\overline{\theta}_N = \hat{\theta}_N + h_N v_N$, we have by (4.2)

$$\begin{aligned} &\mathbb{M}^{\pi}_N(\overline{\theta}_N) - \mathbb{M}^{\pi}_N(\hat{\theta}_N) = \mathbb{P}^{\pi}_N \log \mathrm{lik}(\overline{\theta}_N, \hat{\eta}_{\overline{\theta}_N,N}) - \mathbb{P}^{\pi}_N \log \mathrm{lik}(\hat{\theta}_N, \hat{\eta}_{\hat{\theta}_N,N}) \\ &\begin{cases} \geq \mathbb{P}^{\pi}_N \log \mathrm{lik}\left(\overline{\theta}_N, \underline{\eta}_{\overline{\theta}_N}(\hat{\psi}_{\hat{\theta}_N,N})\right) - \mathbb{P}^{\pi}_N \log \mathrm{lik}\left(\hat{\theta}_N, \underline{\eta}_{\hat{\theta}_N}(\hat{\psi}_{\hat{\theta}_N,N})\right) \\ \leq \mathbb{P}^{\pi}_N \log \mathrm{lik}\left(\overline{\theta}_N, \underline{\eta}_{\overline{\theta}_N}(\hat{\psi}_{\overline{\theta}_N,N})\right) - \mathbb{P}^{\pi}_N \log \mathrm{lik}\left(\hat{\theta}_N, \underline{\eta}_{\hat{\theta}_N}(\hat{\psi}_{\overline{\theta}_N,N})\right) \end{cases} \end{aligned}$$

Both the upper and the lower bounds are differences $\mathbb{P}^{\pi}_N\ell(\overline{\theta}_N, \psi) - \mathbb{P}^{\pi}_N\ell(\hat{\theta}_N, \psi)$ with $\psi = \hat{\psi}_{\overline{\theta}_N,N}$ and $\psi = \hat{\psi}_{\hat{\theta}_N,N}$, respectively. We apply a two-term Taylor expansion to these differences leaving $\psi$ fixed.

For the lower bound, we expand around $\hat{\theta}_N$ and obtain that this is equal to

$$h_N v_N^T \mathbb{P}_N \dot{\ell}(\hat{\theta}_N, \hat{\psi}_{\hat{\theta}_N,N}) + \frac{1}{2} h_N^2 v_N^T \mathbb{P}_N^\pi \ddot{\ell}(\tilde{\theta}_N, \hat{\psi}_{\hat{\theta}_N,N}) v_N,$$

for $\tilde{\theta}_N$, a convex combination of $\overline{\theta}_N$ and $\hat{\theta}_N$. The first term is zero because the map $t \mapsto \mathbb{P}_N^\pi \log \text{lik}(t, \underline{\eta}_t(\hat{\psi}_{\hat{\theta}_N,N}))$ is maximized at $t = \hat{\theta}_N$. The second term is

$$-\frac{1}{2} h_N^2 (v_N^T \tilde{I}_0 v_N + o_P(1)),$$

by (4.4).

For the upper bound, we expand around $\overline{\theta}_N$ and obtain that this is equal to

$$h_N v_N^T \mathbb{P}_N^\pi \dot{\ell}(\overline{\theta}_N, \hat{\psi}_{\overline{\theta}_N,N}) - \frac{1}{2} h_N^2 v_N^T \mathbb{P}_N^\pi \ddot{\ell}(\tilde{\theta}_N, \hat{\psi}_{\overline{\theta}_N,N}) v_N$$

for $\tilde{\theta}_N$, a convex combination of $\overline{\theta}_N$ and $\hat{\theta}_N$. The second term is

$$\frac{1}{2} h_N^2 (v_N^T \tilde{I}_0 v_N + o_P(1)),$$

by (4.4). The first term is equal to

$$\begin{aligned}
&\frac{h_N}{\sqrt{N}} v_N^T \mathbb{G}_N^\pi \dot{\ell}(\overline{\theta}_N, \hat{\psi}_{\overline{\theta}_N,N}) + h_N v_N^T P_0 \dot{\ell}(\overline{\theta}_N, \hat{\psi}_{\overline{\theta}_N,N}) \\
&= \frac{h_N}{\sqrt{N}} (v_N^T \tilde{I}_0 \sqrt{N}(\hat{\theta}_N - \theta_0) + o_P(1)) - h_N \left\{ v_N^T \tilde{I}_0 (\overline{\theta}_N - \theta_0) + o_P(\|\overline{\theta}_N - \theta_0\| + N^{-1/2}) \right\},
\end{aligned}$$

by (4.1), (4.3) and (4.5). This reduces to

$$-h_N^2 (v_N^T \tilde{I}_0 v_N + o_P(1))$$

by the assumption of $h_N$ and the definition of $\overline{\theta}_N$. □

*Proof of Lemma 4.1.1.* The first two statements are proved in [60].

We prove the third statement. Let $\eta > 0$ be arbitrary. Since $\Delta_N = O_{Pr}(1)$, there exists a sequence $M_N$ such that $Pr(|\Delta_N| > M_N) \to 0$ as $N \to \infty$. It follows from Markov's inequality and Fubini's theorem that

$$P^* \left\{ P_W^*(|\Delta_N| > M_N) > \eta \right\} \leq \eta^{-1} E^* P_W^*(|\Delta_N| > M_N) \leq Pr^*(|\Delta_N| > M_N) \to 0.$$

We prove the fourth statement. Let $\eta > 0$ be arbitrary. Since $\Delta_N = O_{P_W}(1)$ in $P^*$-probability, there exists $M_N > 0$ possibly depending on $\eta$ such that $P^*\{P^*_W(|\Delta_N| > M_N) > \eta\} \to 0$ as $N \to \infty$. Then it follows from Fubini's theorem that

$$\begin{aligned}
Pr^*(|\Delta_N| > M_N) &= E^*\{P^*_W(|\Delta_N| > M_N)\} \\
&= E^*[P^*_W(|\Delta_N| > M_N)I\{P^*_W(|\Delta_N| > M_N) > \eta\}] \\
&\quad +E^*[P^*_W(|\Delta_N| > M_N)I\{P^*_W(|\Delta_N| > M_N) \leq \eta\}] \\
&\leq E^*[P^*_W(|\Delta_N| > M_N)I\{P^*_W(|\Delta_N| > M_N) > \eta\}] + \eta \\
&\leq E^*I\{P^*_W(|\Delta_N| > M_N) > \eta\} + \eta \\
&\leq P^*\{P^*_W(|\Delta_N| > M_N) > \eta\} + \eta \to \eta,
\end{aligned}$$

as $N \to \infty$. Since $\eta$ is arbitrary, this completes the proof. □

The following proposition is the bootstrap version of Proposition 3.5.1.

**Proposition 4.3.1.** *Suppose that Condition 4.1.4 holds. (i) Then* $|\hat{\alpha}^{\#}_N - \alpha_0| \to_{P_W} 0$ *with*

*$\# \in \{c, mc, cc\}$ in $P^*$-probability, and*

$$\begin{aligned}
&\sqrt{N}(\hat{\hat{\alpha}}_N^{c} - \alpha_0) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \dot{G}(0)^{-1} \left\{P_0 Z^{\otimes 2}\right\}^{-1} \frac{\xi_i}{\pi_0(V_0)} (W_{Ni} - 1) Z_i + o_{P*_W}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 Z^{\otimes 2}\right\}^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1 - p_j}{p_j}} \mathbb{G}_j Z, \\
&\sqrt{N}(\hat{\hat{\alpha}}_N^{mc} - \alpha_0) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \dot{G}(0)^{-1} \left\{P_0 \frac{1 - \pi_0(V)}{\pi_0(V)} Z^{\otimes 2}\right\}^{-1} \frac{\xi_i}{\pi_0(V_0)} (W_{Ni} - 1) Z_i + o_{P*_W}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 \frac{1 - \pi_0(V)}{\pi_0(V)} Z^{\otimes 2}\right\}^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1 - p_j}{p_j}} \mathbb{G}_j Z, \\
&\sqrt{N}(\hat{\hat{\alpha}}_N^{cc} - \alpha_0) \\
&= -\dot{G}(0)^{-1} \left\{P_0 \frac{1 - \pi_0(V)}{\pi_0(V)} (Z - \mu_Z)^{\otimes 2}\right\}^{-1} \\
&\quad \times \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\xi_i}{\pi_0(V_0)} (W_{Ni} - 1)(Z_i - \mu_Z) + o_{P*_W}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 \frac{1 - \pi_0(V)}{\pi_0(V)} (Z - \mu_Z)^{\otimes 2}\right\}^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1 - p_j}{p_j}} \mathbb{G}_j (Z - \mu_Z),
\end{aligned}$$

*in $P^*$-probability where the $P_{0|j}$-Brownian bridge processes, $\mathbb{G}_j$, are independent.*
*(ii)The same conclusion holds if we replace $\hat{\hat{\alpha}}_N^{c}$, $\hat{\hat{\alpha}}_N^{mc}$, $\hat{\hat{\alpha}}_N^{cc}$ by $\hat{\hat{\alpha}}_N^{dc}$, $\hat{\hat{\alpha}}_N^{dmc}$, and $\hat{\hat{\alpha}}_N^{dcc}$, respectively.*

*Proof.* First we consider modified calibration with $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{mc}$ obtained as the solution to the equation (4.9). The cases for (centered) calibration (i.e., $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{c}$ and $\hat{\hat{\alpha}}_N = \hat{\hat{\alpha}}_N^{cc}$) are similar.

Define $\hat{\Phi}_{N,mc}(\alpha) \equiv \hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\alpha) Z - \mathbb{P}_N^{\pi} Z$ and $\Phi_{mc}(\alpha) \equiv P_0[(G_{mc}(V;\alpha) - 1)Z]$. Note that $\hat{\Phi}_{N,mc}(\hat{\hat{\alpha}}_N) = 0$ and $\Psi_{mc}(0) = 0$. We apply Theorem 5.7 of [56] for a consistency proof.

For the first condition of the theorem, we have

$$\begin{aligned}
&\sup_{\alpha\in\mathbb{R}^k}\left\|\hat{\Phi}_{N,mc}(\alpha)-\Phi_{mc}(\alpha)\right\| \\
&=\sup_{\alpha\in\mathbb{R}^k}\left\|(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\alpha)Z+\mathbb{P}_N^{\pi}\left\{G_{mc}(V;\alpha)-1\right\}Z-P_0\left\{G_{mc}(V;\alpha)-1\right\}Z\right\| \\
&\le\sup_{\alpha\in\mathbb{R}^k}\left\|(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\alpha)Z\right\|+\sup_{\alpha\in\mathbb{R}^k}\left\|(\mathbb{P}_N^{\pi}-P_0)\left\{G_{mc}(V;\alpha)-1\right\}Z\right\|,
\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm. Applying Theorems 7.1.1 and 6.1.1 to the first and second terms, respectively, together with Lemma 4.1.1 yields the above display is $o_{P_W^*}(1)$ in $P^*$-probability. This established the first condition of Theorem 5.7 of [56]. The second condition of Theorem 5.7 of [56] was verified in the proof of Proposition 3.5.1 and hence $\hat{\hat{\alpha}}_N\to_{P_W^*}\alpha_0$ in $P^*$-probability.

We apply Theorem 3.3.1 of [58] to show the asymptotic normality of $\hat{\hat{\alpha}}_N$. For the asymptotic equicontinuity condition, note that

$$\begin{aligned}
\hat{\Phi}_{N,mc}(\alpha_1)-\hat{\Phi}_{N,mc}(\alpha_2)&=\hat{\mathbb{P}}_N^{\pi,S}\{G_{mc}(V;\alpha_1)-G_{mc}(V;\alpha_2)\}Z,\\
\Phi_{mc}(\alpha_1)-\Phi_{mc}(\alpha_2)&=P_0\{G_{mc}(V;\alpha_1)-G_{mc}(V;\alpha_2)\}Z.
\end{aligned}$$

Thus, it follows by Taylor's theorem that

$$\begin{aligned}
&\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\hat{\hat{\alpha}}_N)-\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\alpha_0)\\
&=\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})\{G_{mc}(V;\hat{\hat{\alpha}}_N)-1\}Z+\sqrt{N}(\mathbb{P}_N^{\pi}-P_0)\{G_{mc}(V;\hat{\hat{\alpha}}_N)-1\}Z\\
&=(\hat{\mathbb{G}}_N^{\pi,S}+\mathbb{G}_N^{\pi})\{G_{mc}(V;\hat{\hat{\alpha}}_N)-1\}Z\\
&=(\hat{\mathbb{G}}_N^{\pi,S}+\mathbb{G}_N^{\pi})\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}(\hat{\hat{\alpha}}_N-\alpha_0)\\
&=\{(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})+(\mathbb{P}_N^{\pi}-P_0)\}\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\sqrt{N}(\hat{\hat{\alpha}}_N-\alpha_0),
\end{aligned}$$

where $\alpha^*$ is some convex combination of $\hat{\hat{\alpha}}_N$ and $\alpha_0$. It follows from Theorems 6.1.1 and 7.1.1 that this term is $o_P(1+\sqrt{N}|\hat{\hat{\alpha}}_N-\alpha_0)|)$. Next, we show the weak convergence of the process $\sqrt{N}(\hat{\Phi}_{N,mc}-\Phi_{mc})(\alpha)$ at $\alpha_0=0$. It follows from Theorem 7.3.1 that

$$\sqrt{N}(\hat{\Phi}_{N,mc}-\Phi_{mc})(\alpha_0)=\sqrt{N}\hat{\Phi}_{N,mc}(0)=\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})Z\rightsquigarrow\mathbb{G}^{\pi,S}Z,$$

in $P^*$-probability. Thus, by Theorem 3.3.1 of [58] we obtain

$$\begin{aligned}\sqrt{N}\hat{\alpha}_N &= -\dot{\Phi}_{mc}(0)\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(0)+o_{P_W^*}(1)\\ &\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}\right\}^{-1}\mathbb{G}^{\pi,S}Z\end{aligned}$$

in $P^*$-probability.

Next, we consider modified calibration with $\hat{\hat{\alpha}}_N=\hat{\hat{\alpha}}_N^{dmc}$ obtained as the solution to the equation (4.12). Define $\hat{\Phi}_{N,dmc}(\alpha)\equiv\hat{\mathbb{P}}_N^{\pi,S}G_{mc}(V;\alpha)G_{mc}(V;\hat{\alpha}_N)Z-\mathbb{P}_N^{\pi}G_{mc}(V;\hat{\alpha}_N)Z$ and $\Phi_{mc}(\alpha)\equiv P_0[(G_{mc}(V;\alpha)-1)Z]$. Note that $\hat{\Phi}_{N,dmc}(\hat{\hat{\alpha}}_N)=0$ and $\Psi_{mc}(0)=0$. We apply Theorem 5.7 of [56] for a consistency proof. For the first condition of the theorem, we have

$$\begin{aligned}&\sup_{\alpha\in\mathbb{R}^k}\left\|\hat{\Phi}_{N,dmc}(\alpha)-\Phi_{mc}(\alpha)\right\|\\ &=\sup_{\alpha\in\mathbb{R}^k}\Big\|(\hat{\mathbb{P}}_N^{\pi,S,dmc}-\mathbb{P}_N^{\pi,mc})Z-P_0\left\{G_{mc}(V;\alpha)-1\right\}Z\\ &\qquad -\mathbb{P}_N^{\pi,mc}G_{mc}(V;\alpha)Z+\mathbb{P}_N^{\pi,mc}G_{mc}(V;\alpha)Z\Big\|\\ &=\sup_{\alpha\in\mathbb{R}^k}\left\|(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\alpha)G_{mc}(V;\hat{\alpha}_N)Z+(\mathbb{P}_N^{\pi,mc}-P_0)\{G_{mc}(V;\alpha)-1\}Z\right\|\\ &\le\sup_{\alpha\in\mathbb{R}^k}\left\|(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\alpha)G_{mc}(V;\hat{\alpha}_N)Z\right\|+\sup_{\alpha\in\mathbb{R}^k}\left\|(\mathbb{P}_N^{\pi,mc}-P_0)\{G_{mc}(V;\alpha)-1\}Z\right\|.\end{aligned}$$

Applying Theorems 7.1.1 and 6.1.1 to the first and second terms, respectively, together with Lemma 4.1.1 yields the above display is $o_{P_W^*}(1)$ in $P^*$-probability. Since the second condition of Theorem 5.7 of [56] was verified in the proof of Proposition 3.5.1, $\hat{\hat{\alpha}}_N\to_{P_W^*}\alpha_0$ in $P^*$-probability.

We apply Theorem 3.3.1 of [58] to show the asymptotic normality of $\hat{\hat{\alpha}}_N$. For the asymptotic equicontinuity condition, note that

$$\hat{\Phi}_{N,dmc}(\alpha_1)-\hat{\Phi}_{N,dmc}(\alpha_2)=\hat{\mathbb{P}}_N^{\pi,S}G_{mc}(V;\hat{\alpha}_N)\{G_{mc}(V;\alpha_1)-G_{mc}(V;\alpha_2)\}Z.$$

Thus, it follows by Taylor's theorem that

$$
\begin{aligned}
&\sqrt{N}(\Phi_{N,dmc}-\Phi_{mc})(\hat{\hat{\alpha}}_N)-\sqrt{N}(\Phi_{N,dmc}-\Phi_{mc})(\alpha_0)\\
&=\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\hat{\alpha}_N)\{G_{mc}(V;\hat{\hat{\alpha}}_N)-1\}Z\\
&\quad+\sqrt{N}(\mathbb{P}_N^{\pi,mc}-P_0)\{G_{mc}(V;\hat{\hat{\alpha}}_N)-1\}Z\\
&=(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})G_{mc}(V;\hat{\alpha}_N)\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\sqrt{N}(\hat{\hat{\alpha}}_N-\alpha_0)\\
&\quad-(\mathbb{P}_N^{\pi,mc}-P_0)\}\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\sqrt{N}(\hat{\hat{\alpha}}_N-\alpha_0),
\end{aligned}
$$

where $\alpha^*$ is some convex combination of $\hat{\hat{\alpha}}_N$ and $\alpha_0$. It follows from Theorems 6.1.1 and 7.1.1 that this term is $o_{P_W}(1+\sqrt{N}|\hat{\hat{\alpha}}_N-\alpha_0)|)$ in $P^*$-probability. Next, we show the weak convergence of the process $\sqrt{N}(\hat{\Phi}_{N,dmc}-\Phi_{mc})(\alpha)$ at $\alpha_0=0$. It follows from Theorem 7.3.1 that

$$
\begin{aligned}
&\sqrt{N}(\hat{\Phi}_{N,dmc}-\Phi_{mc})(\alpha_0)=\sqrt{N}\hat{\Phi}_{N,dmc}(0)=\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,}-\mathbb{P}_N^{\pi})G_{mc}(V;\hat{\alpha}_N)Z\\
&=\hat{\mathbb{G}}_N^{\pi,S}Z+\hat{\mathbb{G}}_N^{\pi,S}\{G_{mc}(V;\hat{\alpha}_N)-1\}Z\\
&=\hat{\mathbb{G}}_N^{\pi,S}Z+(\hat{\mathbb{P}}_N^{\pi,S}-\mathbb{P}_N^{\pi})\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z\sqrt{N}(\hat{\alpha}_N-\alpha_0),
\end{aligned}
$$

where $\alpha^*$ is some convex combination of $\hat{\alpha}_N$ and $\alpha_0$. The first term converges to $\mathbb{G}^{\pi}$ in $P^*$-probability by Theorem 7.3.1. Since $\sqrt{N}(\hat{\alpha}_N-\alpha_0)=O_{P_W^*}(1)$ in $P^*$-probability, it follows from Theorem 6.1.1 that the second term is $o_{P_W^*}(1)$ in $P^*$-probability. Thus, by Theorem 3.3.1 of [58] we obtain

$$
\begin{aligned}
\sqrt{N}\hat{\alpha}_N &= -\dot{\Phi}_{mc}(0)\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(0)+o_{P*_W}(1)\\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}\right\}^{-1}\mathbb{G}^{\pi,S}Z
\end{aligned}
$$

in $P^*$-probability.

□

We prove the lemma below to prove Theorem 4.1.2 as its corollary. Let $\hat{\theta}_N$ be an estimator of $\theta$ obtained as a solution of the estimating equations given by

$$
\Psi_N(\theta)\equiv\|\mathbb{P}_N B(\theta)\|_{\mathcal{H}}=o_{P^*}(N^{-1/2}),
$$

where $B(\theta)$ is a map from some index set $\mathcal{H}$ to $\mathbb{R}$ indexed by $\theta$. Under two-phase sampling, let $\hat{\theta}^{\pi}_{N,\#}$ be an estimator of $\theta$ obtained as a solution of the weighted estimating equations given by

$$\Psi^{\pi}_{N,\#}(\theta) \equiv \left\| \mathbb{P}^{\pi,\#}_{N} B(\theta) \right\|_{\mathcal{H}} = o_{P^*}(N^{-1/2}).$$

Also, $\hat{\hat{\theta}}^{\pi}_{N,S,\#}$ be the survey bootstrap estimator of $\theta$ given by

$$\hat{\Psi}^{\pi}_{N,S,\#}(\theta) \equiv \left\| \hat{\mathbb{P}}^{\pi,S,\#}_{N} B(\theta) \right\|_{\mathcal{H}} = o_{P^*_W}(N^{-1/2}),$$

in $P^*$-probability and $\hat{\hat{\theta}}^{\pi}_{N,\#}$ be the two-phase bootstrap estimator of $\theta$ given by

$$\hat{\Psi}^{\pi}_{N,\#}(\theta) \equiv \left\| \hat{\mathbb{P}}^{\pi,\#}_{N} B(\theta) \right\|_{\mathcal{H}} = o_{P^*_W}(N^{-1/2}),$$

in $P^*$-probability where # is null for the case corresponding to the plain (bootstrap) WLE or $\# \in \{c, mc, cc, dc, dmc, dcc\}$. Let $\Psi(\theta) \equiv P_0 B(\theta)$ be a map from $\Theta$ to $\ell^\infty(\mathcal{H})$.

**Condition 4.3.1.** *For the true parameter $\theta_0 \in \Theta$, $\Psi(\theta_0) = 0$ and the set $\{B(\theta_0)h : h \in \mathcal{H}\}$ is Donsker.*

**Condition 4.3.2.** *Suppose that $\Psi$ is Fréchet differentiable at $\theta_0$;*

$$\left\| \Psi(\theta) - \Psi(\theta_0) - \dot{\Psi}_0(\theta - \theta_0) \right\|_{\mathcal{H}} = o\left( \|\theta - \theta_0\| \right).$$

*Moreover, $\dot{\Psi}_0$ is continuously invertible at $\theta_0$ with inverse denoted as $\Psi_0^{-1}$*

**Condition 4.3.3.** *For any $\delta_N \to 0$, the following stochastic equicontinuity condition holds at $\theta_0$;*

$$\sup_{\|\theta - \theta_0\| \le \delta_N} \|\sqrt{N}(\Psi_N - \Psi)(\theta) - \sqrt{N}(\Psi_N - \Psi)(\theta_0)\|_{\mathcal{H}} = o_{P^*}(1 + \sqrt{N}\|\theta - \theta_0\|).$$

We study $\hat{\hat{\theta}}^{\pi}_{N,S}$ and $\hat{\hat{\theta}}^{\pi}_{N,S,\#}$ with $\# \in \{c, mc, cc\}$. The survey bootstrap IPW empirical processes applied to $B(\theta)$ are denoted as $\hat{\mathbb{G}}^{\pi,S,\#}_{N} B(\theta) = \sqrt{N}(\hat{\mathbb{P}}^{\pi,S,\#}_{N} - \mathbb{P}^{\pi}_{N}) B(\theta) = \sqrt{N}(\hat{\Psi}^{\pi}_{N,S,\#} - \Psi^{\pi}_{N})(\theta)$. Note that these estimators are all centered around $\hat{\theta}^{\pi}_{N}$.

**Theorem 4.3.1.** *Suppose that Conditions 4.3.1-4.3.3 hold and that estimators $\hat{\theta}^{\pi}_{N,S}, \hat{\theta}^{\pi}_{N,S,\#}$ with $\# \in \{c, mc, cc\}$ are consistent for $\theta_0$. Then*

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j B(\theta_0),$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,c} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (I - Q_c) B(\theta_0),$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (I - Q_{mc}) B(\theta_0),$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,cc} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (I - Q_{cc}) B(\theta_0),$$

*in $\ell^{\infty}(\mathcal{H})$ in $P^*$-probability.*

*Proof.* We only prove the claim for the modified calibration. Other cases are similar. First, Condition 4.3.1 together with Theorem 7.3.1 implies that

$$\hat{\mathbb{G}}^{\pi,S,mc}_{N} B(\theta_0) \rightsquigarrow \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (I - Q_c) B(\theta_0), \quad \text{in } \ell^{\infty}(\mathcal{H}),$$

in $P^*$-probability.

For a fixed arbitrary sequence $\{\delta_N\}$ with $\delta_N \to 0$, let

$$\begin{aligned} \mathcal{D}_N &\equiv \left\{ \frac{B(\theta)(h) - B(\theta_0)(h)}{1 + \sqrt{N}\|\theta - \theta_0\|} : h \in \mathcal{H}, \|\theta - \theta_0\| \leq \delta_N \right\} \\ &\equiv \left\{ B_N(\theta, \theta_0)(h) : h \in \mathcal{H}, \|\theta - \theta_0\| \leq \delta_N \right\}. \end{aligned}$$

Condition 4.3.3 can be rewritten as $\|\mathbb{G}_N\|_{\mathcal{D}_N} = o_{P^*}(1)$. This implies that $E\|\mathbb{G}_N\|_{\mathcal{D}_N} \to 0$ as $N \to \infty$ by Theorem 6.3.2. It follows by Lemma 7.4.8 that $E\|\hat{\mathbb{G}}^{\pi,S}_{N}\|_{\mathcal{D}_N} \to 0$ as $N \to \infty$ and therefore $\|\hat{\mathbb{G}}^{\pi,S}_{N}\|_{\mathcal{D}_N} = o_{P^*_W}(1)$ in $P^*$-probability. In view of the proof of Lemma 7.3.3,

$$\|\hat{\mathbb{G}}^{\pi,S,mc}_{N}\|_{\mathcal{D}_N} \leq \|\hat{\mathbb{G}}^{\pi,S,mc}_{N} - \hat{\mathbb{G}}^{\pi}_{N}\|_{\mathcal{D}_N} + \|\hat{\mathbb{G}}^{\pi,S,mc}_{N}\|_{\mathcal{D}_N} = o_{P^*_W}(1).$$

Moreover, Theorem 6.3.2 implies that $\|\mathbb{G}^{\pi,mc}_{N}\|_{\mathcal{D}_N} = o_{P^*}(1)$. Thus, consistency of $\hat{\hat{\theta}}^{\pi}_{N,S,mc}$ to $\theta_0$ and Condition 4.3.3 imply that

$$\|\mathbb{G}^{\pi,S,mc}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - B(\theta_0))\|_{\mathcal{H}} = o_{P^*_W}(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|),$$

$$\|\hat{\mathbb{G}}^{\pi,S,mc}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - B(\theta_0))\|_{\mathcal{H}} = o_{P^*_W}(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|),$$

in $P^*$-probability.

We prove $\sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\| = O_{P^*_W}(1)$ in $P^*$-probability. We have in $P^*$-probability that

$$
\begin{aligned}
&\hat{\mathbb{G}}^{\pi,S,mc}_N B(\theta_0) + \mathbb{G}^{\pi}_N B(\theta_0) + \sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0)) \\
&= \hat{\mathbb{G}}^{\pi,S,mc}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + \mathbb{G}^{\pi}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + \hat{\mathbb{G}}^{\pi,S,mc}_N B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) \\
&\quad + \mathbb{G}^{\pi}_N B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) \quad + \sqrt{N}(P_0 B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - P_0 B(\theta_0)) \\
&= \hat{\mathbb{G}}^{\pi,S,mc}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + \mathbb{G}^{\pi}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) \\
&\quad + \sqrt{N}(\hat{\mathbb{P}}^{\pi,S,mc}_N - \mathbb{P}^{\pi}_N) B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) + \sqrt{N}(\mathbb{P}^{\pi}_N - P_0) B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) \\
&\quad + \sqrt{N}(P_0 B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - P_0 B(\theta_0)) \\
&= \hat{\mathbb{G}}^{\pi,S,mc}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + \mathbb{G}^{\pi}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) \\
&\quad + \sqrt{N}\hat{\mathbb{P}}^{\pi,S,mc}_N B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) + \sqrt{N} P_0 B(\theta_0)) \\
&= \hat{\mathbb{G}}^{\pi,S,mc}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + \mathbb{G}^{\pi}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc})) + o_{P^*_W}(1).
\end{aligned}
$$

Here we used $\hat{\mathbb{P}}^{\pi,S,mc}_N(\hat{\hat{\theta}}^{\pi,S,mc}_N) = o_{P^*_W}(N^{-1/2})$ in $P^*$-probability and $P_0 B(\theta_0) = 0$. Thus,

$$
\begin{aligned}
&\left\|\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0))\right\|_{\mathcal{H}} - \left\|\hat{\mathbb{G}}^{\pi,S,mc}_N B(\theta_0)\right\|_{\mathcal{H}} - \|\mathbb{G}^{\pi}_N B(\theta_0)\|_{\mathcal{H}} \\
&\leq \left\|\hat{\mathbb{G}}^{\pi,S,mc}_N B(\theta_0) + \mathbb{G}^{\pi}_N B(\theta_0) + \sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0))\right\|_{\mathcal{H}} \\
&\leq \left\|\hat{\mathbb{G}}^{\pi,S,mc}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,S,mc}))\right\|_{\mathcal{H}} + \left\|\mathbb{G}^{\pi}_N (B(\theta_0) - B(\hat{\hat{\theta}}^{\pi}_{N,mc}))\right\|_{\mathcal{H}} + o_{P^*_W}(1) \\
&= o_{P^*_W}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|) + o_{P^*_W}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|) + o_{P^*_W}(1) \\
&= o_{P^*_W}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|),
\end{aligned}
$$

in $P^*$-probability. Since by the continuous invertibility of $\Psi_0$ at $\theta_0$ implies that there is some constant $c > 0$ such that

$$
c\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\| \leq \|\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0)\|_{\mathcal{H}},
$$

we have

$$
\begin{aligned}
&c\sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\| \\
&\leq \left\|\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0))\right\|_{\mathcal{H}} \\
&\leq \left\|\hat{\mathbb{G}}^{\pi,S,mc}_N B(\theta_0)\right\|_{\mathcal{H}} + \|\mathbb{G}^{\pi}_N B(\theta_0)\|_{\mathcal{H}} + o_{P^*_W}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0\|),
\end{aligned}
$$

in $P^*$-probability. By Condition 4.3.1, $\mathbb{G}_N^\pi B(\theta_0) = O_{P^*}(1)$ and $\hat{\mathbb{G}}_N^{\pi,S,mc} B(\theta_0) = O_{P_W^*}(1)$ in $P^*$-probability. Thus, the claim $\sqrt{N}\|\hat{\hat{\theta}}_{N,S,mc}^\pi - \theta_0\| = O_{P_W^*}(1)$ in $P^*$-probability follows.

Now we prove the asymptotic normality of $\hat{\hat{\theta}}_N^{\pi,c}$. We have

$$
\begin{aligned}
&\sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,mc}^\pi) - \Psi(\hat{\theta}_N^\pi)) + \hat{\mathbb{G}}_N^{\pi,S,mc} B(\theta_0) \\
&= \sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,mc}^\pi) - \Psi(\hat{\theta}_N^\pi)) + \hat{\mathbb{G}}_N^{\pi,S,mc} B(\theta_0) + \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc} B(\hat{\hat{\theta}}_{N,S,mc}^\pi) \\
&\quad - \sqrt{N}\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) + \sqrt{N}\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) - \sqrt{N}\mathbb{P}_N^\pi B(\theta_0) - \sqrt{N}\mathbb{P}_N^\pi B(\hat{\hat{\theta}}_{N,S,mc}^\pi) + \sqrt{N}\mathbb{P}_N^\pi B(\theta_0) \\
&\quad - \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc} B(\hat{\hat{\theta}}_{N,S,mc}^\pi) + \sqrt{N}\mathbb{P}_N^\pi B(\hat{\hat{\theta}}_{N,S,mc}^\pi) \\
&= \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc} B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - \sqrt{N}\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) \\
&\quad + \sqrt{N}\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) - \sqrt{N}P_0 B(\hat{\theta}_N^\pi) - \sqrt{N}\mathbb{P}_N^\pi B(\theta_0) + \sqrt{N}P_0 B(\theta_0) \\
&\quad - \sqrt{N}\mathbb{P}_N^\pi B(\hat{\hat{\theta}}_{N,S,mc}^\pi) + \sqrt{N}P_0 B(\hat{\hat{\theta}}_{N,S,mc}^\pi) + \sqrt{N}\mathbb{P}_N^\pi B(\theta_0) - \sqrt{N}P_0 B(\theta_0) \\
&\quad - \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,mc} - \mathbb{P}_N^\pi) B(\hat{\hat{\theta}}_{N,S,mc}^\pi) + \hat{\mathbb{G}}_N^{\pi,S,mc} B(\theta_0) \\
&= \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc} B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - \sqrt{N}\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) + \mathbb{G}_N^\pi (B(\hat{\theta}_N^\pi) - B(\theta_0)) \\
&\quad - \mathbb{G}_N^\pi (B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - B(\theta_0)) - \hat{\mathbb{G}}_N^{\pi,S,mc} (B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - B(\theta_0))
\end{aligned}
$$

Since $\sqrt{N}\|\hat{\hat{\theta}}_{N,S,mc}^\pi - \theta_0\| = O_{P_W^*}(1)$ in $P^*$-probability, we have

$$
\left\| \hat{\mathbb{G}}_N^{\pi,S,mc} (B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - B(\theta_0)) \right\|_{\mathcal{H}} = o_{P_W^*}(1)(1 + O_{P_W^*}(1)) = o_{P_W^*}(1),
$$

in $P^*$-probability. Similar reasoning implies

$$
\begin{aligned}
\left\| \mathbb{G}_N^\pi (B(\hat{\hat{\theta}}_{N,S,mc}^\pi) - B(\theta_0)) \right\|_{\mathcal{H}} &= o_{P_W^*}(1)(1 + O_{P_W^*}(1)) = o_{P_W^*}(1), \\
\left\| \mathbb{G}_N^\pi (B(\hat{\theta}_N^\pi) - B(\theta_0)) \right\|_{\mathcal{H}} &= o_{P_W^*}(1)(1 + O_{P_W^*}(1)) = o_{P_W^*}(1),
\end{aligned}
$$

in $P^*$-probability. Moreover, $\hat{\mathbb{P}}_N^{\pi,S,mc} B(\hat{\hat{\theta}}_{N,S,mc}^\pi) = o_{P_W^*}(N^{-1/2})$ in $P^*$-probability and $\mathbb{P}_N^\pi B(\hat{\theta}_N^\pi) = o_{P^*}(N^{-1/2})$. Thus,

$$
\sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,mc}^\pi) - \Psi(\hat{\theta}_N^\pi)) = -\hat{\mathbb{G}}_N^{\pi,c} B(\theta_0) + o_{P_W^*}(1) \tag{4.24}
$$

in $P^*$-probability.

By Fréchet differentiability of $\Psi(\theta)$ at $\theta_0$ and $\sqrt{N}$-consistency of $\hat{\theta}_N^\pi$ and $\hat{\hat{\theta}}_{N,S,mc}^\pi$ implies that

$$
\sqrt{N}(\Psi(\hat{\theta}_N^\pi) - \Psi(\theta_0)) = \dot{\Psi}_0 \left( \sqrt{N}(\hat{\theta}_N^\pi - \theta_0) \right) + o_{P^*}(1)
$$

and

$$\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\theta_0)) = \dot{\Psi}_0\left(\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \theta_0)\right) + o_{P^*_W}(1)$$

in $P^*$-probability. Subtraction gives

$$\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,mc}) - \Psi(\hat{\theta}^{\pi}_{N})) = \dot{\Psi}_0\left(\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \hat{\theta}^{\pi}_{N})\right) + o_{P^*}(1) + o_{P^*_W}(1).$$

Combine this with (4.24) and use the invertibility of $\dot{\Psi}(\theta)$ at $\theta_0$ to obtain

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,mc} - \hat{\theta}^{\pi}_{N}) = -\dot{\Psi}_0^{-1}\hat{\mathbb{G}}^{\pi,S,mc}_{N}B(\theta_0) + o_{P^*_W}(1)$$

in $P^*$-probability. This completes the proof. □

Next, we study doubly calibrated bootstrap estimators. Note the difference in centering from the previous theorem. The survey bootstrap IPW empirical processes applied to $B(\theta)$ in these cases are denoted as $\hat{\mathbb{G}}^{\pi,S,\#}_{N}B(\theta) = \sqrt{N}(\hat{\mathbb{P}}^{\pi,S,\#}_{N} - \mathbb{P}^{\pi,\#}_{N})B(\theta) = \sqrt{N}(\hat{\Psi}^{\pi}_{N,S,\#} - \Psi^{\pi}_{N,\#})(\theta)$.

**Theorem 4.3.2.** *Suppose that Conditions 4.3.1-4.3.3 hold and that $\hat{\theta}^{\pi}_{N,\#}$ and $\hat{\hat{\theta}}^{\pi}_{N,d\#}$ are consistent for $\theta_0$ with $\# \in \{c, mc, cc\}$. Then*

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dc} - \hat{\theta}^{\pi}_{N,c}) \rightsquigarrow -\dot{\Psi}_0^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(I - Q_c)B(\theta_0),$$
$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dmc} - \hat{\theta}^{\pi}_{N,mc}) \rightsquigarrow -\dot{\Psi}_0^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(I - Q_{mc})B(\theta_0),$$
$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dcc} - \hat{\theta}^{\pi}_{N,cc}) \rightsquigarrow -\dot{\Psi}_0^{-1}\sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(I - Q_{cc})B(\theta_0),$$

*in $\ell^\infty(\mathcal{H})$ in $P^*$-probability.*

*Proof.* The proof is similar to the previous theorem. We only prove the claim for $\hat{\hat{\theta}}_{N,S,mc}$. Other cases are similar.

First, Condition 4.3.1 implies that

$$\hat{\mathbb{G}}^{\pi,S,dc}_{N}B(\theta_0) \rightsquigarrow \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(I - Q_{mc})B(\theta_0), \quad \text{in } \ell^\infty(\mathcal{H}),$$

in $P^*$-probability.

For a fixed arbitrary sequence $\{\delta_N\}$ with $\delta_N \to 0$, let

$$\begin{aligned} \mathcal{D}_N &\equiv \left\{ \frac{B(\theta)(h) - B(\theta_0)(h)}{1 + \sqrt{N}\|\theta - \theta_0\|} : h \in \mathcal{H}, \|\theta - \theta_0\| \leq \delta_N \right\} \\ &\equiv \{B_N(\theta, \theta_0)(h) : h \in \mathcal{H}, \|\theta - \theta_0\| \leq \delta_N\}. \end{aligned}$$

Condition 4.3.3 can be rewritten as $\|\mathbb{G}_N\|_{\mathcal{D}_N} = o_{P^*}(1)$. This implies that $E\|\mathbb{G}_N\|_{\mathcal{D}_N} \to 0$ as $N \to \infty$ by Theorem 6.3.2. It follows by Lemma 7.4.8 that $E\|\hat{\mathbb{G}}_N^{\pi}\|_{\mathcal{D}_N} \to 0$ as $N \to \infty$ and therefore $\|\hat{\mathbb{G}}_N^{\pi}\|_{\mathcal{D}_N} = o_{P_W^*}(1)$ in $P^*$-probability. In view of the proof of Lemma 7.3.3

$$\|\hat{\mathbb{G}}_N^{\pi,S,dmc}\|_{\mathcal{D}_N} \leq \|\hat{\mathbb{G}}_N^{\pi,S,dmc} - \hat{\mathbb{G}}_N^{\pi,S}\|_{\mathcal{D}_N} + \|\hat{\mathbb{G}}_N^{\pi,S,dmc}\|_{\mathcal{D}_N} = o_{P_W^*}(1).$$

Moreover, Theorem 6.3.2 also implies that $\|\mathbb{G}_N^{\pi,mc}\|_{\mathcal{D}_N} = o_{P^*}(1)$. Thus, consistency of $\hat{\hat{\theta}}_{N,S,mc}^{\pi}$ to $\theta_0$ and Condition 4.3.3 imply that

$$\begin{aligned} \|\mathbb{G}_N^{\pi,mc}(B(\hat{\hat{\theta}}_{N,S,mc}^{\pi}) - B(\theta_0))\|_{\mathcal{H}} &= o_{P_W^*}(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,mc} - \theta_0\|), \\ \|\hat{\mathbb{G}}_N^{\pi,S,dmc}(B(\hat{\hat{\theta}}_{N,S,mc}^{\pi}) - B(\theta_0))\|_{\mathcal{H}} &= o_{P_W^*}(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,mc} - \theta_0\|), \end{aligned}$$

in $P^*$-probability.

