Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 10-018

"Efficient Estimation of Average Treatment Effects under Treatment-Based Sampling", Second Version

by

Kyungchul Song

http://ssrn.com/abstract=1617262

Efficient Estimation of Average Treatment Effects under Treatment-Based Sampling

Kyungchul Song[1]

Department of Economics, University of Pennsylvania

May 24, 2010

Abstract

Nonrandom sampling schemes are often used in program evaluation settings to improve the quality of inference. This paper considers what we call treatment-based sampling, a type of standard stratified sampling where part of the strata are based on treatment status. This paper establishes semiparametric efficiency bounds for estimators of weighted average treatment effects and average treatment effects on the treated. This paper finds that adapting the efficient estimators of Hirano, Imbens, and Ridder (2003) to treatment-based sampling does not always lead to an efficient estimator. This paper proposes efficient estimators that involve a different form of propensity score-weighting. Finally, this paper establishes an optimal design of treatment-based sampling that minimizes the semiparametric efficiency bound over the sampling designs.

Key words and Phrases: treatment-based sampling, standard stratified sampling, semiparametric efficiency, treatment effects, optimal sampling designs

JEL Classifications: C12, C14, C52.

1 Introduction

Program evaluation studies often adopt nonrandom sampling to improve the quality of inference. For example, Ashenfelter and Card (1985) analyzed data from the Comprehensive Employment and Training Act (CETA) training program using a sample constructed by combining subsamples of program participants and a sample of nonparticipants drawn from the Current Population Survey (CPS). Also, the studies of Lalonde (1986), Dehejia and Wahba (1998, 1999) and Smith and Todd (2005) investigated the National Supported Work (NSW) training program where the training group consisted of individuals eligible for the program and the comparison sample were drawn from the CPS and the Panel Study of Income Dynamics (PSID) surveys. Numerous studies focused on

[1] I would like to express my gratitude to Petra Todd who gave me numerous valuable comments and advice. I also thank the Co-editor and two referees of *Econometric Theory* for comments that led to improvements of the paper. All errors are solely mine. Address correspondence to: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297.

the Job Training Partnership Act (JTPA) training program (e.g. Heckman, Ichimura, Smith and Todd (1998), Heckman, Ichimura and Todd (1997)). The participants in these data sets typically represented about 50% in the study sample in comparison to 3% in the population.

The rationale for such nonrandom sampling is often the belief that when the participants constitute a small proportion in the population, sampling relatively more from the participants will improve the quality of inference. However, this is not an accurate description because we need to consider also the contribution of the noise in the subsample to the variance of the estimator. (See Hahn, Hirano, and Karlan (2009) for a similar observation.) This paper makes this point clear by developing an optimal design of treatment-based sampling which is a kind of standard stratified sampling with strata based on the treatment status and other covariates.

The main objects of interest in this paper are the weighted average treatment effects and the average treatment effects on the treated considered by HIR. First, this paper considers observations from treatment-based sampling, and establishes semiparametric efficiency bounds for these parameters. Then, the paper proposes efficient estimators. The main challenge in the development is that it is not *a priori* clear how one can obtain an efficient estimator from the efficiency bounds, because the usual sample analogue principle does not apply. One might consider adapting the efficient estimators of Hahn (1998) or Hirano, Imbens, and Ridder (2003) (HIR, hereafter) to treatment-based sampling using appropriate change of measure as in Tripathi (2008). However, as this paper demonstrates, this naive adaptation does not work in general. This paper proposes efficient estimators that involve propensity score-weighting different from HIR.

Finally, this paper finds an optimal design of treatment-based sampling which minimizes the semiparametric efficiency bound over the sample designs. The analysis makes it clear how the noise from each subsample contributes to the semiparametric efficiency bound. As a corollary, a necessary and sufficient condition for a treatment-based sampling to improve on the random sampling is established when the strata involves only the treatment status. (See Hahn, Hirano and Karlan (2009) for an optimal design of social experiments in a related context.)

Early literatures on nonrandom sampling have assumed that the conditional distribution of observations given a stratum belongs to a parametric family. (Manski and Lerman (1977), Manski and McFadden (1981), Cosslett (1981a, 1981b), Imbens (1992), and Imbens and Lancaster (1996).) Wooldridge (1999, 2001) studied M-estimators under nonrandom sampling which do not rely on this assumption.

Closer to this paper, Breslow, McNeney and Wellner (2003) and Tripathi (2008) investigated the problem of efficient estimation under nonrandom sampling schemes. Tripathi (2008) considered moment-based models under various nonrandom sampling schemes and proved that the empirical likelihood estimators adapted to an appropriate change of measure achieve efficiency. The stratified sampling scheme studied by Tripathi (2008) is different from this paper's set-up because the identification of the counterfactual quantities in this paper cannot be formulated as arising from the moment condition of his paper. Neither does this paper's framework fall into the framework of Breslow, McNeney and Wellner (2003) who considered variable probability sampling which is different from the standard stratified sampling studied here.

In the program evaluations literature, there are surprisingly few researches that deal with inference under treatment-based sampling. Chen, Hong, and Tarozzi (2008) established semiparametric efficiency bounds in a broader context where one has outcome observations with missing values and has auxiliary data that aid identification. While the general approach of Chen, Hong, and Tarozzi (2008) applies to some stratified sampling schemes, it does not here because the event of missing values involves the treatment status, failing the unconfoundedness condition assumed in their paper. A paper by Heckman and Todd (2008) offers a nice, simple idea to estimate treatment effect on the treated under treatment-based sampling without assuming knowledge of aggregate shares. However, their paper does not focus on efficient procedures.

This paper proceeds as follows. Section two introduces treatment-based sampling data designs and Section three presents a general discussion on semiparametric efficiency bound when observations are from treatment-based sampling. Section four establishes semiparametric efficiency bounds for weighted average treatment effects and average treatment effects on the treated. Section five investigates efficient estimation. Section six develops optimal treatment-based sampling. Section seven concludes and the proofs are relegated to the appendix.

2 Treatment-Based Sampling

Treatment-based sampling proceeds as follows. Let D be a random variable that takes values in $\{0,1\}$, where $D=1$ means participation in the program and $D=0$ being left in the control group. Let $X=(V,W)$ be a vector of covariates, where W is a discrete random variable taking values from a finite set $\mathcal{W}$. For example, W may be the vector of dummy variables for the service regions in the JTPA job training program. The random vector V can contain continuous or discrete components. Under treatment-based sampling, a random sample of size N for the discrete vector (D,W) is first collected. Let $N_{d,w}=\sum_{i=1}^{N} 1\{(D_i,W_i)=(d,w)\}$, $(d,w)\in\{0,1\}\times\mathcal{W}$. From each subsample with $(D_i,W_i)=(d,w)$, a random sample $\{Y_i,V_i\}_{i=1}^{n_{d,w}}$ of predetermined size $n_{d,w}$ for a vector (Y,V) is collected, where $Y=\sum_{d\in\{0,1\}} Y_d 1\{D=d\}$ and Y_1 denotes the potential outcome of a person treated in the program and Y_0 the potential outcome of a person not treated. In this paper, we call this type of sampling *treatment-based sampling* as the strata $\{0,1\}\times\mathcal{W}$ involve treatment status. When $W_i=1$ for all i, so that the strata are constructed based only on the treatment status, we call this sampling *pure treatment-based sampling.* Throughout this paper, it is assumed that we do not have individual observations for $(D_i,W_i)_{i=1}^N$ from the original data set, although we may require knowledge of *aggregate shares* $p_{d,w}=P\{(D,W)=(d,w)\}$ for identification of certain parameters. (See the discussions prior to Theorem 1 in the following section.) While the observations in the combined sample $\{(D_i,Y_i,V_i)\}_{i=1}^n$ are independent across i's, the marginals are not identical. Hence inference based on random sampling can be misleading.

For an illustration of treatment-based sampling, consider a job training program implemented in K different service regions. (In the case of the JTPA job training program, there were 16 service regions.) Let $\mathcal{W} = \{1, 2, \cdots, K\}$, the set of index numbers representing the K service regions, and $W \in \mathcal{W}$ the service region index for the worker. Each individual worker has a treatment-region status represented by the pair (D, W). For example a worker with $(D, W) = (0, 3)$ means that the work is not treated and belongs to Service Region 3. When a service region has very few workers eligible for the program in the population, one may want to sample treated workers with a larger proportion than one represented in the population. The extent of the oversampling may differ across different service regions. Then one combines samples obtained by oversampling or undersampling the observations of (Y, V) from each (d, w)-subsample. The resulting total sample is one from treatment-based sampling whose distribution by itself is no longer *representative* of the population.

First, note that a likelihood for observations generated from standard stratified sampling can be viewed as a conditional likelihood from multinomial sampling given $\{n_{d,w}\}_{d,w \in (\{0,1\} \times \mathcal{W})}$. As pointed out by Imbens and Lancaster (1996) (see also Tripathi (2008)), (D, W) is ancillary in both stratified sampling and multinomial sampling, and hence it suffices for semiparametric efficiency to consider only multinomial sampling with design probabilities, say, $\{q_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$. Furthermore, $\{n_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$ is a sufficient statistic for multinomial distributions, and hence as far as semiparametric efficiency is concerned, we can assume that $\{q_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$ are known. We do not require full knowledge of $\{q_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$ for the actual construction of efficient estimators. The multinomial sampling is used only for the computation of semiparametric efficiency bounds.

Let the observations $\{(Y_i, V_i, D_i, W_i)\}_{i=1}^n$ for (Y, V, D, W) be generated by the multinomial sampling scheme using known design probabilities $\{q_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$. In other words, we draw a stratum (d, w) from $\{0, 1\} \times \mathcal{W}$ using the multinomial distribution with known probabilities $\{q_{d,w}\}_{(d,w) \in \{0,1\} \times \mathcal{W}}$, and then draw (Y, V) from the subsample with $(D, W) = (d, w)$. We repeat the procedure until the total sample size becomes n. Unless $q_{d,w} = p_{d,w}$ for all $(d, w) \in \{0, 1\} \times \mathcal{W}$, the observations $\{(Y_i, V_i, D_i, W_i)\}_{i=1}^n$ are not i.i.d. draws from P. The observations $\{(Y_i, V_i, D_i, W_i)\}_{i=1}^n$ are i.i.d., however, under probability Q with density $q_{d,w} f_{Y,V|D,W}(y, v|d, w)$, where $f_{Y,V|D,W}(y, v|d, w)$ is the conditional density of (Y, V) given $(D, W) = (d, w)$ with respect to a σ-finite measure, say, μ. Therefore, the nature of treatment-based sampling is that we have observations that are i.i.d. from Q but the parameter of interest is a functional of P. The notations of expectation and variance without subscripts are assumed to be under P. Expectation $\mathbf{E}_Q$ denotes expectation under Q. Expectation $\mathbf{E}_{d,w}$ denotes the conditional expectation given $(D, W) = (d, w)$. In pure treatment-based sampling, we suppress the notation w from subscripts, for example, writing p_d instead of $p_{d,w}$ and $\mathbf{E}_d$ instead of $\mathbf{E}_{d,w}$.

3 Semiparametric Efficiency under Treatment-Based Sampling

In this section, we explain how we can compute the semiparametric efficiency bound for the parameter, say, $\psi(P)$, under treatment based sampling. The standard theory of efficiency in semiparametric models and methods to compute efficiency bounds are well established in the literature.

(See Newey (1990) and Bickel, Klaassen, Ritov, and Wellner (1993) for a review.) Closely related to this paper, Bickel and Kwon (2001) showed how we can adapt the results based on i.i.d sampling to a multinomial sampling environment. (See Example 1 there.) To save space, we assume basic terminologies and concepts in Bickel, Klassen, Ritov, and Wellner (1993) and highlight how the standard method can be adapted to observations from treatment-based sampling.

Since we know the marginal probabilities $q_{d,w}$, we consider the following form of regular parametric submodels:

$$f_t(z,d,w) = f^t_{Z|D,W}(z|d,w)q_{d,w},\ t \in [0,\varepsilon),\ \varepsilon > 0, \tag{1}$$

where $\{f^t_{Z|D,W}(\cdot|d,w) : t \in [0,\varepsilon)\}$ denotes a regular parametric submodel passing through $f_{Z|D,W}(\cdot|d,w)$, the conditional density of Z given $(D,W) = (d,w)$. Then, the parametric submodel $\{f_t : t \in [0,\varepsilon)\}$ is associated with a score, $s(z,d,w) = s_{d,w}(z) \in L_2(Q)$, where $s_{d,w} = \frac{\partial}{\partial t} \log f^t_{Z|D,W}(\cdot|d,w)|_{t=0}$ denotes the score associated with $\{f^t_{Z|D,W}(\cdot|d,w) : t \in [0,\varepsilon)\}$. Let $\mathcal{T}$ denote the tangent space, i.e., the closed linear span of all such scores s for all regular parametric submodels in the form of (1).