We prove $\sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\| = O_{P_W^*}(1)$ in $P^*$-probability. We have

$$\begin{aligned} &\hat{\mathbb{G}}_N^{\pi,S,dc}B(\theta_0) + \mathbb{G}_N^{\pi,mc}B(\theta_0) + \sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - \Psi(\theta_0)) \\ &= \hat{\mathbb{G}}_N^{\pi,S,dc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) + \mathbb{G}_N^{\pi,mc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) + \hat{\mathbb{G}}_N^{\pi,S,dc}B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) \\ &\quad + \mathbb{G}_N^{\pi,mc}B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) + \sqrt{N}(P_0(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - P_0B(\theta_0)) \\ &= \hat{\mathbb{G}}_N^{\pi,S,dc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,mc}^{\pi})) + \mathbb{G}_N^{\pi,mc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,mc}^{\pi})) \\ &\quad + \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,dc} - \mathbb{P}_N^{\pi,mc})B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) + \sqrt{N}(\mathbb{P}_N^{\pi,mc} - P_0)B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) \\ &\quad + \sqrt{N}(P_0(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - P_0B(\theta_0)) \\ &= \hat{\mathbb{G}}_N^{\pi,S,dc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) + \mathbb{G}_N^{\pi,mc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) \\ &\quad + \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,dc}B(\hat{\hat{\theta}}_{N,S,mc}^{\pi}) + \sqrt{N}P_0B(\theta_0)) \\ &= \hat{\mathbb{G}}_N^{\pi,S,dc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) + \mathbb{G}_N^{\pi,mc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi})) + o_{P_W^*}(1). \end{aligned}$$

Here we used $\hat{\mathbb{P}}_N^{\pi,S,dc}(\hat{\hat{\theta}}_{N,S,mc}^{\pi}) = o_{P_W^*}(N^{-1/2})$ in $P^*$-probability and $P_0 B(\theta_0) = 0$. Thus,

$$
\begin{aligned}
&\|\sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - \Psi(\theta_0))\|_{\mathcal{H}} - \|\hat{\mathbb{G}}_N^{\pi,S,dmc} B(\theta_0)\|_{\mathcal{H}} - \|\mathbb{G}_N^{\pi,mc} B(\theta_0)\|_{\mathcal{H}} \\
&\leq \|\hat{\mathbb{G}}_N^{\pi,S,dmc} B(\theta_0) + \mathbb{G}_N^{\pi,mc} B(\theta_0) + \sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - \Psi(\theta_0))\|_{\mathcal{H}} \\
&\leq \|\hat{\mathbb{G}}_N^{\pi,S,dmc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}))\|_{\mathcal{H}} + \|\mathbb{G}_N^{\pi,mc}(B(\theta_0) - B(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}))\|_{\mathcal{H}} + o_{P_W^*}(1) \\
&= o_{P_W^*}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\|) + o_{P_W^*}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\|) + o_{P_W^*}(1) \\
&= o_{P_W^*}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\|),
\end{aligned}
$$

in $P^*$-probability. Since by the continuous invertibility of $\Psi_0$ at $\theta_0$ implies that there is some constant $c > 0$ such that

$$
c\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\| \leq \|\Psi(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - \Psi(\theta_0)\|_{\mathcal{H}},
$$

we have

$$
\begin{aligned}
&c\sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\| \\
&\leq \|\sqrt{N}(\Psi(\hat{\hat{\theta}}_{N,S,dmc}^{\pi}) - \Psi(\theta_0))\|_{\mathcal{H}} \\
&\leq \|\hat{\mathbb{G}}_N^{\pi,S,dmc} B(\theta_0)\|_{\mathcal{H}} + \|\mathbb{G}_N^{\pi,mc} B(\theta_0)\|_{\mathcal{H}} + o_{P_W^*}(1)(1 + \sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\|),
\end{aligned}
$$

in $P^*$-probability. By Condition 4.3.1, $\mathbb{G}_N^{\pi,mc} B(\theta_0) = O_{P^*}(1)$ and $\hat{\mathbb{G}}_N^{\pi,S,dmc} B(\theta_0) = O_{P_W^*}(1)$ in $P^*$-probability. Thus, the claim $\sqrt{N}\|\hat{\hat{\theta}}_{N,S,dmc}^{\pi} - \theta_0\| = O_{P_W^*}(1)$ in $P^*$-probability follows.

Now we prove the asymptotic normality of $\hat{\hat{\theta}}^{\pi}_{N,S,dmc}$. We have

$$
\begin{aligned}
&\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \Psi(\hat{\theta}^{\pi}_{N,mc})) + \hat{\mathbb{G}}^{\pi,S,dmc}_{N} B(\theta_0) \\
&= \sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \Psi(\hat{\theta}^{\pi}_{N,mc})) + \hat{\mathbb{G}}^{\pi,S,dmc}_{N} B(\theta_0) + \sqrt{N}\hat{\mathbb{P}}^{\pi,S,dc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) \\
&\quad -\sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) + \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) - \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\theta_0) \\
&\quad -\sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) + \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\theta_0) - \sqrt{N}\hat{\mathbb{P}}^{\pi,S,dc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) \\
&\quad +\sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) \\
&= \sqrt{N}\hat{\mathbb{P}}^{\pi,S,dc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) \\
&\quad +\sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) - \sqrt{N}P_0 B(\hat{\theta}^{\pi}_{N,mc}) - \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\theta_0) + \sqrt{N}P_0 B(\theta_0) \\
&\quad -\sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) + \sqrt{N}P_0 B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) + \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\theta_0) - \sqrt{N}P_0 B(\theta_0) \\
&\quad -\sqrt{N}(\hat{\mathbb{P}}^{\pi,S,dc}_{N} - \mathbb{P}^{\pi,mc}_{N}) B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) + \hat{\mathbb{G}}^{\pi,S,dmc}_{N} B(\theta_0) \\
&= \sqrt{N}\hat{\mathbb{P}}^{\pi,S,dc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \sqrt{N}\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) + \mathbb{G}^{\pi,mc}_{N}(B(\hat{\theta}^{\pi}_{N,mc}) - B(\theta_0)) \\
&\quad -\mathbb{G}^{\pi,mc}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - B(\theta_0)) - \hat{\mathbb{G}}^{\pi,S,dmc}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - B(\theta_0))
\end{aligned}
$$

Since $\sqrt{N}\|\hat{\hat{\theta}}^{\pi}_{N,S,dmc} - \theta_0\| = O_{P^*_W}(1)$ in $P^*$-probability, we have

$$
\|\hat{\mathbb{G}}^{\pi,S,dmc}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - B(\theta_0))\|_{\mathcal{H}} = o_{P^*_W}(1)(1 + O_{P^*_W}(1)) = o_{P^*_W}(1),
$$

in $P^*$-probability. Similar reasoning implies

$$
\begin{aligned}
\|\mathbb{G}^{\pi}_{N}(B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - B(\theta_0))\|_{\mathcal{H}} &= o_{P^*_W}(1)(1 + O_{P^*_W}(1)) = o_{P^*_W}(1), \\
\|\mathbb{G}^{\pi}_{N}(B(\hat{\theta}^{\pi}_{N,mc}) - B(\theta_0))\|_{\mathcal{H}} &= o_{P^*_W}(1)(1 + O_{P^*_W}(1)) = o_{P^*_W}(1),
\end{aligned}
$$

in $P^*$-probability. Moreover, $\hat{\mathbb{P}}^{\pi,S,dc}_{N} B(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) = o_{P^*_W}(N^{-1/2})$ in $P^*$-probability and $\mathbb{P}^{\pi,mc}_{N} B(\hat{\theta}^{\pi}_{N,mc}) = o_{P^*}(N^{-1/2})$. Thus,

$$
\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \Psi(\hat{\theta}^{\pi}_{N,mc})) = -\hat{\mathbb{G}}^{\pi,S,dmc}_{N} B(\theta_0) + o_{P^*_W}(1) \tag{4.25}
$$

in $P^*$-probability.

By Fréchet differentiability of $\Psi(\theta)$ at $\theta_0$ and $\sqrt{N}$-consistency of $\hat{\theta}^{\pi}_{N}$ and $\hat{\hat{\theta}}^{\pi}_{N,S,mc}$ implies that

$$
\sqrt{N}(\Psi(\hat{\theta}^{\pi}_{N,mc}) - \Psi(\theta_0)) = \dot{\Psi}_0\left(\sqrt{N}(\hat{\theta}^{\pi}_{N,mc} - \theta_0)\right) + o_{P^*}(1)
$$

and

$$\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \Psi(\theta_0)) = \dot{\Psi}_0 \left(\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dmc} - \theta_0)\right) + o_{P^*_W}(1)$$

in $P^*$-probability. Subtraction gives

$$\sqrt{N}(\Psi(\hat{\hat{\theta}}^{\pi}_{N,S,dmc}) - \Psi(\hat{\theta}^{\pi}_{N,mc})) = \dot{\Psi}_0 \left(\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dmc} - \hat{\theta}^{\pi}_{N,mc})\right) + o_{P^*}(1) + o_{P^*_W}(1).$$

Combine this with (4.25) and use the invertibility of $\dot{\Psi}(\theta)$ at $\theta_0$ to obtain

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,S,dmc} - \hat{\theta}^{\pi}_{N,mc}) = -\dot{\Psi}_0^{-1}\hat{\mathbb{G}}^{\pi,S,dmc}_N B(\theta_0) + o_{P^*_W}(1)$$

in $P^*$-probability. This completes the proof. □

We study doubly calibrated two-phase bootstrap estimators. The two-phase bootstrap IPW empirical processes applied to $B(\theta)$ in these cases are denoted as $\hat{\mathbb{G}}^{\pi,\#}_N B(\theta) = \sqrt{N}(\hat{\mathbb{P}}^{\pi,\#}_N - \mathbb{P}^{\pi,\#}_N)B(\theta) = \sqrt{N}(\hat{\Psi}^{\pi}_{N,\#} - \Psi^{\pi}_{N,\#})(\theta)$.

**Theorem 4.3.3.** *Suppose that Conditions 4.3.1-4.3.3 hold and that $\hat{\theta}_N$, $\hat{\theta}^{\pi}_N$, $\hat{\theta}^{\pi}_{N,\#}$, and $\hat{\hat{\theta}}^{\pi}_{N,d\#}$ with $\# \in \{c, mc, cc\}$ are consistent for $\theta_0$. Then*

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,c} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_c)B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,dc} - \hat{\theta}^{\pi}_{N,c}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_c)B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,mc} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_{mc})B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,dmc} - \hat{\theta}^{\pi}_{N,mc}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_{mc})B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,cc} - \hat{\theta}^{\pi}_{N}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_{cc})B(\theta_0)\right\},$$

$$\sqrt{N}(\hat{\hat{\theta}}^{\pi}_{N,dcc} - \hat{\theta}^{\pi}_{N,cc}) \rightsquigarrow -\dot{\Psi}_0^{-1}\left\{\mathbb{G} + \sum_{j=1}^{J}\sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (I-Q_{cc})B(\theta_0)\right\},$$

*in $\ell^\infty(\mathcal{H})$ in $P^*$-probability.*

*Proof.* The proof is similar to proofs of Theorems 4.3.1 and 4.3.2. □

*Proof of Theorem 4.1.2.* This is a corollary of Theorems 4.3.1 and 4.3.2. □

We give a proof of Theorem 4.1.3.

*Proof of Theorem 4.1.3.* We only consider the WLE with modified calibration, $\hat{\hat{\theta}}_{N,S,mc}$. The other cases are similar.

We evaluate the stochastic order of $\sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc}\dot{\ell}_{\theta_0,\eta_0} + \sqrt{N}P_0\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}}$. Because $\hat{\mathbb{P}}_N^{\pi,S,mc}\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} = o_{P_W^*}(N^{-1/2})$ in $P^*$-probability by assumption and $P_0\dot{\ell}_{\theta_0,\eta_0} = 0$, we have

$$\begin{aligned}
&\sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc}\dot{\ell}_{\theta_0,\eta_0} + \sqrt{N}P_0\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} \\
&= -\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,mc} - P_0)(\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} - \dot{\ell}_{\theta_0,\eta_0}) + o_{P_W^*}(1) \\
&= -(\hat{\mathbb{G}}_N^{\pi,S,mc} + \mathbb{G}_N^{\pi,mc})(\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} - \dot{\ell}_{\theta_0,\eta_0}) + o_{P_W^*}(1),
\end{aligned}$$

in $P^*$-probability. Since $(\hat{\hat{\theta}}_{N,S,mc}, \hat{\hat{\eta}}_{N,S,mc})$ are consistent for $(\theta_0, \eta_0)$, it follows from Lemmas 6.3.2 and 7.3.3 that the above display is $o_{P_W^*}(1)$ in $P^*$-probability in the same way as in the proof of Theorem 3.3.1. Similarly, $\sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc}B_{\theta_0,\eta_0}[\underline{h}^*] + \sqrt{N}P_0B_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}}[\underline{h}^*] = o_{P_W^*}(1)$ in $P^*$-probability. These stochastic orders and Condition 3.3.4 imply that

$$\begin{aligned}
&P_0\left\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\hat{\theta}}_{N,S,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\hat{\eta}}_{N,S,mc} - \eta_0])\right\} \\
&\quad + o\left(|\hat{\hat{\theta}}_{N,S,mc} - \theta_0|\right) + O\left(\|\hat{\hat{\eta}}_{N,S,mc} - \eta_0\|^\alpha\right) + \hat{\mathbb{P}}_N^{\pi,S,mc}\dot{\ell}_{\theta_0,\eta_0} \\
&= P_0\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\hat{\theta}}_{N,S,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\hat{\eta}}_{N,S,mc} - \eta_0]) - \dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} + \dot{\ell}_{\theta_0,\eta_0}\} \\
&\quad + o\left(|\hat{\hat{\theta}}_{N,S,mc} - \theta_0|\right) + O\left(\|\hat{\hat{\eta}}_{N,S,mc} - \eta_0\|^\alpha\right) + P_0\dot{\ell}_{\hat{\hat{\theta}}_{N,S,mc},\hat{\hat{\eta}}_{N,S,mc}} + \hat{\mathbb{P}}_N^{\pi,S,mc}\dot{\ell}_{\theta_0,\eta_0} \\
&= o_{P_W^*}(N^{-1/2})
\end{aligned} \tag{4.26}$$

in $P^*$-probability, and, furthermore, that

$$\begin{aligned}
&P_0\left\{-B_{\theta_0,\eta_0}[\underline{h}^*](\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\hat{\theta}}_{N,S,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\hat{\eta}}_{N,S,mc} - \eta_0])\right\} \\
&\quad + o\left(|\hat{\hat{\theta}}_{N,S,mc} - \theta_0|\right) + O\left(\|\hat{\hat{\eta}}_{N,S,mc} - \eta_0\|^\alpha\right) + \hat{\mathbb{P}}_N^{\pi,S,mc}B_{\theta_0,\eta_0}[\underline{h}^*] \\
&= o_{P_W^*}(N^{-1/2}),
\end{aligned} \tag{4.27}$$

in $P^*$-probability.

By Condition 4.1.6 and $\alpha\beta > 1/2$, $\sqrt{N} O_{P_W^*}\left(\|\hat{\hat{\eta}}_N - \eta_0\|^\alpha\right) = o_{P_W^*}(1)$ in $P^*$-probability. So by Condition 3.3.2 and taking the difference of (4.26) and (4.27), we have

$$\begin{aligned}
&-P_0\left(\left\{\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right\}\dot{\ell}^T_{\theta_0,\eta_0}\right)\left(\hat{\theta}_{N,mc} - \theta_0\right) + \ o\left(|\hat{\theta}_{N,mc} - \theta_0|\right)\\
&\quad + o_{P_W}(N^{-1/2}) - o_{P_W}(N^{-1/2}) + \hat{\mathbb{P}}_N^{\pi,S,mc}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right)\\
&= o_{P_W}(N^{-1/2}) - o_{P_W}(N^{-1/2}),
\end{aligned}$$

in $P^*$-probability or

$$-I_0(\hat{\theta}_{N,mc} - \theta_0) = \hat{\mathbb{P}}_N^{\pi,S,mc}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right) + o_{P_W^*}(N^{-1/2}),$$

in $P^*$-probability. It follows by the invertibility of $I_0$ that

$$\sqrt{N}\left(\hat{\hat{\theta}}_{N,S,mc} - \theta_0\right) = -\sqrt{N}\hat{\mathbb{P}}_N^{\pi,S,mc} I_0^{-1}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right) + o_{P_W^*}(1),$$

in $P^*$-probability. Since we have in the proof of Theorem 3.3.1 that

$$\sqrt{N}\left(\hat{\theta}_N - \theta_0\right) = -\sqrt{N}\mathbb{P}_N^{\pi} I_0^{-1}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right) + o_{P_W^*}(1),$$

in $P^*$-probability, taking a difference yields

$$\sqrt{N}\left(\hat{\hat{\theta}}_{N,S,mc} - \hat{\theta}_N\right) = -\sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,mc} - \mathbb{P}_N^{\pi}) I_0^{-1}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right]\right) + o_{P_W^*}(1),$$

in $P^*$-probability. Apply Theorem 7.3.1 to complete the proof. □

*Proof of Theorem 4.2.1.* This is a corollary of Theorems 4.3.3. □

*Proof of Theorem 4.2.2.* The proof is similar to the proof of Theorem 4.1.3. □

# Chapter 5

# EXAMPLES AND NUMERICAL RESULTS

In this chapter, we apply results developed in the previous chapters to several examples. We apply our $Z$-theorems (Theorems 3.2.1 and 3.3.1) to the Cox proportional hazards model with right censored and interval censored data in Section 5.1. We discuss verification of the required conditions for $Z$-theorems in detail with tools from Chapter 6. These examples illustrate how to develop asymptotic results in a specific model. Next, we study the finite sample properties of the WLE's in simulation studies for the Cox model with right censored data. We compare our bootstrap methods developed in Chapter 4 with the Horvitz-Thompson estimators of the asymptotic variances of WLE's. We also touch the issue of different designs (classical versus exposure stratified, Bernoulli sampling versus sampling without replacement). In Section 5.3, we reanalyze data from the National Wilms Tumor Study (NWTS) [14, 17].

## *5.1 Applications to the Cox Proportional Hazards Model*

To prove asymptotic normality of WLE's, consistency and rate of convergence need to be established in order to apply our $Z$-theorems in the previous chapter. To this end, general results on IPW empirical processes discussed in the next chapter will be useful. We illustrate this in the Cox proportional hazards models with right censoring and interval censoring under two-phase sampling.

Let $T \sim F$ be a failure time, and $X$ be a vector of covariates with bounded supports in the regression model. The Cox model [13] specifies the relationship

$$\Lambda(t|x) = \exp(\theta^T x)\Lambda(t),$$

where $\theta \in \Theta \subset \mathbb{R}^p$ is the regression parameter, and $\Lambda \in H$ is the (baseline) cumulative hazard function. Here the space $H$ for the nuisance parameter $\Lambda$ is the set of nonnegative, nondecreasing cadlag functions defined on the positive line. The true parameters are $\theta_0$ and $\Lambda_0$.

In addition to $X$, let $U$ be a vector of auxiliary variables collected at phase I which are correlated with the covariate vector $X$. For simplicity of notation, we assume that the covariates $X$ are only observed for the subjects sampled at phase II. Thus, if some of the coordinates of $X$ are available at phase I, then we include an identical copy of those coordinates of $X$ in the vector $U$.

### *5.1.1 Cox model with right censored data*

Under right censoring, we only observe the minimum of the failure time $T$ and the censoring time $C \sim G$. Define the observed time $Y = T \wedge C$ and the censoring indicator $\Delta = I(T \leq C)$. The phase I data is $V = (Y, \Delta, U)$, and the observed data is $(Y, \Delta, \xi X, U, \xi)$ where $\xi$ is the sampling indicator.

We assume the following conditions adopted from [57].

**Condition 5.1.1.** *The finite-dimensional parameter space $\Theta$ is compact and contains the true parameter $\theta_0$ as an interior point.*

**Condition 5.1.2.** *The failure time $T$ and the censoring time $C$ are conditionally independent given $X$, and that there is $\tau > 0$ such that $P(T > \tau) > 0$ and $P(C \geq \tau) = P(C = \tau) > 0$. Both $T$ and $C$ have continuous conditional densities given the covariates $X = x$.*

**Condition 5.1.3.** *The covariate vector $X$ has bounded support. For any measurable function $h$, $P(X \neq h(Y)) > 0$.*

Let $\lambda(t) = (d/dt)\Lambda(t)$ be the baseline hazard function. With complete data, the density of $(Y, \Delta, X)$ is

$$p_{\theta,\Lambda}(y, \delta, x) = \left(\lambda(y) e^{\theta^T x - \Lambda(y) e^{\theta^T x}} (1 - G)(y|x)\right)^{\delta} \left(e^{-\Lambda(y) e^{\theta^T x}} g(y|x)\right)^{1-\delta} p_X(x),$$

where $p_X$ is the marginal density of $X$ and $g(\cdot|x)$ is the conditional density of $C$ given $X = x$. The score for $\theta$ is given by

$$\dot{\ell}_{\theta,\Lambda}(y, \delta, x) = x\left(\delta - e^{\theta^T x}\Lambda(y)\right),$$

and the score operator $B_{\theta,\Lambda} : \mathcal{H} \mapsto L_2(P_{\theta,\Lambda})$ is defined on the unit ball $\mathcal{H}$ in the space $BV[0,\tau]$ such that

$$B_{\theta,\Lambda}h(y,\delta,x) = \delta h(y) - e^{\theta^T x} \int_{[0,y]} h d\Lambda.$$

Because the likelihood based on the density above does not yield the MLE with complete data, we define the log likelihood for one observation with complete data by

$$\ell_{\theta,\Lambda}(y,\delta,x) = \log\left\{ \left(e^{\theta^T x}\Lambda\{y\}\right)^{\delta} e^{-\Lambda(y)e^{\theta^T x}} \right\} = \delta\Lambda\{y\} + \delta\theta^T x - e^{\theta^T x}\Lambda(y),$$

where $\Lambda\{t\}$ is the (point) mass of $\Lambda$ at $t$. Then maximizing the weighted log likelihood $\mathbb{P}_N^{\pi}\ell_{\theta,\Lambda}$ reduces to solving the system of equations $\mathbb{P}_N^{\pi}\dot{\ell}_{\theta,\Lambda} = 0$ and $\mathbb{P}_N^{\pi}B_{\theta,\Lambda}h = 0$ for every $h \in \mathcal{H}$. The efficient score for $\theta$ with complete data is given by

$$\ell^*_{\theta_0,\Lambda_0}(y,\delta,x) = \delta\left(x - \frac{M_1}{M_0}(y)\right) - e^{\theta_0^T x}\int_{[0,y]} \delta\left(x - \frac{M_1}{M_0}(t)\right) d\Lambda_0(t),$$

and the efficient information for $\theta$ with complete data is

$$\tilde{I}_{\theta_0,\Lambda_0} = E\left[\left(\ell^*_{\theta_0,\Lambda_0}\right)^{\otimes 2}\right] = E\left\{ e^{\theta_0^T X}\int_0^{\tau}\left(X - \frac{M_1}{M_0}(y)\right)^{\otimes 2}(1-G)(y|X)d\Lambda_0(y)\right\},$$

where

$$M_k(s) = P_{\theta_0,\Lambda_0}X^k e^{\theta_0^T X} I(Y \geq s), \quad k = 0, 1.$$

**Theorem 5.1.1** (Consistency)**.** *Under Conditions 3.1.1, 3.1.2, 5.1.1-5.1.3, the WLE's are consistent for* $(\theta_0, \Lambda_0)$*.*

*Proof.* We only consider the WLE with modified calibration. Proofs for the other four estimators are similar. Our proof closely follows the consistency proof for the MLE with complete data in [57].

Because of the assumption on $\tau$, we restrict our attention to the interval $[0,\tau]$. For a bounded function $h \in L_2(\Lambda)$, define a perturbation $d\hat{\Lambda}_{N,mc,t} = (1+th)d\hat{\Lambda}_{N,mc}$ of $\hat{\Lambda}_{N,mc}$. The weighted log likelihood with modified calibration, $\mathbb{P}_N^{\pi,mc}\ell_{\theta,\Lambda}$, evaluated at $(\hat{\theta}_{N,mc}, \hat{\Lambda}_{N,mc,t})$ viewed as a function of $t$ is maximal at $t = 0$ by the definition of the WLE with modified calibration. Thus, differentiating at $t = 0$ we obtain $\mathbb{P}_N^{\pi,mc}B_{\hat{\theta}_{N,mc},\hat{\Lambda}_{N,mc}}h = 0$, or

$$\begin{aligned}
\mathbb{P}_N^{\pi,mc}\Delta h(Y) &= \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_{N,mc}^T X}\int_{[0,Y]} h d\hat{\Lambda}_{N,mc} \\
&= \int \mathbb{P}_N^{\pi,mc}\left\{e^{\hat{\theta}_{N,mc}^T X} I_{[Y \geq s]}\right\} h(s) d\hat{\Lambda}_{N,mc}(s).
\end{aligned}$$

Let $\hat{M}_{N,0}(s) = \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_{N,mc}^T X} I(Y \geq s)$. Replacing $h$ in the above display by $h/\hat{M}_{N,0}$ yields

$$\hat{\Lambda}_{N,mc} h = \int \frac{h(s)}{\hat{M}_{N,0}(s)} \mathbb{P}_N^{\pi,mc} \left\{ e^{\hat{\theta}_{N,mc}^T X} I(Y \geq s) \right\} d\hat{\Lambda}_{N,mc}(s) = \mathbb{P}_N^{\pi,mc} \frac{\Delta h(Y)}{\hat{M}_{N,0}(Y)}.$$

Similar reasoning via $P_0 B_0 h = 0$ leads to $\Lambda_0 h = P_0 \Delta h(Y)/M_0(Y)$. Let

$$\tilde{\Lambda}_N h = \mathbb{P}_N^{\pi,mc} \Delta h(Y)/M_0(Y).$$

Since $P(T > \tau) > 0$ and $P(C = \tau) > 0$, we have for $s \leq \tau$ that $M_0(s) \geq M_0(\tau) > 0$. The function $(y, \delta) \mapsto \delta h(y)/M_0(y)$ is bounded, and therefore $\{\delta h(y)/M_0(y) : h \in \mathcal{H}\}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) and the fact that $\mathcal{H}$ is Glivenko-Cantelli. Thus, $\|\tilde{\Lambda}_N\|_{\mathcal{H}} \to_{P^*} \|P_{\theta_0,\Lambda_0} \Delta h(Y)/M_0(Y)\|_{\mathcal{H}} = \|\Lambda_0\|_{\mathcal{H}}$. Moreover, since $\hat{\Lambda}_{N,mc}\{Y_i\} = \hat{\Lambda}_{N,mc}\delta_{Y_i} = N^{-1}(\xi_i/\pi_{\hat{\alpha}_N}(V_i))(\Delta_i/\hat{M}_{N,0}(Y_i))$, and similarly

$$\tilde{\Lambda}_N\{Y_i\} = N^{-1}(\xi_i/\pi_{\hat{\alpha}_N}(V_i))(\Delta_i/M_0(Y_i)),$$

it follows that

$$\hat{\Lambda}_{N,mc}\{Y_i\}/\tilde{\Lambda}_N\{Y_i\} = M_0(Y_i)/\hat{M}_{N,0}(Y_i).$$

Since the weighted log likelihood with modified calibration evaluated at $(\hat{\theta}_{N,mc}, \hat{\Lambda}_{N,mc})$ is larger than at $(\theta_0, \tilde{\Lambda}_N)$, we have

$$\begin{aligned} 0 &\leq \mathbb{P}_N^{\pi,mc} \left( \ell_{\hat{\theta}_{N,mc}, \hat{\Lambda}_{N,mc}} - \ell_{\theta_0, \tilde{\Lambda}_N} \right) \\ &= (\hat{\theta}_{N,mc} - \theta_0)^T \mathbb{P}_N^{\pi,mc} \Delta X - \mathbb{P}_N^{\pi,mc} \left( e^{\hat{\theta}_{N,mc}^T X} \hat{\Lambda}_N(Y) - e^{\theta_0^T X} \tilde{\Lambda}_N(Y) \right) \\ &\quad + \mathbb{P}_N^{\pi,mc} \Delta \log \frac{M_0(Y)}{\hat{M}_{N,0}(Y)}. \end{aligned}$$

We take the limit of this on $N$. Because $\Theta$ is compact, there is a subsequence of $\{\hat{\theta}_N\}$ that converges to $\theta_\infty \in \Theta$. It follows by Theorem 6.1.1 that along the convergent subsequence of $\{\hat{\theta}_N\}$

$$(\hat{\theta}_N - \theta_0)^T \mathbb{P}_N^{\pi,mc} \Delta X \to_{P^*} (\theta_\infty - \theta_0)^T P_{\theta_0,\Lambda_0} \Delta X.$$

For the second term, note that $\hat{\Lambda}_N(\tau)$ is uniformly bounded, because $e^{\theta^T X}$ is uniformly bounded in $\theta$ and $X$, and $\hat{\Lambda}_N(\tau) \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_N^T X} I(Y = \tau) \leq \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_{N,mc}^T X} \hat{\Lambda}_N(Y) = \mathbb{P}_N^{\pi,mc} \Delta \leq 1$. Here we use the weighted likelihood equation with $h = 1$ above. Since $\{\hat{\Lambda}_{N,mc}\}$ and $\{\tilde{\Lambda}_N\}$ are both subsets of the class of monotone, bounded cadlag functions that is Glivenko-Cantelli, it follows by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) and Theorem 6.1.1 that

$$\begin{aligned}&\mathbb{P}_N^{\pi,mc}\left(e^{\hat{\theta}_{N,mc}^T X}\hat{\Lambda}_N(Y) - e^{\theta_0^T X}\tilde{\Lambda}_N(Y)\right)\\&\quad = P_{\theta_0,\Lambda_0}\left(e^{\theta_\infty^T X}\hat{\Lambda}_N(Y) - e^{\theta_0^T X}\tilde{\Lambda}_N(Y)\right) + o_{P^*}(1),\end{aligned} \tag{5.1}$$

along a subsequence of $\hat{\theta}_{N,mc}$.

For the third term, note that $\{\hat{M}_{N,0}\}$ is a subset of the class of monotone, bounded, cadlag functions, which is Glivenko-Cantelli, and hence so is it. Note also that $\hat{M}_{N,0}(\tau) = \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_{N,mc}^T X} I(Y = \tau)$ is bounded away from zero with probability tending to 1 since $P(T > \tau) > 0$ and $P(C = \tau) > 0$. Since $\hat{M}_{N,0}(t) \geq \hat{M}_{N,0}(\tau)$ for $t \leq \tau$, the set $\{\delta \log(M_0(y)/\hat{M}_{N,0}(y))\}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) again so that

$$\mathbb{P}_N^{\pi,mc}\Delta \log(M_0(Y)/\hat{M}_{N,0}(Y)) = P_{\theta_0,\Lambda_0}\Delta\log(M_0(Y)/\hat{M}_N(Y)) + o_{P^*}(1) \tag{5.2}$$

by Theorem 6.1.1.

The set $\left\{\delta h(y)/\hat{M}_{N,0}(y) : h \in \mathcal{H}\right\}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) so that $\|\hat{\Lambda}_N\|_{\mathcal{H}} = \|P_{\theta_0,\Lambda_0}\Delta h(Y)/\hat{M}_{N,0}(Y)\|_{\mathcal{H}} + o_{P^*}(1)$ by Theorem 6.1.1. Since we have by Theorem 6.1.1 that

$$\hat{M}_{N,0}(s) = \mathbb{P}_N^{\pi,mc} e^{\hat{\theta}_{N,mc}^T X} I(Y \geq s) \to_{P^*} P_{\theta_0,\Lambda_0} e^{\theta_\infty^T X} I(Y \geq s) \equiv M_{\infty,0}(s)$$

uniformly in $s$, it follows by the dominated convergence theorem that

$$\begin{aligned}\|\hat{\Lambda}_N\|_{\mathcal{H}} &= \|P_{\theta_0,\Lambda_0}\Delta h(Y)/\hat{M}_{N,0}(Y)\|_{\mathcal{H}} + o_{P^*}(1)\\ &\to_{P^*} \|P_{\theta_0,\Lambda_0}\Delta h(Y)/M_{\infty,0}(Y)\|_{\mathcal{H}} \equiv \|\Lambda_\infty\|_{\mathcal{H}},\end{aligned}$$

along a subsequence of $\hat{\theta}_N$.

Apply the dominated convergence theorem to replace $\hat{\Lambda}_{N,mc}$, $\tilde{\Lambda}_N$, and $\hat{M}_{N,0}$ by $\Lambda_\infty$, $\Lambda_0$, and $M_{\infty,0}$ in (5.1) and (5.2) and conclude

$$\begin{aligned}0 \;\leq\; & (\theta_\infty - \theta_0)^T P_{\theta_0,\Lambda_0}\Delta X - P_{\theta_0,\Lambda_0}\left(e^{\theta_\infty^T X}\Lambda_\infty(Y) - e^{\theta_0^T X}\Lambda_0(Y)\right)\\ & + P_{\theta_0,\Lambda_0}\Delta\log\frac{M_0(Y)}{M_\infty(Y)}.\end{aligned} \tag{5.3}$$

Since $M_0/M_\infty = d\Lambda_\infty/d\Lambda_0$, (5.3) is in fact minus one times the Kullback-Leibler divergence

$$K(P_{\theta_0,\Lambda_0}, P_{\theta_\infty,\Lambda_\infty}) \equiv P_{\theta_0,\Lambda_0} \log \{p_{\theta_0,\Lambda_0}/p_{\theta_\infty,\Lambda_\infty}\} \geq 0,$$

for the complete data model. Thus, (5.3) is exactly zero. But since $K(P_{\theta_0,\Lambda_0}, P_{\theta,\Lambda})$ is strictly positive unless $(\theta,\Lambda) = (\theta_0,\Lambda_0)$ by the identifiability of parameters, we must have $(\theta_\infty,\Lambda_\infty) = (\theta_0,\Lambda_0)$. This is true for any subsequence of $\hat{\theta}_{N,mc}$, and the result follows. □

We apply our $Z$-theorem (Theorem 3.2.1) to show the asymptotic normality of the WLE's.

**Theorem 5.1.2** (Asymptotic normality)**.** *Under Conditions 3.1.1, 3.1.2, 5.1.1-5.1.3,*

$$\begin{aligned}
\sqrt{N}(\hat{\theta}_N - \theta_0) &= \sqrt{N}\mathbb{P}_N^{\pi}\tilde{\ell}_{\theta_0,\Lambda_0} + o_{P^*}(1) \to_d N(0,\Sigma),\\
\sqrt{N}(\hat{\theta}_{N,e} - \theta_0) &= \sqrt{N}\mathbb{P}_N^{\pi,e}\tilde{\ell}_{\theta_0,\Lambda_0} + o_{P^*}(1) \to_d N(0,\Sigma_e),\\
\sqrt{N}(\hat{\theta}_{N,c} - \theta_0) &= \sqrt{N}\mathbb{P}_N^{\pi,c}\tilde{\ell}_{\theta_0,\Lambda_0} + o_{P^*}(1) \to_d N(0,\Sigma_c),\\
\sqrt{N}(\hat{\theta}_{N,mc} - \theta_0) &= \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde{\ell}_{\theta_0,\Lambda_0} + o_{P^*}(1) \to_d N(0,\Sigma_{mc}),\\
\sqrt{N}(\hat{\theta}_{N,cc} - \theta_0) &= \sqrt{N}\mathbb{P}_N^{\pi,cc}\tilde{\ell}_{\theta_0,\Lambda_0} + o_{P^*}(1) \to_d N(0,\Sigma_{cc}),
\end{aligned}$$

*where $\tilde{\ell}_{\theta_0,\Lambda_0} = I_{\theta_0,\Lambda_0}^{-1}\ell_{\theta_0,\Lambda_0}^*$ is the efficient influence function for $\theta$ with complete data, and $\Sigma$, $\Sigma_e$, $\Sigma_c$, $\Sigma_{mc}$ and $\Sigma_{cc}$ are given in Theorem 3.2.1.*

*Proof.* We verify the conditions of Theorem 3.2.1. Condition 3.2.1 holds by Theorem 5.1.1. Conditions 3.2.2 and 3.2.3 hold under the present hypotheses as was shown in [56], section 25.12. □

For variance estimation regarding $\hat{\theta}_N$, $\hat{I}_N \equiv \mathbb{P}_N^{\pi}\left\{\ell_{\hat{\theta}_N,\hat{\Lambda}_N}^*\right\}^{\otimes 2}$ can be used to estimate $I_0$. Letting $\hat{\tilde{\ell}}_0 \equiv \hat{I}_N^{-1}\ell_{\hat{\theta}_N,\hat{\Lambda}_N}^*$, we can estimate $\mathrm{Var}_{0|j}\tilde{\ell}_0$ by $\hat{P}_j\tilde{\ell}_0^{\otimes 2} - \left\{\hat{P}_j\tilde{\ell}_0\right\}^{\otimes 2}$ where $\hat{P}_j\tilde{\ell}_0 \equiv \mathbb{P}_N^{\pi}\hat{\tilde{\ell}}_0 I(V \in \mathcal{V}_j)$ and $\hat{P}_j\tilde{\ell}_0^{\otimes 2} \equiv \mathbb{P}_N^{\pi}\hat{\tilde{\ell}}_0^{\otimes 2} I(V \in \mathcal{V}_j)$. The other four cases are similar.

### *5.1.2 Cox Proportional Hazards Model with Interval Censored Data*

Let $Y$ be a censoring time that is assumed to be conditionally independent of a failure time $T$ given a covariate vector $X$. Under the case 1 interval censoring, we do not observe $T$ but

$(Y, \Delta)$ where $\Delta \equiv I(T \le Y)$. The phase I data is $V = (Y, \Delta, U)$ and the observed data is $(Y, \Delta, \xi X, U, \xi)$ where $\xi$ is the sampling indicator.

With complete data, the log likelihood for one observation is given by

$$\begin{aligned} \ell(\theta, F) &\equiv \delta \log \left\{1 - \overline{F}(y)^{\exp(\theta^T x)}\right\} + (1-\delta) \log \overline{F}(y)^{\exp(\theta^T x)} \\ &= \delta \log \left\{1 - \exp(-\Lambda(y) \exp(\theta^T x))\right\} - (1-\delta) \exp(\theta^T x) \Lambda(y) \\ &\equiv \ell(\theta, \Lambda), \end{aligned}$$

where $\overline{F} = 1 - F$. The WLE $(\hat{\theta}_N, \hat{\Lambda}_N)$ of $(\theta, \Lambda)$ maximizes $\mathbb{P}_N^\pi \ell(\theta, \Lambda)$.

The score for $\theta$ and the score operator $B_{\theta,\Lambda}$ for $\Lambda$ with complete data are

$$\begin{aligned} &\dot{\ell}_{\theta,\Lambda} = x \exp(\theta^T x) \Lambda(y) (\delta r(y, x; \theta, \Lambda) - (1-\delta)), \\ &B_{\theta,\Lambda}[htbp][= \exp(\theta^T x) h(y) \left\{\delta r(y, x; \theta, \Lambda) - (1-\delta)\right\}. \end{aligned}$$

where

$$r(y, x; \theta, \Lambda) = \frac{\exp\left(-\exp(\theta^T x)\Lambda(y)\right)}{1 - \exp\left(-\exp(\theta^T x)\Lambda(y)\right)}.$$

The efficient score for $\theta$ with complete data is given by

$$\ell^*_{\theta_0,\Lambda_0} = e^{\theta_0^T x} Q(y, \delta, x; \theta_0, \Lambda_0) \Lambda_0(y) \left\{ x - \frac{E\left[X e^{2\theta_0^T X} O(Y|X) | Y = y\right]}{E\left[e^{2\theta_0^T X} O(Y|X) | Y = y\right]} \right\}$$

where $Q(y, \delta, x; \theta, \Lambda) = \delta r(y, x; \theta, \Lambda) - (1-\delta)$ and $O(y|x) = \overline{F}_0(y|x)/\left[1 - \overline{F}_0(y|x)\right]$. The efficient information for $\theta$ with complete data is

$$\tilde{I}_{\theta_0,\Lambda_0} = E\left[(\ell^*_{\theta_0,\Lambda_0})^{\otimes 2}\right] = E\left[R(Y, X)\left\{X - \frac{E[XR(Y, X)|Y]}{E[R(Y, X)|Y]}\right\}\right]$$

where $R(Y, X) = \Lambda_0^2(Y|X) O(Y|X)$. See [21] for further details.

We impose the same assumptions made for complete data in [21].

**Condition 5.1.4.** *The finite-dimensional parameter space $\Theta$ is compact and contains the true parameter $\theta_0$ as its interior point.*

**Condition 5.1.5.** *(a) The covariate vector $X$ has bounded support; that is, there exists $x_0$ such that $|X| \le x_0$ with probability 1. (b) For any $\theta \ne \theta_0$, the probability $P(\theta^T X \ne \theta_0^T X) > 0$.*

**Condition 5.1.6.** $F_0(0) = 0$. *Let* $\tau_{F_0} = \inf\{t : F_0(t) = 1\}$. *The support of* $Y$ *is an interval* $S[Y] = [l_Y, u_Y]$, *and* $0 < l_Y \leq u_Y < \tau_{F_0}$.

**Condition 5.1.7.** *The cumulative hazard function* $\Lambda_0$ *has strictly positive derivative on* $S[Y]$, *and the joint function* $G(y, x)$ *of* $(Y, X)$ *has bounded second order (partial) derivative with respect to* $y$.

*Characterization of the WLE*

We characterize the WLE's before studying their asymptotic properties. Let $n = \sum_{i=1}^{N} \xi_i$ be the number of observations sampled at phase II. Let $Y_{(1)}, \ldots, Y_{(n)}$ be the order statistics of $Y_1, \ldots, Y_N$ with $\xi_i = 1, i = 1, \ldots, N$. Let $\Delta_{(i)}$, $X_{(i)}$, $U_{(i)}$, and $\xi_{(i)}$ correspond to $Y_{(i)}$; for example, if $Y_{(i)} = Y_j$, then $\Delta_{(i)} = \Delta_j$. Let $\pi_{(i)} = \pi_0(V_{(i)})$. Because only fully observed subjects contribute to the weighted likelihood, $\hat{\Lambda}_N(Y_i)$ for subjects with $\xi_i = 0$ does not matter in the maximization. In fact, $\hat{\Lambda}_N(Y_{(i)}) = \hat{\Lambda}_N(Y_{(i-1)})$ for subjects with $\xi_{(i)} = 0$ for $i \geq 2$. The WLE $\hat{\Lambda}_N$ of $\Lambda$ corresponds to $\underline{x} = (\hat{\Lambda}_{(1)}, \ldots, \hat{\Lambda}_{(N)})$ that maximizes

$$\phi(\theta, \underline{x}) = \sum_{i=1}^{n} \frac{1}{\pi_{(i)}} \left[ \log \left\{ 1 - \exp\left( -e^{\theta^T X_{(i)}} \right) x_i \right\} - (1 - \Delta_{(i)}) e^{\theta^T X_{(i)}} x_i \right]$$

at $\hat{\theta}_N$ subject to $0 \leq x_1 \leq \cdots \leq x_n$. The monotonicity constraint on $\underline{x}$ is imposed to guarantee that an estimate of $\Lambda$ is nondecreasing. Note that $\phi(\theta, \underline{x})$ is concave in $\underline{x}$.

Without loss of generality, we can assume that $\Delta_{(1)} = 1$ and $\Delta_{(n)} = 0$. If $\Delta_{(1)} = 0$ or $\Delta_{(n)} = 1$, then $\hat{\Lambda}_N(Y_{(1)}) = 0$ or $\hat{\Lambda}_N(Y_{(n)}) = \infty$, so that the first or the last summand in $\phi$ is zero. Hence ignoring these terms does not change the maximization of the weighted likelihood.

**Proposition 5.1.1.** *Assume that* $\Delta_{(1)} = 1$ *and* $\Delta_{(n)} = 0$. *Then the WLE* $(\hat{\theta}_N, \hat{\Lambda}_N)$ *satisfies*

$$\begin{aligned}
&\mathbb{P}_N^{\pi} \hat{\Lambda}_N(Y) \exp(\hat{\theta}_N^T X) X Q(Y, \Delta, X; \hat{\theta}_N, \hat{\Lambda}_N(Y)) = 0, \\
&\sum_{j \geq i} \frac{\xi_{(j)}}{\pi_{(j)}} Q(Y_{(j)}, \Delta_{(j)}, X_{(j)}; \hat{\theta}_N, \hat{\Lambda}_N) \exp(\hat{\theta}_N^T X_{(j)}) \leq 0, \ \textit{for } i = 1, \ldots, n, \\
&\mathbb{P}_N^{\pi} Q(Y, \Delta, X; \hat{\theta}_N, \hat{\Lambda}_N) \exp(\hat{\theta}_N^T X) \hat{\Lambda}_N(Y) = 0.
\end{aligned}$$

*Moreover, the corresponding (in)equalities hold for the WLE's with estimated weights and (modified and centered) calibration.*

*Proof.* The first equation is simply the weighted score equation for $\theta$.

For the second inequality, let $1_j$ be the vector which has 1's as its last $j$ components and zeros as its first $n-j$ components. Let $\underline{\hat{\Lambda}}_N = (\hat{\Lambda}_N(Y_{(i)}))_{i=1}^n$. For $\epsilon > 0$, the vector $\underline{\hat{\Lambda}}_N + \epsilon 1_j$ satisfies the monotonicity constraint. It follows by the definition of the WLE that

$$\begin{aligned} 0 &\geq \lim_{\epsilon \downarrow 0} \frac{\phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N + \epsilon 1_j) - \phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N)}{\epsilon} \\ &= \sum_{i=1}^n \frac{1}{\pi_{(i)}} \left[ \Delta_{(i)} \frac{e^{-e^{\hat{\theta}_N^T X_{(i)}} \hat{\Lambda}_N(Y_{(i)}) + \hat{\theta}_N^T X_{(i)}}}{1 - e^{-e^{\hat{\theta}_N^T X_{(i)}} \hat{\Lambda}_N(Y_{(i)})}} - (1 - \Delta_{(i)}) e^{\hat{\theta}_N^T X_{(i)}} \right] I(i \geq j). \end{aligned}$$

Relabeling $i$ and $j$ gives the desired result. Note that the assumption that $\Delta_{(1)} = 1$ and $\Delta_{(n)} = 0$ guarantees that the above derivative is finite.

The last equality follows for the same reason that

$$\lim_{h \to 0} \frac{\phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N + h\underline{\hat{\Lambda}}_N) - \phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N)}{h} = 0.$$

Note that adding terms associated with $\xi_i = 0$ does not contribute to the sum in the above derivative.

For the other four estimators, change weights appropriately. □

*Consistency*

We prove consistency of the WLE's in the metric given by

$$d((\theta_1, \Lambda_1), (\theta_2, \Lambda_2)) \equiv \|\theta_1 - \theta_2\| + \|\Lambda_1 - \Lambda_2\|_{P_Y,2},$$

where $\|\cdot\|$ for $\theta$ is the Euclidean distance,

$$\|\Lambda_1 - \Lambda_2\|_{P_Y,2}^2 = \int (\Lambda_1(y) - \Lambda_2(y))^2 \, dP_Y,$$

and $P_Y$ is the marginal probability measure of the censoring variable $Y$. The idea of our proof is first to show the consistency in the Kullback-Leibler divergence. To this end, we use the Glivenko-Cantelli theorem for the IPW empirical processes (Theorem 6.1.1) in Chapter 6. Then noting that the Kullback-Leibler divergence bounds the Hellinger distance, we apply the inequality of Lemma A5 of [34] which bounds the metric $d$ by the Hellinger distance.

**Theorem 5.1.3** (Consistency)**.** *Under Conditions 3.1.1, 3.1.2, 5.1.4-5.1.7, the WLE's are consistent in the metric d.*

*Proof.* We only prove consistency for the WLE. Proofs for the other four estimators are similar. Instead of directly working on $H$, let $\tilde{H}$ be the set of all subdistribution functions defined on $[0,\infty]$. We denote the WLE of $F$ as $\hat{F}_N = 1 - \exp\left(-\hat{\Lambda}_N\right)$.

Define the set $\mathcal{F}$ of functions by

$$\mathcal{F} \equiv \left\{ f(\theta, F) = \delta(1 - \overline{F}(y)^{\exp(\theta^T x)}) + (1-\delta)\overline{F}(y)^{\exp(\theta^T x)} : \theta \in \Theta, F \in \tilde{H} \right\}.$$

Boundedness of $X$ and compactness of $\Theta \subset \mathbb{R}^p$ imply that the set $\{\exp(\theta^T x) : \theta \in \Theta\}$ is Glivenko-Cantelli. The set $\tilde{H}$ is also Glivenko-Cantelli since it is a subset of the set of bounded monotone functions. Thus, it follows from boundedness of functions in $\mathcal{F}$ and a Glivenko-Cantelli preservation theorem (Theorem 3, [59]) that $\mathcal{F}$ is Glivenko-Cantelli.

Let $0 < \alpha < 1$ be a fixed constant. It follows by concavity of the function $u \mapsto \log u$ and Jensen's inequality that

$$\begin{aligned} &P_0\left[\log\left\{1 + \alpha\left(\frac{f(\theta,F)}{f(\theta_0,F_0)} - 1\right)\right\}\right] \le \log\left(P_0\left[1 + \alpha\left(\frac{f(\theta,F)}{f(\theta_0,F_0)} - 1\right)\right]\right) \\ &= \log\left(1 - \alpha + \alpha P_0 \frac{f(\theta,F)}{f(\theta_0,F_0)}\right) \le 0, \end{aligned}$$

where the first equality holds if and only if $1 + \alpha(f(\theta,F)/f(\theta_0,F_0) - 1)$ is constant on $S[Y]$, in other words, $(\theta, F) = (\theta_0, F_0)$ on $S[Y]$ by the identifiability condition 5.1.5. Note also that by monotonicity of the logarithm

$$P_0\left[\log\left\{1 + \alpha\left(\frac{f(\theta,F)}{f(\theta_0,F_0)} - 1\right)\right\}\right] \ge P_0\left[\log\left\{1 + \alpha\left(0 - 1\right)\right\}\right] = \log(1-\alpha).$$

Thus, the set

$$\mathcal{G} = \left\{ \log\left\{1 + \alpha\left(\frac{f(\theta,F)}{f(\theta_0,F_0)} - 1\right)\right\} : f(\theta,F) \in \mathcal{F} \right\}$$

has an integrable envelope. To see this, form a sequence $(\theta_n, F_n)$ such that

$$\begin{aligned} g_n &\equiv \log\left\{1 + \alpha\left(\frac{f(\theta_n,F_n)}{f(\theta_0,F_0)} - 1\right)\right\} \\ &\nearrow \sup_{\theta\in\Theta, F\in\tilde{H}} \log\left\{1 + \alpha\left(\frac{f(\theta,F)}{f(\theta_0,F_0)} - 1\right)\right\} \equiv G. \end{aligned}$$

Then $\{g_n - \log(1-\alpha)\}_{n\in\mathbb{N}}$ is a monotone increasing sequence of nonnegative functions. By monotone convergence theorem,

$$P_0 g_n - \log(1-\alpha) \to P_0 G - \log(1-\alpha) \le -\log(1-\alpha).$$

Thus we can choose $G \vee -\log(1-\alpha)$ as an integrable envelope. Moreover, the set $\mathcal{G}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]).

Now, by the concavity of the map $u \mapsto \log u$, and the definition of the WLE, we have

$$\begin{aligned}
&\mathbb{P}_N^\pi \log\left\{1+\alpha\left(\frac{f(\hat\theta_N,\hat F_N)}{f(\theta_0,F_0)}-1\right)\right\}\\
&\ge \mathbb{P}_N^\pi\left((1-\alpha)\log(1)+\alpha\log\frac{f(\hat\theta_N,\hat F_N)}{f(\theta_0,F_0)}\right)\\
&=\alpha\left\{\mathbb{P}_N^\pi \log f(\hat\theta_N,\hat F_N)-\mathbb{P}_N^\pi\log f(\theta_0,F_0)\right\}\ge 0.
\end{aligned}$$

Since $\Theta$ and $\tilde H$ are compact, there exists a subsequence of $(\hat\theta_N,\hat F_N)$ that converges to $(\theta_\infty,F_\infty)\in\Theta\times\tilde H$. Along this subsequence it follows by Theorem 6.1.1 that

$$\begin{aligned}
&0\le \mathbb{P}_N^\pi \log\left\{1+\alpha\left(\frac{f(\hat\theta_N,\hat F_N)}{f(\theta_0,F_0)}-1\right)\right\}\\
&\to_{P^*} P_{\theta_0,F_0}\left[\log\left\{1+\alpha\left(\frac{f(\theta_\infty,F_\infty)}{f(\theta_0,F_0)}-1\right)\right\}\right]\le 0.
\end{aligned}$$

Thus, we have

$$P_{\theta_0,F_0}\log\left\{1+\alpha\left(\frac{f(\theta_\infty,F_\infty)}{f(\theta_0,F_0)}-1\right)\right\}=0.$$

This is possible only at $(\theta_\infty,F_\infty)=(\theta_0,F_0)$ because $(\theta,F)\mapsto P[\log\{1+\alpha(f(\theta,F)/f(\theta_0,F_0)-1)\}]$ attains its maximum only at $(\theta_0,F_0)$. Hence conclude that $(\hat\theta_N,\hat F_N)$ converges to $(\theta_0,F_0)$ in the sense of Kullback-Leibler divergence. Since the Kullback-Leibler divergence bounds the Hellinger distance, it follows by Lemma A5 of [34] that $d\left((\hat\theta_N,\hat\Lambda_N),(\theta_0,\Lambda_0)\right)=o_{P^*}(1)$. □

*Rate of Convergence*

We prove the rate of convergence of the WLE is $N^{1/3}$. We apply the rate theorem (Theorem 6.2.1) in Chapter 6. Since we proved the consistency of $(\hat\theta_N,\hat\Lambda_N)$ to $(\theta_0,\Lambda_0)$ on $S[Y]$, under

Condition 5.1.6 we can restrict a parameter space of $\Lambda$ to

$$H_M \equiv \left\{\Lambda \in H : M^{-1} \leq \Lambda \leq M, \text{ on } S[Y]\right\},$$

where $M$ is a positive constant such that $M^{-1} \leq \Lambda_0 \leq M$ on $S[Y]$. Define

$$\mathcal{M} \equiv \{\ell(\theta, \Lambda) : \theta \in \Theta, \Lambda \in H_M\}.$$

**Theorem 5.1.4** (Rate of convergence)**.** *Under Conditions 5.1.4-5.1.7,*

$$d\left((\hat{\theta}_N, \hat{\Lambda}_N), (\theta_0, \Lambda_0)\right) = O_{P^*}\left(N^{-1/3}\right).$$

*This holds if we replace the WLE by the WLE's with estimated weights and (modified and centered) calibration assuming Conditions 3.1.1 and 3.1.2.*

*Proof.* Since the rate of convergence for the WLE is easier to verify than the other four estimators, we only prove the theorem for the WLE with modified calibration. The cases for the WLE's with estimated weights and (centered) calibration are similar.

We proceed by verifying the conditions in Theorem 6.2.1. The bound (6.4) follows by Lemma 6.2.2 in Chapter 6 and Lemma A5 of [34].

For the bound (6.5), we follow the proof of (6.3) in [21]. Since $\hat{\alpha}_N$ is consistent, we can specify the small neighborhood $\mathcal{A}_{mc,0}$ of a zero vector such that $G_{mc}(z; \alpha)$ is contained in a small interval that contains 1 and consists of strictly positive numbers. Thus, multiplying the log likelihood by a uniformly bounded quantity, $G_{mc}(z; \alpha)$ only require a slight modification of Huang's proof of his Lemma 3.1 to obtain

$$\sup_Q \log N_{[]}\left(\epsilon, G\mathcal{M}, L_2(Q)\right) \lesssim \epsilon^{-1},$$

for $\epsilon$ small enough where the supremum is taken over the all discrete probability measures, and $G\mathcal{M} = \{G_{mc}(\cdot; \alpha)\ell(\theta, \Lambda) : \alpha \in \mathcal{A}_{mc,0}, \ell(\theta, \Lambda) \in \mathcal{M}\}$. Thus, it follows by Lemma 3.2.2 of [58] that

$$E^* \|\mathbb{G}_N\|_{G\mathcal{M}_\delta} \lesssim \delta^{1/2}\left(1 + \frac{\delta^{1/2}}{\delta^2\sqrt{N}}M\right) \equiv \phi_N(\delta),$$

where the set $G\mathcal{M}_\delta$ is

$$\{m(\theta, \Lambda, \alpha) - m(\theta_0, \Lambda_0, \alpha) : m(\theta, \Lambda, \alpha) \in G\mathcal{M}, d\left((\theta, \Lambda), (\theta_0, \Lambda_0)\right) \leq \delta\}.$$

Apply Theorem 6.2.1 to conclude $r_N = N^{1/3}$. □

*Asymptotic Normality of the Estimators*

We apply Theorem 3.3.1 to derive the asymptotic distributions of the WLE's.

**Theorem 5.1.5** (Asymptotic normality). *Under Conditions 3.1.1, 3.1.2, 5.1.4-5.1.7,*

$$\sqrt{N}(\hat\theta_N - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi}\tilde\ell_{\theta_0,\Lambda_0} + o_{P^*}(1) \rightsquigarrow N\left(0, \Sigma\right),$$
$$\sqrt{N}(\hat\theta_{N,e} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,e}\tilde\ell_{\theta_0,\Lambda_0} + o_{P^*}(1) \rightsquigarrow N\left(0, \Sigma_e\right),$$
$$\sqrt{N}(\hat\theta_{N,mc} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,c}\tilde\ell_{\theta_0,\Lambda_0} + o_{P^*}(1) \rightsquigarrow N\left(0, \Sigma_c\right),$$
$$\sqrt{N}(\hat\theta_{N,mc} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde\ell_{\theta_0,\Lambda_0} + o_{P^*}(1) \rightsquigarrow N\left(0, \Sigma_{mc}\right),$$
$$\sqrt{N}(\hat\theta_{N,cc} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,cc}\tilde\ell_{\theta_0,\Lambda_0} + o_{P^*}(1) \rightsquigarrow N\left(0, \Sigma_{cc}\right),$$

*where* $\tilde\ell_{\theta_0,\Lambda_0} = I^{-1}_{\theta_0,\Lambda_0}\ell^*_{\theta_0,\Lambda_0}$ *is the efficient influence function with complete data and* $\Sigma$, $\Sigma_e$, $\Sigma_c$, $\Sigma_{mc}$ *and* $\Sigma_{cc}$ *are given in Theorem 3.3.1.*

*Proof.* We give a proof for the WLE with modified calibration by verifying the conditions of Theorem 3.3.1. The cases for the other four estimators are similar.

Condition 3.3.1 is satisfied with $\beta = 1/3$ by Theorems 5.1.3 and 5.1.4. Conditions 3.3.2-3.3.4 are verified by [21] with

$$\underline{h}^*(y) \equiv \frac{\Lambda_0(y)E\left[X\exp(2\theta_0^T X)O(Y|X)|Y=y\right]}{E\left[\exp(2\theta_0^T X)O(Y|X)|Y=y\right]}.$$

Since $\mathbb{P}_N^{\pi,mc}\dot\ell_{\hat\theta_{N,mc},\hat\Lambda_{N,mc}} = 0$ by Proposition 5.1.1, it remains to show that $\mathbb{P}_N^{\pi,mc}B_{\hat\theta_{N,mc},\hat\Lambda_{N,mc}}[\underline{h}^*] = o_{P^*}(N^{-1/2})$. Let $g_0 \equiv \underline{h}^* \circ \Lambda_0^{-1}$ be the composition of $\underline{h}^*$ and the inverse of $\Lambda_0$. Note that $\Lambda_0$ is a strictly increasing continuous function by our assumption. Since $g_0(\hat\Lambda_{N,mc}(y))$ is a right continuous function and has exactly the same jump points as $\hat\Lambda_{N,mc}(y)$, by characterization of $\hat\Lambda_{N,mc}$ in Proposition 5.1.1,

$$\mathbb{P}_N^{\pi,mc}g_0\left(\hat\Lambda_{N,mc}(Y)\right)e^{\hat\theta^T_{N,mc}X}Q(Y,\Delta,X;\hat\theta_{N,mc},\hat\Lambda_{N,mc}) = 0.$$

By Conditions 5.1.5-5.1.7, $h^*$ has bounded derivative. This and the assumption that $\Lambda_0$ has strictly positive derivative by Condition 5.1.7 imply that $g_0$ has bounded derivative, too.

So, noting that $\underline{h}^* = g_0 \circ \Lambda_0$, we have

$$
\begin{aligned}
&\mathbb{P}_N^{\pi,mc} B_{\hat\theta_{N,mc},\hat\Lambda_{N,mc}}\left[\underline{h}^*\right] \\
&= \mathbb{P}_N^{\pi,mc} \underline{h}^*(Y) e^{\hat\theta_{N,mc}^T X} Q(Y,\Delta,X;\hat\theta_{N,mc},\hat\Lambda_{N,mc}) \\
&= \mathbb{P}_N^{\pi,mc} \left\{ g_0 \circ \Lambda_0(Y) - g_0(\hat\Lambda_{N,mc}(Y)) \right\} e^{\hat\theta_{N,mc}^T X} Q(Y,\Delta,X;\hat\theta_{N,mc},\hat\Lambda_{N,mc}) \\
&= (\mathbb{P}_N^{\pi,mc} - P_{\theta_0,\Lambda_0}) \left\{ g_0 \circ \Lambda_0(Y) - g_0(\hat\Lambda_{N,mc}(Y)) \right\} e^{\hat\theta_{N,mc}^T X} Q(Y,\Delta,X;\hat\theta_{N,mc},\hat\Lambda_{N,mc}) \\
&\quad + P_{\theta_0,\Lambda_0} \left\{ g_0 \circ \Lambda_0(Y) - g_0(\hat\Lambda_{N,mc}(Y)) \right\} e^{\hat\theta_{N,mc}^T X} Q(Y,\Delta,X;\hat\theta_{N,mc},\hat\Lambda_{N,mc}).
\end{aligned}
$$

[21] showed that the second term in the display is $o_{P^*}(N^{-1/2})$. We show that the first term in the display is also $o_{P^*}(N^{-1/2})$. Let $C > 0$ be an arbitrary constant. Define for a fixed constant $\eta > 0$

$$
\mathcal{D}(\eta) \equiv \left\{ \psi(y,x;\theta,\Lambda) : d\left((\theta,\Lambda),(\theta_0,\Lambda_0)\right) \le \eta, \Lambda \in H_M \right\},
$$

where $\psi(y,\delta,x;\theta,\Lambda) \equiv \{g_0 \circ \Lambda_0(y) - g_0(\Lambda(y))\} e^{\theta^T x} Q(y,\delta,x;\theta,\Lambda)$. Because [21] showed that $\mathcal{D}(\eta)$ is Donsker for every $\eta > 0$ and that $\|\mathbb{G}_N\|_{\mathcal{D}(CN^{-1/3})} = o_{P^*}(1)$, it follows by Lemma 6.3.2 with $\mathcal{F}_N$ replaced by $\mathcal{D}(CN^{-1/3})$ that $\|\mathbb{G}_N^{\pi,mc}\|_{\mathcal{D}(CN^{-1/3})} = o_{P^*}(1)$. This completes the proof. □

Unlike the previous example, $\ell^*_{\theta,\Lambda}$ depends on additional unknown functions, and the bootstrap-based methods discussed in Chapter 4 will be applied to estimate asymptotic variances.

## *5.2 Simulations*

We study finite sample properties of WLE's in a simulation study. We generated 2000 data sets for each of three different scenarios based on the Cox proportional hazards model with right censored data with various combinations of parameters. In each scenario, we compare four different sampling schemes; the exposure stratified case cohort designs under sampling without replacement or under Bernoulli sampling, or the classical case cohort designs (stratified only by the censoring indicator) under sampling without replacement or under Bernoulli sampling. For the exposure stratified case cohort designs, we generated a single data set and then produced 2000 bootstrap samples based on two-phase bootstrap and survey bootstrap to estimate variances of estimators. All simulations are based on the statistical software R and in particular the *survey* package [29].

### *5.2.1 Model 1*

The first scenario considers the case where time to the event or censoring $Y = \min\{T, C\}$, censoring indicator $\Delta$ and auxiliary variable $V$ are available at the first phase but the exposure $X$ of interest is missing. Specifically, the hazard function is given by

$$\lambda(t|x) = \lambda_0(t)\exp(\theta x)$$

where $t$ is a failure time and $x \in \{0, 1\}$ is the exposure of interest. We chose $\theta = \log 2$. The simulation results for $\theta = 0$ are similar and now shown. The auxiliary variable $V \in \{0, 1\}$ is related to the exposure $X$ by sensitivity and specificity given by

$$P(V = 1|X = 1) = \alpha, \quad P(V = 0|X = 0) = \beta.$$

We chose $\alpha = \beta \in \{.5, .9\}$ in a simulation study. The prevalence of $X$ is $P(X = 1) = .5$. The censoring time $C$ is distributed as Unif$(0, 1.1)$.

For the exposure stratified case cohort design, three strata are formed based on $\Delta$ and $V$. The first stratum consists of observations with $\Delta = 1$. The second and third strata consist of observations with $\Delta = 0$ and $V = 0$ or $\Delta = 0$ and $V = 1$, respectively. The sampling probability for each stratum is

$$\begin{aligned}
&P(\xi = 1|\Delta = 1) = 1,\\
&P(\xi = 1|\Delta = 0, V = 0) = p_{1,e} = .3,\\
&P(\xi = 1|\Delta = 0, V = 1) = p_{2,e} = .3,
\end{aligned}$$

regardless of Bernoulli sampling or sampling without replacement. For the classical case cohort design, there are two strata, consisting of observations with $\Delta = 0$ and $\Delta = 1$. The sampling probability for this design is

$$\begin{aligned}
&P(\xi = 1|\Delta = 1) = 1,\\
&P(\xi = 1|\Delta = 0) = p_{1,c} = .3.
\end{aligned}$$

*5.2.2 Model 2*

The second scenario considers the case where time to the event or censoring $Y = \min\{T, C\}$, censoring indicator $\Delta$ and exposure $X$ of interest are available at the first phase but the confounder $V$ is missing. Specifically, the hazard function is given by

$$\lambda(t|x, v) = \lambda_0(t)\exp(\theta_1 x + \theta_2 v)$$

where $t$ is a failure time, $x \in \{0, 1\}$ is the exposure of interest and and $v \in \{0, 1\}$ is the confounder. We chose $\theta = (\theta_1, \theta_2) = (\log 2, 0)$ or $\theta = (\theta_1, \theta_2) = (\log 2, \log 2)$. We focus on estimation of $\theta_1$ in the following discussion (see Model 3 for comparison). The confounder $V \in \{0, 1\}$ is related to the exposure $X$ by sensitivity and specificity given by

$$P(V = 1|X = 1) = \alpha = .9, \quad P(V = 0|X = 0) = \beta = .9.$$

The prevalence of $X$ is $P(X = 1) = .5$ and the censoring time $C$ is distributed as $\text{Unif}(0, 1.1)$ as before.

For the exposure stratified case cohort design, three strata are formed based on $\Delta$ and $X$. The first stratum consists of observations with $\Delta = 1$. The second and third strata consist of observations with $\Delta = 0$ and $X = 0$ or $\Delta = 0$ and $X = 1$, respectively. The sampling probability for each stratum is

$$\begin{aligned}
&P(\xi = 1|\Delta = 1) = 1, \\
&P(\xi = 1|\Delta = 0, X = 0) = p_{1,e} = .3, \\
&P(\xi = 1|\Delta = 0, X = 1) = p_{2,e} = .3,
\end{aligned}$$

regardless of Bernoulli sampling or sampling without replacement. For the classical case cohort design, there are two strata, consisting of observations with $\Delta = 0$ and $\Delta = 1$. The sampling probability for this design is

$$\begin{aligned}
&P(\xi = 1|\Delta = 1) = 1, \\
&P(\xi = 1|\Delta = 0) = p_{1,c} = .3.
\end{aligned}$$

### 5.2.3 Model 3

The third scenario considers the case where time to the event or censoring $Y = \min\{T, C\}$, censoring indicator $\Delta$ and a confounder $V$ are available at the first phase but the exposure $X$ of interest is missing. Specifically, the hazard function is given by

$$\lambda(t|x, v) = \lambda_0(t) \exp(\theta_1 x + \theta_2 v)$$

where $t$ is a failure time, $x \in \{0, 1\}$ is the exposure of interest and $v \in \{0, 1\}$ is a confounder. The difference from the first scenario is that $V$ is now in the outcome model. We chose $\theta = (\theta_1, \theta_2) = (\log 2, 0)$. We focus on estimation of $\theta_1$ in the following discussion. The confounder $V \in \{0, 1\}$ is related to the exposure $X$ by sensitivity and specificity given by

$$P(V = 1|X = 1) = \alpha = .9, \quad P(V = 0|X = 0) = \beta = .9.$$

The prevalence of $X$ is $P(X = 1) = .5$ and the censoring time $C$ is distributed as Unif$(0, 1.1)$ as before. Strata and corresponding sampling probabilities are defined similarly to the model 1.

### 5.2.4 Results

Estimators compared are the plain WLE, the calibrated WLE with calibration on $Y$, the within-stratum calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $Y$ for the exposure stratified case cohort studies. For the classical case cohort designs, calibrations on $V$ for models 1 and 3 and $X$ for the model 2 are also considered. Note that the information on $V$ is not used for stratum formation in the classical case cohort studies.

Tables 5.1 - 5.5 shows sample sizes at the first and second phases. Tables except Table 5.4 with the regression coefficient $\theta = (\log 2, \log 2)$ are similar.

*Adjusting Weights*

We first consider results for Models 1 and 2 except the Model 1 with $\alpha = \beta = .5$ and the Model 2 with $\theta = (\log 2, 0)$ and $N = 250$ to discuss the advantage of the within-stratum centered calibration. We then cover the rest of results to see its limitations.

Table 5.1: Sample sizes for $\alpha = \beta = .9$ in Model 1. $N$ is a phase I sample size. WR and Ber stand for sampling without replacement and Bernoulli sampling, respectively. The numbers without parentheses denote the numbers $n_j$ of observations sampled at the second phase and the numbers inside parentheses are corresponding numbers $N_j$ to the phase I sample.

| | stratum | | exposure stratified | | classical | |
|---|---|---|---|---|---|---|
| $N$ | $\Delta$ | $V$ | WR | Ber | WR | Ber |
| 250 | 1 | | 37 (37) | 37 (37) | 37 (37) | 37 (37) |
| | 0 | 0 | 31 (102) | 31 (102) | 64 (213) | 64 (213) |
| | 0 | 1 | 33 (111) | 33 (111) | | |
| | total | | 101(250) | 101(250) | 101(250) | 101 (250) |
| 500 | 1 | | 73 (73 ) | 73 (73) | 73 (73) | 73 (73) |
| | 0 | 0 | 61 (205) | 62 (205) | 128 (427) | 128 (427) |
| | 0 | 1 | 67 (222) | 66 (222) | | |
| | total | | 201 (500) | 201 (500) | 201 (500) | 201 (500) |

Table 5.2: Sample sizes for $\alpha = \beta = .5$ in Model 1. $N$ is a phase I sample size. WR and Ber stand for sampling without replacement and Bernoulli sampling, respectively. The numbers without parentheses denote the numbers $n_j$ of observations sampled at the second phase and the numbers inside parentheses are corresponding numbers $N_j$ to the phase I sample.

| | stratum | | exposure stratified | | classical | |
|---|---|---|---|---|---|---|
| $N$ | $\Delta$ | $V$ | WR | Ber | WR | Ber |
| 250 | 1 | | 37 (37) | 37 (37) | 37 (37) | 37 (37) |
| | 0 | 0 | 32 (107) | 32 (107) | 64 (213) | 64 (213) |
| | 0 | 1 | 32 (107) | 32 (107) | | |
| | total | | 101(250) | 101(250) | 101(250) | 101 (250) |
| 500 | 1 | | 73 (73) | 73 (73) | 73 (73) | 73 (73) |
| | 0 | 0 | 64 (213) | 64 (213) | 128 (427) | 128 ( 427 ) |
| | 0 | 1 | 64 (213) | 64 (213) | | |
| | total | | 201 (500) | 201 (500) | 201 (500) | 201 (500) |

Table 5.3: Sample sizes for $\alpha = \beta = .9$ and $\theta = (\log 2, 0)$ in Model 2. $N$ is a phase I sample size. WR and Ber stand for sampling without replacement and Bernoulli sampling, respectively. The numbers without parentheses denote the numbers $n_j$ of observations sampled at the second phase and the numbers inside parentheses are corresponding numbers $N_j$ to the phase I sample.

| | stratum | | exposure stratified | | classical | |
|---|---|---|---|---|---|---|
| $N$ | $\Delta$ | $V$ | WR | Ber | WR | Ber |
| 250 | 1 | | 37 ( 37 ) | 37 (37) | 37 (37) | 37 (37) |
| | 0 | 0 | 34 (112) | 34 (112) | 64 (213) | 64 ( 213 ) |
| | 0 | 1 | 30 (101) | 30 (101) | | |
| | total | | 101 | 101 | 101 | 101 |
| 500 | 1 | | 74 (74) | 74 (74) | 74 (74) | 74 (74) |
| | 0 | 0 | 67 (224) | 67 (224) | 128 (426) | 128 (426) |
| | 0 | 1 | 61 (202) | 61 (202) | | |
| | total | | 202 | 201 | 202 | 201 |

Table 5.4: Sample sizes for $\alpha = \beta = .9$ and $\theta = (\log 2, \log 2)$ in Model 2. $N$ is a phase I sample size. WR and Ber stand for sampling without replacement and Bernoulli sampling, respectively. The numbers without parentheses denote the numbers $n_j$ of observations sampled at the second phase and the numbers inside parentheses are corresponding numbers $N_j$ to the phase I sample.

| | stratum | | exposure stratified | | classical | |
|---|---|---|---|---|---|---|
| $N$ | $\Delta$ | $V$ | WR | Ber | WR | Ber |
| 250 | 1 | | 48 (48) | 48 (48) | 48 (48) | 48 (48) |
| | 0 | 0 | 31 (102) | 31 (102) | 61 (202) | 61 (202) |
| | 0 | 1 | 30 (100) | 30 (100) | | |
| | total | | 109 | 109 | 10 | 109 |
| 500 | 1 | | 97 (97) | 97 (97) | 97 (97) | 97 (97) |
| | 0 | 0 | 61 (204) | 62 (204) | 121 (403) | 121 (403) |
| | 0 | 1 | 60 (199) | 60 (199) | | |
| | total | | 218 | 218 | 218 | 218 |

Table 5.5: Sample sizes for $\alpha = \beta = .9$ in Model 3. $N$ is a phase I sample size. WR and Ber stand for sampling without replacement and Bernoulli sampling, respectively. The numbers without parentheses denote the numbers $n_j$ of observations sampled at the second phase and the numbers inside parentheses are corresponding numbers $N_j$ to the phase I sample.

| | stratum | | exposure stratified | | classical | |
|---|---|---|---|---|---|---|
| $N$ | $\Delta$ | $V$ | WR | Ber | WR | Ber |
| 250 | 1 | | 37 (37) | 37 (37) | 37 (37) | 37 (37) |
| | 0 | 0 | 31 (102) | 31 (102) | 64 (213) | 64 (213) |
| | 0 | 1 | 33 (111) | 33 (111) | | |
| | total | | 101 | 101 | 101 | 101 |
| 500 | 1 | | 73 (73) | 73 (73) | 73 (73) | 73 (73) |
| | 0 | 0 | 61 (205) | 61 (205) | 128 (427) | 128 (427) |
| | 0 | 1 | 67 (222) | 66 (222) | | |
| | total | | 201 | 201 | 201 | 201 |

For Models 1 and 2 except the Model 1 with $\alpha = \beta = .5$ and the Model 2 with $\theta = (\log 2, 0)$ and $N = 250$, the within-stratum centered calibration improved efficiency most among different calibrations. This is particularly seen for the classical case cohort designs in Tables 5.6, 5.7, 5.11, 5.12 and 5.13. For example, within-stratum centered calibration on $V$ and $Y$ achieved variance reduction over the plain WLE by 11% and 13% (.179 to .159 and .0879 to .0764) for $N = 250$ and $N = 500$ respectively in the classical case cohort designs under sampling without replacement (Tables 5.6 and 5.7). Overall variance reductions seen by this method are 1-7% (sampling without replacement) and 2-6% (Bernoulli sampling) for the exposure stratified case cohort design with calibration on $Y$, and 7-13% (sampling without replacement) and 7-15% (Bernoulli sampling) for the classical case cohort designs with calibration on $V$ and $Y$ (Tables 5.6, 5.7, 5.11, 5.12 and 5.13). The variance reductions at the second phase are also important because adjusting weights only improve efficiency on the phase II variances. In Tables 5.6, 5.7 and 5.13 where the phase I variances are reliably estimated, variance reductions at the second phases are 24%, 28% and 1% for the

Table 5.6: Summary of simulations in Model 1 for $N = 250$ and $\alpha = \beta = .9$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.708 | 0.124 | NA | NA | NA |
| (1) | WLE | 0.713 | 0.155 | 0.157 | 0.129 | 0.0274 |
| | CalY | 0.713 | 0.156 | 0.159 | 0.130 | 0.0292 |
| | CalYs | 0.715 | 0.147 | 0.150 | 0.130 | 0.0192 |
| | CalYsc | 0.714 | 0.146 | 0.150 | 0.130 | 0.0206 |
| (2) | WLE | 0.717 | 0.154 | 0.157 | 0.129 | 0.0280 |
| | CalY | 0.717 | 0.157 | 0.160 | 0.130 | 0.0298 |
| | CalYs | 0.716 | 0.145 | 0.150 | 0.130 | 0.0192 |
| | CalYsc | 0.714 | 0.145 | 0.150 | 0.129 | 0.0210 |
| (3) | WLE | 0.711 | 0.179 | 0.179 | 0.130 | 0.0499 |
| | CalY | 0.711 | 0.182 | 0.181 | 0.130 | 0.0509 |
| | CalVY | 0.711 | 0.183 | 0.183 | 0.131 | 0.0525 |
| | CalYs | 0.712 | 0.182 | 0.181 | 0.130 | 0.0509 |
| | CalVYs | 0.714 | 0.164 | 0.163 | 0.131 | 0.0327 |
| | CalYsc | 0.711 | 0.180 | 0.180 | 0.130 | 0.0500 |
| | CalVYsc | 0.709 | 0.159 | 0.157 | 0.130 | 0.0275 |
| (4) | WLE | 0.714 | 0.184 | 0.181 | 0.130 | 0.0507 |
| | CalY | 0.714 | 0.187 | 0.182 | 0.131 | 0.0519 |
| | CalVY | 0.714 | 0.189 | 0.185 | 0.131 | 0.0538 |
| | CalYs | 0.715 | 0.187 | 0.183 | 0.131 | 0.0519 |
| | CalVYs | 0.719 | 0.162 | 0.165 | 0.131 | 0.0336 |
| | CalYsc | 0.715 | 0.186 | 0.181 | 0.130 | 0.0509 |
| | CalVYsc | 0.717 | 0.156 | 0.158 | 0.130 | 0.0282 |