There are two situations for the identification of $\psi(P)$ that this paper considers. The first situation is where we can identify $\psi(P)$ only using the conditional distribution of Z given (D,W). The second situation is where we have knowledge of the aggregate shares $p_{d,w}$ which is needed to identify $\psi(P)$. In both cases, the relevant tangent space is the same $\mathcal{T}$ and $\psi(P)$ is identified from the knowledge of Q and $\{q_{d,w}\}_{(d,w)\in\mathcal{D}\times\mathcal{W}}$. Hence, we can write

$$\psi(P) = \psi_Q(Q),$$

for some functional ψ_Q. The parameter of interest $\psi_Q(Q)$ is assumed to be differentiable in Q in the sense of van der Vaart (1991) and to have $\dot{\psi}_Q \in L_2(Q)$ such that for all regular parametric submodels of the form in (1),

$$\frac{\partial \psi_Q(Q_t)}{\partial t}|_{t=0} = \mathbf{E}_Q\left[\dot{\psi}_Q(Z,D,W)s(Z,D,W)\right].$$

When $\dot{\psi}_Q \in \mathcal{T}$, we call it an efficient influence function and denote it by $\dot{\psi}^e_Q$. Then, the semiparametric efficiency bound is given by the inverse of

$$V_{TS} \equiv Var_Q(\dot{\psi}^e_Q(Z,D,W)) = \sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} q_{d,w}\mathbf{E}_{d,w}\left[\dot{\psi}^e_Q(Z,D,W)^2\right]. \tag{2}$$

In this paper, we find $\dot{\psi}^e_Q(Z,D,W)$ in the following way. First, note that $\mathcal{T}$ can be also viewed as the tangent space at P with parametric submodels P_t having density $f^t_{Z|D,W}(z|d,w)p_{d,w}$. We find $\dot{\psi}_P \in L_2(P)$ such that for all regular parametric submodels with density $f^t_{Z|D,W}(z|d,w)p_{d,w}$,

$$\frac{\partial \psi(P_t)}{\partial t}|_{t=0} = \mathbf{E}\left[\dot{\psi}_P(Z,D,W)s(Z,D,W)\right], \tag{3}$$

for some $s \in \mathcal{T}$. Then, observe that

$$\mathbf{E}\left[\dot{\psi}_P(Z,D,W)s(Z,D,W)\right] = \mathbf{E}_Q\left[\dot{\psi}_Q(Z,D,W)s(Z,D,W)\right],$$

if we take $\dot{\psi}_Q(z,d,w) = \dot{\psi}_P(z,d,w)p_{d,w}/q_{d,w}$. Hence we find an influence function $\dot{\psi}_P^e$ under P such that $\dot{\psi}_Q^e(z,d,w) = \dot{\psi}_P^e(z,d,w)p_{d,w}/q_{d,w}$ falls into $\mathcal{T}$. Thus, $\dot{\psi}_Q^e(z,d,w)$ constructed in this way is an efficient influence function.

4 Semiparametric Efficiency Bounds for Treatment Effects Parameters

The main objects of interest are the weighted average treatment effect, τ_{wate}, and the average treatment effect on the treated, τ_{atet}, defined as follows:

$$\tau_{wate} = \frac{\mathbf{E}\left[g(X)\{Y_1 - Y_0\}\right]}{\mathbf{E}\left[g(X)\right]} \text{ and } \tau_{atet} = \mathbf{E}\left[Y_1 - Y_0 | D = 1\right], \tag{4}$$

where g denotes a weighting function. As pointed out by HIR, τ_{wate} is reduced to τ_{atet} when $g(X) = p_1(X)$, where $p_d(X) = P\{D = d|X\}$, $d \in \{0,1\}$, denotes the propensity score. This paper adopts the unconfoundedness condition:

$$(Y_0, Y_1) \perp\!\!\!\perp D|X, \tag{5}$$

meaning that (Y_0, Y_1) is conditionally independent of D given X. Condition (5) is imposed on the original data set, not on the data from treatment-based sampling.

Under treatment-based sampling, τ_{wate} and τ_{atet} are not identified without knowledge of the aggregate shares $p_{d,w}$, because the marginal distribution of X is not identified from the data. However, under pure treatment-based sampling, we can identify τ_{atet} without knowledge of p_d. In fact, under (5), the design of pure treatment-based sampling (i.e. the choice of q_d) does not play a role in determining the conditional distribution of (Y_1, Y_0) given X. These facts about identification are summarized in the following table:

Table: Identification of Treatment Effects Parameters (TS stands for treatment-based sampling)

	τ_{wate}	τ_{atet} (non-pure TS)	τ_{atet} (pure TS)
Known Aggregate Shares	Yes	Yes	Yes
Unknown Aggregate Shares	No	No	Yes

As Wooldridge (2001) has pointed out, the assumption of known aggregate shares $p_{d,w}$ is motivated by the sampling environment where $N_{d,w}$ is very large relative to the subsample size $n_{d,w}$. Such sampling is reasonable when it is much less costly to gather information about (D, W) than the outcome Y or full covariates X. In this case, a proper large sample theory would be one with

$n_{d,w}/N_{d,w} \to_P 0$. At the level of treatment-based samples, the asymptotic theory implies knowledge of $p_{d,w}$.

We introduce some notations:

$$\begin{aligned}\beta_d(X) &\equiv \mathbf{E}\left[Y_d|X\right],\ \sigma_d^2(X) \equiv \mathbf{E}\left[(Y_d - \beta_d(X))^2|X\right],\ \text{and} \\ \tau(X) &\equiv \mathbf{E}\left[Y_1|X\right] - \mathbf{E}\left[Y_0|X\right].\end{aligned}$$

Theorem 1 : *Suppose that (5) holds, and that $g(\cdot)$ and $p_{d,w}$, $(d,w) \in \{0,1\} \times \mathcal{W}$, are known. Then the semiparametric efficiency bound for τ_{wate} under treatment-based sampling is equal to $V_{TS}^{-1}(\tau_{wate})$, where*

$$V_{TS}(\tau_{wate}) \equiv \frac{1}{\{\mathbf{E}[g(X)]\}^2} \sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \frac{p_{d,w}^2}{q_{d,w}} \mathbf{E}_{d,w}\left[g(X)^2 \frac{\sigma_d^2(X)}{p_d^2(X)} + \zeta_d^2(X)\right],$$

and $\zeta_d(x) \equiv g(x)\{\tau(x) - \tau_{wate}\} - \mathbf{E}_{d,w}[g(X)\{\tau(X) - \tau_{wate}\}]$ with $x \equiv (v,w)$. In particular, when the sampling is pure treatment-based sampling and $p_d = q_d$, $V_{TS}(\tau_{wate}) = V_{RS}(\tau_{wate})$, where

$$V_{RS}(\tau_{wate}) \equiv \frac{1}{\{\mathbf{E}[g(X)]\}^2} \mathbf{E}\left[g(X)^2 \left\{\frac{\sigma_1^2(X)}{p_1(X)} + \frac{\sigma_0^2(X)}{p_0(X)}\right\} + \sum_{d\in\{0,1\}} \zeta_d^2(X) p_d(X)\right].$$

Theorem 1 implies that knowledge of $p_{d,w}$ is not ancillary in general. In the special case of pure treatment-based sampling with $p_d = q_d$, we can compare $V_{TS}(\tau_{wate})$ with the variance bound of HIR:

$$V_{HIR}(\tau_{wate}) \equiv \frac{1}{\{\mathbf{E}[g(X)]\}^2} \mathbf{E}\left[g(X)^2 \left\{\frac{\sigma_1^2(X)}{p_1(X)} + \frac{\sigma_0^2(X)}{p_0(X)}\right\} + g^2(X)(\tau(X) - \tau_{wate})^2\right].$$

Note that $V_{TS}(\tau_{wate}) \leq V_{HIR}(\tau_{wate})$ and the equality holds if and only if

$$\mathbf{E}_d[g(X)\{\tau(X) - \tau_{wate}\}] = 0 \text{ for all } d \in \{0,1\}. \tag{6}$$

Hence knowledge of p_d is not ancillary for τ_{wate}.

Let us turn to τ_{atet}. Although τ_{atet} is reduced to τ_{wate} when $g(X) = p_1(X)$, we treat it separately because when $g(X) = p_1(X)$, the weighting function g is not known.

Theorem 2 : (i) *Suppose that (5) holds and $\{p_{d,w}\}_{(d,w)\in\{0,1\}\times\mathcal{W}}$ are known. Then the semiparametric efficiency bound for τ_{atet} under treatment-based sampling is equal to $V_{TS}^{-1}(\tau_{atet})$, where*

$$V_{TS}(\tau_{atet}) \equiv \sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \frac{p_{d,w}^2}{q_{d,w}} \mathbf{E}_{d,w}\left[\frac{d}{p_1^2}\left\{\sigma_1^2(X) + \tilde{\zeta}_1^2(X)\right\} + \frac{1-d}{p_1^2}\frac{\sigma_0^2(X)p_1^2(X)}{p_0^2(X)}\right]$$

and $\tilde{\zeta}_d(x) \equiv \tau(x) - \tau_{atet} - \mathbf{E}_{d,w}\left[\tau(X) - \tau_{atet}\right]$ with $x \equiv (v,w)$.

(ii) *Suppose that* (5) *holds and the sampling is pure treatment-based sampling. Then, regardless of whether we know* $\{p_d\}_{d\in\{0,1\}}$ *or not, the semiparametric efficiency bound for* τ_{atet} *is given by* $V_{PTS}^{-1}(\tau_{atet})$, *where*

$$V_{PTS}(\tau_{atet}) = \frac{1}{q_1}\mathbf{E}\left[\sigma_1^2(X) + \{\tau(X) - \tau_{atet}\}^2|D=1\right] + \frac{1}{q_0}\mathbf{E}\left[\frac{f(X|1)^2\sigma_0^2(X)}{f(X|0)^2}|D=0\right]. \tag{7}$$

Under random sampling (i.e., $p_{d,w} = q_{d,w}$), $V_{TS}(\tau_{atet})$ is smaller than the variance bound in Hahn (1998) that does not assume knowledge of $p_{d,w}$. Therefore, the aggregate shares are not ancillary in general. However, the situation becomes different when the sampling is pure treatment-based sampling. In this case, the aggregate shares p_d are ancillary. Indeed, in pure treatment-based sampling with $p_d = q_d$, $V_{PTS}(\tau_{atet})$ is reduced to

$$V_{RS}(\tau_{atet}) \equiv \mathbf{E}\left[\left\{\frac{p_1(X)\sigma_1^2(X)}{p_1^2} + \frac{\sigma_0^2(X)p_1^2(X)}{p_0(X)p_1^2}\right\} + \frac{\{\tau(X) - \tau_{atet}\}^2 p_1(X)}{p_1^2}\right]$$

which is identical to the variance bound of Hahn (1998) for τ_{atet}. Therefore, $V_{PTS}(\tau_{atet})$ can be viewed as a generalization of the variance bound of Hahn (1998) to pure treatment-based sampling.

5 Efficient Estimation of Weighted Average Treatment Effects

5.1 Propensity Score Estimation

We begin with propensity score estimation. Let $f_Q(x)$ be the density of X (under Q) with respect to some σ-finite measure, and $f(v|d,w)$ the conditional density function of V given $(D,W) = (d,w)$. By Bayes' rule, the propensity score is identified as

$$p_d(v,w) = \frac{f(v|d,w)p_{d,w}}{\sum_{d\in\{0,1\}} f(v|d,w)p_{d,w}}, \tag{8}$$

where $p_d(v,w) = P\{D = d|V = v, W = w\}$. The identification of $p_d(v,w)$ certainly requires knowledge of $p_{d,w}$.