Table 5.7: Summary of simulations in Model 1 for $N = 500$ and $\alpha = \beta = .9$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.702 | 0.0623 | NA | NA | NA |
| (1) | WLE | 0.709 | 0.0778 | 0.0756 | 0.0623 | 0.01324 |
| | CalY | 0.709 | 0.0777 | 0.0761 | 0.0625 | 0.01365 |
| | CalYs | 0.709 | 0.0725 | 0.0716 | 0.0625 | 0.00907 |
| | CalYsc | 0.708 | 0.0725 | 0.0718 | 0.0623 | 0.00944 |
| (2) | WLE | 0.703 | 0.0762 | 0.0759 | 0.0624 | 0.01352 |
| | CalY | 0.702 | 0.0765 | 0.0766 | 0.0626 | 0.01399 |
| | CalYs | 0.704 | 0.0723 | 0.0719 | 0.0626 | 0.00929 |
| | CalYsc | 0.704 | 0.0723 | 0.0720 | 0.0624 | 0.00963 |
| (3) | WLE | 0.700 | 0.0879 | 0.0869 | 0.0624 | 0.0245 |
| | CalY | 0.700 | 0.0882 | 0.0873 | 0.0625 | 0.0248 |
| | CalVY | 0.700 | 0.0883 | 0.0877 | 0.0626 | 0.0251 |
| | CalYs | 0.700 | 0.0882 | 0.0873 | 0.0626 | 0.0247 |
| | CalVYs | 0.704 | 0.0785 | 0.0783 | 0.0626 | 0.0156 |
| | CalYsc | 0.700 | 0.0880 | 0.0869 | 0.0624 | 0.0245 |
| | CalVYsc | 0.702 | 0.0764 | 0.0757 | 0.0624 | 0.0132 |
| (4) | WLE | 0.700 | 0.0874 | 0.0871 | 0.0624 | 0.0247 |
| | CalY | 0.700 | 0.0877 | 0.0877 | 0.0626 | 0.0251 |
| | CalVY | 0.700 | 0.0880 | 0.0883 | 0.0627 | 0.0256 |
| | CalYs | 0.701 | 0.0878 | 0.0877 | 0.0626 | 0.0250 |
| | CalVYs | 0.703 | 0.0796 | 0.0786 | 0.0627 | 0.0159 |
| | CalYsc | 0.700 | 0.0876 | 0.0873 | 0.0625 | 0.0248 |
| | CalVYsc | 0.703 | 0.0774 | 0.0759 | 0.0625 | 0.0134 |

Table 5.8: Summary of simulations in Model 1 for $N = 250$ and $\alpha = \beta = .5$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.704 | 0.134 | NA | NA | NA |
| (1) | WLE | 0.707 | 0.181 | 0.179 | 0.130 | 0.0498 |
| | CalY | 0.708 | 0.182 | 0.181 | 0.130 | 0.0508 |
| | CalYs | 0.707 | 0.182 | 0.181 | 0.131 | 0.0500 |
| | CalYsc | 0.706 | 0.182 | 0.179 | 0.130 | 0.0493 |
| (2) | WLE | 0.710 | 0.187 | 0.1810 | 0.130 | 0.0513 |
| | CalY | 0.710 | 0.188 | 0.1825 | 0.130 | 0.0521 |
| | CalYs | 0.710 | 0.189 | 0.1819 | 0.131 | 0.0512 |
| | CalYsc | 0.708 | 0.189 | 0.1805 | 0.130 | 0.0506 |
| (3) | WLE | 0.701 | 0.188 | 0.179 | 0.130 | 0.0498 |
| | CalY | 0.700 | 0.191 | 0.181 | 0.130 | 0.0508 |
| | CalVY | 0.698 | 0.193 | 0.182 | 0.131 | 0.0512 |
| | CalYs | 0.701 | 0.191 | 0.181 | 0.131 | 0.0507 |
| | CalVYs | 0.700 | 0.192 | 0.181 | 0.131 | 0.0504 |
| | CalYsc | 0.700 | 0.189 | 0.180 | 0.130 | 0.0499 |
| | CalVYsc | 0.699 | 0.190 | 0.180 | 0.130 | 0.0496 |
| (4) | WLE | 0.714 | 0.194 | 0.180 | 0.130 | 0.0506 |
| | CalY | 0.713 | 0.197 | 0.182 | 0.130 | 0.0515 |
| | CalVY | 0.714 | 0.199 | 0.183 | 0.131 | 0.0518 |
| | CalYs | 0.714 | 0.197 | 0.182 | 0.130 | 0.0515 |
| | CalVYs | 0.715 | 0.197 | 0.182 | 0.131 | 0.0511 |
| | CalYsc | 0.714 | 0.196 | 0.180 | 0.130 | 0.0506 |
| | CalVYsc | 0.714 | 0.196 | 0.180 | 0.130 | 0.0503 |

Table 5.9: Summary of simulations in Model 1 for $N = 500$ and $\alpha = \beta = .5$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.704 | 0.0670 | NA | NA | NA |
| (1) | WLE | 0.705 | 0.0943 | 0.0869 | 0.0624 | 0.0246 |
| | CalY | 0.704 | 0.0949 | 0.0874 | 0.0626 | 0.0248 |
| | CalYs | 0.704 | 0.0951 | 0.0873 | 0.0626 | 0.0247 |
| | CalYsc | 0.704 | 0.0949 | 0.0869 | 0.0625 | 0.0245 |
| (2) | WLE | 0.696 | 0.0962 | 0.0874 | 0.0624 | 0.0250 |
| | CalY | 0.696 | 0.0970 | 0.0879 | 0.0626 | 0.0253 |
| | CalYs | 0.696 | 0.0972 | 0.0878 | 0.0627 | 0.0251 |
| | CalYsc | 0.696 | 0.0968 | 0.0874 | 0.0625 | 0.0249 |
| (3) | WLE | 0.703 | 0.0976 | 0.0870 | 0.0624 | 0.0246 |
| | CalY | 0.702 | 0.0982 | 0.0874 | 0.0626 | 0.0248 |
| | CalVY | 0.702 | 0.0987 | 0.0877 | 0.0628 | 0.0249 |
| | CalYs | 0.702 | 0.0982 | 0.0874 | 0.0626 | 0.0248 |
| | CalVYs | 0.702 | 0.0984 | 0.0874 | 0.0627 | 0.0247 |
| | CalYsc | 0.702 | 0.0978 | 0.0870 | 0.0625 | 0.0246 |
| | CalVYsc | 0.703 | 0.0980 | 0.0870 | 0.0625 | 0.0245 |
| (4) | WLE | 0.708 | 0.0937 | 0.0871 | 0.0624 | 0.0247 |
| | CalY | 0.709 | 0.0938 | 0.0874 | 0.0625 | 0.0249 |
| | CalVY | 0.708 | 0.0937 | 0.0877 | 0.0627 | 0.0250 |
| | CalYs | 0.709 | 0.0938 | 0.0874 | 0.0625 | 0.0249 |
| | CalVYs | 0.708 | 0.0938 | 0.0874 | 0.0626 | 0.0248 |
| | CalYsc | 0.708 | 0.0937 | 0.0871 | 0.0624 | 0.0247 |
| | CalVYsc | 0.708 | 0.0936 | 0.0871 | 0.0624 | 0.0246 |

Table 5.10: Summary of simulations in Model 2 for $N = 250$ and $\theta = (\log 2, 0)$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalXY, CalYs, CalXYs, CalYsc, CalXYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $X$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $X$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $X$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.717 | 0.335 | NA | NA | NA |
| (1) | WLE | 0.782 | 1.404 | 0.706 | 0.589 | 0.117 |
| | CalY | 0.782 | 1.413 | 0.701 | 0.583 | 0.119 |
| | CalYs | 0.777 | 1.394 | 0.692 | 0.585 | 0.108 |
| | CalYsc | 0.775 | 1.398 | 0.708 | 0.600 | 0.107 |
| (2) | WLE | 0.743 | 1.079 | 0.698 | 0.579 | 0.118 |
| | CalY | 0.747 | 1.091 | 0.649 | 0.528 | 0.121 |
| | CalYs | 0.749 | 1.058 | 0.643 | 0.533 | 0.110 |
| | CalYsc | 0.739 | 1.037 | 0.696 | 0.587 | 0.109 |
| (3) | WLE | 0.771 | 1.539 | 0.577 | 0.429 | 0.148 |
| | CalY | 0.771 | 1.547 | 0.577 | 0.428 | 0.149 |
| | CalXY | 0.770 | 1.543 | 0.569 | 0.413 | 0.156 |
| | CalYs | 0.772 | 1.549 | 0.578 | 0.429 | 0.149 |
| | CalXYs | 0.769 | 1.491 | 0.558 | 0.426 | 0.131 |
| | CalYsc | 0.768 | 1.553 | 0.575 | 0.428 | 0.147 |
| | CalXYsc | 0.763 | 1.557 | 0.539 | 0.424 | 0.115 |
| (4) | WLE | 0.760 | 1.156 | 0.649 | 0.496 | 0.153 |
| | CalY | 0.760 | 1.151 | 0.645 | 0.491 | 0.155 |
| | CalXY | 0.760 | 1.161 | 0.643 | 0.479 | 0.164 |
| | CalYs | 0.761 | 1.152 | 0.646 | 0.491 | 0.155 |
| | CalXYs | 0.760 | 1.139 | 0.624 | 0.487 | 0.137 |
| | CalYsc | 0.755 | 1.145 | 0.647 | 0.496 | 0.151 |
| | CalXYsc | 0.763 | 1.149 | 0.609 | 0.488 | 0.121 |

Table 5.11: Summary of simulations in Model 2 for $N = 500$ and $\theta = (\log 2, 0)$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalXY, CalYs, CalXYs, CalYsc, CalXYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $X$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $X$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $X$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.703 | 0.169 | NA | NA | NA |
| (1) | WLE | 0.732 | 0.269 | 0.242 | 0.182 | 0.0608 |
| | CalY | 0.731 | 0.269 | 0.244 | 0.182 | 0.0616 |
| | CalYs | 0.731 | 0.263 | 0.238 | 0.182 | 0.0554 |
| | CalYsc | 0.729 | 0.261 | 0.237 | 0.182 | 0.0554 |
| (2) | WLE | 0.729 | 0.262 | 0.2441 | 0.183 | 0.0616 |
| | CalY | 0.730 | 0.263 | 0.2459 | 0.184 | 0.0623 |
| | CalYs | 0.730 | 0.258 | 0.2398 | 0.184 | 0.0559 |
| | CalYsc | 0.729 | 0.258 | 0.2394 | 0.183 | 0.0560 |
| (3) | WLE | 0.736 | 0.294 | 0.265 | 0.187 | 0.0784 |
| | CalY | 0.737 | 0.295 | 0.267 | 0.188 | 0.0789 |
| | CalXY | 0.737 | 0.299 | 0.269 | 0.187 | 0.0814 |
| | CalYs | 0.737 | 0.295 | 0.267 | 0.188 | 0.0789 |
| | CalXYs | 0.733 | 0.281 | 0.257 | 0.188 | 0.0687 |
| | CalYsc | 0.736 | 0.295 | 0.266 | 0.187 | 0.0784 |
| | CalXYsc | 0.729 | 0.270 | 0.249 | 0.187 | 0.0618 |
| (4) | WLE | 0.730 | 0.285 | 0.259 | 0.181 | 0.0780 |
| | CalY | 0.730 | 0.287 | 0.261 | 0.183 | 0.0784 |
| | CalXY | 0.729 | 0.290 | 0.264 | 0.183 | 0.0809 |
| | CalYs | 0.730 | 0.288 | 0.261 | 0.183 | 0.0784 |
| | CalXYs | 0.729 | 0.275 | 0.251 | 0.183 | 0.0679 |
| | CalYsc | 0.730 | 0.285 | 0.259 | 0.182 | 0.0777 |
| | CalXYsc | 0.726 | 0.265 | 0.243 | 0.182 | 0.0611 |

Table 5.12: Summary of simulations in Model 2 for $N = 250$ and $\theta = (\log 2, \log 2)$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalXY, CalYs, CalXYs, CalYsc, CalXYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $X$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $X$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $X$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.699 | 0.252 | NA | NA | NA |
| (1) | WLE | 0.780 | 0.484 | 0.401 | 0.295 | 0.1059 |
| | CalY | 0.780 | 0.492 | 0.404 | 0.298 | 0.1065 |
| | CalYs | 0.787 | 0.482 | 0.395 | 0.299 | 0.0966 |
| | CalYsc | 0.784 | 0.477 | 0.393 | 0.296 | 0.0967 |
| (2) | WLE | 0.783 | 0.546 | 0.413 | 0.308 | 0.1053 |
| | CalY | 0.782 | 0.551 | 0.411 | 0.304 | 0.1065 |
| | CalYs | 0.787 | 0.538 | 0.404 | 0.307 | 0.0966 |
| | CalYsc | 0.784 | 0.533 | 0.404 | 0.308 | 0.0965 |
| (3) | WLE | 0.781 | 0.538 | 0.439 | 0.308 | 0.131 |
| | CalY | 0.783 | 0.541 | 0.440 | 0.308 | 0.131 |
| | CalXY | 0.782 | 0.545 | 0.446 | 0.306 | 0.140 |
| | CalYs | 0.784 | 0.543 | 0.441 | 0.310 | 0.131 |
| | CalXYs | 0.783 | 0.521 | 0.418 | 0.304 | 0.114 |
| | CalYsc | 0.783 | 0.542 | 0.440 | 0.310 | 0.130 |
| | CalXYsc | 0.798 | 0.495 | 0.404 | 0.302 | 0.102 |
| (4) | WLE | 0.783 | 0.559 | 0.427 | 0.296 | 0.130 |
| | CalY | 0.783 | 0.564 | 0.426 | 0.296 | 0.130 |
| | CalXY | 0.781 | 0.576 | 0.430 | 0.292 | 0.138 |
| | CalYs | 0.784 | 0.565 | 0.427 | 0.297 | 0.130 |
| | CalXYs | 0.784 | 0.543 | 0.408 | 0.297 | 0.111 |
| | CalYsc | 0.782 | 0.565 | 0.425 | 0.296 | 0.129 |
| | CalXYsc | 0.803 | 0.500 | 0.399 | 0.298 | 0.101 |

Table 5.13: Summary of simulations in Model 2 for $N = 500$ and $\theta = (\log 2, \log 2)$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalXY, CalYs, CalXYs, CalYsc, CalXYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $X$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $X$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $X$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.691 | 0.114 | NA | NA | NA |
| (1) | WLE | 0.736 | 0.205 | 0.178 | 0.117 | 0.0614 |
| | CalY | 0.737 | 0.206 | 0.179 | 0.117 | 0.0616 |
| | CalYs | 0.738 | 0.203 | 0.174 | 0.117 | 0.0564 |
| | CalYsc | 0.736 | 0.201 | 0.174 | 0.117 | 0.0565 |
| (2) | WLE | 0.725 | 0.194 | 0.180 | 0.117 | 0.0631 |
| | CalY | 0.725 | 0.195 | 0.180 | 0.117 | 0.0633 |
| | CalYs | 0.728 | 0.189 | 0.175 | 0.117 | 0.0578 |
| | CalYsc | 0.726 | 0.188 | 0.175 | 0.117 | 0.0580 |
| (3) | WLE | 0.738 | 0.218 | 0.194 | 0.117 | 0.0764 |
| | CalY | 0.738 | 0.219 | 0.194 | 0.118 | 0.0765 |
| | CalXY | 0.738 | 0.220 | 0.196 | 0.118 | 0.0781 |
| | CalYs | 0.739 | 0.220 | 0.195 | 0.118 | 0.0766 |
| | CalXYs | 0.739 | 0.209 | 0.184 | 0.118 | 0.0661 |
| | CalYsc | 0.737 | 0.218 | 0.194 | 0.118 | 0.0762 |
| | CalXYsc | 0.738 | 0.203 | 0.179 | 0.118 | 0.0614 |
| (4) | WLE | 0.738 | 0.219 | 0.195 | 0.117 | 0.0776 |
| | CalY | 0.737 | 0.219 | 0.195 | 0.118 | 0.0776 |
| | CalXY | 0.737 | 0.221 | 0.197 | 0.118 | 0.0791 |
| | CalYs | 0.737 | 0.219 | 0.195 | 0.118 | 0.0776 |
| | CalXYs | 0.732 | 0.203 | 0.185 | 0.118 | 0.0672 |
| | CalYsc | 0.736 | 0.217 | 0.195 | 0.117 | 0.0773 |
| | CalXYsc | 0.735 | 0.195 | 0.180 | 0.118 | 0.0623 |

Table 5.14: Summary of simulations in Model 3 for $N = 250$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.699 | 0.316 | NA | NA | NA |
| (1) | WLE | 0.748 | 0.988 | 0.746 | 0.600 | 0.146 |
| | CalY | 0.745 | 0.996 | 0.700 | 0.552 | 0.148 |
| | CalYs | 0.747 | 1.000 | 0.702 | 0.555 | 0.147 |
| | CalYsc | 0.745 | 1.002 | 0.740 | 0.597 | 0.143 |
| (2) | WLE | 0.765 | 0.923 | 0.660 | 0.509 | 0.151 |
| | CalY | 0.767 | 0.935 | 0.630 | 0.478 | 0.153 |
| | CalYs | 0.768 | 0.940 | 0.633 | 0.482 | 0.151 |
| | CalYsc | 0.767 | 0.938 | 0.652 | 0.504 | 0.147 |
| (3) | WLE | 0.717 | 1.200 | 0.690 | 0.544 | 0.147 |
| | CalY | 0.720 | 1.214 | 0.680 | 0.532 | 0.147 |
| | CalVY | 0.719 | 1.225 | 0.672 | 0.519 | 0.153 |
| | CalYs | 0.720 | 1.216 | 0.679 | 0.531 | 0.147 |
| | CalVYs | 0.722 | 1.220 | 0.676 | 0.527 | 0.148 |
| | CalYsc | 0.717 | 1.217 | 0.690 | 0.545 | 0.145 |
| | CalVYsc | 0.723 | 1.309 | 0.669 | 0.524 | 0.145 |
| (4) | WLE | 0.795 | 1.911 | 0.880 | 0.732 | 0.148 |
| | CalY | 0.794 | 1.911 | 0.746 | 0.595 | 0.151 |
| | CalVY | 0.793 | 1.920 | 0.827 | 0.667 | 0.159 |
| | CalYs | 0.795 | 1.912 | 0.777 | 0.626 | 0.151 |
| | CalVYs | 0.798 | 1.934 | 0.843 | 0.691 | 0.152 |
| | CalYsc | 0.793 | 1.923 | 0.886 | 0.739 | 0.147 |
| | CalVYsc | 0.816 | 2.178 | 0.946 | 0.800 | 0.146 |

Table 5.15: Summary of simulations in Model 3 for $N = 500$. (1), (2), (3), (4) stand for the exposure stratified case cohort designs under sampling without replacement and Bernoulli sampling, and the classical case cohort design under sampling without replacement and Bernoulli sampling, respectively. WLE, CalY, CalVY, CalYs, CalVYs, CalYsc, CalVYsc stand for the WLE, the calibrated WLE on $Y$, the calibrated WLE on $V$ and $Y$, the within-stratum calibrated WLE on $Y$, the within-stratum calibrated WLE on $V$ and $Y$, the within-stratum centered calibrated WLE on $Y$ and the within-stratum centered calibrated WLE on $V$ and $Y$, respectively. Var (emp), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) stand for the empirical variance, the average of the reported variance, phase I variance and phase II variance from the *survey* package [29].

| Sampling | Estimator | Mean | Var (emp) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|---|
| | Full | 0.702 | 0.162 | NA | NA | NA |
| (1) | WLE | 0.721 | 0.288 | 0.259 | 0.181 | 0.0781 |
| | CalY | 0.722 | 0.290 | 0.260 | 0.181 | 0.0786 |
| | CalYs | 0.722 | 0.291 | 0.260 | 0.182 | 0.0784 |
| | CalYsc | 0.722 | 0.293 | 0.259 | 0.181 | 0.0776 |
| (2) | WLE | 0.723 | 0.296 | 0.2600 | 0.181 | 0.0791 |
| | CalY | 0.723 | 0.298 | 0.2615 | 0.182 | 0.0796 |
| | CalYs | 0.724 | 0.298 | 0.2617 | 0.182 | 0.0794 |
| | CalYsc | 0.724 | 0.298 | 0.2605 | 0.182 | 0.0787 |
| (3) | WLE | 0.740 | 0.284 | 0.262 | 0.184 | 0.0776 |
| | CalY | 0.740 | 0.285 | 0.263 | 0.185 | 0.0780 |
| | CalVY | 0.741 | 0.287 | 0.264 | 0.184 | 0.0793 |
| | CalYs | 0.740 | 0.285 | 0.263 | 0.185 | 0.0780 |
| | CalVYs | 0.741 | 0.286 | 0.264 | 0.185 | 0.0783 |
| | CalYsc | 0.741 | 0.284 | 0.262 | 0.185 | 0.0774 |
| | CalVYsc | 0.746 | 0.283 | 0.261 | 0.183 | 0.0776 |
| (4) | WLE | 0.718 | 0.287 | 0.262 | 0.184 | 0.0786 |
| | CalY | 0.717 | 0.288 | 0.263 | 0.184 | 0.0790 |
| | CalVY | 0.716 | 0.291 | 0.265 | 0.185 | 0.0803 |
| | CalYs | 0.717 | 0.289 | 0.263 | 0.184 | 0.0790 |
| | CalVYs | 0.718 | 0.290 | 0.263 | 0.184 | 0.0793 |
| | CalYsc | 0.719 | 0.287 | 0.262 | 0.184 | 0.0784 |
| | CalVYsc | 0.721 | 0.288 | 0.261 | 0.182 | 0.0785 |

exposure stratified case cohort designs under sampling without replacement and 45% 46% and 19% for the classical case cohort designs, respectively. Note that large improvements on efficiency at the second phase are required in order to see noticeable differences in the overall variances because the phase II variances are much smaller than the phase I variances in general.

Efficiency improvements are greatly influenced by different designs. First, the magnitudes of variance reductions differ between the exposure stratified and the classical cohort designs. As we discussed, the best possible efficiency improvements are larger in the classical case cohort designs. On the other hand, the best estimators in the classical case cohort designs are, at best, as efficient as the plain WLE in the exposure stratified case cohort designs. For instance, we see in Table 5.6 that the within-stratum centered calibration on $V$ and $Y$ in the classical case cohort design has empirical variance .159 while the plain WLE in the exposure stratified case cohort design has empirical variance .155.

Second, calibrations that gain efficiency the in one design may not improve efficiency in another design even if the same auxiliary variables are chosen for calibrations. Within-stratum calibrations on $Y$ achieved variance reduction over the plain WLE in the exposure case cohort designs while they gained no efficiency at all in the classical case cohort design. Instead, within-stratum (centered) calibrations on $V$ and $Y$ led to efficiency gains. Note that the usual calibration did not improve efficiency at all for all cases.

Third, adjusting weights may have no effects on variance reduction when strata formation based on an uncorrelated auxiliary variables. Adjusting weights did not affect the performance of estimators in Tables 5.8 and 5.9 where the auxiliary variable $V$ is independent of $X$, $Y$, $\Delta$. This result indicates that it is important to consider which variables to be collected for stratification at the first phase when planning a two-phase design.

Lastly, when the exposure of interest is collected at the second phase and a stratifying variable is a confounder in the outcome model (Cox model), adjusting weights may not influence efficiency gain. This is a result from Model 3 (Tables 5.14 and 5.14).

Table 5.16: R outputs and bootstrap results based one data set in Model 1 of the exposure stratified case cohort designs for $\alpha = \beta = .9$. Mean (R), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) are outputs from the R *survey* package [29] for mean, variance, the phase I and phase II variances. Mean (tp), Var(tp), Mean (svy), Var(svy) are mean and variance from the two-phase and survey bootstraps, respectively.

| | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|
| N=250 | WLE | 1.43 | 0.167 | 0.132 | 0.0352 |
| | CalY | 1.43 | 0.167 | 0.132 | 0.0348 |
| | CalYs | 1.34 | 0.159 | 0.132 | 0.0268 |
| | CalYsc | 1.33 | 0.159 | 0.132 | 0.0271 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 1.47 | 0.1777 | 1.44 | 0.0368 |
| | CalY | 1.47 | 0.1793 | 1.44 | 0.0375 |
| | CalYs | 1.38 | 0.1628 | 1.35 | 0.0289 |
| | CalYsc | 1.38 | 0.1618 | 1.34 | 0.0295 |
| N=500 | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
| | WLE | 0.629 | 0.0705 | 0.0597 | 0.01072 |
| | CalY | 0.629 | 0.0705 | 0.0597 | 0.01076 |
| | CalYs | 0.543 | 0.0660 | 0.0598 | 0.00626 |
| | CalYsc | 0.547 | 0.0663 | 0.0600 | 0.00633 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 0.638 | 0.0721 | 0.633 | 0.01131 |
| | CalY | 0.638 | 0.0728 | 0.633 | 0.01142 |
| | CalYs | 0.554 | 0.0608 | 0.548 | 0.00687 |
| | CalYsc | 0.558 | 0.0646 | 0.552 | 0.00680 |

Table 5.17: R outputs and bootstrap results based one data set in Model 1 of the exposure stratified case cohort designs for $\alpha = \beta = .5$. Mean (R), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) are outputs from the R *survey* package [29] for mean, variance, the phase I and phase II variances. Mean (tp), Var(tp), Mean (svy), Var(svy) are means and variances from the two-phase and survey bootstraps, respectively.

| | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|
| N=250 | WLE | 1.03 | 0.183 | 0.132 | 0.0514 |
| | CalY | 1.02 | 0.189 | 0.133 | 0.0551 |
| | CalYs | 1.04 | 0.189 | 0.134 | 0.0548 |
| | CalYsc | 1.05 | 0.188 | 0.134 | 0.0537 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 1.07 | 0.187 | 1.04 | 0.0547 |
| | CalY | 1.05 | 0.193 | 1.03 | 0.0598 |
| | CalYs | 1.07 | 0.198 | 1.05 | 0.0604 |
| | CalYsc | 1.09 | 0.191 | 1.06 | 0.0599 |
| N=500 | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
| | WLE | 0.831 | 0.0845 | 0.0598 | 0.0247 |
| | CalY | 0.830 | 0.0848 | 0.0599 | 0.0249 |
| | CalYs | 0.832 | 0.0846 | 0.0599 | 0.0247 |
| | CalYsc | 0.836 | 0.0843 | 0.0599 | 0.0244 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 0.843 | 0.0878 | 0.829 | 0.0248 |
| | CalY | 0.841 | 0.0884 | 0.828 | 0.0252 |
| | CalYs | 0.845 | 0.0882 | 0.830 | 0.0253 |
| | CalYsc | 0.851 | 0.0849 | 0.835 | 0.0253 |

Table 5.18: R outputs and bootstrap results based one data set in Model 2 of the exposure stratified case cohort designs for $\theta = (\log 2, 0)$. Mean (R), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) are outputs from the R *survey* package [29] for mean, variance, the phase I and phase II variances. Mean (tp), Var(tp), Mean (svy), Var(svy) are means and variances from the two-phase and survey bootstraps, respectively.

| | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|
| N=250 | WLE | 1.59 | 0.871 | 0.536 | 0.334 |
| | CalY | 1.69 | 0.774 | 0.474 | 0.300 |
| | CalYs | 1.62 | 0.761 | 0.479 | 0.282 |
| | CalYsc | 1.53 | 0.813 | 0.523 | 0.290 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 1.86 | 1.146 | 1.86 | 0.670 |
| | CalY | 1.92 | 0.979 | 1.92 | 0.541 |
| | CalYs | 1.87 | 0.968 | 1.87 | 0.541 |
| | CalYsc | 1.86 | 1.008 | 1.81 | 0.610 |
| N=500 | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
| | WLE | 0.863 | 0.242 | 0.190 | 0.0528 |
| | CalY | 0.869 | 0.244 | 0.191 | 0.0527 |
| | CalYs | 0.842 | 0.244 | 0.193 | 0.0512 |
| | CalYsc | 0.854 | 0.245 | 0.194 | 0.0506 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 0.852 | 0.260 | 0.891 | 0.0630 |
| | CalY | 0.857 | 0.265 | 0.897 | 0.0636 |
| | CalYs | 0.828 | 0.257 | 0.873 | 0.0621 |
| | CalYsc | 0.831 | 0.264 | 0.882 | 0.0616 |

Table 5.19: R outputs and bootstrap results based one data set in Model 2 of the exposure stratified case cohort designs for $\theta = (\log 2, \log 2)$. Mean (R), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) are outputs from the R *survey* package [29] for mean, variance, the phase I and phase II variances. Mean (tp), Var(tp), Mean (svy), Var(svy) are mean and variance from the two-phase and survey bootstraps, respectively.

| | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|
| N=250 | WLE | 1.18 | 0.361 | 0.161 | 0.199 |
| | CalY | 1.18 | 0.362 | 0.163 | 0.200 |
| | CalYs | 1.17 | 0.397 | 0.176 | 0.221 |
| | CalYsc | 1.13 | 0.416 | 0.181 | 0.234 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 1.31 | 0.401 | 1.33 | 0.270 |
| | CalY | 1.31 | 0.407 | 1.33 | 0.276 |
| | calYs | 1.33 | 0.450 | 1.36 | 0.327 |
| | calYsc | 1.32 | 0.473 | 1.33 | 0.356 |
| N=500 | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
| | WLE | 0.578 | 0.192 | 0.117 | 0.0747 |
| | CalY | 0.668 | 0.203 | 0.120 | 0.0828 |
| | CalYs | 0.464 | 0.196 | 0.123 | 0.0733 |
| | CalYsc | 0.505 | 0.204 | 0.128 | 0.0753 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 0.598 | 0.254 | 0.615 | 0.108 |
| | CalY | 0.690 | 0.270 | 0.704 | 0.122 |
| | calYs | 0.490 | 0.252 | 0.499 | 0.115 |
| | calYsc | 0.532 | 0.257 | 0.543 | 0.113 |

Table 5.20: R outputs and bootstrap results based one data set in Model 3 of the exposure stratified case cohort designs. Mean (R), $\text{Var}_{\text{total}}$(R), $\text{Var}_1$(R), $\text{Var}_2$(R) are outputs from the R *survey* package [29] for mean, variance, the phase I and phase II variances. Mean (tp), Var(tp), Mean (svy), Var(svy) are means and variances from the two-phase and survey bootstraps, respectively.

| | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
|---|---|---|---|---|---|
| N=250 | WLE | 0.879 | 0.674 | 0.533 | 0.140 |
| | CalY | 0.770 | 0.728 | 0.601 | 0.127 |
| | CalYs | 0.768 | 0.740 | 0.610 | 0.129 |
| | CalYsc | 0.854 | 0.691 | 0.555 | 0.136 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | 0.888 | 0.953 | 0.945 | 0.212 |
| | CalY | 0.763 | 0.953 | 0.826 | 0.180 |
| | CalYs | 0.760 | 0.969 | 0.825 | 0.185 |
| | CalYsc | 0.855 | 0.863 | 0.918 | 0.206 |
| N=500 | Estimator | Mean (R) | $\text{Var}_{\text{total}}$(R) | $\text{Var}_1$(R) | $\text{Var}_2$(R) |
| | WLE | -0.029 | 0.203 | 0.147 | 0.0561 |
| | CalY | -0.029 | 0.203 | 0.147 | 0.0561 |
| | CalYs | -0.002 | 0.202 | 0.148 | 0.0536 |
| | CalYsc | -0.012 | 0.203 | 0.150 | 0.0527 |
| | | Mean (tp) | Var(tp) | Mean (svy) | Var(svy) |
| | WLE | -0.036 | 0.244 | -0.043 | 0.0768 |
| | CalY | -0.040 | 0.244 | -0.050 | 0.0771 |
| | CalYs | -0.005 | 0.234 | -0.018 | 0.0753 |
| | CalYsc | -0.006 | 0.230 | -0.024 | 0.0748 |

Table 5.21: Bootstrap results based on four different data sets in Model 2 of the exposure stratified case cohort designs for with $\theta = (\log 2, 0)$. $\text{Var}_{\text{total}}(\text{R})$, $\text{Var}_1(\text{R})$, $\text{Var}_2(\text{R})$ are outputs from the R *survey* package [29] for variance, the phase I and phase II variances. Var(tp), Mean (svy), Var(svy) are variances from the two-phase and survey bootstraps, respectively.

| | | Original | | | Bootstrap | |
|---|---|---|---|---|---|---|
| | Estimator | $\text{Var}_{\text{total}}(\text{R})$ | $\text{Var}_1(\text{R})$ | $\text{Var}_2(\text{R})$ | Var (tp) | Var(svy) |
| (1) | WLE | 0.117 | 0.0729 | 0.0437 | 0.139 | 0.0567 |
| | CalY | 0.121 | 0.0732 | 0.0480 | 0.146 | 0.0626 |
| | CalYs | 0.105 | 0.0735 | 0.0319 | 0.114 | 0.0441 |
| | CalYsc | 0.106 | 0.0735 | 0.0325 | 0.117 | 0.0439 |
| (2) | WLE | 0.186 | 0.135 | 0.0511 | 0.212 | 0.0688 |
| | CalY | 0.186 | 0.132 | 0.0549 | 0.217 | 0.0762 |
| | CalYs | 0.171 | 0.132 | 0.0388 | 0.189 | 0.0569 |
| | CalYsc | 0.170 | 0.133 | 0.0367 | 0.191 | 0.0539 |
| (3) | WLE | 0.408 | 0.320 | 0.0881 | 0.452 | 0.150 |
| | CalY | 0.406 | 0.312 | 0.0938 | 0.466 | 0.165 |
| | CalYs | 0.401 | 0.320 | 0.0815 | 0.455 | 0.150 |
| | CalYsc | 0.402 | 0.320 | 0.0815 | 0.446 | 0.141 |
| (4) | WLE | 0.804 | 0.766 | 0.0378 | 0.782 | 0.0396 |
| | CalY | 0.676 | 0.634 | 0.0418 | 0.664 | 0.0457 |
| | CalYs | 0.675 | 0.633 | 0.0417 | 0.660 | 0.0439 |
| | CalYsc | 0.738 | 0.696 | 0.0420 | 0.696 | 0.0455 |

*Separate Variance Estimation*

The efficient score for the complete data can be estimated from the derivative of the weighted partial likelihood with respect to the regression coefficients as it is estimated from the partial likelihood in the complete data model. Thus, the efficient information for the complete data can be estimated from the Horvitz-Thompson estimator of the square of the estimated efficient score for the complete data. This method is implemented in the R *survey* package [29] and is justified by Theorem 4.1.1.

The performance of this method can be measured by comparison of the variance of the MLE for the complete data and the R output for the phase I variance in Tables 5.6 - 5.15. These two are in good agreement in the Model 1 and the Model 2 with $N = 500$ and $\theta = (\log 2, \log 2)$. However, the method overestimated the phase I variance in other simulations. For example, the estimated phase I variances for the plain WLE for the exposure stratified case cohort design in Table 5.10 are .582 for $N = 250$ and .182 for $N = 500$ while the variance of the MLE for the complete data are .335 for $N = 250$ and .169 for $N = 500$. These bias became smaller as sample sizes increased, but sufficient sample sizes vary greatly from model to model.

The phase II variances for the exposure stratified case cohort design under sampling without replacement are reported in Tables 5.16-5.21. Since the empirical variances and the estimated variances from the R outputs are in good agreement in the Model 1 as well as the empirical variances of the MLE and the estimated phase I variances are, the estimated phase II variances in the Model 1 are reliable. Tables 5.16 and 5.17 show that estimated phase II variances from the survey bootstrap are in good agreement with the outputs for a single data set on which the survey bootstrap is based and the results in Tables 5.6-5.9.

For other cases, we observe that the estimated total variances from the R outputs are generally smaller than the empirical variances (Tables 5.10- 5.15) while the estimated phase I variance from the R outputs tend to be larger than the truth (Tables 5.10- 5.15 except Table 5.13). This indicates that the estimated phase II variances are greatly underestimated. Tables 5.18-5.20 show that the estimated phase II variances from the survey bootstrap are larger than the phase II variances reported from the R outputs. This suggests that the survey bootstrap corrected the reported underestimated phase II variances but how precise the corrections are is unknown. To study this question is difficult since the bootstrap depends on the original data. That is, if the original data where the bootstrap is based are extreme enough by chance, a perfect correction of the bias by bootstrap is unlikely. We address this questions below after discussing results of the two-phase bootstrap.

*Simultaneous Variance Estimation*

The two-phase bootstrap estimated variances sufficiently close to the reported variances from the R output and the results in Tables 5.6-5.9. Recall that total variances, and phase I and phase II variances are well estimated from the Horvitz-Thompson estimators in these cases. In other words, $N = 250$ is sufficiently large for the Model 1. Thus, good performance of the two-phase bootstrap in the Model 1 means that this method performs very well for sufficiently large sample sizes.

As mentioned above, the total variances reported from the R outputs are smaller than the empirical ones in other cases, or for insufficient sample sizes. As the survey bootstrap, the two-phase bootstrap yielded the estimates larger than the reported ones as seen in Tables 5.18-5.20. Thus, the two-phase bootstrap also captured the enough variances not reported from the R output, too.

Table 5.21 shows the results for our bootstrap methods based on four different original data sets in the Model 2 with $N = 500$ and $\theta = (\log 2, \log 2)$. We observe from Table 5.18 that if the original data set is a good representative of the population (see the R outputs in Table 5.18 and results in Table 5.11 are in good agreement), results from the both bootstrap methods yielded reliable variance estimates. The two-phase bootstrap did not provide reliable estimates for all four cases in Table 5.21. This method tried to correct the bias upward for the cases where reported variances are (1) too small, (2) moderately small, (3) moderately large, while estimates from the two-phase bootstrap are smaller than reported ones for the case where the reported variance is too large (4). The survey bootstrap reliably estimated the phase II variances for the cases where the reported variances are small (1 and 2) while it overestimated the phase II variances for the case of (3) and (4). This method yielded larger estimates than reported ones for all four cases. As the sample sizes increases, the probability of getting a data set of “good representative” of the population becomes larger and our bootstrap method became more likely to perform reliably as the “in probability” statement of bootstrap $Z$-theorems (Theorems 4.1.2 and 4.1.3) expect. However, the finite sample performance greatly depends on the original data set for bootstrap.

### 5.3 Data Analysis

We analyze data from the National Wilms Tumor Study [14, 17]. The data set is available online from the website of Dr. Norman E. Breslow (http://faculty.washington.edu/norm/IEA08.html). In this study, 3915 patients with Wilms tumor diagnosed during 1980-1994 were followed until the disease progression or death. The baseline covariates are age at diagnosis, stage of disease (I-IV), histology (favorable versus unfavorable) from the registering institution and the central reference laboratory, and tumor diameter. Although information on outcomes and baseline covariates are known for all patients, we take a subsample from this study to create an artificial two-phase design as considered in [5, 6]. Specifically, we create nine strata based on age (less than or greater than one year of age), severity of stage (I-II versus III-IV), and institutional histology in addition to a censoring indicator. Moreover, we treat histology from the central reference laboratory as the gold standard (sensitivity 74% and specificity 98%) and known for patients sampled at the second phase. Though there are nine strata, all patients are sampled except three strata. For the first stratum, 120 patients are sampled from 452 patients with favorable institutional histology, stage I or II and less than one year of age. For the second stratum, 160 patients are sampled from 1620 patients with favorable institutional histology, stage I or II and greater than one year of age. For the third stratum, 120 patients are sampled fro-rm 914 patients with favorable institutional histology, stage III or IV and greater than one year of age. The overall phase II sample size is 1329.

The estimators we consider are the plain WLE and the within-stratum centered calibrated WLE on the time to event or censoring, stage (I-IV), and age (continuous). We compare these two estimators with the calibrated WLE with imputation proposed by [27] which was considered in [5, 6]. The statistical model considered is the same as that considered in [5, 6].