We consider two consistent estimators of the propensity score that are based on the identification in (8). Let $X = (V_1, V_2, W) \in \mathbf{R}^L$, where $V_1 \in \mathcal{X}_1$ is continuous and $V_2 \in \mathcal{X}_2$ is discrete with supports $\mathcal{X}_1 \subset \mathbf{R}^{L_1}$ and $\mathcal{X}_2 \subset \mathbf{R}^{L_2}$ respectively for V_1 and V_2. Define $\mathcal{X}$ to be the support of X_i. Let $S_{d,w} = \{1 \le i \le n : (D_i, W_i) = (d,w)\}$. Define $\hat{f}(v_1, v_2|d,w) = \frac{1}{q_{d,w}n}\sum_{i\in S_{d,w}} K_h(V_{1i} - v_1)1\{V_{2i} = v_2\}$, where $K_h(s_1,\cdot\cdot\cdot,s_{L_1}) = K(s_1/h,\cdot\cdot\cdot,s_{L_1}/h)/h^{L_1}$ and $K(\cdot)$ is a multivariate kernel function. Then, we define

$$\hat{p}_d(v,w) = \frac{\hat{f}(v_1,v_2|d,w)p_{d,w}}{\sum_{d\in\{0,1\}}\hat{f}(v_1,v_2|d,w)p_{d,w}}. \tag{9}$$

Letting $L_{d,w,i} \equiv \frac{p_{d,w}}{q_{d,w}} 1\{(D_i, W_i) = (d,w)\}$, and $L_{w,i} \equiv L_{0,w,i} + L_{1,w,i}$, we can rewrite $\hat{p}_d(v,w)$ as

$$\hat{p}_d(v,w) = \frac{\hat{\lambda}_d(v,w)}{\hat{\lambda}_1(v,w) + \hat{\lambda}_0(v,w)},$$

where $\hat{\lambda}_d(v,w) = \frac{1}{n}\sum_{i=1}^n L_{d,w,i} K_h(V_{1i} - v_1) 1\{V_{2i} = v_2\}$. Therefore, the propensity score estimator is a weighted Nadaraya-Watson estimator. This is intuitive because the probability under treatment-based sampling is the average of conditional probabilities using different weights.

Alternatively, we can estimate the propensity score using the estimated fraction

$$\hat{q}_{d,w} = \frac{1}{n}\sum_{i=1}^n 1\{(D_i, W_i) = (d,w)\} = n_{d,w}/n$$

in place of $q_{d,w}$. Using this, we define $\hat{L}_{d,w,i} \equiv \frac{p_{d,w}}{\hat{q}_{d,w}} 1\{(D_i, W_i) = (d,w)\}$, $\hat{L}_{w,i} \equiv \hat{L}_{0,w,i} + \hat{L}_{1,w,i}$, and

$$\tilde{p}_d(v,w) \equiv \frac{\tilde{\lambda}_d(v,w)}{\tilde{\lambda}_1(v,w) + \tilde{\lambda}_0(v,w)}, \tag{10}$$

where $\tilde{\lambda}_d(v,w) = \frac{1}{n}\sum_{i=1}^n \hat{L}_{d,w,i} K_h(V_{1i} - v_1) 1\{V_{2i} = v_2\}$.

5.2 Efficient Estimation of Average Treatment Effects

Let us first search for an efficient estimator of τ_{wate}. The first idea will be adapting the estimator of HIR to treatment-based sampling:

$$\hat{\tau}_{wate} \equiv \frac{\sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}} \frac{1}{n} \sum_{i\in S_{1,w}} \hat{1}_{n,i} g(V_i,w) Y_i/\hat{p}_1(V_i,w) - \frac{p_{0,w}}{q_{0,w}} \frac{1}{n} \sum_{i\in S_{0,w}} \hat{1}_{n,i} g(V_i,w) Y_i/\hat{p}_0(V_i,w) \right\}}{\sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}} \frac{1}{n} \sum_{i\in S_{1,w}} g(V_i,w) + \frac{p_{0,w}}{q_{0,w}} \frac{1}{n} \sum_{i\in S_{0,w}} g(V_i,w) \right\}},$$

where $\hat{p}_d(v,w)$ is estimated by (9) and $\hat{1}_{n,i} = 1\left\{\hat{\lambda}_d(V_i,w) \geq \delta_n : d \in \{0,1\}\right\}$ for a positive sequence $\delta_n \to 0$. When we are under pure treatment-based sampling and $p_d = q_d$, $\hat{\tau}_{wate}$ is reduced to the estimator of HIR except with a different nonparametric estimator for the propensity score. In Theorem 2 below, we show that this estimator is consistent and asymptotically normal, but inefficient in general.

Alternatively, we suggest the following estimator:

$$\tilde{\tau}_{wate} \equiv \frac{\sum_{w\in\mathcal{W}} \frac{p_{1,w}}{n_{1,w}} \sum_{i\in S_{1,w}} \tilde{1}_{n,i} g(V_i,w) Y_i/\tilde{p}_1(V_i,w)}{\sum_{w\in\mathcal{W}} \frac{p_{1,w}}{n_{1,w}} \sum_{i\in S_{1,w}} \tilde{1}_{n,i} g(V_i,w)/\tilde{p}_1(V_i,w)} - \frac{\sum_{w\in\mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i\in S_{0,w}} \tilde{1}_{n,i} g(V_i,w) Y_i/\tilde{p}_0(V_i,w)}{\sum_{w\in\mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i\in S_{0,w}} \tilde{1}_{n,i} g(V_i,w)/\tilde{p}_0(V_i,w)},$$

where $\tilde{p}_d(v,w)$ is as in (10) and $\tilde{1}_{n,i} = 1\left\{\tilde{\lambda}_d(V_i,w) \geq \delta_n : d \in \{0,1\}\right\}$. The estimator $\tilde{\tau}_{wate}$ involves a further weighting of $g(V_i,w)$ by $\tilde{p}_d(V_i,w)$. Note that $\hat{\tau}_{wate}$ uses $q_{d,w}$ while $\tilde{\tau}_{wate}$ uses $\hat{q}_{d,w} = n_{d,w}/n$.

Assumption 1 : There exist $\alpha_1 \in \mathbf{R}$ and $\alpha_2 \in \mathbf{R}$ such that $0 < \alpha_1 \leq p_1(x) \leq \alpha_2 < 1$ for all $x \in \mathcal{X}$

such that $g(x) \neq 0$.

Assumption 2 : For each $(d, w, v_2) \in \{0,1\} \times \mathcal{W} \times \mathcal{X}_2$, the following holds.
(i) $f(v_1, v_2|d, w)||v_1||$, $\beta_d(\cdot, v_2, w)$, and $g(\cdot, v_2, w)$ are bounded and $L_1 + 1$ times continuously differentiable with bounded derivatives on $\mathbf{R}^{L_1}$ and uniformly continuous (L_1+1)-th derivatives.
(ii) $\mathbf{E}_{d,w} Y_1^r < \infty$, $\mathbf{E}_{d,w} Y_0^r < \infty$, $\mathbf{E}_{d,w}||V_{1i}||^r < \infty$, for some $r \geq 4$.
(iii) $p_{d,w},\ q_{d,w} \in (0,1)$ and $\Sigma_{(d,w)\in\{0,1\}\times\mathcal{W}} p_{d,w} = \Sigma_{(d,w)\in\{0,1\}\times\mathcal{W}} q_{d,w} = 1$.
(iv) For some $\bar{a} \geq 4$, $\mathbf{E}_{d,w}\left[f^{-\bar{a}}(X_i)\right] < \infty$ for all $a \in [0, \bar{a}]$.

Assumption 3 : (i) K is zero outside an interior of a bounded set, $L_1 + 1$ times continuously differentiable with bounded derivatives, $\int K(s)ds = 1$, and $\int s_1^{l_1} \cdots s_{L_1}^{l_{L_1}} K(s)ds = 0$ for all nonnegative integers $l_1, \cdots, l_{L_1}$ such that $l_1 + \cdots + l_{L_1} \leq L_1$ and $\int |s_1^{l_1} \cdots s_{L_1}^{l_{L_1}} K(s)|ds < \infty$ for all nonnegative integers $l_1, \cdots, l_{L_1}$ such that $l_1 + \cdots + l_{L_1} = L_1 + 1$.
(ii) $\sqrt{n}\{\varepsilon_n^2 \delta_n^{-1} + \delta_n^{\bar{a}}\} \to 0$, and $\delta_n^{-1}\varepsilon_n \to 0$, where $\varepsilon_n = n^{-1/2} h^{-L_1/2}\sqrt{\log n} + h^{L_1+1}$.

Assumption 1 is the condition of sample overlap needed for the identification of τ_{wate}. This is violated when part of X is only observed among the treated or untreated subsamples. (See Heckman, Ichimura, and Todd (1997) for a discussion in this regard.) See Khan and Tamer (2009) for situations where Assumption 1 is violated with $p_1(x)$ being arbitrarily close to 0 or 1. Assumption 2 requires that $f(\cdot, v_2|d, w)$ is continuous on $\mathbf{R}^{L_1}$. While HIR requires that the density of V_1 is bounded away from zero, our Assumption 2 excludes such a case. Assumption 2 (iv) is the tail condition for the density of V_{1i}. (See, e.g. Assumption NP7 of Andrews (1995).) Assumption 3(i) is a standard assumption for higher order kernels. The following theorem establishes the asymptotic distribution of $\hat{\tau}_{wate}$ and $\tilde{\tau}_{wate}$.

Theorem 3 : *Suppose that the condition* (5) *and* Assumptions 1-3 *hold. Then*

$$\begin{aligned} \sqrt{n}(\hat{\tau}_{wate} - \tau_{wate}) &\to_d N(0, V_1), \text{ and} \\ \sqrt{n}(\tilde{\tau}_{wate} - \tau_{wate}) &\to_d N(0, V_{TS}(\tau_{wate})), \end{aligned}$$

where

$$V_1 \equiv \frac{1}{\{\mathbf{E}[g(X)]\}^2} \sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \frac{p_{d,w}^2}{q_{d,w}} \mathbf{E}_{d,w}\left[g(X)^2 \left\{\frac{\sigma_d^2(X)}{p_d^2(X)} + (\tau(X) - \tau_{wate})^2\right\}\right].$$

When the sampling is random sampling, the asymptotic variance of $\hat{\tau}_{wate}$ is reduced to V_{HIR} which is greater than $V_{TS}(\tau_{wate})$ in general. Therefore, $\hat{\tau}_{wate}$ is inefficient. The efficiency is achieved by an alternative estimator $\tilde{\tau}_{wate}$. The efficient estimator can be used when only $\hat{q}_{d,w} = n_{d,w}/n$ (not $q_{d,w}$) is available in the data.

Let us turn to the efficient estimation of τ_{atet}. In this case, the identification of τ_{atet} allows us to formulate Assumption 1 differently:

Assumption 1P : There exist $\alpha_1 \in \mathbf{R}$ and $\alpha_2 \in \mathbf{R}$ such that $0 < \alpha_1 \leq p_1(x) \leq \alpha_2 < 1$ for all $x \in \mathcal{X}$.

We suggest the following estimator:

$$\tilde{\tau}_{atet} = \frac{1}{p_1} \sum_{w \in \mathcal{W}} \frac{p_{1,w}}{n_{1,w}} \sum_{i \in S_{1,w}} Y_i - \frac{\sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \tilde{p}_1(V_i, w) Y_i / \tilde{p}_0(V_i, w)}{\sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \tilde{p}_1(V_i, w) / \tilde{p}_0(V_i, w)},$$

where $\tilde{p}_d(v, w)$ is estimated by (10). Theorem 4 below establishes that this estimator is efficient.