Table 5.22: R outputs for one data set for the MLE for the complete data, the plain WLE, the within-stratum centered calibrated WLE on the time to event or censoring, stage, and age, and the calibrated WLE with imputation of [27]. UH stands for unfavorable histology from the central reference laboratory,age1 and age2, piecewise linear terms for age at diagnosis (years) before and after 1 year, respectively, stg34 is a indicator of the stage III-IV, tumdiam is tumor diameter, stgdiam is a stage times tumor diameter, and UH:age1 and UH:age2 are interaction between UH and age1 or age2.

| | Full | WLE | Cal | dfbeta |
|---|---|---|---|---|
| UH | 4.042 (0.413) | 4.136 (0.532) | 4.264 (0.516) | 4.190 (0.525) |
| age1 | -0.661 (0.326) | -0.742 (0.372) | -0.640 (0.333) | -0.603 (0.337) |
| age2 | 0.104 (0.017) | 0.095 (0.026) | 0.107 (0.018) | 0.100 (0.017) |
| stg34 | 1.346 (0.244) | 1.028 (0.350) | 1.019 (0.347) | 1.399 (0.277) |
| tumdiam | 0.069 (0.014) | 0.058 (0.020) | 0.057 (0.020) | 0.074 (0.015) |
| stgdiam | -0.076 (0.019) | -0.050 (0.029) | -0.050 (0.028) | -0.082 (0.022) |
| UH:age1 | -2.635 (0.46) | -2.965 (0.617) | -3.097 (0.601) | -3.023 (0.605) |
| UH:age2 | -0.058 (0.034) | -0.006 (0.056) | -0.012 (0.054) | -0.010 (0.053) |

Table 5.22 summarizes results from three estimators for one simulated data set. All point estimates from three estimators are similar to each other and their 95% confidence intervals all include point estimates of the MLE based on the complete data. The within-stratum centered calibration and the method of [27] improved efficiency over the plain WLE. Improvements by these methods are similar for estimation of most of coefficients, but our method sometimes gained better efficiency gain (UH), though improvement on estimation of stg34 is large for the method of [27]. Tables 5.23 and 5.24 are results for the variance estimation for the plain WLE and the within-stratum centered calibrated WLE based on two-phase and survey bootstraps. Our bootstrap results agree well with the R outputs. Although the method of [27] was implemented via calibration as in [5, 6], a variable for calibration is estimated by imputation of a missing variable (in this case, UH). Thus, variance estimation for this method is outside the scope of our bootstrap-based variance

Table 5.23: Variance estimation of the plain WLE from two-phase and survey bootstraps. sd(tp) and sd(svy) denote standard deviations of estimators based on two-phase and survey bootstrap methods, respectively. $\text{sd}_{\text{total}}$(R) and $\text{sd}_2$(R) stand for the reported total and phase II standard deviations from the R *survey* package [29], respectively. The bootstrap samples were generated 2000 times.

| | sd(tp) | $\text{sd}_{\text{total}}$(R) | sd(svy) | $\text{sd}_2$(R) |
|---|---|---|---|---|
| UH | 0.549 | 0.532 | 0.169 | 0.165 |
| age1 | 0.381 | 0.372 | 0.176 | 0.169 |
| age2 | 0.027 | 0.026 | 0.021 | 0.020 |
| stg34 | 0.356 | 0.350 | 0.222 | 0.222 |
| tumdiam | 0.020 | 0.020 | 0.014 | 0.014 |
| stgdiam | 0.029 | 0.029 | 0.020 | 0.020 |
| UH:age1 | 0.638 | 0.617 | 0.277 | 0.275 |
| UH:age2 | 0.055 | 0.056 | 0.045 | 0.045 |

Table 5.24: Variance estimation of the calibrated WLE from two-phase and survey bootstraps. sd(tp) and sd(svy) denote standard deviations of estimators based on two-phase and survey bootstrap methods, respectively. $\text{sd}_{\text{total}}$(R) and $\text{sd}_2$(R) stand for the reported total and phase II standard deviations from the R *survey* package [29], respectively. The bootstrap samples were generated 2000 times.

| | sd(tp) | $\text{sd}_{\text{total}}$(R) | sd(svy) | $\text{sd}_2$(R) |
|---|---|---|---|---|
| UH | 0.503 | 0.516 | 0.116 | 0.119 |
| age1 | 0.329 | 0.333 | 0.080 | 0.080 |
| age2 | 0.014 | 0.018 | 0.0010 | 0.009 |
| stg34 | 0.347 | 0.347 | 0.222 | 0.220 |
| tumdiam | 0.020 | 0.020 | 0.014 | 0.014 |
| stgdiam | 0.029 | 0.028 | 0.019 | 0.019 |
| UH:age1 | 0.595 | 0.601 | 0.243 | 0.246 |
| UH:age2 | 0.052 | 0.054 | 0.042 | 0.043 |

estimation methods. In fact, bootstrap-based methods did not yield enough variances in our simulations (not shown).

Chapter 6

# GENERAL RESULTS FOR IPW EMPIRICAL PROCESSES

The IPW empirical measure and IPW empirical process inherit important properties from the empirical measure and empirical process, respectively. We emphasize the similarity between empirical processes and IPW empirical processes. All proofs are presented in the end of this chapter.

## *6.1 Glivenko-Cantelli theorem*

The next theorem states that the Glivenko-Cantelli property for complete data is preserved under two-phase sampling.

**Theorem 6.1.1.** *Suppose that $\mathcal{F}$ is a $P_0$-Glivenko-Cantelli class. Then*

$$\|\mathbb{P}_N^\pi - P_0\|_{\mathcal{F}} \to_{P^*} 0 \tag{6.1}$$

*where $\|\cdot\|_{\mathcal{F}}$ is the supremum norm. This also holds if we replace $\mathbb{P}_N^\pi$ by $\mathbb{P}_N^{\pi,e}$, $\mathbb{P}_N^{\pi,c}$, $\mathbb{P}_N^{\pi,mc}$ or $\mathbb{P}_N^{\pi,cc}$, assuming Conditions 3.1.1 and 3.1.2.*

## *6.2 Rate of convergence*

The rate of convergence of an $M$-estimator with complete data is often established via maximal inequalities for the empirical processes. If we follow the same line of reasoning, it is natural to derive maximal inequalities for IPW empirical processes, though this may require some efforts. Fortunately, these maximal inequalities for empirical processes (or slight modifications of them) suffice to establish the same rate of convergence under two-phase sampling.

**Theorem 6.2.1.** *Let $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ be the set of criterion functions and define $\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}$ for some fixed $\delta > 0$ where $d$ is a semimetric on the parameter space $\Theta$.*

*(1) Suppose that for every $\theta$ in a neighborhood of $\theta_0$,*

$$P_0(m_\theta - m_{\theta_0}) \lesssim -d^2(\theta, \theta_0); \tag{6.2}$$

*here $a \lesssim b$ means $a \leq Kb$ for some constant $K \in (0, \infty)$. Assume that there exists a function $\phi_N$ such that $\delta \mapsto \phi_N(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ (not depending on $N$) and for every $N$,*

$$E^*\|\mathbb{G}_N\|_{\mathcal{M}_\delta} \lesssim \phi_N(\delta), \tag{6.3}$$

*where $\mathbb{G}_N$ is the empirical process. If an estimator $\hat{\theta}_N$ satisfying $\mathbb{P}_N^\pi m_{\hat{\theta}_N} \geq \mathbb{P}_N^\pi m_{\theta_0} - O_{P^*}(r_N^{-2})$ converges in outer probability to $\theta_0$, then $r_N d(\hat{\theta}_N, \theta_0) = O_{P^*}(1)$ for every sequence $r_N$ such that $r_N^2 \phi_N(1/r_N) \leq \sqrt{N}$ for every $N$.*

*(2) Suppose Condition 3.1.2 holds. Suppose also that for every $\theta \in \Theta$ in a neighborhood of $\theta_0$,*

$$P_0\{G_{mc}(V; \alpha)(m_\theta - m_{\theta_0})\} \lesssim -d^2(\theta, \theta_0) + |\alpha - \alpha_0|^2. \tag{6.4}$$

*Assume that*

$$E^*\|\mathbb{G}_N\|_{G\mathcal{M}_\delta} \lesssim \phi_N(\delta), \tag{6.5}$$

*where $G\mathcal{M}_\delta \equiv \{G_{mc}(\cdot; \alpha)f : |\alpha| \leq \delta, \alpha \in \mathcal{A}_N, f \in \mathcal{M}_\delta\}$ for some $\mathcal{A}_N \subset \mathcal{A}_{mc}$. If an estimator with modified calibration, $\hat{\theta}_{N,mc}$, satisfying $\mathbb{P}_N^{\pi,mc} m_{\hat{\theta}_{N,mc}} \geq \mathbb{P}_N^{\pi,mc} m_{\theta_0} - O_{P^*}(r_N^{-2})$ converges in outer probability to $\theta_0$, then $r_N d(\hat{\theta}_{N,mc}, \theta_0) = O_{P^*}(1)$ for every sequence $r_N$ such that $r_N^2 \phi_N(1/r_N) \leq \sqrt{N}$ for every $N$.*

*(3) Suppose Conditions 3.1.1 and 3.1.2 hold. Under the same conditions of (2) with $G_{mc}$ replaced by $\pi_0/G_e$, $G_c$ or $G_{cc}$, the same conclusions hold for an estimator with estimated weights, $\hat{\theta}_{N,e}$, satisfying $\mathbb{P}_N^{\pi,e} m_{\hat{\theta}_{N,e}} \geq \mathbb{P}_N^{\pi,e} m_{\theta_0} - O_{P^*}(r_N^{-2})$, an estimator with calibration, $\hat{\theta}_{N,c}$, satisfying $\mathbb{P}_N^{\pi,c} m_{\hat{\theta}_{N,c}} \geq \mathbb{P}_N^{\pi,c} m_{\theta_0} - O_{P^*}(r_N^{-2})$, and an estimator with centered calibration, $\hat{\theta}_{N,cc}$, satisfying $\mathbb{P}_N^{\pi,cc} m_{\hat{\theta}_{N,cc}} \geq \mathbb{P}_N^{\pi,cc} m_{\theta_0} - O_{P^*}(r_N^{-2})$, respectively.*

**Remark 6.2.1.** *The key to establishing a general theorem for the rate of convergence is to make use of the boundedness of the weights in the IPW empirical process and also deal with the dependence of the weights. In treating independent bootstrap weights in the weighted bootstrap [31], Lemmas 1-3, require the boundedness of bootstrap weights, because the product of an unbounded weight and a bounded function is no longer bounded. Our theorem* *exploits the boundedness of sampling indicators in the IPW empirical processes by applying a multiplier inequality for the case of bounded weights (Lemma 6.2.1) to cover more general cases.*

The following is a multiplier inequality for bounded exchangeable weights. Note that the sum of stochastic processes in the second term is divided by $n^{1/2}$ rather than $k^{1/2}$.

**Lemma 6.2.1.** *(1) For i.i.d. stochastic processes $Z_1, \ldots, Z_n$, every bounded, exchangeable random vector $(\xi_1, \ldots, \xi_n)$ with each $\xi_i \in [l, u]$ that is independent of $Z_1, \ldots, Z_n$, and any $1 \le n_0 \le n$,*

$$E\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_i\right\|_{\mathcal{F}}^* \le \frac{2(n_0-1)}{n}\sum_{i=1}^n E^*\|Z_i\|_{\mathcal{F}} E\max_{1\le i\le n}\frac{\xi_i}{\sqrt{n}} + 2(u-l)\max_{n_0\le k\le n} E\left\|\frac{1}{\sqrt{n}}\sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^*.$$

*(2) Let $Z_1, \ldots, Z_n$ be i.i.d. stochastic processes with $E\|Z_i\|_{\mathcal{F}} < \infty$ independent of the Rademacher variables $\epsilon_1, \ldots, \epsilon_n$. Then for every i.i.d. sample $\xi_1, \ldots, \xi_n$ of bounded mean zero random variables independent of $Z_1, \ldots, Z_n$ and any $1 \le n_0 \le n$ with $l \le \xi_i \le u$*

$$E^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_i\right\|_{\mathcal{F}} \le 2(n_0-1)E^*\|Z_1\|_{\mathcal{F}} E\max_{1\le i\le n}\frac{|\xi_i|}{\sqrt{n}} + 2(u-l)\max_{n_0\le k\le n} E^*\left\|\frac{1}{\sqrt{n}}\sum_{i=n_0}^k \epsilon_i Z_i\right\|_{\mathcal{F}}.$$

The bound (6.5) is not difficult to verify in the presence of the bound (6.3) since $G_{mc}(\cdot;\alpha)$ is a bounded monotone function indexed by a finite dimensional parameter. The bound (6.4) may be verified through the lemma below for some applications including the Cox model with current status data.

**Lemma 6.2.2.** *Suppose Conditions 3.1.1 and 3.1.2 hold. Let $m_\theta$ be the log likelihood $\log p_\theta$ where $p_\theta$ is the density with dominating measure $\mu$, and $d$ is the Hellinger distance. Then the bound (6.4) and the corresponding bounds for the WLE's with estimated weights and (centered) calibration hold.*

### 6.3 Donsker theorem

The next theorem yields weak convergence of the IPW empirical processes under sampling without replacement.

**Theorem 6.3.1** (Donsker theorem for two-phase sampling)**.** *Suppose that $\mathcal{F}$ with $\|P_0\|_{\mathcal{F}} < \infty$ is a $P_0$-Donsker class and Conditions 3.1.1 and 3.1.2 hold. Then,*

$$\mathbb{G}_N^{\pi} \rightsquigarrow \mathbb{G}^{\pi} \equiv \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j, \tag{6.6}$$

$$\mathbb{G}_N^{\pi,e} \rightsquigarrow \mathbb{G}^{\pi,e} \equiv \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_e \cdot), \tag{6.7}$$

$$\mathbb{G}_N^{\pi,c} \rightsquigarrow \mathbb{G}^{\pi,c} \equiv \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_c \cdot), \tag{6.8}$$

$$\mathbb{G}_N^{\pi,mc} \rightsquigarrow \mathbb{G}^{\pi,mc} \equiv \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_{mc} \cdot), \tag{6.9}$$

$$\mathbb{G}_N^{\pi,cc} \rightsquigarrow \mathbb{G}^{\pi,cc} \equiv \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_{cc} \cdot), \tag{6.10}$$

*in $\ell^\infty(\mathcal{F})$ where the $P_0$-Brownian bridge process, $\mathbb{G}$, indexed by $\mathcal{F}$ and the $P_{0|j}$-Brownian bridge processes, $\mathbb{G}_j$, indexed by $\mathcal{F}$ are all independent.*

**Remark 6.3.1.** *The integrability hypothesis $\|P_0\|_{\mathcal{F}} < \infty$ is only required for the IPW empirical processes with estimated weights and (modified and centered) calibration.*

**Remark 6.3.2.** *An immediate application of Theorem 6.3.1 is establishing that the weighted Kaplan-Meier estimator and the weighted Nelson-Aalen estimator under two-phase sampling are asymptotically Gaussian. To see this, establish weak convergence of $\mathbb{P}_N^{\pi,\#} I(T \leq t)$ and $\mathbb{P}_N^{\pi,\#} \Delta I(T \leq t)$ via Theorem 6.3.1 with $\#$ is null or $\# \in \{c, mc, cc\}$ and apply the functional delta method. Here $T$ is time to event and $\Delta$ is a consoring indicator.*

If the index set $\mathcal{F}$ is Donsker, then it follows by the previous theorem and Lemma 2.3.11 of [58] that asymptotic equicontinuity in probability and in mean follows for the metric that depends on the limit process. In applications, it is of interest to have these results for the original metric $\rho_{P_0}(f, g) = \sigma_{P_0}(f - g)$.

**Theorem 6.3.2.** *Let $\mathcal{F}$ be Donsker and define $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \rho_{P_0}(f,g) < \delta\}$ for some fixed $\delta > 0$. Then, for every sequence $\delta_N \downarrow 0$,*

$$E^* \left\|\mathbb{G}_N^\pi\right\|_{\mathcal{F}_{\delta_N}} \to 0,$$

*and consequently, $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$. Moreover, $\left\|\mathbb{G}_N^{\pi,\#}\right\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$ for $\# \in \{e, c, mc, cc\}$ assuming Conditions 3.1.1 and 3.1.2.*

We end this chapter with two important lemmas. The first lemma is an extension of Lemma 3.3.5 of [58] and will be used in our proof of Theorem 3.2.1 to verify asymptotic equicontinuity.

**Lemma 6.3.1.** *Suppose $\mathcal{F} = \{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\}$ is $P_0$-Donsker for some $\delta > 0$ and that $\sup_{h\in\mathcal{H}} P_0(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0$, as $\theta \to \theta_0$. If $\hat{\theta}_N$ converges in outer probability to $\theta_0$, then*

$$\left\|\mathbb{G}_N^\pi(\psi_{\hat{\theta}_N,h} - \psi_{\theta_0,h})\right\|_{\mathcal{H}} = o_{P^*}(1).$$

*This also holds if we replace $\mathbb{G}_N^\pi$ by $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$, $\mathbb{G}_N^{\pi,mc}$ or $\mathbb{G}_N^{\pi,cc}$ assuming Conditions 3.1.1 and 3.1.2. hold and $\|P_0\|_{\mathcal{F}} < \infty$.*

The second lemma is used to verify asymptotic equicontinuity in the proof of Theorem 3.3.1, the first part for the IPW empirical process and the second part for the other four IPW empirical processes with estimated weights and (modified and centered) calibration.

**Lemma 6.3.2.** *Let $\mathcal{F}_N$ be a sequence of decreasing classes of functions such that $\|\mathbb{G}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$. Assume that there exists an integrable envelope for $\mathcal{F}_{N_0}$ for some $N_0$. Then $E\|\mathbb{G}_N\|_{\mathcal{F}_N} \to 0$ as $N \to \infty$. As a consequence, $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_N} = o_{P^*}(1)$.*

*Suppose, moreover, that $\mathcal{F}_N$ is $P_0$-Glivenko-Cantelli with $\|P_0\|_{\mathcal{F}_{N_1}} < \infty$ for some $N_1$, and that every $f = f_N \in \mathcal{F}_N$ converges to zero either pointwise or in $L_1(P_0)$ as $N \to \infty$. Then $\|\mathbb{G}_N^{\pi,e}\|_{\mathcal{F}_N} = o_{P^*}(1)$, $\|\mathbb{G}_N^{\pi,c}\|_{\mathcal{F}_N} = o_{P^*}(1)$, $\|\mathbb{G}_N^{\pi,mc}\|_{\mathcal{F}_N} = o_{P^*}(1)$ and $\|\mathbb{G}_N^{\pi,cc}\|_{\mathcal{F}_N} = o_{P^*}(1)$, assuming Conditions 3.1.1 and 3.1.2.*

### *6.4 Proofs*

*Proof of Theorem 6.1.1.* First consider $\mathbb{P}_N^\pi$. By the decomposition (2.1) of the IPW empirical processes (see also [8]), we have

$$\|\mathbb{P}_N^\pi - P_0\|_{\mathcal{F}} \leq \|\mathbb{P}_N - P_0\|_{\mathcal{F}} + \sum_{j=1}^J \frac{N_j}{N}\frac{N_j}{n_j}\left\|\mathbb{P}_{j,N_j}^\xi - \frac{n_j}{N_j}\mathbb{P}_{j,N_j}\right\|_{\mathcal{F}}.$$

The first term is $o_{P^*}(1)$ since $\mathcal{F}$ is Glivenko-Cantelli. Since $(N_j/N)(N_j/n_j) \to_{P^*} \nu_j/p_j$, each summand in the second term is $o_{P^*}(1)$ by the bootstrap Glivenko-Cantelli theorem, which is an easy corollary to Lemma 3.6.16 of [58].

Consider $\mathbb{P}_N^{\pi,e}$. Because $\hat{\alpha}_N \to_{P^*} \alpha_0$ by Proposition 3.5.1, it suffices to consider a compact neighborhood $K \subset \mathbb{R}^{J+k}$ of $\alpha_0$. Since $Z$ is bounded and $G_e$ is continuous, $\{\pi_\alpha(V)\}^{-1} = \{G_e(\alpha^T Z)\}^{-1}$ is bounded in this neighborhood. Because $\alpha$ is a vector in $\mathbb{R}^{J+k}$ and $G_e$ is monotone, $\{\{G_e(\alpha)\}^{-1} : \alpha \in K\}$ is a VC subgraph class by Lemmas 2.6.15 and 2.6.18 of [58]. Boundedness of $G_e$ implies that the set

$$\left\{\pi_0\{G_e(\cdot\alpha)\}^{-1}f : f \in \mathcal{F}, \alpha \in K\right\}$$

is $P_0$-Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). Since $\hat{\alpha}_N \to_{P^*} \alpha_0$, we have by (6.1) that
$\left\|\mathbb{P}_N^{\pi,e} - P_0\right\|_{\mathcal{F}} \to_{P^*} 0$, by recognizing that

$$\mathbb{P}_N^{\pi,e} = \frac{1}{N}\sum_{i=1}^N \frac{\xi_i}{\pi_0(V_i)}\left\{\frac{\pi_0(V_i)}{G_e(\hat{\alpha}_N^T Z_i)}\delta_{X_i}\right\}.$$

Consider $\mathbb{P}_N^{\pi,mc}$. The cases for $\mathbb{P}_N^{\pi,c}$ and $\mathbb{P}_N^{\pi,cc}$ are similar. We verified in the proof of Proposition 3.5.1 that $\{G_{mc}(\cdot;\alpha) : \alpha \in \mathbb{R}^k\}$ is a VC subgraph class. Boundedness of $G$ implies that the set

$$\left\{G_{mc}(\cdot;\alpha)f : f \in \mathcal{F}, \alpha \in \mathbb{R}^k\right\}$$

is $P_0$-Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). Since $\hat{\alpha}_N$ converges to zero in probability by Proposition 3.5.1, the result follows by (6.1). $\square$

Several lemmas are required for the proof of Theorem 6.2.1.

**Lemma 6.4.1.** *Let $\mathcal{F}$ be a class of functions with $P_0|f| < \infty$ for every $f \in \mathcal{F}$. Then,*

$$E^* \left\| \sqrt{\frac{N_j}{N}} I(N_j > 0) \mathbb{G}_{j,N_j} \right\|_{\mathcal{F}} \lesssim E^* \left\| \mathbb{G}_N \right\|_{\mathcal{F}}, \quad \textit{for each} \;\; j = 1, \ldots, J.$$

*Proof.* Let $\epsilon_i$, $i = 1, \ldots, N$, be independent Rademacher variables, independent of $X_i, i = 1, \ldots, N$, and $N_j$. It follows from the symmetrization inequality (Lemma 2.3.6) of [58]

$$E^* \|\mathbb{G}_N\|_{\mathcal{F}} \;\; \gtrsim \;\; E^* \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i f(X_i) \right\|_{\mathcal{F}}.$$

Rewrite this and use Jensen's inequality again with $E[\epsilon f(X)] = 0$ to obtain

$$\begin{aligned} & E^* \left\| \sum_{j=1}^{J} I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \epsilon_{j,i} f(X_{j,i}) \right\|_{\mathcal{F}} \\ & \gtrsim E^* \left\| I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \epsilon_{j,i} f(X_{j,i}) \right\|_{\mathcal{F}}. \end{aligned}$$

Here we implicitly change the law. This can be justified by Proposition A.1 of [8].

Now applying the Lemma 2.3.6 of [58] to the $j$th stratum, this is further bounded below, up to some constant, by

$$\begin{aligned} & E^* \left\| I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} (f(X_{j,i}) - P_{0|j} f) \right\|_{\mathcal{F}} \\ & = E^* \left\| I(N_j > 0) \sqrt{N_j/N} \mathbb{G}_{j,N_j} \right\|_{\mathcal{F}}. \end{aligned}$$

□

The following is a multiplier inequality for bounded exchangeable weights. Note that the sum of stochastic processes in the second term is divided by $n^{1/2}$ rather than $k^{1/2}$.

*Proof of Lemma 6.2.1.* (1) This follows the proof of Lemma 3.6.7 of [58] up to the last line. Since the $\xi_i$'s can be split into their positive and negative parts, we only consider the case where they are nonnegative. Thus for any $1 \le n_0 \le n$,

$$\begin{aligned} E \left\| \sum_{i=1}^{n} \xi_i Z_i \right\|_{\mathcal{F}}^* & \le E \left\| \sum_{i=1}^{n_0 - 1} \xi_{(i)} Z_i \right\|_{\mathcal{F}}^* + E \left\| \sum_{j=n_0}^{n} \xi_{(i)} Z_i \right\|_{\mathcal{F}}^* \\ & \le E \left( \max_{1 \le i \le n} \xi_i \right) \frac{n_0 - 1}{n} \sum_{i=1}^{n} E^* \|Z_i\|_{\mathcal{F}} + E \left\| \sum_{i=n_0}^{n} \xi_{(i)} Z_i \right\|_{\mathcal{F}}^*, \end{aligned}$$

where $\xi_{(i)}, i = 1, \ldots, n$, are the reverse order statistics of $\xi_i, i = 1, \ldots, n$. To bound the second term, we substitute $\xi_{(i)} = \sum_{k=i}^n (\xi_{(k)} - \xi_{(k+1)})$ with $\xi_{(n+1)} = 0$, and change the order of summation to obtain

$$\begin{aligned} E\left\|\sum_{i=n_0}^n \xi_{(i)} Z_i\right\|_{\mathcal{F}}^* &= E\left\|\sum_{i=n_0}^n \sum_{k=i}^n (\xi_{(k)} - \xi_{(k+1)}) Z_i\right\|_{\mathcal{F}}^* \\ &= E\left\|\sum_{k=n_0}^n (\xi_{(k)} - \xi_{(k+1)}) \sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^*. \end{aligned}$$

It follows from the triangle inequality and the independence of the $\xi$'s and $Z_i$'s that this is bounded by

$$\begin{aligned} &\sum_{k=n_0}^n E^*\left\|(\xi_{(k)} - \xi_{(k+1)}) \sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^* \\ &= \sum_{k=n_0}^n E^*\left\{(\xi_{(k)} - \xi_{(k+1)}) \left\|\sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^*\right\} \\ &= \sum_{k=n_0}^n E^*(\xi_{(k)} - \xi_{(k+1)}) E^*\left\|\sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^* \\ &\leq \sum_{k=n_0}^n E^*(\xi_{(k)} - \xi_{(k+1)}) \max_{n_0 \leq k \leq n} E^*\left\|\sum_{i=n_0}^k Z_i\right\|_{\mathcal{F}}^* \\ &= E^* \sum_{k=n_0}^n (\xi_{(k)} - \xi_{(k+1)}) \max_{n_0 \leq k \leq n} E^*\left\|\sum_{i=n_0}^k Z_{R_i}\right\|_{\mathcal{F}}^* \\ &\leq (u - l) \max_{n_0 \leq k \leq n} E^*\left\|\sum_{i=n_0}^k Z_{R_i}\right\|_{\mathcal{F}}^* \end{aligned}$$

using the boundedness of the $\xi_i$'s in the last line. The proof for the negative parts of the $\xi_i$'s is similar and the inequality follows.

The case (2) is similar. □

The following lemma is a key to establishing a rate of convergence of $Z$-estimators under two-phase sampling.

**Lemma 6.4.2.** *For an arbitrary set $\mathcal{F}$ of integrable functions,*

$$E^* \|\mathbb{G}_N^\pi\|_{\mathcal{F}} \lesssim E^* \|\mathbb{G}_N\|_{\mathcal{F}}.$$

*Proof.* We decompose $\mathbb{G}_N^\pi$ as in (2.1): thus

$$\begin{aligned} E^*\|\mathbb{G}_N^\pi\|_{\mathcal{F}} &= E^*\left\|\mathbb{G}_N + \sum_{j=1}^J \sqrt{\frac{N_j}{N}}\left(\frac{N_j}{n_j}\right)\mathbb{G}_{j,N_j}^\xi\right\|_{\mathcal{F}} \\ &\leq E^*\left\|\mathbb{G}_N\right\|_{\mathcal{F}} + \sum_{j=1}^J E^*\left\|\sqrt{\frac{N_j}{N}}\left(\frac{N_j}{n_j}\right)\mathbb{G}_{j,N_j}^\xi\right\|_{\mathcal{F}}. \end{aligned}$$

It therefore suffices to show that each $E^*\left\|m_j\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}$ is bounded up to some constant by $E^*\left\|\mathbb{G}_N\right\|_{\mathcal{F}}$ where $m_j \equiv (N_j/N)^{1/2}(N_j/n_j)$.

Rewrite $\mathbb{G}_{j,N_j}^\xi$ as

$$\mathbb{G}_{j,N_j}^\xi = \sqrt{N_j}\left(\mathbb{P}_{j,N_j}^\xi - \frac{n_j}{N_j}\mathbb{P}_{j,N_j}\right) = \frac{1}{\sqrt{N_j}}\sum_{i=1}^{N_j}\xi_{j,i}\left(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j}\right).$$

Now we condition on $\underline{N} \equiv (N_1, \ldots, N_J)$, and write $E_{\underline{N}}$ for $E(\cdot|\underline{N})$. Since $\xi_{j,i} \in \{0,1\}$, it follows by the multiplier inequality of Lemma 6.2.1 applied conditionally with $n_0 = 1$ and $Z_i = m_j(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})$ that $E_{\underline{N}}\|m_j\mathbb{G}_{j,N_j}^\xi\|_{\mathcal{F}}$ is bounded by

$$\begin{aligned} &(1-0)\max_{1\leq k\leq N_j} E_{\underline{N}}\left\|\frac{1}{\sqrt{N_j}}\sum_{i=1}^k m_j(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})\right\|_{\mathcal{F}}^* \\ &= \max_{1\leq k\leq N_j} E_{\underline{N}}\left[\frac{N_j}{n_j}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^k(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})\right\|_{\mathcal{F}}^*\right]. \end{aligned}$$

Note that $N_j/n_j \leq \sigma^{-1}$ for some $\sigma > 0$ by assumption so that we can replace $N_j/n_j$ by $\sigma^{-1}$ in the last display to obtain an upper bound. Then, apply the triangle inequality to further bound this by

$$\max_{1\leq k\leq N_j} E_{\underline{N}}^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^k(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}}^* + \max_{1\leq k\leq N_j} E_{\underline{N}}^*\left[\frac{k}{\sqrt{N}}\left\|(\mathbb{P}_{j,N_j} - P_{0|j})\right\|_{\mathcal{F}}\right].$$

Since $\delta_{X_{j,i}} - P_{0|j}$ has mean zero, it follows by Jensen's inequality that the first term is bounded by

$$E_{\underline{N}}^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}}^* = E_{\underline{N}}^*\left\|\sqrt{\frac{N_j}{N}}\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}.$$

The second term is bounded by $E_{\underline{N}}^*\left\|\sqrt{N_j/N}\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}$. Now compute unconditionally and apply Lemma 6.4.1 to find that both terms are bounded by $E^*\|\mathbb{G}_N\|_{\mathcal{F}}$. □

*Proof of Theorem 6.2.1.* It follows by Lemma 6.4.2 and the assumption on $E^* \|\mathbb{G}_N\|_{\mathcal{M}_\delta}$ that

$$E^* \|\mathbb{G}_N^\pi\|_{\mathcal{M}_\delta} \lesssim E^* \|\mathbb{G}_N\|_{\mathcal{M}_\delta} \leq \phi_N(\delta).$$

By application of Theorem 3.2.5 of [58], we conclude that the conclusion of (1) of the theorem holds.

For the second statement, note that Theorem 3.2 of [35] holds in a general setting where $P_0 m_{\theta,\eta}$ and $\mathbb{P}_n m_{\theta,\eta}$ are replaced by the deterministic function $\mathbb{M}(\theta,\eta)$ and the stochastic process $\mathbb{M}_n(\theta,\eta)$, respectively. Our parameters $\alpha$ and $\theta$ play roles of their $\theta$ and $\eta$, respectively. Our choice of $\mathbb{M}$ and $\mathbb{M}_N$ is $P_0 G_{mc}(V;\alpha) m_\theta$ and $\mathbb{P}_N^{\pi,mc} m_\theta$. The condition 6.4 corresponds to (3.5) of [35]. The condition 6.5 together with Lemma 6.4.2 verifies their (3.6). Apply their Theorem 3.2 to obtain $d(\hat{\theta}_{N,mc},\theta_0) \leq O_{P^*}(\delta_N^{-1} + |\hat{\alpha}_N - \alpha_0|) = O_{P^*}(\delta_N^{-1})$. The cases for $\hat{\theta}_{N,e}$, $\hat{\theta}_{N,c}$ and $\hat{\theta}_{N,cc}$ are similar. □

*Proof of Lemma 6.2.2.* We consider modified calibration. Other three cases are similar. Because $G(0) = 1$ and $Z$ is bounded, consistency of $\hat{\alpha}_N$ implies that there exists $\mathcal{A}_{mc,2} \subset \mathcal{A}_{mc}$ such that for some fixed constant $C > 0$, $G_{mc}(v;\alpha) \geq C$ and $\dot{G}_{mc}(v;\alpha) \geq C$ for every $\alpha \in \mathcal{A}_{mc,2}$ and $P(\hat{\alpha}_N \in \mathcal{A}_{mc,2}) \to 1$. Then, for arbitrary $\alpha \in \mathcal{A}_{mc,2}$,

$$\begin{aligned}
&P_0 G_{mc}(V;\alpha)(m_\theta - m_{\theta_0}) = P_0 G_{mc}(V;\alpha) \log \frac{p_\theta}{p_{\theta_0}} \\
&\leq 2 P_0 G_{mc}(V;\alpha) \left( \sqrt{\frac{p_\theta}{p_{\theta_0}}} - 1 \right) \\
&= \int G_{mc}(v;\alpha) \left\{ -(p_\theta^{1/2} - p_{\theta_0}^{1/2})^2 + p_\theta - p_{\theta_0} \right\} d\mu \\
&\leq -C \int (p_\theta^{1/2} - p_{\theta_0}^{1/2})^2 d\mu + \int \{G_{mc}(v;\alpha) - 1\}(p_\theta - p_{\theta_0}) d\mu \\
&= -C h^2(p_\theta, p_{\theta_0}) + \int \dot{G}_{mc}(v;\alpha^*)(\pi_0^{-1}(v) - 1) v^T (p_\theta - p_{\theta_0}) d\mu (\alpha - \alpha_0),
\end{aligned}$$

where $\alpha^*$ is some convex combination of $\alpha$ and $\alpha_0$. Because the integral in the last display is a bounded row vector, the second term in the last display is bounded by $|\alpha - \alpha_0|^2$ up to some constant. Thus, the condition (6.4) holds. □

The following lemma is useful when showing asymptotic equicontinuity of processes involving $\mathbb{P}_N^{\pi,e}$, $\mathbb{P}_N^{\pi,c}$, $\mathbb{P}_N^{\pi,mc}$ and $\mathbb{P}_N^{\pi,cc}$.

**Lemma 6.4.3.** *Suppose Conditions 3.1.2 and 3.1.1 hold. Let $\mathcal{F}$ be a Glivenko-Cantelli class. Then*

$$\sup_{f\in\mathcal{F}}\left|\sqrt{N}(\mathbb{P}_N - P_0)\left\{\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}f - \frac{\xi}{\pi_{\alpha_0}(V)}f\right\}\right| = o_{P^*}(1), \tag{6.11}$$

*where $\pi_{\hat{\alpha}_N}$ is either an estimated or calibrated probability (with modified or centered calibration).*

*Proof.* We only consider modified calibration. The cases for estimated weights and (centered) calibration are similar. It follows by Taylor's theorem that

$$\begin{aligned}&\sup_{f\in\mathcal{F}}\left|\sqrt{N}(\mathbb{P}_N - P_0)\left\{\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}f - \frac{\xi}{\pi_{\alpha_0}(V)}f\right\}\right|\\&=\sup_{f\in\mathcal{F}}\left|(\mathbb{P}_N^\pi - P_0)\left((\pi_0^{-1}(V)-1)Z^T\dot{G}_{mc}(Z;\alpha^*)f\right)\right|\sqrt{N}|\hat{\alpha}_N-\alpha_0|,\end{aligned}$$

for some $\alpha^*$ with $|\alpha^*-\alpha_0|\le|\hat{\alpha}_N-\alpha_0|$. Because $\sqrt{N}(\hat{\alpha}_N-\alpha_0)=O_{P^*}(1)$ by Proposition 3.5.1, it follows that (6.11) is $o_{P^*}(1)$ by Theorem 6.1.1 and Proposition 3.5.1 if the set $\{(\pi_0(V)^{-1}-1)Z^T\dot{G}\{(\pi_0^{-1}(V)-1)Z^T\alpha\} : \alpha\in\mathcal{A}_{mc,3}, f\in\mathcal{F}\}$ is $P_0$-Glivenko-Cantelli where $\mathcal{A}_{mc,3}\subset\mathcal{A}_{mc}$ is some compact set containing $\alpha_0=0$. This is easily verified in the same way as in the proof of Proposition 3.5.1. □

*Proof of Theorem 6.3.1.* The result (6.6) follows from [8]. Consider the IPW empirical process with modified calibration. It follows by Taylor's theorem that

$$\begin{aligned}&\mathbb{G}_N^{\pi,mc}f - \mathbb{G}_N^\pi f\\&=\mathbb{G}_N\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}-\frac{\xi}{\pi_{\alpha_0}(V)}\right)f + \sqrt{N}P_0\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}-\frac{\xi}{\pi_{\alpha_0}(V)}\right)f\\&=\mathbb{G}_N\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}-\frac{\xi}{\pi_{\alpha_0}(V)}\right)f\\&\quad+P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha^*)f\right)\sqrt{N}(\hat{\alpha}_N-\alpha_0),\end{aligned} \tag{6.12}$$

where $\alpha^*$ is some convex combination of $\hat{\alpha}_N$ and $\alpha_0$. The first term is $o_{P^*}(1)$ by Lemma 6.4.3. Since $(\pi_0(V)^{-1}-1)Z^T\dot{G}_{mc}$ is bounded and $f$ is integrable, it follows from the dominated convergence theorem that

$$P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha^*)f\right)\to P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}(0)f\right).$$

Apply the result (6.6) and Proposition 3.5.1 to conclude the finite-dimensional convergence

$$\begin{aligned}
&\mathbb{G}_N^{\pi,mc} f = \mathbb{G}_N^{\pi} f + (\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi}) f \\
&\to_d \mathbb{G} f + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j f \\
&\quad - P_0 \left( \frac{1-\pi_0(V)}{\pi_0(V)} Z^T \dot{G}(0) f \right) \dot{G}(0)^{-1} \left\{ P_0 \frac{1-\pi_0(V)}{\pi_0(V)} Z^{\otimes 2} \right\}^{-1} \\
&\quad \times \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Z \\
&= \mathbb{G} f + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j f - \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Q_{mc} f \\
&= \mathbb{G} f + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (f - Q_{mc} f).
\end{aligned}$$

Next, we prove the asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc}$ with respect to the metric $\rho_{mc}$ defined by

$$\rho_{mc}^2(f,g) = P_0(f-g)^2 + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(f-g).$$

First recall that $\mathbb{G}_N^{\pi}$ is asymptotically equicontinuous with respect to the metric $\rho$ defined by

$$\rho^2(f,g) = \sigma_{P_0}^2(f-g) + \sum_{j=1}^{J} \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(f-g).$$

The part $\sigma_{P_0}^2(f-g)$ corresponds to the empirical process $\mathbb{G}_N \equiv \sqrt{N}(\mathbb{P}_N - P_0)$ in the decomposition (2.1) of the IPW empirical process. However, this empirical process $\mathbb{G}_N$ is asymptotically equicontinuous with respect to the $L_2(P)$-metric with an assumption $\|P_0\|_{\mathcal{F}} < \infty$ in view of Problem 2.1.2 of [58]. Thus, $\mathbb{G}_N^{\pi}$ is asymptotically equicontinuous with respect to $\rho_{mc}$. Now, it remains to verify the asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi}$. Let $h_N \in \mathcal{F}_{\delta_N} \equiv \{f - g : f, g \in \mathcal{F}, \rho_{mc}(f,g) \le \delta_N\}$ for an arbitrary sequence $\delta_N \downarrow 0$. In view of (6.12)

$$(\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi}) h_N = o_{P^*}(1) + P_0 \left( \frac{1-\pi_0(V)}{\pi_0(V)} Z^T \dot{G}_{mc}(V;\alpha^*) h_N \right) O_{P^*}(1),$$

where $\alpha^*$ is some convex combination of $\hat{\alpha}_N$ and $\alpha_0$. Because each element of a vector $(\pi_0(V)^{-1} - 1) Z^T \dot{G}_{mc}(V;\alpha^*)$ is bounded, it follows from the Cauchy-Schwarz inequality that each element of $P_0\{(\pi_0(V)^{-1}-1)Z^T\dot{G}_{mc}(V;\alpha^*)h_N\}$ is bounded up to some constant by $P_0(h_N^2)$. Since $\rho_{mc}(f,g)\to 0$ implies $P_0(f-g)^2\to 0$, we have $P_0h_N^2\to 0$ as $N\to\infty$. This verifies the asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc}$ and hence completes showing its weak convergence.