We saw that in the case of pure treatment-based sampling, the knowledge of p_d is ancillary. One might consider alternatively the estimator of HIR that is adapted to pure treatment-based sampling:

$$\hat{\tau}_{atet,p} = \frac{\frac{p_1}{q_1 n} \sum_{i \in S_1} Y_i - \frac{p_0}{q_0 n} \sum_{i \in S_0} \hat{1}_{n,i} \hat{p}_1(X_i) Y_i / \hat{p}_0(X_i)}{\frac{p_0}{q_0 n} \sum_{i \in S_0} \hat{1}_{n,i} \hat{p}_1(X_i) + \frac{p_1}{q_1 n} \sum_{i \in S_1} \hat{1}_{n,i} \hat{p}_1(X_i)}.$$

While this estimator is efficient (see Theorem 4 below), it requires knowledge of p_d. Instead, we suggest the following estimator that does not require knowledge of the aggregate shares p_d:

$$\begin{aligned}
\tilde{\tau}_{atet,p} &= \frac{1}{n_1} \sum_{i \in S_1} Y_i - \frac{\sum_{i \in S_0} \tilde{1}_{n,i} \tilde{p}_1(X_i) Y_i / \tilde{p}_0(X_i)}{\sum_{i \in S_0} \tilde{1}_{n,i} \tilde{p}_1(X_i) / \tilde{p}_0(X_i)} \\
&= \frac{1}{n_1} \sum_{i \in S_1} Y_i - \frac{\sum_{i \in S_0} Y_i \tilde{1}_{n,i} \left(\{\frac{1}{n_1} \sum_{j \in S_1} K_{ji}\} / \{\frac{1}{n_0} \sum_{j \in S_0} K_{ji}\} \right)}{\sum_{i \in S_0} \tilde{1}_{n,i} \left(\{\frac{1}{n_1} \sum_{j \in S_1} K_{ji}\} / \{\frac{1}{n_0} \sum_{j \in S_0} K_{ji}\} \right)},
\end{aligned}$$

where $K_{ji} = K_h(V_{1j} - V_{1i}) 1\{V_{2j} = V_{2i}\}$. The estimator $\tilde{\tau}_{atet,p}$ is in fact an estimator $\tilde{\tau}_{atet}$ that is specialized to pure treatment-based sampling. Hence the estimator is efficient.

Theorem 4 : *Suppose that the condition* (5) *and* Assumptions 1P, 2-3 *hold. Then,*

$$\sqrt{n}(\tilde{\tau}_{atet} - \tau_{atet}) \to_d \quad N(0, V_{TS}(\tau_{atet})).$$

Suppose further that we are under pure treatment-based sampling. Then

$$\begin{aligned}
\sqrt{n}(\tilde{\tau}_{atet,p} - \tau_{atet}) &\to_d N(0, V_{PTS}(\tau_{atet})) \text{ and} \\
\sqrt{n}(\hat{\tau}_{atet,p} - \tau_{atet}) &\to_d N(0, V_{PTS}(\tau_{atet})).
\end{aligned}$$

6 Optimal Design of Treatment-Based Sampling

In this section, we develop an optimal design of treatment-based sampling. Let $\dot{\psi}_P^e(y, v, d, w)$ be the efficient influence function of a generic parameter such as τ_{wate} or τ_{atet}. Then we can design an optimal treatment-based sampling as follows. Let

$$J_{d,w} = p_{d,w}^2 \mathbf{E}_{d,w} \left[\dot{\psi}_P^e(Y, V, D, W)^2 \right].$$

We can write the variance bound (under treatment-based sampling) as

$$V_{TS} = \sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \frac{J_{d,w}}{q_{d,w}}.$$

We can view $J_{d,w}/q_{d,w}$ as the contribution of the (d,w)-subsample to the variance bound.

We define the optimal design to be those $\{q_{d,w}\}_{(d,w)\in\{0,1\}\times\mathcal{W}}$ such that minimize V_{TS} under the constraint that $q_{d,w} \geq 0$ and $\sum_{(d,w)\in\{0,1\}\times\mathcal{W}} q_{d,w} = 1$. It is easy to see that the optimal design is given by

$$q^*_{d,w} = \frac{\sqrt{J_{d,w}}}{\sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \sqrt{J_{d,w}}}. \tag{11}$$

The optimal design suggests that we sample from the (d,w)-subsample precisely according to the "noise" proportion $\sqrt{J_{d,w}}$ of the subsample (d,w) in $\sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \sqrt{J_{d,w}}$. In other words, we sample more from a subsample that induces more sampling variability to the efficient estimator. When we have some pilot sample obtained from a two-stage sampling scheme or other data sources that can be used to draw information about $J_{d,w}$, the result here may serve as a guide for optimally choosing the size of the sampling fractions $q_{d,w}$.[2]

Using $q^*_{d,w}$ yields *the minimum semiparametric efficiency bound* as

$$\left\{ \sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \sqrt{J_{d,w}} \right\}^2. \tag{12}$$

The variance in (12) is the minimum variance bound over all the choices of the sampling probabilities $q_{d,w}$. The variance (12) can be used to compare different choices of additional stratum variables W_i.

In the case of pure treatment-based sampling, we can make precise the condition for treatment-based sampling to yield improved inference than random sampling. Let V_{RS} be the variance bound under random sampling, which is equal to V_{TS} with $p_d = q_d$. Then it is not hard to see that $V_{RS} \geq V_{TS}$ if and only if

$$\min\left\{p_1, \frac{J_1}{J_1 + J_0}\right\} \leq q_1 \leq \max\left\{p_1, \frac{J_1}{J_1 + J_0}\right\}. \tag{13}$$

Therefore, it is not always true that sampling more from a subsample of low population proportion leads to a better result. The improvement hinges on the noise proportion $J_1/(J_1 + J_0)$ as well. When p_1 happens to coincide with $J_1/(J_1 + J_0)$, there is no way for treatment-based sampling to improve upon random sampling. Theorems 1 and 2 allow us to identify $J_{d,w}$ in (11) for τ_{wate} and τ_{atet}.

[2] When it is less costly to sample from a specific subsample from others, we can incorporate an appropriate differential cost consideration into the optimal design by turning the optimization problem into one subject to certain inequality constraints.

Corollary 1 : *Under the conditions of* Theorems 1 *and* 2 *respectively for* τ_{wate} *and* τ_{atet}, *the optimal choice of* $q_{d,w}$ *is given as follows:*

$$q_{d,w} = \frac{\sqrt{J^*_{d,w}}}{\sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \sqrt{J^*_{d,w}}} \text{ (for } \tau_{wate}\text{) and } q_{d,w} = \frac{\sqrt{\tilde{J}_{d,w}}}{\sum_{(d,w)\in\{0,1\}\times\mathcal{W}} \sqrt{\tilde{J}_{d,w}}} \text{ (for } \tau_{atet}\text{)},$$

where

$$\begin{aligned} J^*_{d,w} &= \frac{p^2_{d,w}}{\{\mathbf{E}[g(X)]\}^2}\mathbf{E}_{d,w}\left[g(X)^2\frac{\sigma^2_d(X)}{p^2_d(X)} + \zeta^2_d(X)\right] \text{ and} \\ \tilde{J}_{d,w} &= p^2_{d,w}\mathbf{E}_{d,w}\left[\frac{d}{p^2_1}\left\{\sigma^2_1(X) + \tilde{\zeta}^2_1(X)\right\} + \frac{1-d}{p^2_1}\frac{\sigma^2_0(X)p^2_1(X)}{p^2_0(X)}\right]. \end{aligned}$$

In the case of pure treatment-based sampling, the estimation of the optimal design does not require knowledge of p_d. Indeed, we define

$$\bar{J}_1 = \mathbf{E}\left[\sigma^2_1(X) + \{\tau(X) - \tau_{atet}\}^2 | D = 1\right] \text{ and } \bar{J}_0 = \mathbf{E}\left[\frac{f(X|1)^2\sigma^2_0(X)}{f(X|0)^2}|D = 0\right].$$

Then, $V_{PTS}(\tau_{atet}) = \bar{J}_1/q_1 + \bar{J}_0/q_0$. The optimal design of q_1 in Corollary 1 is given by

$$q_1(\tau_{atet}) = \frac{\sqrt{\bar{J}_1}}{\sqrt{\bar{J}_1} + \sqrt{\bar{J}_0}}$$

and a necessary and sufficient condition for $V_{PTS}(\tau_{atet}) \leq V_{PRS}(\tau_{atet})$ is given by the condition in (13) with J_1 and J_0 replaced by $\bar{J}_1$ and $\bar{J}_0$. Note that estimation of $\bar{J}_d$ does not require knowledge of the aggregate shares p_d.

7 Conclusion

This paper has established semiparametric efficiency bounds for certain average treatment effect parameters under treatment-based sampling. This paper also proposes efficient estimators for the parameters. This paper's finding suggests that under treatment-based sampling, tailoring the estimators of HIR to treatment-based sampling does not work when the aggregate shares are not ancillary. An optimal design of treatment-based sampling is also derived. The theory of optimal design illuminates the role of treatment-based sampling in improving the quality of inference.

8 Appendix: Mathematical Proofs

Proof of Theorem 1 : Let $f(y, v, d, w)$ be the density of (Y, V, D, W) with respect to a σ-finite measure μ under P. We use the notations $\int \cdot d\mu(w)$, $\int \cdot d\mu(v)$, $\int \cdot d\mu(y)$, etc., to denote the integrations with respect to the marginals of μ for the coordinates of w, v, y, etc. Let $\mathcal{Q} =$

$\{f_{Y,V|D,W}(\cdot|\cdot)q_\cdot : f_{Y,V|D,W} \in \mathcal{P}_{d,w}, (d,w) \in \{0,1\} \times \mathcal{W}\}$ and fix $Q \in \mathcal{Q}$. Let $f(y,v|d,w)$ be the conditional density of (Y,V) given $(D,W) = (d,w)$. We use subscripts P and Q for densities to make it explicit under which probability they are defined when they differ. We do not use the subscripts for the conditional densities given $(D,W) = (d,w)$ or given $(D,W,V) = (d,w,v)$ because they remain the same both under P and under Q.

We write the density $f_Q(y,v,d,w)$ of (Y,V,D,W) under Q as

$$\begin{aligned} f_Q(y,v,d,w) &= f(y|v,d,w)f(v|d,w)q_{d,w} \\ &= f_d(y|x)f(v|d,w)q_{d,w}, \end{aligned}$$

where $f_{d,P}(y|x)$ is the conditional density of Y_{di} given $X_i = x$ under P. The second equality follows by the unconfoundedness condition. Hence the score $s(y,v,d,w)$ is written as $s_d(y|x) + s(v|d,w)$, where $\int s_d(y|x)f_{d,P}(y|x)d\mu(y) = 0$ and $\int s(v|d,w)f(v|d,w)d\mu(v,w) = 0$. The closed linear span of such scores constitutes the tangent space $\mathcal{T}$.

Take a regular parametric submodel $f_Q^t(y,v,d,w) = f^t(y,v|d,w)q_{d,w}$ and let P_t be the parametric submodel with density $f^t(y,v|d,w)p_{d,w}$. We need to find $\dot{\psi}_P$. The weighted average treatment effect under P_t is written as

$$\begin{aligned} \tau_{wate}(t) &= \frac{\sum_{w\in\mathcal{W}} \int\int g(v,w)y\left\{f_t(y|v,1,w) - f_t(y|v,0,w)\right\}d\mu(y)f_t(v,w)d\mu(v)}{\sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} p_{d,w}\int g(v,w)f_t(v|d,w)d\mu(v)} \\ &= \frac{\sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} \int g(v,w)\left\{\int yf_{1,t}(y|v,w)d\mu(y) - \int yf_{0,t}(y|v,w)d\mu(y)\right\}p_{d,w}f_t(v|d,w)d\mu(v)}{\sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} p_{d,w}\int g(v,w)f_t(v|d,w)d\mu(v)}. \end{aligned}$$

The first order derivative of $\tau_{wate}(t)$ with respect to t at $t = 0$ is equal to

$$\begin{aligned} &\frac{1}{\mathbf{E}[g(X)]}\mathbf{E}\left[g(X)\left(\mathbf{E}\left[Ys_1(Y|X)|X\right] - \mathbf{E}\left[Ys_0(Y|X)|X\right]\right)\right] \\ &-\frac{1}{\mathbf{E}[g(X)]}\mathbf{E}[s(V|D,W)g(X)\{\tau(X) - \tau_{wate}\}]. \end{aligned}$$

Let

$$\begin{aligned} \dot{\psi}_P(y,v,d,w) &= \frac{1}{\mathbf{E}[g(X)]}g(v,w)\left(\frac{d(y-\beta_1(v,w))}{p_1(v,w)} - \frac{(1-d)(y-\beta_0(v,w))}{p_0(v,w)}\right) \\ &\quad -\frac{1}{\mathbf{E}[g(X)]}\zeta_d(v,w). \end{aligned} \tag{14}$$