The cases for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ follow analogously. □

*Proof of Theorem 6.3.2.* Since $\mathcal{F}$ is Donsker, it follows by Lemma 2.3.11 of [58] that $E^*\|\mathbb{G}_N\|_{\mathcal{F}_{\delta_N}}\to 0$ for every sequence $\delta_N\downarrow 0$. Thus, the result follows from Lemma 6.4.2. Apply Markov's inequality to obtain $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_{\delta_N}}=o_{P^*}(1)$. For the second statement, consider the expansion (6.12) of $\mathbb{G}_N^{\pi,mc}f-\mathbb{G}_N^\pi f$ with $f\in\mathcal{F}_{\delta_N}$. The first term is $o_{P^*}(1)$ by Lemma 6.4.3. Since $f$ converges to zero in $L_2(P_0)$, the second term is $o_{P^*}(1)$ by the dominated convergence theorem and Proposition 3.5.1. Apply the triangle inequality to conclude $\left\|\mathbb{G}_N^{\pi,mc}\right\|_{\mathcal{F}_{\delta_N}}=o_{P^*}(1)$.

The proofs for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ are similar. □

*Proof of Lemma 6.3.1.* Without loss of generality, assume that $\hat{\theta}_N$ takes its values in $\Theta_\delta\equiv\{\theta\in\Theta:\|\theta-\theta_0\|<\delta\}$ because of consistency of $\hat{\theta}_N$ to $\theta_0$. Define a function $f:\ell^\infty(\Theta_\delta\times\mathcal{H})\times\Theta_\delta\mapsto\ell^\infty(\mathcal{H})$ by $f(z,\theta)h=z(\theta,h)$. Note that $f$ is continuous at every point $(z,\theta_0)$ such that $\|z(\theta,h)-z(\theta_0,h)\|_{\mathcal{H}}\to 0$, as $\theta\to\theta_0$. To see this, suppose $z_N\to z$ and $\theta_N\to\theta_0$. Then, for a fixed $\epsilon>0$, there exists $n_0$ such that $\|z_N-z\|<\epsilon$ and $\|\theta_N-\theta_0\|<\epsilon$ for $N\geq N_0$. For $N\geq N_0$, we have

$$
\begin{aligned}
&\|f(z_N,\theta_N)-f(z,\theta_0)\|_{\mathcal{H}}\\
&\leq\|f(z_N,\theta_N)-f(z_0,\theta_N)\|_{\mathcal{H}}+\|f(z_0,\theta_N)-f(z_0,\theta_0)\|\\
&\leq\sup_{\theta\in\Theta_\delta,h\in\mathcal{H}}|z_N(\theta,h)-z(\theta,h)|+\|z(\theta_N,h)-z(\theta_0,h)\|_{\mathcal{H}}\\
&<2\epsilon.
\end{aligned}
$$

Define a stochastic process $\mathbb{Z}_N$ indexed by $\Theta_\delta\times\mathcal{H}$ by

$$
\mathbb{Z}_N(\theta,h)=\mathbb{G}_N^\pi\left(\psi_{\theta,h}-\psi_{\theta_0,h}\right).
$$

Because $\{\psi_{\theta,h}-\psi_{\theta_0,h}:\|\theta-\theta_0\|<\delta,\theta\in\Theta,h\in\mathcal{H}\}$ is Donsker, Theorem 6.3.1 implies that the sequence $\mathbb{Z}_N$ converges in $\ell^\infty(\Theta_\delta \times \mathcal{H})$ to a tight Gaussian process $\mathbb{Z}$ given by

$$\mathbb{Z} = \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j.$$

This process has continuous sample paths with respect to the semimetric $\rho$ given by

$$\rho^2\left((\theta_1, h_1), (\theta_2, h_2)\right) = P\left(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2}\right)^2$$

because $(\Theta_\delta \times \mathcal{H}, \rho)$ is totally bounded and $\mathbb{Z}$ is uniformly $\rho$-continuous. To see the latter, note that

$$\rho^2\left((\theta_1, h_1), (\theta_2, h_2)\right) \geq P\left\{\left(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2}\right)^2 \middle| V \in \mathcal{V}_j\right\}\nu_j$$

for each $j = 1, \ldots, J$. By assumption

$$\sup_{h\in\mathcal{H}} \rho^2\left((\theta, h), (\theta_0, h)\right) = \sup_{h\in\mathcal{H}} P\left(\psi_{\theta,h} - \psi_{\theta_0,h} + 0\right)^2 \to 0,$$

as $\theta \to \theta_0$. Thus, $f$ is continuous at almost all sample paths of $\mathbb{Z}$.

By Slutsky's theorem, $(\mathbb{Z}_N, \hat{\theta}_N) \rightsquigarrow (\mathbb{Z}, \theta_0)$. By the continuous mapping theorem, $\mathbb{Z}_N(\hat{\theta}_N) = f(\mathbb{Z}_N, \hat{\theta}_N) \rightsquigarrow f(\mathbb{Z}, \theta_0) = 0$ in $\ell^\infty(\mathcal{H})$.

The other cases for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$, $\mathbb{G}_N^{\pi,mc}$ and $\mathbb{G}_N^{\pi,cc}$ follow analogously; see the proof of Theorem 6.3.1. □

**Lemma 6.4.4.** *Let $\mathbb{Z}_1, \mathbb{Z}_2, \ldots$ be i.i.d. stochastic processes indexed by $\mathcal{F}_N$ with $E^*\|\mathbb{Z}_1\|_{\mathcal{F}_N}$ uniformly bounded in $N$. Suppose that $\|\mathbb{S}_N\|_{\mathcal{F}_N}$ $\equiv \|\sum_{i=1}^N \mathbb{Z}_i\|_{\mathcal{F}_N} = o_{P^*}(1)$. Then*

$$E^*\|\mathbb{S}_N\|_{\mathcal{F}_N} \to 0, \quad N \to \infty.$$

*Proof.* Fix $\epsilon > 0$. Let $\mathbb{Y}_i$ be independent copies of $\mathbb{Z}_i$ and define $\mathbb{T}_N = \sum_{i=1}^N \mathbb{Y}_i$, and $\mathbb{U}_N = \mathbb{T}_N - \mathbb{S}_N$. Since $\|\mathbb{U}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$, $\limsup_N P(\|\mathbb{U}_N\|_{\mathcal{F}_N} \geq x\sqrt{N}) \leq \limsup_N P(\|\mathbb{U}_N\|_{\mathcal{F}_N} \geq x) = 0$ by the portmanteau theorem. This implies that there exists $N_0$ such that for $N \geq N_0$

$$P^*(\|\mathbb{U}_N\|_{\mathcal{F}_N} > x\sqrt{N}) \leq \epsilon/x^2.$$

Since $\mathbb{U}_N$ is a sum of independent symmetric processes, we can apply Lévy's inequality to obtain

$$P^*\left(\max_{1\le i\le n}\|\mathbb{Z}_i-\mathbb{Y}_i\|_{\mathcal{F}_N}>x\sqrt{N}\right)\le 2P^*(\|\mathbb{U}_N\|_{\mathcal{F}_N}>x\sqrt{N})\le 2\epsilon/x^2.$$

In view of Problem 2.3.2 of [58], for every $N\ge N_0$,

$$x^2NP^*(\|\mathbb{Z}_1-\mathbb{Y}_1\|_{\mathcal{F}_N}>x\sqrt{N})\le 4\epsilon.$$

Note that on the event that $\|\mathbb{Z}_1\|_{\mathcal{F}_N}>x$, we have

$$\beta_N(x)\equiv P_Y^*(\|\mathbb{Y}_1\|_{\mathcal{F}_N}<x/2)\le P_Y^*(\|\mathbb{Z}_1-\mathbb{Y}_1\|_{\mathcal{F}_N}>x/2).$$

Integrating both sides with respect to $\mathbb{Z}$ gives

$$\beta_N(x)P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N}>x)\le P^*(\|\mathbb{Z}_1-\mathbb{Y}_1\|_{\mathcal{F}_N}>x/2).$$

By Markov's inequality,

$$\beta_N(x)=1-P^*(\|\mathbb{Y}_1\|_{\mathcal{F}_N}\ge x/2)\ge 1-2x^{-1}E\|\mathbb{Y}_1\|_{\mathcal{F}_N}$$

Since $E\|\mathbb{Y}_1\|_{\mathcal{F}_N}$ is uniformly bounded in $N$, it follows that, for $x$ sufficiently large, $\beta_N(x)^{-1}$ is uniformly bounded in $N$ and, therefore, $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N}>x\sqrt{N})$ is bounded by $P^*(\|\mathbb{Z}_1-\mathbb{Y}_1\|_{\mathcal{F}_N}>x\sqrt{N})$ up to some constant for every $N$. Hence this proves that $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N}>x)=o(x^{-2})$.

Now we apply the Hoffmann-Jørgensen inequality to obtain

$$E^*\|\mathbb{S}_N\|_{\mathcal{F}_N}\lesssim E^*\max_{i\le N}\|\mathbb{Z}_i\|_{\mathcal{F}_N}+G_N^{-1}(u)$$

for an absolute constant $u$ where

$$G_N(t)=P^*(\|\mathbb{S}_N\|_{\mathcal{F}_N}\le t).$$

Since $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N}>x)=o(x^{-2})$, $E^*\max_{i\le N}\|\mathbb{Z}_i\|_{\mathcal{F}_N}\to 0$ in view of Problem 2.3.3 of [58]. The second term goes to zero since $\|\mathbb{S}_N\|_{\mathcal{F}_N}=o_{P^*}(1)$. This completes the proof. □

*Proof of Lemma 6.3.2.* Define $\mathcal{G}_N=\{N^{-1/2}f:f\in\mathcal{F}_N\}$. We apply Lemma 6.4.4 with $\mathbb{Z}_i$ and $\mathcal{F}_N$ in Lemma 6.4.4 replaced by $\delta_{X_i}-P_0$ and $\mathcal{G}_N$, respectively. The uniform boundedness condition of Lemma 6.4.4 is satisfied, because $E^*\|\delta_{X_1} - P_0\|_{\mathcal{F}_N} < \infty$ for $N \geq N_0$, and this expectation is decreasing in $N \geq N_0$. Thus, $E^*\|\mathbb{G}_N\|_{\mathcal{F}_N} = E^*\|\sum_{i=1}^N(\delta_{X_i} - P)\|_{\mathcal{G}_N} \to 0$. Apply Lemma 6.4.2, and Markov's inequality to obtain $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_N} = o_{P^*}(1)$.

For the IPW empirical process with modified calibration, consider the expansion (6.12) of $(\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^\pi)f$. Then the first term is $o_{P^*}(1)$ by Lemma 6.4.3. Suppose that $f = f_N \in \mathcal{F}_N$ converges to zero pointwise. Since $(\pi_0(V)^{-1} - 1)Z\dot{G}_{mc}$ is bounded, the second term in the expansion (6.12) is $o_{P^*}(1)$ by the dominated convergence theorem and Proposition 3.5.1. Suppose instead that $f = f_N \in \mathcal{F}_N$ converges to zero in $L_1(P_0)$. Then the same conclusion that the second term in the expansion (6.12) is $o_{P^*}(1)$ follows directly. Apply the triangle inequality to conclude $\left\|\mathbb{G}_N^{\pi,mc}\right\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$.

The proofs for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ are similar. □

# Chapter 7

# GENERAL RESULTS FOR BOOTSTRAP IPW EMPIRICAL PROCESSES

We present asymptotic results on several bootstrap schemes proposed in Chapter 4. All proofs are presented at the end of this chapter.

## *7.1 Glivenko-Cantelli Theorem*

The following theorem is a survey bootstrap version of of a Glivenko-Cantelli Theorem for bootstrapped two-phase sampling (Theorem 6.1.1).

**Theorem 7.1.1** (Glivenko-Cantelli theorem for bootstrapped two-phase sampling). *Suppose that $\mathcal{F}$ is a $P_0$-Glivenko-Cantelli class.*
*(1) (a) Then*

$$\left\|\hat{\mathbb{P}}_N^{\pi,S} - \mathbb{P}_N^{\pi}\right\|_{\mathcal{F}} \to_{P_W^*} 0, \tag{7.1}$$

*in $P^*$-probability where $\|\cdot\|_{\mathcal{F}}$ is the supremum norm. This also holds if we replace $\hat{\mathbb{P}}_N^{\pi,S}$ by $\hat{\mathbb{P}}_N^{\pi,S,c}$, $\hat{\mathbb{P}}_N^{\pi,S,mc}$ or $\hat{\mathbb{P}}_N^{\pi,S,cc}$, assuming Condition 3.1.2 with 3.1.2(a) replaced by Condition 4.1.4(a-1).*
*(b) Under Condition 3.1.2 with 3.1.2(a) replaced by Condition 4.1.4(a-2), it holds that*

$$\left\|\hat{\mathbb{P}}_N^{\pi,S,d\#} - \mathbb{P}_N^{\pi,\#}\right\|_{\mathcal{F}} \to_{P_W^*} 0,$$

*in $P^*$-probability where $\# \in \{c, mc, cc\}$.*
*(2)*

$$\left\|\hat{\mathbb{P}}_N^{\pi} - \mathbb{P}_N^{\pi}\right\|_{\mathcal{F}} \to_{P_W^*} 0, \tag{7.2}$$

*in $P^*$-probability. This also holds if we replace $\hat{\mathbb{P}}_N^{\pi}$ by $\hat{\mathbb{P}}_N^{\pi,c}$, $\hat{\mathbb{P}}_N^{\pi,mc}$ or $\hat{\mathbb{P}}_N^{\pi,cc}$, assuming Condition 3.1.2 with 3.1.2(a) replaced by Condition 4.1.4(a-1).*
*(b) Under Condition 3.1.2 with 3.1.2(a) replaced by Condition 4.1.4(a-2), it holds that*

$$\left\|\hat{\mathbb{P}}_N^{\pi,d\#} - \mathbb{P}_N^{\pi,\#}\right\|_{\mathcal{F}} \to_{P_W^*} 0,$$

*in $P^*$-probability where $\# \in \{c, mc, cc\}$.*

## 7.2 Rate of Convergence

The following theorem is a bootstrap version of Theorem 6.2.1. It follows from this theorem (Theorem 7.2.1) and Theorem 6.2.1 that rates of convergence all agree for $M$-estimators with complete data, IPW $M$-estimators under two-phase sampling, and bootstrap IPW $M$-estimators under two-phase sampling. For notational simplicity, some subscripts and superscripts indicating either of survey or two-phase bootstrap are suppressed because these differences do not play a role in the theorem.

**Theorem 7.2.1.** *Let $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ be the set of criterion functions and define $\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}$ for some fixed $\delta > 0$ where $d$ is a semimetric on the parameter space $\Theta$.*

*(1) Suppose that for every $\theta$ in a neighborhood of $\theta_0$,*

$$P_0(m_\theta - m_{\theta_0}) \lesssim -d^2(\theta, \theta_0);$$

*here $a \lesssim b$ means $a \leq Kb$ for some constant $K \in (0, \infty)$. Assume that there exists a function $\phi_N$ such that $\delta \mapsto \phi_N(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ (not depending on $N$) and for every $N$,*

$$E^* \|\mathbb{G}_N\|_{\mathcal{M}_\delta} \lesssim \phi_N(\delta),$$

*where $\mathbb{G}_N$ is the empirical process. If the bootstrap estimator $\hat{\hat{\theta}}_N$ satisfying $\hat{\mathbb{P}}_N^\pi m_{\hat{\hat{\theta}}_N} \geq \hat{\mathbb{P}}_N^\pi m_{\theta_0} - O_{P_W^*}(r_N^{-2})$ in $P^*$-probability converges in outer $P_W^*$-probability to $\theta_0$ in $P^*$-probability, then $r_N d(\hat{\hat{\theta}}_N, \theta_0) = O_{P_W^*}(1)$ in $P^*$-probability for every sequence $r_N$ such that $r_N^2 \phi_N(1/r_N) \leq \sqrt{N}$ for every $N$.*

*(2) Suppose Condition 3.1.2 holds. Suppose also that for every $\theta \in \Theta$ in a neighborhood of $\theta_0$,*

$$P_0\{G_{mc}(V; \alpha)(m_\theta - m_{\theta_0})\} \lesssim -d^2(\theta, \theta_0) + |\alpha - \alpha_0|^2.$$

*Assume that*

$$E^* \|\mathbb{G}_N\|_{G\mathcal{M}_\delta} \lesssim \phi_N(\delta),$$

*where* $G\mathcal{M}_\delta \equiv \{G_{mc}(\cdot;\alpha)f : |\alpha| \leq \delta, \alpha \in \mathcal{A}_N, f \in \mathcal{M}_\delta\}$ *for some* $\mathcal{A}_N \subset \mathcal{A}_{mc}$. *If a bootstrap estimator with modified calibration,* $\hat{\hat{\theta}}_{N,mc}$*, satisfying* $\hat{\mathbb{P}}_N^{\pi,mc} m_{\hat{\hat{\theta}}_{N,mc}} \geq \hat{\mathbb{P}}_N^{\pi,mc} m_{\theta_0} - O_{P_W^*}(r_N^{-2})$ *in* $P^*$*-probability converges in outer* $P_W^*$*-probability to* $\theta_0$ *in* $P^*$*-probability, then* $r_N d(\hat{\hat{\theta}}_{N,mc}, \theta_0) = O_{P_W^*}(1)$ *in* $P^*$*-probability for every sequence* $r_N$ *such that* $r_N^2 \phi_N(1/r_N) \leq \sqrt{N}$ *for every* $N$.

*(3) Suppose Condition 3.1.2 holds. Under the same conditions of (2) with* $G_{mc}$ *replaced by* $G_c$ *or* $G_{cc,S}$*, the same conclusions hold for a bootstrap estimator with calibration,* $\hat{\hat{\theta}}_{N,c}$*, satisfying* $\hat{\mathbb{P}}_N^{\pi,c} m_{\hat{\hat{\theta}}_{N,c}} \geq \hat{\mathbb{P}}_N^{\pi,c} m_{\theta_0} - O_{P_W^*}(r_N^{-2})$ *in* $P^*$*-probability, and a bootstrap estimator with centered calibration,* $\hat{\hat{\theta}}_{N,cc}$*, satisfying* $\hat{\mathbb{P}}_N^{\pi,cc} m_{\hat{\hat{\theta}}_{N,cc}} \geq \hat{\mathbb{P}}_N^{\pi,cc} m_{\theta_0} - O_{P_W^*}(r_N^{-2})$ *in* $P^*$*-probability, respectively.*

## 7.3 Donsker Theorem

The next theorem yields weak convergence of the survey bootstrap IPW empirical processes (Section 4.1.2). Compare this theorem with a Donsker theorem for two-phase sampling (Theorem 6.3.1) to see that $\mathbb{G}$ is lacking in the limiting processes for the survey bootstrap IPW empirical processes. This means that the survey bootstrap yields the phase II variances only as expected.

**Theorem 7.3.1** (Donsker theorem for survey bootstrap)**.** *Suppose that* $\mathcal{F}$ *with* $\|P_0\|_{\mathcal{F}} < \infty$ *is a* $P_0$*-Donsker class and Condition 3.1.2 holds with Condition 3.1.2(a) replaced by*

*Condition 4.1.4(a-2). Then,*

$$\hat{\mathbb{G}}_N^{\pi,S} \rightsquigarrow \mathbb{G}^{\pi,S} \equiv \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j, \tag{7.3}$$

$$\hat{\mathbb{G}}_N^{\pi,S,c} \rightsquigarrow \mathbb{G}^{\pi,S,c} \equiv \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_c \cdot), \tag{7.4}$$

$$\hat{\mathbb{G}}_N^{\pi,S,dc} \rightsquigarrow \mathbb{G}^{\pi,S,c}, \tag{7.5}$$

$$\hat{\mathbb{G}}_N^{\pi,S,mc} \rightsquigarrow \mathbb{G}^{\pi,S,mc} \equiv \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_{mc} \cdot), \tag{7.6}$$

$$\hat{\mathbb{G}}_N^{\pi,S,dmc} \rightsquigarrow \mathbb{G}^{\pi,S,mc}, \tag{7.7}$$

$$\hat{\mathbb{G}}_N^{\pi,S,cc} \rightsquigarrow \mathbb{G}^{\pi,S,cc} \equiv \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j(\cdot - Q_{cc} \cdot), \tag{7.8}$$

$$\hat{\mathbb{G}}_N^{\pi,S,dcc} \rightsquigarrow \mathbb{G}^{\pi,S,cc}, \tag{7.9}$$

*in $\ell^\infty(\mathcal{F})$ in $P^*$-probability where the $P_0$-Brownian bridge process, $\mathbb{G}$, indexed by $\mathcal{F}$ and the $P_{0|j}$-Brownian bridge processes, $\mathbb{G}_j$, indexed by $\mathcal{F}$ are all independent and $Q_\#$ are defined in Theorem 3.2.1.*

**Remark 7.3.1.** *The result (7.3) is a partial extension of [3] to the process level. [3] considered asymptotics for the weighted sample mean but under more general sampling schemes including ours.*

The following lemma is the uncentered conditional multiplier central limit theorem. This lemma and a Donsker theorem for survey bootstrap (7.3.1) are used to prove a Donsker theorem for two-phase bootstrap (Theorem 7.3.2). The lemma itself is also of interest because it provides a rigorous justification of the weighted bootstrap of [31] by choosing $\tilde{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)$ with $\hat{\mathbb{P}}_n = n^{-1}\sum_{i=1}^n w_i \delta_{X_i}$ where $w_i$ are positive i.i.d. bootstrap weights, independent of $X_i$'s, with $Ew_1 = 1$ and $\mathrm{Var}(w_1) = 1$. For related results, the conditional multiplier central limit theorem and the uncentered unconditional multiplier central limit theorem are given in Theorem 2.9.6 and Corollary 2.9.4 of [58], respectively.

**Lemma 7.3.1** (uncentered conditional multiplier CLT)**.** *Suppose that a class of measurable functions $\mathcal{F}$ is Donsker with $\|P_0\|_{\mathcal{F}} < \infty$. Let $X_1, \ldots, X_n$ be i.i.d. $P_0$. Let $\xi_1, \ldots, \xi_n$*

*be i.i.d. random variables with mean zero, variance* $c^2$ *and* $\|\xi_1\|_{2,1} < \infty$*, independent of* $X_1, \ldots, X_n$*. Let* $\tilde{\mathbb{G}}_n \equiv n^{-1/2} \sum_{i=1}^n \xi_i \delta_{X_i}$ *and* $\tilde{\mathbb{G}} = \mathbb{G} + ZP_0$ *where* $\mathbb{G}$ *is a* $P_0$*-Brownian bridge process independent of the standard normal random variable* $Z$*. Then,* $\sup_{h \in BL_1} |E_\xi h(\tilde{\mathbb{G}}_n) - h(c\tilde{\mathbb{G}})| \to 0$ *in outer probability and the sequence* $\tilde{\mathbb{G}}_n$ *is asymptotically measurable. Moreover, if* $P_0\|f - P_0 f\|_{\mathcal{F}}^2 < \infty$*, then* $\sup_{h \in BL_1} |E_\xi h(\tilde{\mathbb{G}}_n) - h(c\tilde{\mathbb{G}})| \to 0$ *outer almost surely, and the sequence* $|E_\xi h(\tilde{\mathbb{G}}_n)^* - h(c\tilde{\mathbb{G}})_*| \to 0$ *almost surely for every* $h \in BL_1$*. Here* $h(\tilde{\mathbb{G}}_n)^*$ *and* $h(\tilde{\mathbb{G}}_n)_*$ *denote measurable majorants and minorants with respect to* $(\xi_1, \ldots, \xi_n, X_1, \ldots, X_n)$ *jointly.*

The next theorem is a Donsker theorem for the two-phase bootstrap IPW empirical processes (Section 4.2). The process $\tilde{\mathbb{G}}$ appearing in the limiting processes is the $P_0$-Brownian motion process, not the $P_0$-Brownian bridge process appearing in Theorem 6.3.1. Because of this difference, the two-phase bootstrap has a distributional consistency only when the index set $\mathcal{F}$ is a set of mean zero functions. This issue is not a problem when considering bootstrap WLE's (see Theorems 4.2.1 and 4.2.2).

**Theorem 7.3.2** (Donsker theorem for two-phase bootstrap)**.** *Let* $\mathcal{F}$ *be a Donsker class with* $\|P_0\|_{\mathcal{F}} < \infty$*. Then,*

$$\begin{aligned}
&\hat{\mathbb{G}}_N^{\pi} \rightsquigarrow \mathbb{G}^{\pi,tp} \equiv \tilde{\mathbb{G}} + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j, \\
&\hat{\mathbb{G}}_N^{\pi,c} \rightsquigarrow \mathbb{G}^{\pi,tp,c} \equiv \tilde{\mathbb{G}} + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(\cdot - Q_c\cdot), \\
&\hat{\mathbb{G}}_N^{\pi,dc} \rightsquigarrow \mathbb{G}^{\pi,tp,c}, \\
&\hat{\mathbb{G}}_N^{\pi,mc} \rightsquigarrow \mathbb{G}^{\pi,tp,mc} \equiv \tilde{\mathbb{G}} + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(\cdot - Q_{mc}\cdot), \\
&\hat{\mathbb{G}}_N^{\pi,dmc} \rightsquigarrow \mathbb{G}^{\pi,tp,mc}, \\
&\hat{\mathbb{G}}_N^{\pi,cc} \rightsquigarrow \mathbb{G}^{\pi,tp,cc} \equiv \tilde{\mathbb{G}} + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(\cdot - Q_{cc}\cdot), \\
&\hat{\mathbb{G}}_N^{\pi,dcc} \rightsquigarrow \mathbb{G}^{\pi,tp,cc},
\end{aligned}$$

*in* $\ell^\infty(\mathcal{F})$ *in* $P^*$*-probability where* $P_0$*- Brownian motion process* $\tilde{\mathbb{G}}$ *and* $P_{0|j}$*-Brownian bridge processes* $\mathbb{G}_j$ *are all independent and* $Q_\#$ *are defined in Theorem 3.2.1.*

**Remark 7.3.2.** *Theorem 7.3.2 holds when replacing the boundedness condition* $W_{Ni}^{(1)} \leq M$ *for the phase I bootstrap weights by the* $L_{2,1}$*-condition* $\|W_{Ni}\|_{2,1} < \infty$.

The next two lemmas are bootstrap versions of Lemmas 6.3.1 and 6.3.2. These are useful in establishing the asymptotic equicontinuity of the bootstrap IPW empirical processes.

**Lemma 7.3.2.** *Suppose* $\mathcal{F} = \{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\}$ *is* $P_0$*-Donsker for some* $\delta > 0$ *and that* $\sup_{h\in\mathcal{H}} P_0(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0$, *as* $\theta \to \theta_0$. *If* $\hat{\theta}_N$ *converges in* $P_W$*-outer probability to* $\theta_0$, *then*

$$\left\|\hat{\mathbb{G}}_N^{\pi,S}(\psi_{\hat{\theta}_N,h} - \psi_{\theta_0,h})\right\|_{\mathcal{H}} = o_{P_W^*}(1),$$

*in* $P^*$*-probability. This also holds if we replace* $\hat{\mathbb{G}}_N^{\pi,S}$ *by* $\hat{\mathbb{G}}_N^{\pi,S,c}$, $\hat{\mathbb{G}}_N^{\pi,S,mc}$, $\hat{\mathbb{G}}_N^{\pi,S,cc}$,$\hat{\mathbb{G}}_N^{\pi,S,dc}$, $\hat{\mathbb{G}}_N^{\pi,S,dmc}$ *or* $\hat{\mathbb{G}}_N^{\pi,S,dcc}$ *assuming Conditions 3.1.1 and 3.1.2. hold and* $\|P_0\|_{\mathcal{F}} < \infty$.

**Lemma 7.3.3.** *Let* $\mathcal{F}_N$ *be a sequence of decreasing classes of functions such that* $\|\mathbb{G}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$. *Assume that there exists an integrable envelope for* $\mathcal{F}_{N_0}$ *for some* $N_0$. *Then* $\|\hat{\mathbb{G}}_N^{\pi,S}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$ *in* $P^*$*-probability.*

*Suppose, moreover, that* $\mathcal{F}_N$ *is* $P_0$*-Glivenko-Cantelli with* $\|P_0\|_{\mathcal{F}_{N_1}} < \infty$ *for some* $N_1$, *and that every* $f = f_N \in \mathcal{F}_N$ *converges to zero either pointwise or in* $L_1(P_0)$ *as* $N \to \infty$. *Then* $\|\hat{\mathbb{G}}_N^{\pi,S,c}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$, $\|\hat{\mathbb{G}}_N^{\pi,S,mc}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$ *and* $\|\hat{\mathbb{G}}_N^{\pi,S,cc}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$, $\|\hat{\mathbb{G}}_N^{\pi,S,dc}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$, $\|\hat{\mathbb{G}}_N^{\pi,S,dmc}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$ *and* $\|\hat{\mathbb{G}}_N^{\pi,S,dcc}\|_{\mathcal{F}_N} = o_{P_W^*}(1)$ *in* $P^*$*-probability, assuming Condition 4.1.4.*

## 7.4 Proofs

### 7.4.1 Multivariate Hypergeometric Distribution

Denote, as $MH(N, n, (m_1, \ldots, m_s))$, the multivariate hypergeometric distribution [24], pages 171 - 177, where $n$ balls are sampled without replacement from the population consisting of the disjoint subgroups of size $m_i, i = 1, \ldots, s$, with $\sum_{i=1}^s m_i = N$. Let $(X_1, \ldots, X_s) \sim MH(N, n, (m_1, \ldots, m_s))$. Note that $X_i$ is the number of balls sampled from the $i$th subgroup. Namely,

$$P(X_1 = x_1, \ldots, X_s = x_s) = \frac{\prod_{i=1}^s \binom{m_i}{x_i}}{\binom{N}{n}}$$

where $\sum_{i=1}^{s} x_i = n$, $\sum_{i=1}^{s} m_i = N$. Since $X_i$ marginally follows the hypergeometric distribution,

$$EX_i = \frac{nm_i}{N}, \quad \mathrm{Var}(X_i) = \frac{m_i}{N}\frac{N-m_i}{N}n\frac{N-n}{N-1}, \quad i = 1, \ldots, s.$$

Consider drawing a ball $s$ times from the population described above. For a fixed $i, j \in \{1, \ldots, s\}$, let $Y_l$ and $Z_l$, $l = 1, \ldots, n$, be indicators of sampling from the $i$th and $j$th subgroups in the $l$th draw, respectively. We can then treat $X_i = \sum_{l=1}^{s} Y_l$ and $X_j = \sum_{l=1}^{n} Z_l$. Note that

$$\begin{aligned}
&EY_l = EY_1 = \frac{m_i}{N}, \quad EZ_k = EZ_1 = \frac{m_j}{N},\\
&EY_l^2 = EY_1^2 = EY_1 = \frac{m_i}{N},\\
&EY_{l_1}Y_{l_2} = EY_1Y_2 = \frac{m_i}{N}\frac{m_i-1}{N-1}, \quad l_1 \neq l_2\\
&EZ_{l_1}Z_{l_2} = EZ_1Z_2 = \frac{m_j}{N}\frac{m_j-1}{N-1}, \quad l_1 \neq l_2\\
&EY_lZ_l = 0,\\
&EY_{l_1}Z_{l_2} = EY_1Z_2 = P(Z_2 = 1|Y_1 = 1)P(Y_1 = 1) = \frac{m_j}{N-1}\frac{m_i}{N}, \quad l_1 \neq l_2.
\end{aligned}$$

Then

$$\begin{aligned}
EX_i^2 &= E\left(\sum_{l=1}^{n} Y_l\right)^2 = \sum_{l=1}^{n} EY_l + \sum_{l_1=1}^{n}\sum_{l_2=1, l_1\neq l_2}^{n} EY_{l_1}Y_{l_2}\\
&= n\frac{m_i}{N} + n(n-1)\frac{m_i(m_i-1)}{N(N-1)}\\
&= \frac{nm_i}{N}\left(1 + \frac{(m_i-1)(n-1)}{N-1}\right),
\end{aligned}$$

and

$$\begin{aligned}
EX_iX_j &= E\sum_{l_1=1}^{n} Y_k \sum_{l_2=1}^{n} Z_{l_2} = \sum_{l_1=1}^{n}\sum_{l_2=1}^{n} EY_{l_1}Z_{l_2}\\
&= \sum_{k=1}^{n}\left(0 + (n-1)\frac{m_im_j}{N(N-1)}\right) = \frac{n(n-1)m_im_j}{N(N-1)}, \quad i \neq j.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= EX_iX_j - EX_iEX_j = \frac{n(n-1)m_im_j}{N(N-1)} - \frac{nm_i}{N}\frac{nm_j}{N} \\
&= \frac{m_im_j}{N^2(N-1)}\left(Nn(n-1) - n^2(N-1)\right) \\
&= -\frac{m_im_jn(N-n)}{N^2(N-1)}.
\end{aligned}$$

In relation to the survey bootstrap weights and phase II bootstrap weights, we consider a special case where a vector of random variables $(W_{n1}, \ldots, W_{nn})$ follows the multivariate hypergeometric distribution $MH(nr, n, (r, \ldots, r))$. Note that $(W_{n1}, \ldots, W_{nn})$ is exchangeable. In general, a random vector from the multivariate hypergeometric distribution does not satisfy exchangeability. However, exchangeability holds when the sizes of the subgroups are all equal; $m_1 = m_2 = \ldots = m_s$. We have

$$W_{ni} \in \{0, 1, \ldots, r\}, \quad \overline{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_{ni} = \frac{1}{n}n = 1.$$

Since each $W_{ni}, i = 1, \ldots, n$, marginally follows the hypergeometric distribution,

$$\begin{aligned}
EW_{ni} &= \frac{nk}{nk} = 1, \\
\mathrm{Var}(W_{ni}) &= \frac{k}{nk}\frac{nk-k}{nk}n\frac{nk-n}{nk-1} = \frac{1}{n}\frac{n-1}{n}n^2\frac{k-1}{nk-1}.
\end{aligned}$$

It follows from the result above that

$$EW_{ni}^2 = \frac{nk}{nk}\left(1 + \frac{(k-1)(n-1)}{nk-1}\right) = \frac{2nk-n-k}{nk-1}$$

and that

$$\mathrm{Cov}(W_{ni}, W_{nj}) = -\frac{k^2n(nk-n)}{n^2k^2(nk-1)} = -\frac{k-1}{nk-1}, \quad i \neq j.$$

### *7.4.2 Bootstrap IPW Empirical Processes*

As a foundational building block for our bootstrap methods, we consider the case where there is one stratum where $N$ and $n$ are sample sizes at phase I and phase II, respectively, with $n/N \to p > 0$. To focus on the phase II observations, we define $Y_i, i = 1, \ldots, n$, as the observations $X_i$ with $\xi_i = 1$. The phase II empirical measure is defined by $\mathbb{P}_n \equiv n^{-1}\sum_{i=1}^{n}\delta_{Y_i}$. Note that

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{n/N}\delta_{X_i} = \mathbb{P}_N^{\pi}. \tag{7.10}$$

Note also that $Y_i$'s are independent since they do not involve the sampling indicators $\xi_i$ (see Remark 4.3 of [48]) and hence the usual Donsker theorem applies to obtain $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_0) \rightsquigarrow \mathbb{G}$ in $\ell^{\infty}(\mathcal{F})$ where $\mathcal{F}$ is a Donsker class, $\mathbb{G}$ is the limit process of the empirical process for complete data given by

$$\mathbb{G}_N \equiv \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{X_i} - P_0\right),$$

and $P_0$ is the law for $X_i$. Let $W_{Ni}$, $i = 1, \ldots, N$, be bootstrap weights from the survey bootstrap. That is, a vector of $W_{Ni}$ with $\xi_i = 1$ follows the mixture of the multivariate hypergeometric distributions $MH(nk, n, (k, \ldots, k))$ with probability $s = (1 - r/n)(1 - r/(N-1))$ and $MH(n(k+1), n, (k+1, \ldots, k+1))$ with probability $1 - s$ where $N = kn + r$, $k, r \in \mathbb{N}$ with $0 \le r < n$. We define the bootstrap empirical measure by

$$\hat{\mathbb{P}}_n \equiv \hat{\mathbb{P}}_N^{\pi,S} = \frac{1}{N}\sum_{i=1}^{N}W_{Ni}\frac{\xi_i}{n/N}\delta_{X_i} \equiv \frac{1}{n}\sum_{i=1}^{n}W_{ni,\xi}\delta_{Y_i}. \tag{7.11}$$

and the bootstrap empirical process by $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)$. Here we write $W_{ni,\xi}$ for $W_{Nj}$ which corresponds to $Y_i$ through $X_j$ with $\xi_j = 1$.

The following lemma is a Glivenko-Cantelli theorem and a Donsker theorem for the survey bootstrap IPW empirical process for one stratum.

**Lemma 7.4.1.** *Let $\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)$ where $\mathbb{P}_n$ and $\hat{\mathbb{P}}_n$ are defined in (7.10) and (7.11), respectively. Let $\mathcal{F}$ be a Glivenko-Cantelli class. Then*

$$\left\|\hat{\mathbb{P}}_n - \mathbb{P}_n\right\|_{\mathcal{F}} \to_{P_W} 0, \quad P^* - a.s. \tag{7.12}$$

*Let $\mathcal{F}$ be a Donsker class. Then*

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_0) \rightsquigarrow \mathbb{G},$$

*in* $\ell(\mathcal{F})$, *and*

$$\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n) \rightsquigarrow \sqrt{1-p}\mathbb{G},$$

*in* $\ell(\mathcal{F})$ *in* $P^*$*-probability.*

*Proof.* We first prove the weak convergence of $\mathbb{G}_n$. As discussed in Remark 4.3 of [48], $Y_i, i = 1, \ldots, n$, are i.i.d. $P_0$ and hence the Donsker theorem applies. Here we provide a different proof using the same argument as [8]. Noting that there is only one stratum ($N_j = N, n_j = n$), it follows from the decomposition (2.1) of the IPW empirical process (see also [8]) that

$$\begin{aligned}
\mathbb{G}_n &= \sqrt{\frac{n}{N}}\mathbb{G}_N^\pi = \sqrt{\frac{n}{N}}\mathbb{G}_N + \sqrt{\frac{n}{N}}\sum_{j=1}^{1}\sqrt{\frac{N}{N}}\frac{N}{n}\mathbb{G}_{1,N}^\xi \\
&= \sqrt{\frac{n}{N}}\mathbb{G}_N + \sqrt{\frac{N}{n}}\mathbb{G}_{1,N}^\xi \\
&\rightsquigarrow \sqrt{p}\mathbb{G} + p^{-1/2}\sqrt{p(1-p)}\mathbb{G}_1 \sim \mathbb{G},
\end{aligned}$$

where $\mathbb{G}_1$ is independent of $\mathbb{G}$ and $\mathbb{G}_1 \sim \mathbb{G}$.

Next, we consider asymptotic results for the bootstrap empirical measure. Recall that $N = kn + r$, $k, r \in \mathbb{N}$, $r < n$. Note that $k$, which implicitly depends on $N$ and $n$, is uniformly bounded in $N$ and $n$ because $n/N \to p > 0$. Suppose, to the contrary, that $k \to \infty$ as $N \to \infty$. Then,

$$1 = N/N = k(n/N) + r/N \geq k(n/N) \to \infty,$$

as $N \to \infty$, which is a contradiction. Thus, $W_{ni,\xi} \in \{0, 1, \ldots, k\}, i = 1, \ldots, n$, is bounded by some $M > 0$ uniformly in $n$. This implies that $\max_i W_{ni,\xi}/n \to 0$ and Lemma 3.6.16 of [58] yields (7.12).

For weak convergence of the bootstrap empirical process, we take advantage of the fact that $Y_i, i = 1, \ldots, n$, are i.i.d. $P_0$ so that the theory of [41] easily applies. Uniform boundedness of exchangeable bootstrap weights implies the first two of the conditions of

(3.6.8) of [58]. For the third condition of (3.6.8) of [58], note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(W_{ni,\xi}-1)^2 &= \frac{1}{n}\sum_{i=1}^{n}W_{ni,\xi}^2 - 2\frac{1}{n}\sum_{i=1}^{n}W_{ni,\xi} + \frac{1}{n}n \\ &= \frac{1}{n}\sum_{i=1}^{n}W_{ni,\xi}^2 - 1.\end{aligned}$$

Thus, it reduces to obtaining the limit of $n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2$ in $P_W$-probability. We first obtain $\lim_{n\to\infty} E|n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2|$. Note that $N = kn + r$ implies that $nk-1 = N-r-1$, and $n(k+1)-1 = N+n-r-1$. Note also that $s = \{(n-r)(N-1-r)\}/\{n(N-1)\}$ and $1-s = r(N+n-r-1)/\{n(N-1)\}$. Then we have

$$\begin{aligned}&E\left[\frac{1}{n}\sum_{i=1}^{n}W_{ni,\xi}^2\right] = EW_{n1,\xi}^2 \\ &= s\frac{2nk-n-k}{nk-1} + (1-s)\frac{2n(k+1)-n-(k+1)}{n(k+1)-1} \\ &= \frac{2-n/N-1/n}{1-1/N} \\ &\to 2-p.\end{aligned}$$

Since $0 \le n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2 < M^2$, $\{n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2\}$ is uniformly integrable and there exists a subsequence $\{n_l\}$ of $\{n\}$ such that $n_l^{-1}\sum_{i=1}^{n_l}W_{(n_l)i,\xi}^2 \to w \in [0, M^2]$ as $l\to\infty$. It follows from Vitali's theorem that $E|n_l^{-1}\sum_{i=1}^{n_l}W_{(n_l)i,\xi}^2| \to E|w| = w$ as $l\to\infty$. But this $w$ must be $2-p$. Since this is true for any subsequence of $\{n\}$, we have $n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2 \to_{P_W^*} 2-p$. This implies that $n^{-1}\sum_{i=1}^{n}W_{ni,\xi}^2 - 1 \to_{P_W^*} 1-p \equiv c^2$. Thus, it follows from Theorem 3.6.13 of [58] that

$$\hat{\mathbb{G}}_n \rightsquigarrow c\mathbb{G} = \sqrt{1-p}\,\mathbb{G}, \quad \text{in } \ell^\infty(\mathcal{F}),$$

in $P^*$-probability. □

We redefine $\hat{\mathbb{P}}_{j,n_j}^{\xi,S}$ and $\mathbb{P}_{j,n_j}^{\xi}$ in order to use Lemma 7.4.1. Let

$$\begin{aligned}\hat{\mathbb{P}}_{j,n_j}^{\xi,S} &= \frac{1}{n_j}\sum_{i=1}^{N_j}W_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} \equiv \frac{1}{n_j}\sum_{i=1}^{n_j}W_{n_j,j,i,\xi}\delta_{X_{j,i,\xi}}, \quad j=1,\ldots,J, \\ \mathbb{P}_{j,n_j}^{\xi} &= \frac{1}{n_j}\sum_{i=1}^{N_j}\xi_{j,i}\delta_{X_{j,i}} \equiv \frac{1}{n_j}\sum_{i=1}^{n_j}\delta_{X_{j,i,\xi}}, \quad j=1,\ldots,J.\end{aligned}$$

Here $W_{n_j,j,i,\xi}$ and $X_{j,i,\xi}$ are $W_{n_j,j,k}$ and $X_{j,k}$ respectively where $k$ is the $i$th smallest index among the observations in the $j$th stratum with $\xi = 1$. Recall that $\hat{\mathbb{P}}^{\xi,S}_{j,n_j}$ and $\mathbb{P}^{\xi}_{j,n_j}$ are first defined in Section 4.1.2.

We prove a Glivenko-Cantelli theorem for bootstrapped two-phase sampling.

*Proof of Theorem 7.1.1.* It follows from Lemma 7.4.1 applied to each $\hat{\mathbb{P}}^{\xi,S}_{j,n_j} - \mathbb{P}^{\xi}_{j,n_j}$ in the decomposition (4.7) of the survey bootstrap IPW empirical process that

$$\left\|\hat{\mathbb{P}}^{\pi,S}_N - \mathbb{P}^{\pi}_N\right\|_{\mathcal{F}} = \left\|\sum_{j=1}^{J} \frac{N_j}{N}(\hat{\mathbb{P}}^{\xi,S}_{j,n_j} - \mathbb{P}^{\xi}_{j,n_j})\right\|_{\mathcal{F}} \leq \sum_{j=1}^{J} \frac{N_j}{N} \left\|\hat{\mathbb{P}}^{\xi,S}_{j,n_j} - \mathbb{P}^{\xi}_{j,n_j}\right\|_{\mathcal{F}} \to_{P^*_W} 0,$$

in $P^*$-probability. Note that we used $N_j/N = O_{P_W}(1)$ in $P^*$-probability.

For $\left\|\hat{\mathbb{P}}^{\pi,S,mc}_N - \mathbb{P}^{\pi}_N\right\|_{\mathcal{F}}$, note that $\tilde{\mathcal{F}}_1 \equiv \{G_{mc}(\cdot;\alpha)f : f \in \mathcal{F}, \alpha \in \mathbb{R}^k\}$ and $\tilde{\mathcal{F}}_2 \equiv \{\dot{G}_{mc}(z;\alpha)(1/\pi_0(v) - 1)z^T f(x) : f \in \mathcal{F}, \alpha \in \mathbb{R}^k\}$ are $P_0$-Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). It follows from the fact that $\hat{\mathbb{P}}^{\pi,S,mc}_N f = \hat{\mathbb{P}}^{\pi,S}_N G_{mc}(V;\hat{\hat{\alpha}}_N)f$ and Taylor's theorem applied to $G_{mc}(V;\hat{\hat{\alpha}}_N) - 1$ that

$$\begin{aligned}
&\left\|\hat{\mathbb{P}}^{\pi,S,mc}_N - \mathbb{P}^{\pi}_N\right\|_{\mathcal{F}} \\
&= \left\|\hat{\mathbb{P}}^{\pi,S,mc}_N f - \mathbb{P}^{\pi}_N G_{mc}(V;\hat{\hat{\alpha}}_N)f + \mathbb{P}^{\pi}_N G_{mc}(V;\hat{\hat{\alpha}}_N)f - \mathbb{P}^{\pi}_N f\right\|_{\mathcal{F}} \\
&\leq \left\|(\hat{\mathbb{P}}^{\pi,S}_N - \mathbb{P}^{\pi}_N)G_{mc}(V;\hat{\hat{\alpha}}_N)f\right\|_{\mathcal{F}} + \left\|\mathbb{P}^{\pi}_N \dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^T f(\hat{\hat{\alpha}}_N - \alpha_0)\right\|_{\mathcal{F}} \\
&\leq \left\|\hat{\mathbb{P}}^{\pi,S}_N - \mathbb{P}^{\pi}_N\right\|_{\tilde{\mathcal{F}}_1} + \|\mathbb{P}^{\pi}_N\|_{\tilde{\mathcal{F}}_2}(\hat{\hat{\alpha}}_N - \alpha_0),
\end{aligned}$$

where $\alpha^*$ is some convex combination of $\hat{\hat{\alpha}}_N$ and $\alpha_0$. The first term is $o_{P^*_W}(1)$ by the result we just established above. For the second term, $\|\mathbb{P}^{\pi}_N\|_{\tilde{\mathcal{F}}_2}$ is $O_{Pr^*}(1)$ so that it is of order $O_{P^*_W}(1)$ in $P^*$-probability by Lemma 4.1.1, Since $\hat{\hat{\alpha}}_N - \alpha_0 = o_{P^*_W}(1)$ in $P^*$-probability by Proposition 4.3.1, the second term is $o_{P^*_W}(1)$ in $P^*$-probability. The arguments for $\hat{\mathbb{P}}^{\pi,S,c}_N$ and $\hat{\mathbb{P}}^{\pi,S,cc}_N$ are similar.

We consider $\left\|\hat{\mathbb{P}}^{\pi,S,dmc}_N - \mathbb{P}^{\pi,mc}_N\right\|_{\mathcal{F}}$. Note that $\tilde{\mathcal{F}}_3 \equiv \{G_{mc}(\cdot;\alpha_1)G_{mc}(\cdot;\alpha_2)f : f \in \mathcal{F}, \alpha_1, \alpha_2 \in$

$\mathbb{R}^k\}$ is $P_0$-Glivenko-Cantelli. As in the above, we have

$$\begin{aligned}
&\left\|\hat{\mathbb{P}}_N^{\pi,S,dmc} - \mathbb{P}_N^{\pi,mc}\right\|_{\mathcal{F}} \\
&= \left\|\hat{\mathbb{P}}_N^{\pi,S,dmc} f - \mathbb{P}_N^{\pi,mc} G_{mc}(V;\hat{\hat{\alpha}}_N) f + \mathbb{P}_N^{\pi,mc} G_{mc}(V;\hat{\hat{\alpha}}_N) f - \mathbb{P}_N^{\pi,mc} f\right\|_{\mathcal{F}} \\
&\leq \left\|\hat{\mathbb{P}}_N^{\pi,S} - \mathbb{P}_N^{\pi}\right\|_{\tilde{\mathcal{F}}_3} + \left\|\mathbb{P}_N^{\pi,mc}\right\|_{\tilde{\mathcal{F}}_2} (\hat{\hat{\alpha}}_N - \alpha_0) \\
&= o_{P_W^*}(1),
\end{aligned}$$

in $P^*$-probability. The cases for $\hat{\mathbb{P}}_N^{\pi,S,dc}$ and $\hat{\mathbb{P}}_N^{\pi,S,dcc}$ are similar.

We consider $\left\|\hat{\mathbb{P}}_N^{\pi,dmc} - \mathbb{P}_N^{\pi,mc}\right\|_{\mathcal{F}}$. The cases for $\hat{\mathbb{P}}_N^{\pi}, \hat{\mathbb{P}}_N^{\pi,\#}, \hat{\mathbb{P}}_N^{\pi,d\#}$ with $\# \in \{c, mc, cc\}$ are similar. We decompose $\hat{\mathbb{P}}_N^{\pi,dmc} - \mathbb{P}_N^{\pi,mc}$;

$$\begin{aligned}
\left\|\hat{\mathbb{P}}_N^{\pi,dmc} - \mathbb{P}_N^{\pi,mc}\right\|_{\mathcal{F}} &\leq \left\|\hat{\mathbb{P}}_N^{\pi,(1),dmc} - \mathbb{P}_N^{\pi,mc}\right\|_{\mathcal{F}} + \left\|\hat{\mathbb{P}}_N^{\pi,dmc} - \hat{\mathbb{P}}_N^{\pi,(1),dmc}\right\|_{\mathcal{F}} \\
&= \left\|(\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^{\pi}) G_{mc}(V;\hat{\alpha}_N) f\right\|_{\mathcal{F}} + \left\|\hat{\mathbb{P}}_N^{\pi,dmc} - \hat{\mathbb{P}}_N^{\pi,(1),dmc}\right\|_{\mathcal{F}}.
\end{aligned}$$

Note that $\{G_{mc}(v;\alpha) f : \alpha \in \mathcal{A}_{mc}, f \in \mathcal{F}\}$ is Glivenko-Cantelli by a Glivenko-Cantelli preservation theorem (Theorem 3, [59]). It follows from Lemma 7.4.5 that the first term in the display is $o_{P_W^*}(1)$ in $P^*$-probability. For the second term, we make use of the observation made in Chapter 4 that $\sqrt{n_j}(\hat{\mathbb{P}}_N^{\pi,dmc} - \hat{\mathbb{P}}_N^{\pi,(1),dmc})$ can be viewed as the survey bootstrap IPW empirical process with modified calibration conditional on $(X_{j,1}, W_{n_j,j,1}^{(1)}), (X_{j,2}, W_{n_j,j,2}^{(1)}), \ldots$, $\hat{\mathbb{G}}_{j,N_j}^{\xi,(2)}$, $j = 1, \ldots, J, i = 1, \ldots, N_j$ (see also a discussion on the probability space in the proof of Lemma 7.4.7). Thus the result just established above implies that the second term is also $o_{P_W^*}(1)$ in $P^*$-probability. This completes the proof. □

Rate of convergence for bootstrap IPW $M$-estimators are easily established since boundedness of bootstrap weights allows us to use the multiplier inequality for bounded weights (Lemma 6.2.1).

*Proof of Theorem 7.2.1.* The proof is similar to the proof of Theorem 6.2.1. □

We establish the weak convergence of the survey bootstrap IPW empirical processes based on the theory of exchangeably weighted bootstrap [41].

*Proof of Theorem 7.3.1.* Recall the decomposition (4.7) of the survey bootstrap IPW empirical process. For $\hat{\mathbb{G}}_N^{\pi,S}$, note that $\sqrt{N_j/N} \to \sqrt{\nu_j}$, $P^*$-almost surely by the strong law of large numbers and $\sqrt{N_j/n_j} \to p_j^{-1/2}$ by assumption, It follows from Lemma 7.4.1 that

$$\hat{\mathbb{G}}_N^{\pi,S} = \sum_{j=1}^{J} \sqrt{\frac{N_j}{N}} \sqrt{\frac{N_j}{n_j}} \hat{\mathbb{G}}_{j,n_j}^{\xi} \rightsquigarrow \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j,$$

in $P^*$-probability.

For $\hat{\mathbb{G}}_N^{\pi,S,mc}$, we have

$$\begin{aligned} \hat{\mathbb{G}}_N^{\pi,S,mc} f &= \hat{\mathbb{G}}_N^{\pi,S,mc} f - \hat{\mathbb{G}}_N^{\pi,S} f + \hat{\mathbb{G}}_N^{\pi,S} f = \hat{\mathbb{G}}_N^{\pi,S} f + \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,mc} - \hat{\mathbb{P}}_N^{\pi,S}) f \\ &= \hat{\mathbb{G}}_N^{\pi,S} f + \hat{\mathbb{P}}_N^{\pi,S} \dot{G}_{mc}(V;\alpha^*) \frac{1-\pi_0(V)}{\pi_0(V)} Z^T f \sqrt{N}(\hat{\alpha}_N - \alpha_0) \end{aligned}$$

Apply the bootstrap Glivenko-Cantelli theorem (Theorem 7.1.1) and Proposition 4.3.1 to obtain the finite-dimensional convergence. For asymptotic equicontinuity, proceed in the same way as in the proof of Theorem 6.3.1. The cases for $\hat{\mathbb{G}}_N^{\pi,S,c}$ and $\hat{\mathbb{G}}_N^{\pi,S,cc}$ are similar.

For $\hat{\mathbb{G}}_N^{\pi,S,dmc}$, we have

$$\begin{aligned} \hat{\mathbb{G}}_N^{\pi,S,dmc} f &= \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,dmc} - \mathbb{P}_N^{\pi,mc}) f - \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\hat{\alpha}_N) f + \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\hat{\alpha}_N) f \\ &= \sqrt{N}\hat{\mathbb{P}}_N^{\pi,S} G_{mc}(V;\hat{\alpha}_N)\{G_{mc}(V;\hat{\hat{\alpha}}_N) - 1\} f + \sqrt{N}\mathbb{P}_N^{\pi,mc}(W-1) f. \end{aligned} \tag{7.13}$$

Applying Taylor's theorem, the first term can be written as

$$\hat{\mathbb{P}}_N^{\pi,S} \left\{ \dot{G}_{mc}(V;\alpha^*) G_{mc}(V;\hat{\alpha}_N) \frac{1-\pi_0(V)}{\pi_0(V)} Z^T f \right\} \sqrt{N}(\hat{\hat{\alpha}}_N - \alpha_0),$$

where $\alpha^*$ is some convex combination of $\hat{\hat{\alpha}}_N$ and $\alpha_0$. It follows from the bootstrap Glivenko-Cantelli theorem with $\alpha^*, \hat{\alpha}_N \to_{P_W^*} \alpha_0$ in $P^*$-probability and Proposition 4.3.1 that this converges to

$$\begin{aligned} &-P_0 \left\{ \dot{G}_{mc}(V;\alpha_0) \frac{1-\pi_0(V)}{\pi_0(V)} Z^T f \right\} \dot{G}(0)^{-1} \left\{ P_0 \frac{1-\pi_0(V)}{\pi_0(V)} Z^{\otimes 2} \right\}^{-1} \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Z \\ &= -\sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Q_{mc} f, \end{aligned}$$

in $P^*$-probability. The second term of (7.13) can be written as

$$\sqrt{N}\mathbb{P}_N^{\pi} G_{mc}(V;\hat{\alpha}_N)(W-1) f = \hat{\mathbb{G}}_N^{\pi,S} f + \sqrt{N}\mathbb{P}_N^{\pi}\{G_{mc}(V;\hat{\alpha}_N) - 1\}(W-1) f.$$

The first term converges to $\hat{\mathbb{G}}_N^{\pi,S} f \rightsquigarrow \mathbb{G}^{\pi,S} f$ from the result above. Applying Taylor's theorem to the second term to obtain

$$\begin{aligned}
&\sqrt{N}\mathbb{P}_N^\pi\{G_{mc}(V;\hat{\alpha}_N)-1\}(W-1)f \\
&= \mathbb{P}_N^\pi(W-1)\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^T f\sqrt{N}(\hat{\alpha}_N-\alpha_0) \\
&= (\hat{\mathbb{P}}_N^\pi-\mathbb{P}_N^\pi)\dot{G}_{mc}(V;\alpha^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^T f\sqrt{N}(\hat{\alpha}_N-\alpha_0).
\end{aligned}$$

Since $\sqrt{N}(\hat{\alpha}_N-\alpha_0)=O_{P_W^*}(1)$ in $P^*$-probability by Proposition 3.5.1 and Lemma 4.1.1, the bootstrap Glivenko-Cantelli theorem (Theorem 7.1.1) implies that

$$\sqrt{N}\mathbb{P}_N^\pi\{G_{mc}(V;\hat{\alpha}_N)-1\}(W-1)f = o_{P_W^*}(1)O_{P_W^*}(1)=o_{P_W^*}(1),$$

in $P^*$-probability. Asymptotic equicontinuity can be verified in the same way as in the proof of Theorem 6.3.1. This completes the proof. □

We consider the uncentered version of the conditional multiplier central limit theorem. We first prove our version of Lemma 2.9.5 of [58]. This establishes the (conditional) finite dimensional convergence of $n^{-1/2}\sum_{i=1}^n \xi_i\delta_{X_i}$.

**Lemma 7.4.2.** *Let $Y_1, Y_2, \ldots,$ be i.i.d. Euclidean random vectors with $E\|Y_i\|^2<\infty$ independent of the i.i.d. $\xi_1,\xi_2,\ldots,$ with $E\xi_i=0$ and $E\xi_i^2=c^2>0$. Then, conditionally on $Y_1, Y_2\ldots,$*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Y_i \rightsquigarrow N(0, c^2EY_1^{\otimes 2}),$$

*for almost every sequence $Y_1, Y_2, \ldots$.*

*Proof.* We apply the Lindeberg central limit theorem with $S_n = n^{-1/2}\sum_{i=1}^n Z_i \equiv n^{-1/2}\sum_{i=1}^n \xi_i Y_i$. Note that $\mu_i \equiv E_\xi Z_i = E_\xi \xi_i Y_i = 0$ and $\sigma_i^2 \equiv \mathrm{Var}_\xi(Y_i) = E_\xi\xi_i^2 Y_i^{\otimes 2} - 0 = c^2 Y_i^{\otimes 2}$. Thus,

$$n^{-1}sd_n^2 \equiv n^{-1}\sum_{i=1}^n \sigma_i^2 = c^2 n^{-1}\sum_{i=1}^n Y_i^{\otimes 2} \to c^2 EY_1^{\otimes 2},$$

for almost all sequences by the strong law of large numbers. For every $\epsilon > 0$,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\|Y_i\|^2 E_\xi \xi_i^2\{|\xi_i|\|Y_i\| > \epsilon\sqrt{n}\} \\
& \leq n^{-1}\sum_{i=1}^{n}\|Y_i\|^2 E_\xi \xi_i^2\{|\xi_i| \max_{1\leq j\leq n}\|Y_j\| > \epsilon\sqrt{n}\} \\
& = n^{-1}\sum_{i=1}^{n}\|Y_i\|^2 E_\xi \xi_1^2\{|\xi_1| \max_{1\leq j\leq n}\|Y_j\| > \epsilon\sqrt{n}\} \\
& = \left(n^{-1}\sum_{i=1}^{n}\|Y_i\|^2\right) E_\xi \xi_1^2\{|\xi_1| \max_{1\leq j\leq n}\|Y_j\| > \epsilon\sqrt{n}\} \\
& \to 0,
\end{aligned}
$$

for almost all sequences, because $E\|Y_i\|^2 < \infty$ implies $\max_{1\leq i\leq n}\|Y_i\|/\sqrt{n} \to 0$ for almost all sequences. □

The next lemma concerns integrability of the empirical process when the $L_2(P_0)$-metric is used. This lemma is used to prove the uncentered conditional multiplier central limit theorem (Lemma 7.3.1).

**Lemma 7.4.3.** *Suppose that $\mathcal{F}$ is a Donsker class with $\|P_0\|_{\mathcal{F}} < \infty$. Let $X_1, X_2, \ldots$ be i.i.d. $P_0$, independent of i.i.d. Rademacher variables $\epsilon_1, \epsilon_2, \ldots$. Define the process $\tilde{\mathbb{G}}'_n = n^{1/2}\sum_{i=1}^{n}\epsilon_i\delta_{X_i}$. Let $\rho(f,g) = \{P_0(f-g)^2\}^{1/2}$ and $\mathcal{F}_\delta = \{f - g : \rho(f,g) < \delta, f, g \in \mathcal{F}\}$. Then $E^*\|\tilde{\mathbb{G}}'_n\|_{\mathcal{F}_\delta} \to 0$ for every $\delta_n \downarrow 0$.*

*Proof.* Since $\mathcal{F}$ is Donsker with $\|P_0\|_{\mathcal{F}} < \infty$, it follows from Corollary 2.9.4 of [58] that $\tilde{\mathbb{G}}'_n$ weakly converges to the Brownian motion process in $\ell^\infty(\mathcal{F})$ and $\tilde{\mathbb{G}}'_n$ is asymptotically equicontinuous in probability with respect to the $L_2(P_0)$-metric $\rho$. Moreover, $\mathcal{F}$ possesses an envelope $F$ with $P(F > x) = o(x^{-2})$ by Corollary 2.3.13 of [58]. This implies that $P(\|\epsilon_1\delta_{X_1}\|_{\mathcal{F}} > x) = P(F > x) = o(x^{-2})$. In view of Problem 2.3.3 of [58],

$$E^*\max_{1\leq i\leq n}\frac{\|\epsilon_i\delta_{X_i}\|_{\mathcal{F}}}{\sqrt{n}} \to 0.$$

It follows from the triangle inequality that the same is true with $\mathcal{F}$ is replaced by $\mathcal{F}_{\delta_n}$. Because asymptotic equicontinuity in probability implies $\|\tilde{\mathbb{G}}'_n\|_{\mathcal{F}_{\delta_n}} \to_P 0$ for every $\delta_n \downarrow 0$, the sequence of quantile functions of $\tilde{\mathbb{G}}'_n$ converges to zero pointwise. We can apply the Hoffmann-Jørgensen inequality to obtain the desired result. □

We prove the uncentered conditional multiplier central limit theorem.

*Proof of Lemma 7.3.1.* The sequence $\tilde{\mathbb{G}}_n$ converges to a $c$ times a $P_0$-Brownian motion process $\tilde{\mathbb{G}}$ in $\ell^\infty(\mathcal{F})$ by Corollary 2.9.4 of [58], and thus it is asymptotically measurable.

A Donsker class $\mathcal{F}$ is totally bounded for the $L_2(P_0)$ metric since $\|P_0\|_{\mathcal{F}} < \infty$ (Problem 2.1.1 of [58]). For each fixed $\delta > 0$ and $f \in \mathcal{F}$, let $\prod_\delta f$ denote a closest element in a given finite $\delta$-net for $\mathcal{F}$. By continuity of the limit process $\tilde{\mathbb{G}}$, we have $\tilde{\mathbb{G}} \circ \prod_\delta \mapsto \tilde{\mathbb{G}}$ almost surely as $\delta \downarrow 0$. Hence it follows that

$$\sup_{h \in BL_1} \left| Eh\left(c\tilde{\mathbb{G}} \circ \prod_\delta\right) - Eh(c\tilde{\mathbb{G}})\right| \to 0, \quad \delta \downarrow 0.$$

Also, it follows from Lemma 7.4.2 that for every fixed $\delta > 0$

$$\sup_{h \in BL_1} \left| E_\xi h\left(\tilde{\mathbb{G}}_n \circ \prod_\delta\right) - Eh(c\tilde{\mathbb{G}})\right| \to 0, \quad n \to \infty,$$

for almost every sequence $X_1, X_2, \ldots$ as in a proof of Theorem 2.9.6 of [58]. For completeness of the proof, we present details in the following. Define $A : \mathbb{R}^p \mapsto \ell^\infty(\mathcal{F})$ by $(Ay)f = y_i$ if $\prod_\delta f = f_i$. Then $h(c\tilde{\mathbb{G}} \circ \prod_\delta) = g(c\tilde{\mathbb{G}}f_1, \ldots, c\tilde{\mathbb{G}}f_p)$, for the function defined by $g(y) = h(Ay)$. If $h$ is bounded Lipschitz on $\ell^\infty(\mathcal{F})$, then $g$ is bounded Lipschitz on $\mathbb{R}^p$ with a smaller bounded Lipschitz norm. Since $BL_1(\mathbb{R}^p)$ is separable for the topology of uniform convergence on compacta, the supremum in the preceding display can be replaced by a countable supremum. It follows that the variable in the display is measurable because $h(c\tilde{\mathbb{G}} \circ \prod_\delta)$ is measurable. Next,

$$\begin{aligned}
\sup_{h \in BL_1} \left| E_\xi h\left(\tilde{\mathbb{G}}_n \circ \prod_\delta\right) - E_\xi h(\tilde{\mathbb{G}}_n)\right| &\le \sup_{h \in BL_1} E_\xi \left| h\left(\tilde{\mathbb{G}}_n \circ \prod_\delta\right) - h(\tilde{\mathbb{G}}_n)\right| \\
\le E_\xi \left\| \tilde{\mathbb{G}}_n \circ \prod_\delta - \tilde{\mathbb{G}}_n \right\|_{\mathcal{F}} &\le E_\xi \|\tilde{\mathbb{G}}_n\|_{\mathcal{F}_\delta},
\end{aligned}$$

where $\mathcal{F}_\delta = \{f - g : P_0(f-g)^2 < \delta^2\}$. Thus, the outer expectation of the left side is bounded above by $E^*\|\tilde{\mathbb{G}}_n\|_{\mathcal{F}_\delta}$.

Since the assumption $\|\xi_1\|_{2,1} < \infty$ implies $E\xi_1^2 < \infty$, we have $E \max_{1 \le i \le n} |\xi_i|/\sqrt{n} \to 0$. Thus, taking a limit on $n$ on both sides of the multiplier inequality (Lemma 2.9.1 of [58]) yields

$$\lim_{n\to\infty} E^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \delta_{X_i} \right\|_{\mathcal{F}_\delta} \leq 2\sqrt{2}\|\xi_1\|_{2,1} \sup_{n_0 \leq k} E^* \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \epsilon_i \delta_{X_i} \right\|_{\mathcal{F}_\delta},$$

for every $n_0$ and $\delta > 0$ where $\epsilon_i$ are i.i.d. Rademacher random variables independent of $\xi_i$ and $X_i$. The left hand side of the inequality converges to zero as $n_0 \to \infty$ followed by $\delta \downarrow 0$ because $\lim_{k\to\infty} E\|\mathbb{G}'_k\|_{\mathcal{F}_\delta} \to 0$ as $\delta \downarrow 0$ by Lemma 7.4.3 where $\tilde{\mathbb{G}}'_n = n^{-1/2}\sum_{i=1}^n \epsilon_i \delta_{X_i}$. Combining this with the previous display with the triangle inequality yields the first part of the claim.

For the second part of the claim, the proof of the first part applies except that it must be argued that $E_\xi \|\tilde{\mathbb{G}}\|^*_{\mathcal{F}_\delta}$ converges to zero outer almost surely as $n \to \infty$ followed by $\delta \downarrow 0$. Since $P_0\|f - P_0 f\|^2_{\mathcal{F}} < \infty$ and $\|P_0\|_{\mathcal{F}} < \infty$ implies

$$\begin{aligned}
&P_0\|f(X_1)\|^2_{\mathcal{F}} \leq P_0\|f(X_1) - P_0 f + P_0 f\|^2_{\mathcal{F}} \\
&\leq P_0\{\|f(X_1) - P_0 f\|^2_{\mathcal{F}} + \|P_0 f\|^2_{\mathcal{F}} + 2\|f - P_0 f\|_{\mathcal{F}}\|P_0 f\|_{\mathcal{F}}\} < \infty,
\end{aligned}$$

it follows from Corollary 2.9.9 that

$$\limsup_{n\to\infty} E_\xi \|\tilde{\mathbb{G}}_n\|^*_{\mathcal{F}_\delta} \leq 6\sqrt{2} \limsup_{n\to\infty} E^* \|\tilde{\mathbb{G}}_n\|_{\mathcal{F}_\delta},$$

almost surely. The right-hand side decreases to zero as $\delta \downarrow 0$ as shown above. To see that the sequence $E_\xi h(\tilde{\mathbb{G}}_n)$ is strongly asymptotically measurable, obtain first by the same proof, but with a star added, that

$$|E_\xi h(\tilde{\mathbb{G}}_n)^* - Eh(c\tilde{\mathbb{G}})| \to_{as^*} 0.$$

The same proof also shows that this is true with a lower star. Thus, the sequence $E_\xi h(\tilde{\mathbb{G}}_n)^* - h(\tilde{\mathbb{G}}_n)_*$ converges to zero almost surely. □

Now, we consider the i.i.d. weights $w_1, \ldots, w_n$ with $Ew_1 = 1$ and $\mathrm{Var}(w_1) = c^2$ that are independent of $X_1, \ldots, X_n$. Since we later consider the weak convergence of the two-phase bootstrap IPW empirical processes conditionally on $X_1, X_2, \ldots$, and the phase I bootstrap weights $W_{n_j,j,i}$, $j = 1, \ldots, J$, $i = 1, \ldots, N_j$, the following lemma allows us to easily apply the weak convergence results in Theorem 7.3.1.

**Lemma 7.4.4.** *Let the class of functions $\mathcal{F}$ be Donsker with $\|P_0\|_{\mathcal{F}} < \infty$. Let $X_1, \ldots, X_n$ be i.i.d. $P_0$ and define the empirical process $\mathbb{G}_n = n^{-1/2}\sum_{i=1}^n(\delta_{X_i} - P_0)$. Let $w_1, \ldots, w_n$ be i.i.d. $P_W$ with $Ew_1 = 1$, $\mathrm{Var}(w_1) = c^2 < \infty$ and $\|w_1\|_{2,1} < \infty$ that are independent of $X_1, \ldots, X_n$. Then the class of functions $\mathcal{WF} = \{g : g(x, w) = wf(x), f \in \mathcal{F}\}$ is $P_0 \times P_W$-Donsker.*

*Proof.* Note that for $g(x, w) = wf(x) \in \mathcal{WF}$,

$$\begin{aligned}
\mathbb{G}_n g &= \sqrt{n}(\mathbb{P}_n wf - P_0(wf)) = \sqrt{n}(\mathbb{P}_n wf - P_0 f) \\
&= n^{-1/2}\sum_{i=1}^n(\delta_{X_i} - P_0)f + \ n^{-1/2}\sum_{i=1}^n(w_i - 1)(\delta_{X_i} - P_0)f \\
&\quad + n^{-1/2}\sum_{i=1}^n(w_i - 1)P_0 f.
\end{aligned}$$

Thus in view of Corollary 2.9.4 of [58],

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} + c\mathbb{G}' + cZP_0, \quad \text{in } \ell^\infty(\mathcal{WF}),$$

where $\mathbb{G}$ and $\mathbb{G}'$ are independent Brownian bridge process that are independent of the standard normal random variable $Z$. □

Several results (Lemmas 7.4.2-7.4.4) regarding the uncentered conditional multiplier central limit theorem provide useful tools to study the phase I bootstrap IPW empirical process. We first prove a Glivenko-Cantelli theorem for the phase I bootstrap IPW empirical process.

**Lemma 7.4.5.** *Let $\mathcal{F}$ be a Glivenko-Cantelli class. Then $\|\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^\pi\|_{\mathcal{F}} \to_{P_W} 0$ in $P^*$-probability.*

*Proof.* Note the decomposition (4.20) of the phase I bootstrap IPW empirical process to obtain $\mathbb{P}_N^\pi = \sum_{j=1}^J (N_j/N)(\mathbb{P}_{j,n_j}^{\xi,(1)} - \mathbb{P}_{j,n_j}^\xi)$. It follows from the triangle inequality that

$$\left\|\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^\pi\right\|_{\mathcal{F}} \le \sum_{j=1}^J \frac{N_j}{N}\left\|\hat{\mathbb{P}}_{j,n_j}^{\xi,(1)} - \mathbb{P}_{j,n_j}^\xi\right\|_{\mathcal{F}}.$$

Note that $N_j/N = O_{P_W}(1)$ in $P^*$-probability. , we can apply Lemma 3.6.16 of [58] to each $\hat{\mathbb{P}}^{\xi,(1)}_{j,n_j} - \mathbb{P}^{\xi}_{j,n_j}$. Let $\overline{W}^{(1)}_j = n_j^{-1}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}\xi_{j,i}$. We have

$$\begin{aligned}
\hat{\mathbb{P}}^{\xi,(1)}_{j,n_j} - \hat{\mathbb{P}}^{\xi}_{j,n_j} &= \frac{1}{n_j}\sum_{i=1}^{N_j} W^{(1)}_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} - \frac{1}{n_j}\sum_{i=1}^{N_j}\xi_{j,i}\delta_{X_{j,i}} \\
&= \overline{W}^{(1)}_j \frac{1}{n_j}\sum_{i=1}^{N_j}\frac{W^{(1)}_{n_j,j,i}\xi_{j,i}}{\overline{W}^{(1)}_j}(\delta_{X_{j,i}} - P_{0|j}) + (\overline{W}^{(1)}_j - 1)P_{0|j}.
\end{aligned}$$

By assumption, $\overline{W}^{(1)}_j \to_{P_W} 1$. Because $n_j^{-1}\sum_{i=1}^{N_j}(W_{n_j,j,i}\xi_{j,i}/\overline{W}^{(1)}_j) = 1$, applying Lemma 3.6.16 of [58] yields that

$$\begin{aligned}
\left\|\overline{W}^{(1)}_j \frac{1}{n_j}\sum_{i=1}^{N_j}\frac{W^{(1)}_{n_j,j,i}\xi_{j,i}}{\overline{W}^{(1)}_j}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} &= |\overline{W}^{(1)}_j| \left\|\frac{1}{n_j}\sum_{i=1}^{N_j}\frac{W^{(1)}_{n_j,j,i}\xi_{j,i}}{\overline{W}^{(1)}_j}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} \\
&= O_{P_W}(1)o_{P_W}(1) = o_{P_W}(1),
\end{aligned}$$

in $P^*$-probability. Thus, it suffices to show $\|P_{0|j}\|_{\mathcal{F}} < \infty$. To see this, notice that

$$\begin{aligned}
&\|P_{0|j}\|_{\mathcal{F}} = \nu_j^{-1}E\|\delta_X I_{\mathcal{V}_j}(V) - \nu_j P_{0|j} - \delta_X I_{\mathcal{V}_j}(V)\|_{\mathcal{F}} \\
&\le \nu_j^{-1}E\|\delta_X I_{\mathcal{V}_j}(V) - \nu_j P_{0|j}\|_{\mathcal{F}} + \nu_j^{-1}E\|\delta_X I_{\mathcal{V}_j}(V)\|_{\mathcal{F}} \\
&\le \nu_j^{-1}E\left\|\sum_{j=1}^{J}(\delta_X I_{\mathcal{V}_j}(V) - \nu_j P_{0|j})\right\|_{\mathcal{F}} + \nu_j^{-1}E\|\delta_X I_{\mathcal{V}_j}(V)\|_{\mathcal{F}} \\
&= \nu_j^{-1}E\|f - P_0 f\|_{\mathcal{F}} + \nu_j^{-1}E\|\delta_X I_{\mathcal{V}_j}(V)\|_{\mathcal{F}} \\
&\le \nu_j^{-1}E\|f - P_0 f\|_{\mathcal{F}} + \nu_j^{-1}E\|\delta_X\|_{\mathcal{F}}.
\end{aligned}$$

Because $\mathcal{F}$ is Donsker, and hence Glivenko-Cantelli, the first term is bounded. Moreover, since $\|P\|_{\mathcal{F}} < \infty$, the second term is bounded by $EF^* < \infty$ for some envelope function $F$ in view of Problem 2.4.1 of [58]. This completes the proof. □

Next, we prove the conditional weak convergence of the phase I bootstrap IPW empirical process.

**Lemma 7.4.6.** *Let $\mathcal{F}$ be a Donsker class with $\|P_0\|_{\mathcal{F}} < \infty$. Then,*

$$\hat{\mathbb{G}}^{\pi,(1)}_N \rightsquigarrow \sum_{j=1}^{J}\sqrt{\frac{\nu_j}{2-p_j}}\tilde{\mathbb{G}}^{(1)}_j, \quad \textit{in } \ell^\infty(\mathcal{F}),$$

*where the $P_{0|j}$-Brownian motion processes $\tilde{\mathbb{G}}_j^{(1)}$ are all independent.*

*The same holds when $\hat{\mathbb{G}}_N^{\pi,(1)}$ is replaced by $\hat{\mathbb{G}}_N^{\pi,(1),\#}$ with $\# \in \{dc, dmc, dcc\}$.*

*Proof.* First, we prove the claim for $\hat{\mathbb{G}}_N^{\pi,(1)}$. Recall the decomposition (4.20) of the phase I bootstrap IPW empirical process;

$$\hat{\mathbb{G}}_N^{\pi,(1)} = \sum_{j=1}^{J} \sqrt{\frac{N_j}{N}} \sqrt{\frac{N_j}{n_j}} \hat{\mathbb{G}}_{j,n_j}^{\xi,(1)},$$

where

$$\hat{\mathbb{G}}_{j,n_j}^{\xi,(1)} \equiv \sqrt{n_j}\left(\hat{\mathbb{P}}_{j,n_j}^{\xi,(1)} - \mathbb{P}_{j,n_j}^{\xi}\right) = \sqrt{n_j}\frac{1}{n_j}\sum_{i=1}^{N_j}(W_{n_j,j,i}^{(1)} - 1)\xi_{j,i}\delta_{X_{j,i}}.$$

Note that $X_{j,i}$, $i = 1, \dots, N_j$, with $\xi_{j,i} = 1$ are independent. Note also that $\tilde{W}_{n_j,j,i}^{(1)} = W_{n_j,j,i}^{(1)} - 1$ has mean zero and variance $c_j^2$ satisfying $\|\tilde{W}_{n_j,j,i}^{(1)}\|_{2,1} < \infty$ in view of Problem 2.9.2 of [58]. Because we showed $\|P_{0|j}\|_{\mathcal{F}} < \infty$ for $j = 1, \dots, J$, in the proof of Lemma 7.4.5, it follows from the uncentered conditional multiplier central limit theorem (Lemma 7.3.1) applied to each of the phase I bootstrap IPW empirical processes $\hat{\mathbb{G}}_{j,n_j}^{\xi,(1)}$ for the $j$th stratum $j = 1, \dots, J$ that

$$\hat{\mathbb{G}}_{j,n_j}^{\xi,(1)} \rightsquigarrow c_j(\mathbb{G}_j^{(1)} + Z_j P_{0|j}) \quad \text{in } \ell^\infty(\mathcal{F}),$$

in $P^*$-probability where $\mathbb{G}_j^{(1)}$ is a $P_{0|j}$-Brownian bridge process independent of the standard normal random variables $Z_j$. Note that $\mathbb{G}_j^{(1)}$ and $Z_j$, $j = 1, \dots, J$, are all independent. Hence

$$\hat{\mathbb{G}}_N^{\pi,(1)} \rightsquigarrow \sum_{j=1}^{J} \sqrt{\frac{\nu_j}{p_j}} \sqrt{\frac{p_j}{2 - p_j}} (\mathbb{G}_j^{(1)} + Z_j P_{0|j}) = \sum_{j=1}^{J} \sqrt{\frac{\nu_j}{2 - p_j}} (\mathbb{G}_j^{(1)} + Z_j P_{0|j}),$$

in $\ell^\infty(\mathcal{F})$ in $P^*$-probability, where the $P_{0|j}$-Brownian bridge processes $\mathbb{G}_j$ and the standard normal random variables $Z_j$ are all independent. Note that $\mathbb{G}_j^{(1)} + Z_j P_{0|j}$ are $P_{0|j}$-Brownian motion processes.