We can write

$$
\begin{aligned}
\frac{\partial \tau_{wate}(t)}{\partial t} &= \mathbf{E}[\dot{\psi}_P(Y,V,D,W)s(Y,V,D,W)] \qquad (15)\\
&= \sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} \mathbf{E}\left[\dot{\psi}_P(Y,V,D,W)s(Y,V,D,W)|(D,W)=(d,w)\right]p_{d,w}\\
&= \sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} \mathbf{E}\left[\dot{\psi}_Q(Y,V,D,W)s(Y,V,D,W)|(D,W)=(d,w)\right]q_{d,w}\\
&= \mathbf{E}_Q\left[\dot{\psi}_Q(Y,V,D,W)s(Y,V,D,W)\right],
\end{aligned}
$$

where $\dot{\psi}_Q(y,v,d,w) = \dot{\psi}_P(y,v,d,w)p_{d,w}/q_{d,w}$. Now, observe that $\dot{\psi}_Q$ belongs to the tangent space $\mathcal{T}$. (This follows from the unconfounded condition.) Therefore, it is an efficient influence function. Since it is the projection of an influence function on $\mathcal{T}$ which is a closed linear space of scores, the efficient influence function is unique. (e.g. van der Vaart (1998), p.363.) Hence the variance bound is given by its $L_2(Q)$-norm:

$$
\begin{aligned}
&\sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} \mathbf{E}[\dot{\psi}_Q^2(Y,V,D,W)|(D,W)=(d,w)]q_{d,w}\\
&= \sum_{(d,w)\in\mathcal{D}\times\mathcal{W}} \frac{p_{d,w}^2}{q_{d,w}}\mathbf{E}[\dot{\psi}_P^2(Y,V,D,W)|(D,W)=(d,w)].
\end{aligned}
$$

■

Proof of Theorem 2 : The tangent space in the proof of Theorem 1 remains the same. The only needed change from Theorem 1 is the computation of the influence function because now $g(x) = p_1(x)$ is not assumed to be known. Let P_t be the submodel as in the proof of Theorem 1. The weighted average treatment effect under P_t is written as

$$
\tau_{atet}(t) = \sum_{w\in\mathcal{W}} \int\int y\left\{f_t(y|v,1,w) - f_t(y|v,0,w)\right\}d\mu(y)f_t(v|1,w)p_{w|1}d\mu(v),
$$

where $p_{w|1} = p_{1,w}/\{\Sigma_{w\in\mathcal{W}}p_{1,w}\}$. The first order derivative of $\tau_{atet}(t)$ with respect to t is equal to

$$
\begin{aligned}
&\mathbf{E}\left[s(V|D,W)\{\tau(X)-\tau_{atet}\}|D=1\right]\\
&+\mathbf{E}\left[\mathbf{E}\left[Ys_1(Y|X)|X,D=1\right] - \mathbf{E}\left[Ys_0(Y|X)|X,D=0\right]|D=1\right].
\end{aligned}
$$

Therefore, we take

$$
\dot{\psi}_P(y,v,d,w) = \frac{1}{p_1}\left\{d(y-\beta_1(v,w)-\tilde{\zeta}_1(v,w)) - \frac{p_1(v,w)(1-d)(y-\beta_0(v,w))}{p_0(v,w)}\right\}.
$$

As shown in the proof of Theorem 1, this yields the semiparametric efficiency bound for τ_{atet}.

Let us turn to the situation with pure treatment-based sampling. The tangent space is the closed linear span of scores of the form $s_d(y|x) + s(v|d)$, where $\int s_d(y|x)f_{d,P}(y|x)d\mu(y) = 0$ and $\int s(v|d)f(v|d)d\mu(x) = 0$. Write

$$\tau_{atet}(t) = \int \int y \left\{ f_t(y|x,1) - f_t(y|x,0) \right\} d\mu(y) f_t(x|1) d\mu(x).$$

The first order derivative of $\tau_{atet}(t)$ with respect to t is equal to

$$\begin{aligned}
&\mathbf{E}\left[s(X|D)\{\tau(X) - \tau_{atet}\}|D=1\right] \\
&+\mathbf{E}\left[\{\mathbf{E}\left[Ys_1(Y|X)|X, D=1\right] - \mathbf{E}\left[Ys_0(Y|X)|X, D=0\right]\}|D=1\right].
\end{aligned}$$

Therefore, we take

$$\dot{\psi}_P(y,x,d) = \left\{ \frac{d(y - \beta_1(x) - \{\tau(x) - \tau_{atet}\})}{p_1} - \frac{p_1(x)(1-d)(y - \beta_0(x))}{p_0(x)p_1} \right\}$$

because $\mathbf{E}\left[\tau(X) - \tau_{atet}|D=1\right] = 0$. Let $\dot{\psi}_Q(y,x,d) = \dot{\psi}_P(y,x,d)p_d/q_d$. Now

$$\begin{aligned}
\sum_{d \in \mathcal{D}} q_d \mathbf{E}\left[\dot{\psi}_Q^2(Y,X,D)|D=d\right] &= \frac{p_1^2}{q_1}\mathbf{E}\left[\frac{(Y_1 - \beta_1(X) - \{\tau(X) - \tau_{atet}\})^2}{p_1^2}|D=1\right] \\
&\quad + \frac{p_0^2}{q_0}\mathbf{E}\left[\frac{p_1(X)^2(Y_0 - \beta_0(X))^2}{p_0(X)^2 p_1^2}|D=0\right] \\
&= \frac{1}{q_1}\mathbf{E}\left[(Y_1 - \beta_1(X) - \{\tau(X) - \tau_{atet}\})^2|D=1\right] \\
&\quad + \frac{1}{q_0}\mathbf{E}\left[\frac{p_0^2 p_1(X)^2}{p_0(X)^2 p_1^2}(Y_0 - \beta_0(X))^2|D=0\right].
\end{aligned}$$

Note that by Bayes' rule,

$$\frac{p_0 p_1(X)}{p_1 p_0(X)} = \frac{p_0 f(X|1) p_1}{p_0 f(X|0) p_1} = \frac{f(X|1)}{f(X|0)}.$$

By plugging in this, we obtain the wanted result. ■

Lemma A1: *Suppose that* Assumptions 1-3 *hold. Then, for each* $w \in \mathcal{W}$,

$$\begin{aligned}
\max_{1 \le i \le n} \hat{1}_{n,i} |p_1(V_i,w) - \hat{p}_1(V_i,w)| &= O_P(\varepsilon_n) \text{ and} \\
\max_{1 \le i \le n} \tilde{1}_{n,i} |p_1(V_i,w) - \tilde{p}_1(V_i,w)| &= O_P(\varepsilon_n).
\end{aligned}$$

Proof: We only consider the first statement. For simplicity, we assume that $V = V_1$ and define

$\mathbf{E}_{Q,w,i}[L_{1,w,i}] = \mathbf{E}_Q[L_{1,w,i}|V_i, W_i = w]f_Q(V_i, w)$ for simplicity. Let

$$\begin{aligned}\hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}] &= \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{h,ji} \text{ and} \\ \hat{\mathbf{E}}_{Q,w,i}[L_{w,i}] &= \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} L_{w,j}K_{h,ji}.\end{aligned}$$

Hence we can write

$$\begin{aligned}p_1(V_i, w) - \hat{p}_1(V_i, w) &= \frac{\mathbf{E}_{Q,w,i}[L_{1,w,i}]}{\mathbf{E}_{Q,w,i}[L_{w,i}]} - \frac{\hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]}{\hat{\mathbf{E}}_{Q,w,i}[L_{w,i}]} \\ &= \frac{\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]}{\mathbf{E}_{Q,w,i}[L_{w,i}]\hat{\mathbf{E}}_{Q,w,i}[L_{w,i}]}.\end{aligned}$$

By applying Theorem 6 of Hansen (2008), we find that uniformly over $i \in \{1, \cdots, n\}$,

$$\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}] = O_P(\varepsilon_n). \tag{16}$$

Furthermore, observe that

$$\begin{aligned}&\hat{1}_{n,i}\frac{\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]}{\mathbf{E}_{Q,w,i}[L_{w,i}]\hat{\mathbf{E}}_{Q,w,i}[L_{w,i}]} \qquad (17)\\ = \;&\hat{1}_{n,i}\frac{\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]}{\mathbf{E}_{Q,w,i}[L_{w,i}]^2} + \hat{1}_{n,i}\frac{\left\{\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]\right\}^2}{\mathbf{E}_{Q,w,i}[L_{w,i}]^2\hat{\mathbf{E}}_{Q,w,i}[L_{w,i}]}.\end{aligned}$$

The absolute value of the last term is bounded by

$$\delta_n^{-1}\hat{1}_{n,i}\left|\mathbf{E}_{Q,w,i}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,w,i}[L_{1,w,i}]\right|^2 \left|\mathbf{E}_{Q,w,i}[L_{w,i}]^{-2}\right|$$

Using Bayes' rule, we deduce that

$$\mathbf{E}_Q[L_{w,i}|V_i, W_i = w] = \frac{f_P(V_i, w)}{f_Q(V_i, w)}$$

and hence

$$\begin{aligned}&\mathbf{E}_Q\left[\mathbf{E}_Q^{-\bar{a}}(L_{w,i}|V_i, W_i = w)f_Q^{-\bar{a}}(V_i, w)\right] \qquad (18)\\ = \;&\mathbf{E}_Q\left[f_P^{-\bar{a}}(V_i, w)\right] = \sum_{d\in\{0,1\}}\mathbf{E}_{d,w}\left[f_P^{-\bar{a}}(V_i, w)\right]q_{d,w} < \infty\end{aligned}$$

by Assumption 2(iv). Hence we find that $\mathbf{E}_Q\left|\mathbf{E}_{Q,w,i}[L_{w,i}]^{-2}\right| < \infty$. The last term of (17) is, therefore, $O_P(\delta_n^{-1}\varepsilon_n^2) = o_P(\varepsilon_n)$. Combining this with (16),

$$\hat{1}_{n,i}\left\{p_1(v,w) - \hat{p}_1(v,w)\right\} = \hat{1}_{n,i}\frac{\mathbf{E}_{Q,v,w}[L_{1,w,i}] - \hat{\mathbf{E}}_{Q,v,w}[L_{1,w,i}]}{\mathbf{E}_{Q,v,w}[L_{w,i}]^2} + o_P(\varepsilon_n) = O_P(\varepsilon_n).$$

Hence we obtain the wanted result. ■

Lemma A2 : *Suppose that S_i is a random variable such that $\mathbf{E}_Q\left[|S_i|^r\right] < \infty$, $r \geq 4$, and $\mathbf{E}_Q[S_i|V_{1i} = \cdot, (V_{2i}, W) = (v_2, w)]$ is L_1+1 times continuously differentiable with bounded derivatives and uniformly continuous (L_1+1)-th derivatives.*
(i) *Suppose that the assumptions of* Theorem 3 *hold. Then, for $d = 0, 1$,*

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i\hat{1}_{n,i}\left(p_d(V_i,w) - \hat{p}_d(V_i,w)\right) \\
= \; & -\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q[S_i|V_i, W_i = w]\mathcal{J}_{d,w,i}}{\mathbf{E}_Q[L_{w,i}|V_i, W_i = w]} \\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q[S_i|V_i, W_i = w]p_d(V_i,w)\mathcal{J}_{w,i}}{\mathbf{E}_Q[L_{w,i}|V_i, W_i = w]} + o_P(n^{-1/2}),
\end{aligned}$$

where $\mathcal{J}_{d,w,i} \equiv L_{d,w,i} - \mathbf{E}_Q\left[L_{d,w,i}|V_i, W_i = w\right]$ and $\mathcal{J}_{w,i} = \mathcal{J}_{1,w,i} + \mathcal{J}_{0,w,i}$.