Next, we prove the claim for $\hat{\mathbb{G}}_N^{\pi,(1),dc}$. Other cases are similar. Note from definitions of $\hat{\mathbb{P}}_N^{\pi,(1),dc}$ and $\mathbb{P}_N^{\pi,c}$ that

$$\hat{\mathbb{G}}_N^{\pi,(1),dc} f = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,(1),dc} - \mathbb{P}_N^{\pi,c}) f = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^{\pi}) G_c(V; \hat{\alpha}_N^c) f = \mathbb{G}_N^{\pi,(1)} G_c(V; \hat{\alpha}_N^c) f.$$

For a finite-dimensional convergence, we have for $f \in \mathcal{F}$ that

$$\begin{aligned}\hat{\mathbb{G}}_N^{\pi,(1),dc} f &= \hat{\mathbb{G}}_N^{\pi,(1)} f + (\hat{\mathbb{G}}_N^{\pi,(1),dc} - \hat{\mathbb{G}}_N^{\pi,(1)}) f \\ &= \hat{\mathbb{G}}_N^{\pi,(1)} f + (\hat{\mathbb{P}}_N^{\pi,(1)} - \mathbb{P}_N^{\pi}) \dot{G}_c(V;\alpha^*) Z^T \sqrt{N}(\hat{\alpha}_N^c - \alpha_0)\end{aligned}$$

where $\alpha^*$ is some convex combination of $\hat{\alpha}_N^c$ and $\alpha_0$. It follows from Lemma 7.4.5 together with Proposition 3.5.1 that the second term is $o_{P_W}(1)O_{P_W}(1) = o_{P_W}(1)$ in $P^*$-probability. Asymptotic equicontinuity can be verified in the same way as in the proof of Theorem 6.3.1.

□

For the weak convergence of the phase II bootstrap IPW empirical processes, we condition on both the data and the phase I bootstrap weights.

**Lemma 7.4.7.** *Let $\mathcal{F}$ be a Donsker class with $\|P_0\|_{\mathcal{F}} < \infty$. Then,*

$$\hat{\mathbb{G}}_N^{\pi,(2)} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}(W_j \cdot),$$

$$\hat{\mathbb{G}}_N^{\pi,(2),c} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_c)W_j \cdot\},$$

$$\hat{\mathbb{G}}_N^{\pi,(2),dc} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_c)W_j \cdot\},$$

$$\hat{\mathbb{G}}_N^{\pi,(2),mc} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_{mc})W_j \cdot\},$$

$$\hat{\mathbb{G}}_N^{\pi,(2),dmc} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_{mc})W_j \cdot\},$$

$$\hat{\mathbb{G}}_N^{\pi,(2),cc} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_{cc})W_j \cdot\},$$

$$\hat{\mathbb{G}}_N^{\pi,(2),dcc} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j^{(2)}\{(I - Q_{mc})W_j \cdot\},$$

*in $\ell^\infty(\mathcal{F})$ in $P^* \times P_W^{(1)}$-probability where $P_{0|j}$-Brownian bridge processes $\mathbb{G}_j^{(2)}$ and $\mathbb{G}_j^{(1)}$ and $Z_j$ defined in Lemma 7.4.6 are all independent, $W_j$ are independent with mean 1 and variance $c_j^2$ that are independent of $(X, V)$, $Q_\#$ with $\# \in \{c, mc, cc\}$ are defined in Theorem 3.2.1.*

*Proof.* We first prove the claim for $\hat{\mathbb{G}}_N^{\pi,(2)}$. Recall the decomposition (4.21) of the phase II bootstrap IPW empirical process;

$$\hat{\mathbb{G}}_N^{\pi,(2)} = \sum_{j=1}^J \sqrt{\frac{N_j}{N}}\sqrt{\frac{N_j}{n_j}}\hat{\mathbb{G}}_{j,n_j}^{\xi,(2)},$$

where

$$\hat{\mathbb{G}}_{j,n_j}^{\xi,(2)} \equiv \sqrt{n_j}\left(\hat{\mathbb{P}}_{j,n_j}^{\xi} - \hat{\mathbb{P}}_{j,n_j}^{\xi,(1)}\right) = n_j^{-1/2}\sum_{i=1}^{N_j}(W_{n_j,j,i}^{(2)} - 1)\xi_{j,i}\{W_{n_j,j,i}^{(1)}\delta_{X_{j,i}}\}.$$

Note that $(W_{n_j,j,i}^{(1)}, X_{j,i})$, $i = 1, \ldots, N_j$, with $\xi_{j,i} = 1$ are independent. Since $\mathcal{F}_j = \{g(x,w) = wf(x) : f \in \mathcal{F}\}$ is $P_{0|j} \times P_{W_j}^{(1)}$-Donsker by Lemma 7.4.4, it follows from Lemma 7.4.1 that

$$\hat{\mathbb{G}}_{j,N_j}^{\xi,(2)} \rightsquigarrow \sqrt{1-p_j}\mathbb{G}_j^{(2)} \quad \text{in } \ell^\infty(\mathcal{F}_j),$$

conditionally on $(X_{j,1}, W_{n_j,j,1}^{(1)}), (X_{j,2}, W_{n_j,j,2}^{(1)}), \ldots$. Note that $\mathbb{G}_j^{(2)}$ are independent of $\mathbb{G}_j^{(1)}$ and $Z_j$ for $j = 1, \ldots, J$. Hence it follows that

$$\hat{\mathbb{G}}_N^{\pi,(2)} \rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j(W_j^{(1)}\cdot) \qquad \text{in } \ell^\infty(\mathcal{F}),$$

in $P^* \times P_W^{(1)}$-probability where $W_j^{(1)}$ is independent of $X$ with mean 1 and variance $c_j^2$.

Another way to derive this result is that conditional on $(X_{j,1}, W_{n_j,j,1}^{(1)}), (X_{j,2}, W_{n_j,j,2}^{(1)}), \ldots$, $\hat{\mathbb{G}}_{j,N_j}^{\xi,(2)}$ can be viewed as the survey bootstrap IPW empirical process indexed by the set $\{g(x,w) = wf(x) : f \in \mathcal{F}\}$ as discussed in Chapter 4. Although we only define the conditional distributions of $W_N^{(1)}$ given the stratum membership, there exists a probability distribution, say, $P_{W,V}$ such that $W_{Ni}^{(1)}$, $i = 1, \ldots, N$ are i.i.d. $P_{W,V}$. See the Appendix A of [8] for details. Thus, the result follows from Theorem 7.3.1. From this viewpoint, the claims for other phase II bootstrap IPW empirical processes, which can also be viewed as the survey bootstrap IPW empirical processes, are proved by Theorem 7.3.1. Of course, we can prove those statements based on Lemma 7.4.1 by closely following the argument in the proof of Theorem 7.3.1 without considering the probability distribution $P_{W,V}$. □

*Proof of Theorem 7.3.2.* We first prove the claim for $\hat{\mathbb{G}}_N^{\pi}$. Other cases are similar. Decompose $\hat{\mathbb{G}}_N^{\pi}$ into $\hat{\mathbb{G}}_N^{\pi,(1)} + \hat{\mathbb{G}}_N^{\pi,(2)}$ and apply Lemmas 7.4.6 and 7.4.7 to obtain a mean-zero

Gaussian process as a limit process. Recall that $c_j^2 = p_j/(2-p_j)$ and $W_j^{(1)}$ is independent of $X$. Its covariance function evaluated at $f, g \in \mathcal{F}$ is given by

$$
\begin{aligned}
&\sum_{j=1}^J \left\{ \frac{\nu_j}{2-p_j} P_{0|j}(f-g)^2 + \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(W_j^{(1)} f - W_j^{(1)} g) \right\} \\
&= \sum_{j=1}^J \left\{ \frac{\nu_j}{2-p_j} P_{0|j}(f-g)^2 + \nu_j \frac{1-p_j}{p_j} \left[ E(W_j^{(1)})^2 P_{0|j}(f-g)^2 - \{E W_j^{(1)} P_{0|j}(f-g)\}^2 \right] \right\} \\
&= \sum_{j=1}^J \left\{ \frac{\nu_j}{2-p_j} P_{0|j}(f-g)^2 + \nu_j \frac{1-p_j}{p_j} \left[ (c_j^2+1) P_{0|j}(f-g)^2 - \{P_{0|j}(f-g)\}^2 \right] \right\} \\
&= \sum_{j=1}^J \left\{ \nu_j P_{0|j}(f-g)^2 + \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(f-g) \right\} \\
&= P_0(f-g)^2 + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j} \mathrm{Var}_{0|j}(f-g).
\end{aligned}
$$

Since this covariance function is the same as that for the process of our claim, the result follows.

Next, we consider the claim for $\hat{\mathbb{G}}_N^{\pi,mc}$. Arguments for other two-phase bootstrap IPW empirical processes are similar. Recall the definition of $Q_{mc} f$ in Theorem 3.2.1. It follows from the independence of $W_j^{(1)}$ and $(X, V)$ and $E W_j^{(1)} = 1$ that

$$
\begin{aligned}
Q_{mc} W_j^{(1)} f &= P_0[(\pi_0^{-1}(V)-1) W_N^{(1)} f Z^T] \{P_0[(\pi_0^{-1}(V)-1) Z^{\otimes 2}]\}^{-1} Z \\
&= E W_j^{(1)} P_0[(\pi_0^{-1}(V)-1) f Z^T] \{P_0[(\pi_0^{-1}(V)-1) Z^{\otimes 2}]\}^{-1} Z \\
&= P_0[(\pi_0^{-1}(V)-1) f Z^T] \{P_0[(\pi_0^{-1}(V)-1) Z^{\otimes 2}]\}^{-1} Z \\
&= Q_{mc} f.
\end{aligned}
$$

Thus, we have $Q_{mc} W_j^{(1)} f - Q_{mc} W_j^{(1)} g = Q_{mc}(f-g)$. We compute $\mathrm{Var}_{0|j}((I - Q_{mc}) W_j^{(1)} f -$

$I - Q_{mc})W_j^{(1)}g)$;

$$
\begin{aligned}
&\mathrm{Var}_{0|j}((I - Q_{mc})W_j^{(1)}f - (I - Q_{mc})W_j^{(1)}g)\\
&= \mathrm{Var}_{0|j}(W_j^{(1)}(f-g) - Q_{mc}(f-g)))\\
&= \mathrm{Var}_{0|j}(W_j^{(1)}(f-g)) + \mathrm{Var}(Q_{mc}(f-g)) - 2\mathrm{Cov}(W_j^{(1)}(f-g), Q_{mc}(f-g))\\
&= c_j^2 P_{0|j}(f-g)^2 + \mathrm{Var}_{0|j}(f-g) + \mathrm{Var}(Q_{mc}(f-g))\\
&\quad -2[EWP_{0|j}\{(f-g)Q_{mc}(f-g)\} - EWP_{0|j}(f-g)P_{0|j}Q_{mc}(f-g)]\\
&= c_j^2 P_{0|j}(f-g)^2 \mathrm{Var}_{0|j}(f-g) + \mathrm{Var}(Q_{mc}(f-g))\\
&\quad -2[P_{0|j}\{(f-g)Q_{mc}(f-g)\} - P_{0|j}(f-g)P_{0|j}Q_{mc}(f-g)]\\
&= c_j^2 P_{0|j}(f-g)^2 \mathrm{Var}_{0|j}(f-g) + \mathrm{Var}(Q_{mc}(f-g)) - 2\mathrm{Cov}(f-g, Q_{mc}(f-g))\\
&= c_j^2 P_{0|j}(f-g)^2 + \mathrm{Var}_{0|j}((I - Q_{mc})f - (I - Q_{mc})g).
\end{aligned}
$$

Note that we used the independence of $W_j^{(1)}$ and $(X, V)$ and $EW_j^{(1)} = 1$. Now, proceed similarly to the case for $\hat{\mathbb{G}}_N^{\pi}$ to compute the covariance function of the limiting process for $\hat{\mathbb{G}}_N^{\pi,dmc}$ evaluated at $f$ and $g$, and verify that the covariance function is the same as that of $\mathbb{G}^{\pi,tp,mc}$ as desired. □

The following is the bootstrap version of Lemma 6.4.2. Because the survey and two-phase bootstrap weights are bounded, we can make use of the multiplier inequality for bounded weights (Lemma 6.2.1).

**Lemma 7.4.8.** *For an arbitrary set $\mathcal{F}$ of integrable functions,*

$$
E^* \left\| \hat{\mathbb{G}}_N^{\pi,S} \right\|_{\mathcal{F}} \lesssim E^* \left\| \mathbb{G}_N \right\|_{\mathcal{F}}, \qquad E^* \left\| \hat{\mathbb{G}}_N^{\pi} \right\|_{\mathcal{F}} \lesssim E^* \left\| \mathbb{G}_N \right\|_{\mathcal{F}}.
$$

*Proof.* We first prove the claim on $\hat{\mathbb{G}}_N^{\pi,S}$. We have

$$
\begin{aligned}
E_W^* \| \hat{\mathbb{G}}_N^{\pi,S} \|_{\mathcal{F}} &\leq \sum_{j=1}^{J} \sqrt{\frac{N_j}{N}} \sqrt{\frac{N_j}{n_j}} E_W^* \left\| \hat{\mathbb{G}}_{j,n_j}^{\xi,S} \right\|_{\mathcal{F}}\\
&\leq \sigma^{-1/2} \sum_{j=1}^{J} E_W^* \left\| \hat{\mathbb{G}}_{j,n_j}^{\xi,S} \right\|_{\mathcal{F}},
\end{aligned}
$$

where $E_W$ denotes the conditional expectation $E(\cdot | (V_i, \xi_i X_i, \xi)_{i=1}^N)$ given all data $(V_i, \xi_i X_i, \xi)_{i=1}^N$. Here we used $N_j/N \leq 1$ and $N_j/n_j \leq \sigma^{-1}$. Taking expectation with respect to $P$, it suffices to show that each $E^* \left\| \hat{\mathbb{G}}_{j,n_j}^{\xi,S} \right\|_{\mathcal{F}}$ is bounded up to some constant by $E^* \left\| \mathbb{G}_N \right\|_{\mathcal{F}}$.

Rewrite $\hat{\mathbb{G}}^{\xi,S}_{j,n_j} = n_j^{-1/2}\sum_{i=1}^{n_j} W_{n_j,j,i,\xi}(\delta_{X_{j,i,\xi}} - \mathbb{P}^{\xi}_{j,n_j})$. Note the conditional exchangeability of $W_{n_j,j,i,\xi}, i = 1,\ldots,n_j$ given all data $(V_i, \xi_i X_i, \xi)_{i=1}^N$ with $n_j^{-1}\sum_{i=1}^{n_j} W_{n_j,j,i,\xi} = 1$. Note also that $0 \le W_{n_j,j,i,\xi} \le M_j$ for some constant $M_j > 0$ uniformly in $n_j$ because of the same argument in the proof of Lemma 7.4.1. Apply the multiplier inequality of Lemma 6.2.1 conditionally to obtain

$$E_W \|\hat{\mathbb{G}}^{\xi,S}_{j,n_j}\|_{\mathcal{F}} \le (M_j - 0) \max_{1\le k\le n_j} E_W \left\| \frac{1}{\sqrt{n_j}} \sum_{i=1}^{k} (\delta_{X_{j,i,\xi}} - \mathbb{P}^{\xi}_{j,n_j}) \right\|^*_{\mathcal{F}}.$$

Then, apply Jensen's inequality and the triangle inequality to further bound this by

$$\begin{aligned} &M_j E_W \left\| \frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} (\delta_{X_{j,i,\xi}} - P_{0|j}) \right\|^*_{\mathcal{F}} + M_j E_W \left\| \sqrt{n_j}(\mathbb{P}^{\xi}_{j,n_j} - \mathbb{P}_{j,n_j}) \right\|^*_{\mathcal{F}} \\ &+ M_j E_W \left\| \sqrt{n_j}(\mathbb{P}_{j,n_j} - P_{0|j}) \right\|^*_{\mathcal{F}}. \end{aligned}$$

Taking expectation unconditionally, it follows from Jensen's inequality that the first term is bounded by

$$M_j E^* \left\| \sqrt{\frac{N_j}{n_j}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} (\delta_{X_{j,i}} - P_{0|j}) \right\|^*_{\mathcal{F}} \le M_j \sigma^{-1/2} E^* \left\| \mathbb{G}_{j,N_j} \right\|_{\mathcal{F}},$$

since $\delta_{X_{j,i}} - P_{0|j}$ has mean zero. Now, apply Lemma 6.4.1 to find that this is bounded by $E^*\|\mathbb{G}_N\|_{\mathcal{F}}$. The third term is handled exactly the same way as the first term. For the second term, noting that $\mathbb{P}^{\xi}_{j,n_j} = (N_j/n_j)\mathbb{P}^{\xi}_{j,N_j}$, the multiplier inequality of Lemma 6.2.1 yields

$$M_j E_W \left\| \sqrt{n_j}(\mathbb{P}^{\xi}_{j,n_j} - \mathbb{P}_{j,n_j}) \right\|^*_{\mathcal{F}} \le M_j \max_{1\le k\le N_j} E_W \left[ \sqrt{\frac{N_j}{n_j}} \left\| \frac{1}{\sqrt{N_j}} \sum_{i=1}^{k} (\delta_{X_{j,i}} - \mathbb{P}_{j,n_j}) \right\|_{\mathcal{F}} \right]$$

Now, following the proof of Lemma 6.4.2, the unconditional expectation of this term is bounded by $E^*\|\mathbb{G}_N\|_{\mathcal{F}}$ up to some constant. This completes the proof for $\hat{\mathbb{G}}^{\pi,S}_N$.

Next, we prove the claim for $\hat{\mathbb{G}}^{\pi}_N$. Note that $W_{n_j,j,i} = W^{(1)}_{n_j,j,i} W^{(2)}_{n_j,j,i}$ are bounded uniformly in $N$ so that the multiplier inequality for bounded weights (Lemma 6.2.1) can apply.

Recall that

$$\begin{aligned}
\hat{\mathbb{G}}_N^\pi &= \sqrt{N}\left(\hat{\mathbb{P}}_N^\pi - \mathbb{P}_N^\pi\right) \\
&= \sum_{j=1}^J \frac{N_j}{\sqrt{N}}\left(\frac{1}{n_j}\sum_{i=1}^{N_j} W_{n_j,j,i}\xi_{j,i}\delta_{X_{j,i}} - \frac{1}{n_j}\sum_{i=1}^{N_j}\xi_{j,i}\delta_{X_{j,i}}\right) \\
&= \sum_{j=1}^J \frac{N_j}{\sqrt{N}}\left(\frac{1}{n_j}\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(W^{(1)}_{n_j,j,i}\delta_{X_{j,i}} - P_{0|j}) - \frac{1}{n_j}\sum_{i=1}^{N_j}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j})\right) \\
&= \sum_{j=1}^J \frac{N_j}{n_j\sqrt{N}}\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(W^{(1)}_{n_j,j,i}\delta_{X_{j,i}} - P_{0|j}) - \sum_{j=1}^J \frac{N_j}{n_j\sqrt{N}}\sum_{i=1}^{N_j}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j}).
\end{aligned}$$

Here we used the fact that $n_j^{-1}\sum_{i=1}^{N_j}\xi_{j,i}W^{(2)}_{n_j,j,i} = 1$ and $n_j^{-1}\sum_{i=1}^{N_j}\xi_{j,i} = 1$. Thus, we have

$$\begin{aligned}
\left\|\hat{\mathbb{G}}_N^\pi\right\|_{\mathcal{F}} &\le \sum_{j=1}^J \left\|\frac{N_j}{n_j\sqrt{N}}\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(W^{(1)}_{n_j,j,i}\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} \\
&\quad + \sum_{j=1}^J \frac{N_j}{n_j\sqrt{N}}\left\|\sum_{i=1}^{N_j}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} \\
&\le \sigma^{-1}\sum_{j=1}^J \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(W^{(1)}_{n_j,j,i}\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} \\
&\quad + \sigma^{-1}\sum_{j=1}^J \frac{1}{\sqrt{N}}\left\|\sum_{i=1}^{N_j}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}}. \qquad (7.14)
\end{aligned}$$

For the $j$th summand in the first term in (7.14), note that, with a slight abuse of notation, $(W^{(2)}_{n_j,j,1}\xi_{j,1},\ldots,W^{(2)}_{n_j,j,N_j}\xi_{j,N_j})$ is exchangeable bounded weights among observations with $\xi_{j,i} = 1$. Note also that we can rewrite the summand as

$$\left\|\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(W^{(1)}_{n_j,j,i}\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}} = \left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j} W^{(2)}_{n_j,j,i}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{W}_j\mathcal{F}},$$

where $\mathcal{W}_j\mathcal{F} = \{g(x, w_j^{(1)}) = w_j^{(1)}f(x) : f \in \mathcal{F}\}$ since $P_{0|j}W_j^{(1)}f(X) = P_{0|j}f(X)$ since $EW^{(1)}_{n_j,j,i} = 1$. Apply the multiplier inequality for bounded weights (Lemma 6.2.1) with $n_0 = 1$ and $Z_{ni} = \delta_{X_{j,i}} - P_{0|j}$ conditionally on the phase I bootstrap weights $W^{(1)}$ to obtain

$$E_{-W^{(1)}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j} W_{n_j,j,i}\xi_{j,i}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{W}_j\mathcal{F}} \lesssim \max_{1\le k\le N_j} E_{-W^{(1)}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{k}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{W}_j\mathcal{F}},$$

where $E_{-W^{(1)}}$ denotes the conditional expectation given the phase I bootstrap weights. By Jensen's inequality, this term is bounded by

$$\max_{1\le k\le N_j} E_{-W^{(1)}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j}(\delta_{X_{j,i}}-P_{0|j})\right\|_{\mathcal{W}_j\mathcal{F}} = E_{-W^{(1)}}\left\|\sqrt{\frac{N_j}{N}}\mathbb{G}_{j,N_j}\right\|_{\mathcal{W}_j\mathcal{F}}.$$

Since $N_j/N \le 1$, the expectation of this term is further bounded by

$$\begin{aligned} E\left\|\sqrt{\frac{N_j}{N}}\mathbb{G}_{j,N_j}\right\|_{\mathcal{W}_j\mathcal{F}} &\le E\|\mathbb{G}_{j,N_j}(W_j^{(1)}-1)f+\mathbb{G}_{j,N_j}f\|_{\mathcal{F}} \\ &\le E\|\mathbb{G}_{j,N_j}(W_j^{(1)}-1)f\|_{\mathcal{F}}+E\|\mathbb{G}_{j,N_j}f\|_{\mathcal{F}}. \end{aligned}$$

The second term in the last display is bounded by $E\|\mathbb{G}_N\|_{\mathcal{F}}$ by Lemma 6.4.1. Since $W_{n_j,j,i}$ are bounded and exchangeable, it follows from the multiplier inequality for the bounded exchangeable weights (Lemma 6.2.1) with $n_0 = 1$ that the first term in the last display is bounded, up to some constant, by

$$\max_{1\le k\le N_j} E\left\|\frac{1}{\sqrt{N_j}}\sum_{i=1}^{k}(\delta_{X_{j,i}}-P_{0|j})\right\|_{\mathcal{F}}.$$

This term is bounded by $E\|\mathbb{G}_N\|_{\mathcal{F}}$ by Jensen's inequality and Lemma 6.4.1. Apply the multiplier inequality and Jensen's inequality to the $j$th summand in the second term in (7.14) as above to obtain

$$\begin{aligned} E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j}\xi_{j,i}(\delta_{X_{j,i}}-P_{0|j})\right\|_{\mathcal{F}} &\lesssim \max_{1\le k\le N_j} E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{k}(\delta_{X_{j,i}}-P_{0|j})\right\|_{\mathcal{F}} \\ &\le \max_{1\le k\le N_j} E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j}(\delta_{X_{j,i}}-P_{0|j})\right\|_{\mathcal{F}} \\ &\le E\left\|\sqrt{\frac{N_j}{N}}\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}. \end{aligned}$$

This completes the proof. □

The following lemma is useful when showing the asymptotic equicontinuity of bootstrap IPW empirical processes as Lemma 7.4.9 for the arguments of the asymptotic equicontinuity of IPW empirical processes.

**Lemma 7.4.9.** *Suppose that Condition 4.1.4 holds. Let $\mathcal{F}$ be a Glivenko-Cantelli class. Then*

$$\sup_{f\in\mathcal{F}}\left|\sqrt{N}(\mathbb{P}_N - P_0)\left\{\frac{W\xi}{\pi_{\hat{\hat{\alpha}}_N}(V)}f - \frac{W\xi}{\pi_{\alpha_0}(V)}f\right\}\right| = o_{P_W^*}(1), \tag{7.15}$$

*in $P^*$-probability where a (doubly) calibrated probability (with modified or centered calibration) and weights $W$ can be either survey or two-phase bootstrap weights.*

*Proof.* The proof is similar to the proof of Lemma 6.4.3. Note that

$$\begin{aligned}
&G_{mc}(V;\alpha_1)G_{mc}(V;\alpha_2) - 1 = G_{mc}(V;\alpha_1)G_{mc}(V;\alpha_2) - G_{mc}(V;\alpha_1) + G_{mc}(V;\alpha_1) - 1\\
&= G_{mc}(V;\alpha_1)\dot{G}_{mc}(V;\alpha_2^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^T(\alpha_2-\alpha_0) + \dot{G}_{mc}(V;\alpha_1^*)\frac{1-\pi_0(V)}{\pi_0(V)}Z^T(\alpha_1-\alpha_0),
\end{aligned}$$

where $\alpha_j^*$ is some convex combination of $\alpha_j$ and $\alpha_0$. □

*Proof of Lemma 7.3.2.* The proof is similar to the proof of Lemma 6.3.1. □

*Proof of Lemma 7.3.3.* Apply Lemma 7.4.8, and Markov's inequality to obtain $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_N} = o_{P_W^*}(1)$ in $P^*$-probability.

For the IPW empirical process with modified calibration, we have by Taylor's theorem that

$$\begin{aligned}
&\hat{\mathbb{G}}_N^{\pi,S,mc}f - \hat{\mathbb{G}}_N^{\pi,S}f = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S,mc} - \hat{\mathbb{P}}_N^{\pi,S})f = \sqrt{N}(\hat{\mathbb{P}}_N^{\pi,S}G_{mc}(V;\hat{\hat{\alpha}}_N)f - \hat{\mathbb{P}}_N^{\pi,S}f)\\
&= \hat{\mathbb{P}}_N^{\pi,S}\frac{1-\pi_0(V)}{\pi_0(V)}\dot{G}_{mc}(V;\alpha^*)fZ^T\sqrt{N}(\hat{\hat{\alpha}}_N - \alpha_0)\\
&= \hat{\mathbb{P}}_N^{\pi,S}\frac{1-\pi_0(V)}{\pi_0(V)}\dot{G}_{mc}(V;\alpha^*)fZ^TO_{P_W^*}(1)\\
&= (\hat{\mathbb{P}}_N^{\pi,S} - \mathbb{P}_N^{\pi})\frac{1-\pi_0(V)}{\pi_0(V)}\dot{G}_{mc}(V;\alpha^*)fZ^TO_{P_W^*}(1)\\
&\quad +(\mathbb{P}_N^{\pi} - P_0)\frac{1-\pi_0(V)}{\pi_0(V)}\dot{G}_{mc}(V;\alpha^*)fZ^TO_{P_W^*}(1)\\
&\quad +P_0\frac{1-\pi_0(V)}{\pi_0(V)}\dot{G}_{mc}(V;\alpha^*)fZ^TO_{P_W^*}(1),
\end{aligned}$$

in $P^*$-probability. The first and second term in the last display are $o_{P_W^*}(1)$ in $P^*$-probability by Theorems 7.1.1 and 6.1.1, respectively. The last term is $o_P^*(1)$ by the same argument as in the proof of Lemma 6.3.2 and hence $o_{P_W^*}(1)$ in $P^*$-probability.

The proofs for other cases are similar. □

# BIBLIOGRAPHY

[1] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982. With Discussion.

[2] Erika Antal and Yves Tillé. A direct bootstrap method for complex sampling designs from a finite population. *J. Amer. Statist. Assoc.*, 106(494):534–543, 2011.

[3] P. J. Bickel and D. A. Freedman. Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.*, 12(2):470–482, 1984.

[4] Ørnulf Borgan, Bryan Langholz, Sven Ove Samuelsen, Larry Goldstein, and Janice Pogoda. Exposure stratified case-cohort designs. *Lifetime Data Anal.*, 6(1):39–58, 2000.

[5] Norman E. Breslow, Thomas Lumley, Christie M. Ballantyne, Lloyd E. Chambless, and Michal Kulich. Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Stat. Biosc.*, 1:32–49, 2009.

[6] Norman E. Breslow, Thomas Lumley, Christie M. Ballantyne, Lloyd E. Chambless, and Michal Kulich. Using the whole cohort in the analysis of case-cohort data. *American J. Epidemiol.*, 169:1398–1405, 2009.

[7] Norman E. Breslow, Brad McNeney, and Jon A. Wellner. Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Ann. Statist.*, 31(4):1110–1139, 2003.

[8] Norman E. Breslow and Jon A. Wellner. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Statist.*, 34(1):86–102, 2007.

[9] Norman E. Breslow and Jon A. Wellner. A $Z$-theorem with estimated nuisance parameters and correction note for: "Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression" [Scand. J. Statist. **34** (2007), no. 1, 86–102]. *Scand. J. Statist.*, 35(1):186–192, 2008.

[10] Kwun Chuen Gary Chan. Personal communication, 2010.

[11] Hua Yun Chen and Roderick J. A. Little. Proportional hazards regression with missing covariates. *J. Amer. Statist. Assoc.*, 94(447):896–908, 1999.

[12] Jiahua Chen and J. N. K. Rao. Asymptotic normality under two-phase sampling designs. *Statist. Sinica*, 17(3):1047–1064, 2007.

[13] D. R. Cox. Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972. With discussion.

[14] G. J. D'Angio, Norman Breslow, J. B. Beckwith, A. Evans, H. Baum, A. deLorimier, D. Fernbach, E. Hrabovsky, B. Jones, and P. Kelalis. Treatment of Wilms' tumor. Results of the Third National Wilms' Tumor Study. *Cancer*, 64(2):349–360, Jul 1989.

[15] Jean-Claude Deville and Carl-Erik Särndal. Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, 87(418):376–382, 1992.

[16] V. P. Godambe and M. E. Thompson. Parameters of superpopulation and survey population: their relationships and estimation. *Internat. Statist. Rev.*, 54(2):127–138, 1986.

[17] D. M. Green, N. E. Breslow, J. B. Beckwith, J. Z. Finklestein, P. E. Grundy, P. R. Thomas, T. Kim, S. J. Shochat, G. M. Haase, M. L. Ritchey, P. P. Kelalis, and G. J. D'Angio. Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: a report from the National Wilms' Tumor Study Group. *J. Clin. Oncol.*, 16(1):237–245, Jan 1998.

[18] S. Gross. Median estimation in sample surveys. Paper presented at 1980 Amer. Statist. Assoc. meeting, 1980.

[19] BF Haynes, PB Gilbert, JM McElrath, ZollaPazner S, Tomaras GD, DT Alam SM, Evans, DC Montefiori, Karnasuta C, Sutthent R, DeVico AL Liao HX, GK Lewis, C Williams, Y Fong, H Janes, A DeCamp, Y Huang, M Rao, E Billings, N Karasavvas, ML Robb, V Ngauy, MS deSouza, R Paris, G Ferrari, RT Bailer, K Soderberg, C Andrews, PW Berman, N Frahm, SC DeRosa, MD Alpert, NL Yates, X Shen, RA Koup, P Pitisuttithum, J Kaewkungwal, S Nitayaphan, S RerksNgarm, NL Michael, and JH. Kim. Immune correlates analysis of the ALVAC-AIDSVAX HIV-1 vaccine efficacy trial. 2012. submitted.

[20] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47:663–685, 1952.

[21] Jian Huang. Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.*, 24(2):540–568, 1996.

[22] Jian Huang and Jon A. Wellner. Regression models with interval censoring. In *Probability theory and mathematical statistics (St. Petersburg, 1993)*, pages 269–296. Gordon and Breach, Amsterdam, 1996.

[23] Cary T. Isaki and Wayne A. Fuller. Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.*, 77(377):89–96, 1982.

[24] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Discrete multivariate distributions*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[25] Leslie Kish and Martin Richard Frankel. Inference from complex samples. *J. Roy. Statist. Soc. Ser. B*, 36:1–37, 1974. With discussion.

[26] Edward L. Korn and Barry I. Graubard. *Analysis of Health Surveys*. Wiley, 1999.

[27] Michal Kulich and D. Y. Lin. Improving the efficiency of relative-risk estimation in case-cohort studies. *J. Amer. Statist. Assoc.*, 99(467):832–844, 2004.

[28] D. Y. Lin. On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87(1):37–47, 2000.

[29] Thomas Lumley. Analysis of complex survey samples. 2008. R package version 3.10.1.

[30] Thomas Lumley, Pamela A. Shaw, and James Y. Dai. Connections between survey calibration estimators and semiparametric models for incomplete data. *International Statistical Review*, 79:200–232, 2011.

[31] Shuangge Ma and Michael R. Kosorok. Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.*, 96(1):190–217, 2005.

[32] Brad McNeney and Jon A. Wellner. Application of convolution theorems in semiparametric models with non-i.i.d. data. *J. Statist. Plann. Inference*, 91(2):441–480, 2000. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998).

[33] S. A. Murphy, A. J. Rossini, and Aad W. van der Vaart. Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.*, 92(439):968–976, 1997.

[34] S. A. Murphy and Aad W. van der Vaart. Semiparametric likelihood ratio inference. *Ann. Statist.*, 25(4):1471–1509, 1997.

[35] Susan A. Murphy and Aad W. van der Vaart. Observed information in semi-parametric models. *Bernoulli*, 5(3):381–412, 1999.

[36] Bin Nan. Efficient estimation for case-cohort studies. *Canad. J. Statist.*, 32(4):403–419, 2004.

[37] Bin Nan, Mary J. Emond, and Jon A. Wellner. Information bounds for Cox regression models with missing data. *Ann. Statist.*, 32(2):723–753, 2004.

[38] Jerzy Neyman. Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.*, 33:101–116, 1938.

[39] Gert G. Nielsen, Richard D. Gill, Per Kragh Andersen, and Thorkild I. A. Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.*, 19(1):25–43, 1992.

[40] S. W. Overton and Stephen V. Stehman. The Horvitz-Thompson theorem as a unifying perspective for probability sampling: With examples from natural resource sampling. *American Statistician*, 49:261–268, 1995.

[41] Jens Præstgaard and Jon A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, 21(4):2053–2086, 1993.

[42] Ross L. Prentice. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73:pp. 1–11, 1986.

[43] J. N. K. Rao and C.-F. J. Wu. Resampling inference with complex survey data. *J. Amer. Statist. Assoc.*, 83(401):231–241, 1988.

[44] S. Rerks-Ngarm, P. Pitisuttithum, S. Nitayaphan, J. Kaewkungwal, J. Chiu, R. Paris, N. Premsri, C. Namwat, M. de Souza, E. Adams, M. Benenson, S. Gurunathan, J. Tartaglia, J. G. McNeil, D. P. Francis, D. Stablein, D. L. Birx, S. Chunsuttiwat, C. Khamboonruang, P. Thongcharoen, M. L. Robb, N. L. Michael, P. Kunasol, J. H. Kim, S. Rerks-Ngarm, S. Chunsuttiwat, N. Premsri, C. Namwat, P. Kunasol, P. Thongcharoen, C. Khamboonruang, P. Pitisuttithum, V. Bussaratid, W. Maek-a nantawat, J. Dhitavat, P. Suntharasamai, S. Pungpak, S. Vanijanonta, J. Kaewkunwal, A. Khamsiriwatchara, P. Jarujareet, S. Nitayaphan, C. Easmila, S. Tabprasit, J. Chiu, R. Paris, M. Benenson, A. Brown, P. Morgan, M. de Souza, R. Trichavaroj, A. Schuetz, N. Thaitawat, S. Gurunathan, J. Tartaglia, J. G. McNeil, R. Harkness, C. Meric, R. El Habib, L. Baglyos, D. Francis, C. Lee, E. Adams, J. H. Kim, M. L. Robb, N. L. Michael, M. Milazzo, A. Bolen, B. Wessner, S. R. Kim, M. Marovich, J. Currier, D. L. Birx, D. Stablein, T. Germanson, L. Dally, R. Wiley, J. L. Excler, and J. Berenberg. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.*, 361:2209–2220, Dec 2009.

[45] James M. Robins, Fu Shing Hsieh, and Whitney Newey. Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *J. Roy. Statist. Soc. Ser. B*, 57(2):409–424, 1995.

[46] James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, 89(427):846–866, 1994.

[47] Susana Rubin-Bleuer. The proportional hazards model for survey data from independent and clustered super-populations. *Journal of Multivariate Analysis*, 101:884–895, 2011.

[48] Susana Rubin-Bleuer and Ioana Schiopu Kratina. On the two-phase framework for joint model and design-based inference. *Ann. Statist.*, 33(6):2789–2810, 2005.

[49] Takumi Saegusa and Jon A. Wellner. Weighted likelihood estimation under two-phase sampling. Technical report, 2011. available at arXiv:1112.4951.

[50] Takumi Saegusa and Jon A. Wellner. Supplementary materials for "Weighted likelihood estimation under two-phase sampling". 2012.

[51] Carl-Erik Särndal, Bengt Swensson, and Jan Wretman. *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer-Verlag, New York, 1992.

[52] Alastair J. Scott and Chris J. Wild. Fitting logistic models under case-control or choice based sampling. *J. Roy. Statist. Soc. Ser. B*, 48(2):170–182, 1986.

[53] Alastair J. Scott and Chris J. Wild. Hypothesis testing in case-control studies. *Biometrika*, 76(4):806–808, 1989.

[54] Alastair J. Scott and Chris J. Wild. On the robustness of weighted methods for fitting models to case-control data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(2):207–219, 2002.

[55] Steven G. Self and Ross L. Prentice. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.*, 16(1):64–81, 1988.

[56] Aad W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[57] Aad W. van der Vaart. Semiparametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1999)*, volume 1781 of *Lecture Notes in Math.*, pages 331–457. Springer, Berlin, 2002.

[58] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996.

[59] Aad W. van der Vaart and Jon A. Wellner. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 115–133. Birkhäuser Boston, Boston, MA, 2000.

[60] Jon A. Wellner and Yihui Zhan. Bootstrapping $Z$- estimators. Technical Report # 308, Department of Statistics, University of Washington, 1996.

[61] J. E. White. A two stage design for the study of the relationship between a rare exposure and and a rare disease. *Am. J. Epidemiol.*, 115(1):119–128, 1986.

[62] Y. Xie and C. F. Manski. The logit model and response-based samples. *Sociol. Meth. Res.*, 17:283–302, 1989.

# VITA

Takumi Saegusa was born to Yasuhiro and Kiyoko Saegusa on April 3 1978 in Takamatsu, Kagawa, Japan. He earned a Bachelor of Laws in Political Science at the University of Tokyo in 2002 and a Master of Science in Mathematics at the California State University, Northridge, in 2005. In 2012, he earned a Doctor of Philosophy at the University of Washington in Biostatistics.