(ii) *Suppose that the assumptions of* Theorem 4 *hold. Then,*

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i\hat{1}_{n,i}\left(\hat{p}_1(V_i,w) - \tilde{p}_1(V_i,w)\right) \\
= \; & \mathbf{E}_Q\left[p_0(V_i,w)p_1(V_i,w)S_i\right]\left(\frac{\hat{q}_{1,w} - q_{1,w}}{q_{1,w}} - \frac{\hat{q}_{0,w} - q_{0,w}}{q_{0,w}}\right) + o_P(n^{-1/2}).
\end{aligned}$$

Proof of Lemma A2 : (i) Observe that by Bayes' rule,

$$f(V_i|1,w) = q_{1,w}(V_i)f_Q(V_i)/q_{1,w} = q_1(V_i,w)q_w(V_i)f_Q(V_i)/q_{1,w},$$

where $q_{1,w}(V_i) = \mathbf{E}_Q[1\{(D_i, W_i) = (d,w)\}|V_i]$, $q_w(V_i) = \mathbf{E}_Q[1\{W_i = w\}|V_i]$ and $f_Q(\cdot)$ is the density of V_i under Q. Hence

$$\begin{aligned}
p_1(V_i,w) &= \frac{f(V_i|1,w)p_{1,w}}{f(V_i|1,w)p_{1,w} + f(V_i|0,w)p_{0,w}} \qquad (19)\\
&= \frac{(q_1(V_i,w)/q_{1,w})p_{1,w}}{\sum_{d\in\mathcal{D}}(q_d(V_i,w)/q_{d,w})p_{d,w}} = \frac{\mathbf{E}_Q[L_{1,w,i}|V_i, W_i = w]}{\mathbf{E}_Q[L_{w,i}|V_i, W_i = w]}.
\end{aligned}$$

Let $K_{ji} = K_h\left(V_{1j} - V_{1i}\right)1\{V_{2j} = V_{2i}\}$ for brevity. By adding and subtracting the sum:

$$\frac{1}{n}\sum_{i=1}^{n} S_i\frac{\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\mathbf{E}_Q[L_{w,i}|V_i, W_i = w]\sum_{j=1,j\neq i}^{n} K_{ji}},$$

and noting (19), we write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i}\left(p_1(V_i,w)-\hat{p}_1(V_i,w)\right) \qquad (20)\\
=\ & \frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i}\left\{\frac{\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w]}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}-\frac{\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}\right\}\\
=\ & \frac{1}{n}\sum_{i=1}^{n} \frac{S_i \hat{1}_{n,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w]-1_{n,i}^{*}\frac{\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} K_{ji}}\right\}\\
&+\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i}\left\{\frac{1_{n,i}^{*}\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]\sum_{j=1,j\neq i}^{n} K_{ji}}-\frac{\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}\right\},
\end{aligned}$$

where $1_{n,i}^{*}=1\{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n} K_{ji}\geq \delta_n\}$. We write the last sum as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i \frac{\hat{1}_{n,i}\sum_{j=1,j\neq i}^{n} L_{1,w,j}K_{ji}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}\left\{1_{n,i}^{*}\frac{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} K_{ji}}-\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]\right\}\\
=\ & -\frac{1}{n}\sum_{i=1}^{n} \frac{S_i \hat{1}_{n,i}\mathbf{E}_Q[L_{1,w,j}|V_i,W_i=w]}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]^2}\left\{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]-1_{n,i}^{*}\frac{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} K_{ji}}\right\}+o_P(n^{-1/2})\\
=\ & -\frac{1}{n}\sum_{i=1}^{n} \frac{S_i \hat{1}_{n,i}p_1(V_i,w)}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]-1_{n,i}^{*}\frac{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} K_{ji}}\right\}+o_P(n^{-1/2}).
\end{aligned}$$

The first equality uses Lemma A1 and the second (19). Let

$$K_{n,i}=\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]-1_{n,i}^{*}\frac{\sum_{j=1,j\neq i}^{n} L_{w,j}K_{ji}}{\sum_{j=1,j\neq i}^{n} K_{ji}}$$

and write the last sum as

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n} \frac{S_i \hat{1}_{n,i}p_1(V_i,w)K_{n,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\\
=\ & -\frac{1}{n}\sum_{i=1}^{n} \frac{S_i p_1(V_i,w)K_{n,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}-\frac{1}{n}\sum_{i=1}^{n} \frac{S_i\left\{1-\hat{1}_{n,i}\right\}p_1(V_i,w)K_n}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}.
\end{aligned}$$

With large probability, the last sum is bounded by

$$K_n\times\frac{1}{n}\sum_{i=1}^{n}\left|\frac{S_i p_1(V_i,w)}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\right| 1\left\{\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w]f_Q(V_i,w)\leq \delta_n+\varepsilon_n\right\}$$

by (16), where $K_n=\max_{1\leq i\leq n}|K_{n,i}|$. It is not hard to see that $K_n=O_P(1)$. Note that the expectation $\mathbf{E}_Q$ of the above sum is bounded by

$$\begin{aligned} & C\left(\delta_n+\varepsilon_n\right)^{\bar{a}} \mathbf{E}_Q\left[\mathbf{E}_Q^{-\bar{a}}[L_{1,w,i}|V_i,W_i=w] f_Q^{-\bar{a}}(V_i,w)\right] \\ = \; & O\left(\left(\delta_n+\varepsilon_n\right)^{\bar{a}}\right)=o(n^{-1/2}). \end{aligned}$$

Hence we conclude that

$$\frac{1}{n}\sum_{i=1}^n \frac{S_i \hat{1}_{n,i} p_1(V_i,w) K_{n,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]} = \frac{1}{n}\sum_{i=1}^n \frac{S_i p_1(V_i,w) K_{n,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]} + o_P(n^{-1/2}). \tag{21}$$

Applying the similar argument to the second to the last sum of (20) to eliminate $\hat{1}_{n,i}$, we finally write

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n S_i \hat{1}_{n,i}\left(p_1(V_i,w)-\hat{p}_1(V_i,w)\right) \\ = \; & \frac{1}{n}\sum_{i=1}^n \frac{S_i}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w] - 1_{n,i}^* \frac{\sum_{j=1,j\neq i}^n L_{1,w,j}K_{ji}}{\sum_{j=1,j\neq i}^n K_{ji}}\right\} \\ & -\frac{1}{n}\sum_{i=1}^n \frac{S_i p_1(V_i,w)}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w] - 1_{n,i}^* \frac{\sum_{j=1,j\neq i}^n L_{w,j}K_{ji}}{\sum_{j=1,j\neq i}^n K_{ji}}\right\} + o_P(n^{-1/2}) \end{aligned}$$

By Lemma B1 below, the difference of the last two terms is asymptotically equivalent to (up to $o_P(n^{-1/2})$)

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{E}_Q[S_i|V_i,W_i=w]}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w] - L_{1,w,i}\right\} \\ & -\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{E}_Q[S_i|V_i,W_i=w]p_1(V_i,w)}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}\left\{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w] - L_{w,i}\right\} \\ = \; & -\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{E}_Q[S_i|V_i,W_i=w]\mathcal{J}_{1,w,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]} + \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{E}_Q[S_i|V_i,W_i=w]p_1(V_i,w)\mathcal{J}_{w,i}}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]} \end{aligned}$$

using the definitions of $\mathcal{J}_{1,w,i}$ and $\mathcal{J}_{w,i}$.

(ii) First, we let $1_{n,i} = 1\left\{\mathbf{E}_Q(L_{w,i}|V_i,W_i=w) f_Q(V_i,w) \geq \delta_n\right\}$, and write

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n S_i \hat{1}_{n,i}\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right) \\ = \; & \frac{1}{n}\sum_{i=1}^n S_i \hat{1}_{n,i} 1_{n,i}\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right) + \frac{1}{n}\sum_{i=1}^n S_i \hat{1}_{n,i}\left(1-1_{n,i}\right)\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right). \end{aligned} \tag{22}$$

By Lemma A1, $\max_{1\leq i\leq n}|\hat{p}(V_i,w)-\tilde{p}(V_i,w)|\,\hat{1}_{n,i}=O_P(\varepsilon_n)$ and hence

$$\begin{aligned}\mathbf{E}_Q\left[\frac{1}{n}\sum_{i=1}^{n}|S_i|\,|1-1_{n,i}|\right] &\leq C\mathbf{E}_Q\left[1\left\{\mathbf{E}_Q(L_{w,i}|V_i,W_i=w)f_Q(V_i,w)\leq\delta_n\right\}\right] \\ &\leq C\delta_n^{\bar{a}}\mathbf{E}_Q\left[\mathbf{E}_Q^{-\bar{a}}(L_{w,i}|V_i,W_i=w)f_Q^{-\bar{a}}(V_i,w)\right]\end{aligned} \tag{23}$$

by Markov's inequality. By (18), the last expectation is finite. Since $\delta_n^{\bar{a}}\varepsilon_n=o(n^{-1/2})$ (Assumption 3(ii)), we conclude that

$$\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right)=\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right)+o_P(n^{-1/2}).$$

As for the leading sum, note that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\left(\hat{p}(V_i,w)-\tilde{p}(V_i,w)\right)\\ =\;&\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\left\{\frac{\sum_{j=1}^{n}L_{1,w,j}K_{ji}}{\sum_{j=1}^{n}L_{w,j}K_{ji}}-\frac{\sum_{j=1}^{n}\hat{L}_{1,w,j}K_{ji}}{\sum_{j=1}^{n}\hat{L}_{w,j}K_{ji}}\right\}\\ =\;&\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\frac{\sum_{j=1}^{n}\left\{L_{1,w,j}-\hat{L}_{1,w,j}\right\}K_{ji}}{\sum_{j=1}^{n}L_{w,j}K_{ji}}\\ &+\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\sum_{j=1}^{n}\hat{L}_{1,w,j}K_{ji}\left\{\frac{1}{\sum_{j=1}^{n}L_{w,j}K_{ji}}-\frac{1}{\sum_{j=1}^{n}\hat{L}_{w,j}K_{ji}}\right\}.\end{aligned}$$

Now, note that as for the second term,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\sum_{j=1}^{n}\hat{L}_{1,w,j}K_{ji}\left\{\frac{1}{\sum_{j=1}^{n}L_{w,j}K_{ji}}-\frac{1}{\sum_{j=1}^{n}\hat{L}_{w,j}K_{ji}}\right\}\\ =\;&\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}1_{n,i}\frac{\sum_{j=1}^{n}\hat{L}_{1,w,j}K_{ji}}{\sum_{j=1}^{n}\hat{L}_{w,j}K_{ji}}\left\{\frac{\sum_{j=1}^{n}\left\{\hat{L}_{w,j}-L_{w,j}\right\}K_{ji}}{\sum_{j=1}^{n}L_{w,j}K_{ji}}\right\}.\end{aligned}$$

Using Lemma A1, we can write the last sum as

$$\frac{1}{n}\sum_{i=1}^{n}S_i\hat{1}_{n,i}\frac{\sum_{j=1}^{n}L_{1,w,j}K_{ji}}{\sum_{j=1}^{n}L_{w,j}K_{ji}}\left\{\frac{\sum_{j=1}^{n}\left\{\hat{L}_{w,j}-L_{w,j}\right\}K_{ji}}{\sum_{j=1}^{n}L_{w,j}K_{ji}}\right\}+o_P(n^{-1/2}).$$

Therefore, we can write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i}\{\hat{p}(X_i)-\tilde{p}(X_i)\} \\
= \quad & -\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i} 1_{n,i} \frac{\sum_{j=1}^{n} L_{0,w,j}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \frac{\sum_{j=1}^{n}\left\{\hat{L}_{1,w,j}-L_{1,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \\
& +\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i} 1_{n,i} \frac{\sum_{j=1}^{n} L_{1,w,j}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \frac{\sum_{j=1}^{n}\left\{\hat{L}_{0,w,j}-L_{0,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} + o_P(n^{-1/2}) \\
= \quad & -\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i} 1_{n,i} \frac{p_0(V_i,w)\sum_{j=1}^{n}\left\{\hat{L}_{1,w,j}-L_{1,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \\
& +\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i} 1_{n,i} \frac{p_1(V_i,w)\sum_{j=1}^{n}\left\{\hat{L}_{0,w,j}-L_{0,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} + o_P(n^{-1/2})
\end{aligned}
$$

As for the first term, observe that

$$
\begin{aligned}
& \hat{1}_{n,i} 1_{n,i} \frac{\frac{1}{n}\sum_{j=1}^{n}\left\{\hat{L}_{1,w,j}-L_{1,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \\
= \quad & \left(\frac{p_{1,w}}{\hat{q}_{1,w}}-\frac{p_{1,w}}{q_{1,w}}\right)\hat{1}_{n,i} 1_{n,i} \frac{\sum_{j=1}^{n} 1\{(D_j,W_j)=(1,w)\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \\
= \quad & \left(\frac{p_{1,w}}{\hat{q}_{1,w}}-\frac{p_{1,w}}{q_{1,w}}\right)\frac{q_1(V_i,w)}{q_1(V_i,w)p_{1,w}/q_{1,w}+q_0(V_i,w)p_{0,w}/q_{0,w}} + o_P(n^{-1/2}) \\
= \quad & \left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)\frac{q_1(V_i,w)p_{1,w}/q_{1,w}}{q_1(V_i,w)p_{1,w}/q_{1,w}+q_0(V_i,w)p_{0,w}/q_{0,w}} + o_P(n^{-1/2}) \\
= \quad & \left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)p_1(V_i,w) + o_P(n^{-1/2}).
\end{aligned}
$$

The last equality follows by (19) and the second equality follows because

$$
\begin{aligned}
& \hat{1}_{n,i} 1_{n,i} \frac{\sum_{j=1}^{n} 1\{(D_i,W_i)=(1,w)\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}} \\
= \quad & \hat{1}_{n,i} 1_{n,i} \frac{\sum_{j=1}^{n} 1\{(D_i,W_i)=(1,w)\}K_{ji}/\sum_{j=1}^{n} 1\{W_i=w\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}/\sum_{j=1}^{n} 1\{W_i=w\}K_{ji}} \\
= \quad & \frac{q_1(V_i,w)}{q_1(V_i,w)p_{1,w}/q_{1,w}+q_0(V_i,w)p_{0,w}/q_{0,w}} + o_P(n^{-1/4}).
\end{aligned}
$$

Applying the same step to the term

$$
\frac{p_1(V_i,w)\sum_{j=1}^{n}\left\{\hat{L}_{0,w,j}-L_{0,w,j}\right\}K_{ji}}{\sum_{j=1}^{n} L_{w,j}K_{ji}},
$$

we conclude that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} S_i \hat{1}_{n,i} 1_{n,i}\{\hat{p}(X_i) - \tilde{p}(X_i)\} \\
=\;& \frac{1}{n}\sum_{i=1}^{n} S_i p_0(V_i,w) p_1(V_i,w)\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}} - \frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) + o_P(n^{-1/2}).
\end{aligned}$$

Finally, we write the last sum as

$$\mathbf{E}_Q\left[p_0(V_i,w)p_1(V_i,w)S_i\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}} - \frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) + o_P(n^{-1/2})$$

and complete the proof. ■

Lemma A3 : (i) *Suppose that the assumptions of* Theorem 3 *hold, and let* $\varepsilon_{d,w,i} = Y_{di} - \beta_d(V_i,w)$. *Then,*

$$\begin{aligned}
&\frac{p_{1,w}}{q_{1,w}n}\sum_{i\in S_{1,w}} \hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}} \hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \\
=\;& \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)} \\
&+\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\tau(V_i,w)L_{w,i} + o_P(n^{-1/2}).
\end{aligned}$$

(ii) *Suppose that the assumptions of* Theorem 4 *hold, and let* $\varepsilon_{d,w,i} = Y_{di} - \beta_d(V_i,w)$. *Then,*

$$\begin{aligned}
&\frac{p_{1,w}}{n_{1,w}}\sum_{i\in S_{1,w}} \tilde{1}_{n,i}\frac{g(V_i,w)Y_i}{\tilde{p}_1(V_i,w)} - \frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}} \tilde{1}_{n,i}\frac{g(V_i,w)Y_i}{\tilde{p}_0(V_i,w)} \\
=\;& \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)} \\
&+\frac{1}{n}\sum_{i=1}^{n}\left\{g(V_i,w)\tau(V_i,w) - \mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]\right\}L_{1,w,i} \\
&+\frac{1}{n}\sum_{i=1}^{n}\left\{g(V_i,w)\tau(V_i,w) - \mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]\right\}L_{0,w,i} \\
&+\mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]p_{1,w} + \mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]p_{0,w} + o_P(n^{-1/2}).
\end{aligned}$$

Proof of Lemma A3 : (i) We first write

$$\begin{aligned}
& \frac{p_{1,w}}{q_{1,w}n} \sum_{i \in S_{1,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i \in S_{0,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \\
= & \frac{1}{n} \sum_{i=1}^{n} \hat{1}_{n,i} \frac{g(V_i,w)Y_i L_{1,w,i}}{\hat{p}_1(V_i,w)} - \frac{1}{n} \sum_{i=1}^{n} \hat{1}_{n,i} \frac{g(V_i,w)Y_i L_{0,w,i}}{\hat{p}_0(V_i,w)} = A_{1n} + A_{2n}, \text{ say.}
\end{aligned}$$

We first write

$$A_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(V_i,w)Y_i \hat{1}_{n,i} L_{1,w,i}}{p_1(V_i,w)} + \tilde{A}_{1n},$$

where

$$\tilde{A}_{1n} = \frac{1}{n} \sum_{i=1}^{n} g(V_i,w)Y_i L_{1,w,i} \hat{1}_{n,i} \left(\frac{1}{\hat{p}_1(V_i,w)} - \frac{1}{p_1(V_i,w)} \right).$$

As for $\tilde{A}_{1n}$, note that

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} g(V_i,w)Y_i L_{1,w,i} \hat{1}_{n,i} \left(\frac{p_1(V_i,w) - \hat{p}_1(V_i,w)}{\hat{p}_1(V_i,w) p_1(V_i,w)} \right) \\
= & \frac{1}{n} \sum_{i=1}^{n} g(V_i,w)Y_i L_{1,w,i} \hat{1}_{n,i} \left(\frac{p_1(V_i,w) - \hat{p}_1(V_i,w)}{p_1^2(V_i,w)} \right) \\
& + \frac{1}{n} \sum_{i=1}^{n} g(V_i,w)Y_i L_{1,w,i} \hat{1}_{n,i} \frac{p_1(V_i,w) - \hat{p}_1(V_i,w)}{p_1(V_i,w)} \left(\frac{1}{\hat{p}_1(V_i,w)} - \frac{1}{p_1(V_i,w)} \right).
\end{aligned}$$

The absolute value of the last sum is bounded by

$$\frac{1}{n} \sum_{i=1}^{n} |g(V_i,w)Y_i L_{1,w,i}| \, \hat{1}_{n,i} \frac{(p_1(V_i,w) - \hat{p}_1(V_i,w))^2}{p_1(V_i,w)^2 \delta_n}. \tag{24}$$

On the other hand, observe that for any $q > 0$,

$$\begin{aligned}
\mathbf{E}_Q \left[p_1^{-q}(V_i,w) \right] &= \sum_{(d,w) \in \{0,1\} \times \mathcal{W}} \mathbf{E}_{d,w} \left[\left\{ \frac{f(V_i|1,w)p_{1,w}}{f(V_i|1,w)p_{1,w} + f(V_i|0,w)p_{0,w}} \right\}^{-q} \right] q_{d,w} \qquad (25) \\
&\le C \sum_{(d,w) \in \{0,1\} \times \mathcal{W}} \mathbf{E}_{d,w} \left[f(V_i|1,w)^{-q} \right] q_{d,w} < \infty
\end{aligned}$$

by Assumption 2(iv). Hence by Lemma A1, we find that the sum in (24) is $o_P(n^{-1/2})$. We conclude that

$$\tilde{A}_{1n} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(V_i,w)Y_i L_{1,w,i}}{p_1^2(V_i,w)} \hat{1}_{n,i} \left(p_1(V_i,w) - \hat{p}_1(V_i,w) \right) + o_P(n^{-1/2}). \tag{26}$$

Let $S_i = g(V_i,w)Y_i L_{1,w,i}/p_1^2(V_i,w)\bar{1}_{n,i}$. Then,

$$\mathbf{E}_Q \left[|S_i|^r \right] \le \sqrt{\mathbf{E}_Q Y^2} \times \sqrt{\mathbf{E}_Q \left[p_1^{-4}(V_i,w) \right]}.$$

Therefore, $\mathbf{E}_Q[|S_i|^r] < \infty$. We apply Lemma A2(i) to obtain that the leading sum in (26) is asymptotically equivalent to (up to $o_P(n^{-1/2})$)

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\mathbf{E}_Q[Y_iL_{1,w,i}|V_i,W_i=w]\mathcal{J}_{1,w,i}}{p_1^2(V_i,w)\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]} \tag{27}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\mathbf{E}_Q[Y_iL_{1,w,i}|V_i,W_i=w]\mathcal{J}_{w,i}}{p_1(V_i,w)\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}.$$

Using the fact that

$$\begin{aligned}\mathbf{E}_Q[Y_iL_{1,w,i}|V_i,W_i &= w]=\mathbf{E}[Y_{1i}|V_i,(D_i,W_i)=(1,w)]q_1(V_i,w)p_{1,w}/q_{1,w}\\ &= \beta_1(V_i,w)q_1(V_i,w)p_{1,w}/q_{1,w}\end{aligned}$$

and $q_1(V_i,w)p_{1,w}/\{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]q_{1,w}\}=p_1(V_i,w)$ from (19), we write

$$\frac{\mathbf{E}_Q[Y_iL_{1,w,i}|V_i,W_i=w]}{\mathbf{E}_Q[L_{w,i}|V_i,W_i=w]}=\beta_1(V_i,w)p_1(V_i,w). \tag{28}$$

Using this, we write the first term in (27) as

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\beta_1(V_i,w)\mathcal{J}_{1,w,i}}{p_1(V_i,w)}$$

and the second term as (using (19))

$$\frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\beta_1(V_i,w)\mathcal{J}_{w,i}=\frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\beta_1(V_i,w)\left\{\mathcal{J}_{1,w,i}+\mathcal{J}_{0,w,i}\right\}.$$

Hence the difference in (27) is equal to

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\beta_1(V_i,w)p_0(V_i,w)}{p_1(V_i,w)}\mathcal{J}_{1,w,i}+\frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\beta_1(V_i,w)\mathcal{J}_{0,w,i}.$$

Therefore, we conclude that

$$\begin{aligned}\tilde{A}_{1n}&=-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\beta_1(V_i,w)p_0(V_i,w)}{p_1(V_i,w)}\mathcal{J}_{1,w,i}\\ &+\frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\beta_1(V_i,w)\mathcal{J}_{0,w,i}+o_P(n^{-1/2}).\end{aligned}$$

We turn to A_{2n} which we write as

$$A_{2n}=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{1}_{n,i}g(V_i,w)Y_iL_{0,w,i}}{p_0(V_i,w)}+\tilde{A}_{2n}+o_P(n^{-1/2}),$$

where

$$\tilde{A}_{2n} = \frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\hat{1}_{n,i} Y_i L_{0,w,i}\left(\frac{1}{\hat{p}_0(V_i,w)} - \frac{1}{p_0(V_i,w)}\right).$$

Similarly as before, we write

$$\begin{aligned}\tilde{A}_{2n} &= \frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\beta_0(V_i,w)\mathcal{J}_{1,w,i} \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)\beta_0(V_i,w)p_1(V_i,w)}{p_0(V_i,w)}\mathcal{J}_{0,w,i} + o_P(n^{-1/2}).\end{aligned}$$

Using the arguments employed to show (21) and combining the two results for $\tilde{A}_{1n}$ and $\tilde{A}_{2n}$, we deduce that

$$\begin{aligned}\tilde{A}_{1n} - \tilde{A}_{2n} &= -\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\frac{\beta_1(V_i,w)p_0(V_i,w)}{p_1(V_i,w)} + \beta_0(V_i,w)\right)\mathcal{J}_{1,w,i} \\ &\quad +\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\beta_1(V_i,w) + \frac{\beta_0(V_i,w)p_1(V_i,w)}{p_0(V_i,w)}\right)\mathcal{J}_{0,w,i} + o_P(n^{-1/2}) \\ &= -\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\frac{\beta_1(V_i,w) - \tau(V_i,w)p_1(V_i,w)}{p_1(V_i,w)}\right)\mathcal{J}_{1,w,i} \\ &\quad +\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\frac{\tau(V_i,w)p_0(V_i,w) + \beta_0(V_i,w)}{p_0(V_i,w)}\right)\mathcal{J}_{0,w,i} + o_P(n^{-1/2}),\end{aligned}$$

using the fact that $\tau(X) = \beta_1(X) - \beta_0(X)$.

Therefore,

$$\begin{aligned}&\frac{p_{1,w}}{q_{1,w}n}\sum_{i\in S_{1,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \\ =\ &\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)} \\ &+\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{1,w,i}\beta_1(V_i,w)}{p_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{0,w,i}\beta_0(V_i,w)}{p_0(V_i,w)} \\ &-\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\frac{\beta_1(V_i,w) - \tau(V_i,w)p_1(V_i,w)}{p_1(V_i,w)}\right)\mathcal{J}_{1,w,i} \\ &+\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\left(\frac{\tau(V_i,w)p_0(V_i,w) + \beta_0(V_i,w)}{p_0(V_i,w)}\right)\mathcal{J}_{0,w,i} + o_P(n^{-1/2}).\end{aligned}$$

By rearranging the terms, we rewrite

$$\begin{aligned}
& \frac{p_{1,w}}{q_{1,w}n}\sum_{i\in S_{1,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)}-\frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \\
= & \frac{1}{n}\sum_{i=1}^n\frac{g(V_i,w)L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^n\frac{g(V_i,w)L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)} \\
& +\frac{1}{n}\sum_{i=1}^n g(V_i,w)\tau(V_i,w)L_{1,w,i}+\frac{1}{n}\sum_{i=1}^n g(V_i,w)\tau(V_i,w)L_{0,w,i} \\
& +\frac{1}{n}\sum_{i=1}^n g(V_i,w)\left(\frac{\beta_1(V_i,w)-\tau(V_i,w)p_1(V_i,w)}{p_1(V_i,w)}\right)(\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w]) \\
& -\frac{1}{n}\sum_{i=1}^n g(V_i,w)\left(\frac{\tau(V_i,w)p_0(V_i,w)+\beta_0(V_i,w)}{p_0(V_i,w)}\right)\{\mathbf{E}_Q[L_{0,w,i}|V_i,W_i=w]\}+o_P(n^{-1/2}).
\end{aligned}$$

As for the last two terms, observe that

$$\begin{aligned}
H_n & \equiv \left\{\frac{\beta_1(V_i,w)}{p_1(V_i,w)}-\tau(V_i,w)\right\}\mathbf{E}_Q[L_{1,w,i}|V_i,W_i=w] \\
& \quad -\left\{\frac{\beta_0(V_i,w)}{p_0(V_i,w)}+\tau(V_i,w)\right\}\mathbf{E}_Q[L_{0,w,i}|V_i,W_i=w] \\
& = \left\{\frac{\beta_1(V_i,w)}{p_1(V_i,w)}-\tau(V_i,w)\right\}\frac{q_1(V_i,w)p_{1,w}}{q_{1,w}} \\
& \quad -\left\{\frac{\beta_0(V_i,w)}{p_0(V_i,w)}+\tau(V_i,w)\right\}\frac{q_0(V_i,w)p_{0,w}}{q_{0,w}}.
\end{aligned}$$

However, by Bayes' rule,

$$\frac{p_{1,w}q_1(V_i,w)}{q_{1,w}}=\frac{p_{1,w}q_1(V_i,w)f_Q(V_i,w)}{q_{1,w}f_Q(V_i,w)}=\frac{p_{1,w}f(V_i|1,w)}{f_Q(V_i,w)}=\frac{p_1(V_i,w)f_P(V_i,w)}{f_Q(V_i,w)}. \tag{29}$$

Therefore,

$$\begin{aligned}
H_n & = \frac{f_P(V_i,w)}{f_Q(V_i,w)}\left\{\left\{\frac{\beta_1(V_i,w)}{p_1(V_i,w)}-\tau(V_i,w)\right\}p_1(V_i,w)-\left\{\frac{\beta_0(V_i,w)}{p_0(V_i,w)}+\tau(V_i,w)\right\}p_0(V_i,w)\right\} \\
& = 0
\end{aligned}$$

by the definition of $\tau(V_i,w)$. Hence we obtain the wanted result.

(ii) We write

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\hat{L}_{1,w,i}}{\tilde{p}_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\hat{L}_{0,w,i}}{\tilde{p}_0(V_i,w)} \tag{30}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}L_{1,w,i}}{\tilde{p}_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}L_{0,w,i}}{\tilde{p}_0(V_i,w)}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\{\hat{L}_{1,w,i}-L_{1,w,i}\}}{\tilde{p}_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\{\hat{L}_{0,w,i}-L_{0,w,i}\}}{\tilde{p}_0(V_i,w)}.$$

We write the first difference as

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{1,w,i}}{\hat{p}_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{0,w,i}}{\hat{p}_0(V_i,w)}\right\}$$
$$+\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{1,w,i}}{p_1^2(V_i,w)}A_i-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{0,w,i}}{p_0^2(V_i,w)}B_i\right\}+o_P(n^{-1/2})$$
$$= J_{1n}+J_{2n}+o_P(n^{-1/2}),\text{ say,}$$

where $A_i=\hat{p}_1(X_i)-\tilde{p}_1(X_i)$ and $B_i=\hat{p}_0(X_i)-\tilde{p}_0(X_i)$. One can check that the normalized sums with trimming factor $\tilde{1}_{n,i}$ can be replaced by the same sums but with $\hat{1}_{n,i}$ (with the resulting discrepancy confined to $o_P(n^{-1/2})$), because $\hat{q}_{d,w}$ is consistent for $q_{d,w}>0$. As for J_{2n}, by applying Lemma A2(ii), we have

$$J_{2n} = \mathbf{E}_Q\left[\frac{g(V_i,w)Y_iL_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}-\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right)$$
$$-\mathbf{E}_Q\left[\frac{g(V_i,w)Y_iL_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}-\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2})$$
$$= \mathbf{E}_Q\left[g(V_i,w)Y_i\left\{\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}+\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)$$
$$-\mathbf{E}_Q\left[g(V_i,w)Y_i\left\{\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}+\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2}).$$

On the other hand, as for the last difference in (30), it is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\{\hat{L}_{1,w,i}-L_{1,w,i}\}}{p_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\{\hat{L}_{0,w,i}-L_{0,w,i}\}}{p_0(V_i,w)}+o_P(n^{-1/2})$$
$$= -\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_iL_{1,w,i}}{p_1(V_i,w)}\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)+\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_iL_{0,w,i}}{p_0(V_i,w)}\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2})$$
$$= -\mathbf{E}_Q\left[\frac{g(V_i,w)Y_iL_{1,w,i}}{p_1(V_i,w)}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)+\mathbf{E}_Q\left[\frac{g(V_i,w)Y_iL_{0,w,i}}{p_0(V_i,w)}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2}).$$

Combining these results, we conclude that

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\hat{L}_{1,w,i}}{\tilde{p}_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\tilde{1}_{n,i}\hat{L}_{0,w,i}}{\tilde{p}_0(V_i,w)} \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{1,w,i}}{\hat{p}_1(V_i,w)} - \frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)Y_i\hat{1}_{n,i}L_{0,w,i}}{\hat{p}_0(V_i,w)} \\
& +\mathbf{E}_Q\left[g(V_i,w)Y_i\left\{-L_{1,w,i}+\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right) \\
& -\mathbf{E}_Q\left[g(V_i,w)Y_i\left\{\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}-L_{0,w,i}\right\}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2}).
\end{aligned} \tag{31}$$

The last difference is written as

$$\begin{aligned}
& \mathbf{E}_Q\left[g(V_i,w)\left\{-Y_{1i}L_{1,w,i}+Y_{0,i}\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right) \\
& -\mathbf{E}_Q\left[g(V_i,w)\left\{Y_{1i}\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}-Y_{0,i}L_{0,w,i}\right\}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) \\
= \quad & \mathbf{E}_Q\left[g(V_i,w)\left\{-\{Y_{1i}-Y_{0,i}\}L_{1,w,i}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right) \\
& +\mathbf{E}_Q\left[g(V_i,w)Y_{0,i}\left\{\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}-L_{1,w,i}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right) \\
& -\mathbf{E}_Q\left[g(V_i,w)\{Y_{1i}-Y_{0,i}\}L_{0,w,i}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) \\
& +\mathbf{E}_Q\left[g(V_i,w)Y_{1i}\left\{L_{0,w,i}-\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right).
\end{aligned}$$

The second and the fourth expectations vanish because

$$\begin{aligned}
& \mathbf{E}_Q\left[g(V_i,w)Y_{0,i}\left\{-L_{1,w,i}+\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right] \\
= \quad & \mathbf{E}\left[g(V_i,w)\beta_0(V_i,w)\left\{-1\{(D_i,W_i)=(1,w)\}+\frac{1\{(D_i,W_i)=(0,w)\}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right] \\
= \quad & \mathbf{E}\left[g(V_i,w)\beta_0(V_i,w)\left\{p_1(V_i,w)-p_1(V_i,w)\right\}\right]=0
\end{aligned}$$

and similarly,

$$\begin{aligned}
& \mathbf{E}_Q\left[g(V_i,w)Y_{1i}\left\{L_{0,w,i}-\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}\right\}\right] \\
= \quad & \mathbf{E}\left[g(V_i,w)\beta_1(V_i,w)\left\{p_0(V_i,w)-p_0(V_i,w)\right\}\right]=0.
\end{aligned}$$

Furthermore, observe that

$$\begin{aligned}
& \mathbf{E}_Q\left[g(V_i,w)\left\{-\{Y_{1i}-Y_{0,i}\}L_{1,w,i}\right\}\right] \\
= \;& -\mathbf{E}\left[g(V_i,w)\{Y_{1i}-Y_{0,i}\}1\{(D_i,W_i)=(1,w)\}\right] \\
= \;& -\mathbf{E}\left[g(V_i,w)\{\beta_1(V_i,w)-\beta_0(V_i,w)\}p_1(V_i,w)\right] \\
= \;& -\mathbf{E}\left[g(V_i,w)\tau(V_i,w)p_1(V_i,w)\right]
\end{aligned}$$

and similarly,

$$-\mathbf{E}_Q\left[g(V_i,w)\{Y_{1i}-Y_{0,i}\}L_{0,w,i}\right] = -\mathbf{E}\left[g(V_i,w)\tau(V_i,w)p_0(V_i,w)\right].$$

Hence, as for the last two terms in (31), we find that

$$\begin{aligned}
& \mathbf{E}_Q\left[g(V_i,w)Y_i\left\{-L_{1,w,i}+\frac{L_{0,w,i}p_1(V_i,w)}{p_0(V_i,w)}\right\}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right) \\
& -\mathbf{E}_Q\left[g(V_i,w)Y_i\left\{\frac{L_{1,w,i}p_0(V_i,w)}{p_1(V_i,w)}-L_{0,w,i}\right\}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) \\
= \;& -\mathbf{E}_Q\left[g(V_i,w)\tau(V_i,w)L_{1,w,i}\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)-\mathbf{E}_Q\left[g(V_i,w)\tau(V_i,w)L_{0,w,i}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) \\
= \;& -p_{1,w}\mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]\left(\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)-p_{0,w}\mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}\right) \\
= \;& -\mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]\frac{1}{n}\sum_{i=1}^{n}\left(L_{1,w,i}-p_{1,w}\right)-\mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]\frac{1}{n}\sum_{i=1}^{n}\left(L_{0,w,i}-p_{0,w}\right).
\end{aligned}$$

Applying the result of (i) to the first difference of (31), we conclude that the difference in (ii) is equal to

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)}-\frac{1}{n}\sum_{i=1}^{n}\frac{g(V_i,w)L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)}+\Gamma_{n,w}+o_P(n^{-1/2}),$$

where

$$\begin{aligned}
\Gamma_{n,w} \equiv \;& \frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\tau(V_i,w)L_{1,w,i}+\frac{1}{n}\sum_{i=1}^{n}g(V_i,w)\tau(V_i,w)L_{0,w,i} \\
& -\mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]\frac{1}{n}\sum_{i=1}^{n}\left(L_{1,w,i}-p_{1,w}\right) \\
& -\mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]\frac{1}{n}\sum_{i=1}^{n}\left(L_{0,w,i}-p_{0,w}\right).
\end{aligned}$$

The proof is complete because

$$
\begin{aligned}
\Gamma_{n,w} &= \left\{\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\tau(V_i,w) - \mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right]\right\} L_{1,w,i} \\
&\quad + \left\{\frac{1}{n}\sum_{i=1}^{n} g(V_i,w)\tau(V_i,w) - \mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right]\right\} L_{0,w,i} \\
&\quad + \mathbf{E}_{1,w}\left[g(V_i,w)\tau(V_i,w)\right] p_{1,w} + \mathbf{E}_{0,w}\left[g(V_i,w)\tau(V_i,w)\right] p_{0,w},
\end{aligned}
$$

rearranging the terms. ■

Proof of Theorem 3 : Note that

$$
\begin{aligned}
\hat{\tau}_{wate} - \tau_{wate} &= \frac{1}{\mathbf{E}g(X_i)} \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}n} \sum_{i\in S_{1,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i\in S_{0,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \right\} \quad (32) \\
&\quad + R_n - \tau_{wate},
\end{aligned}
$$

where

$$
\begin{aligned}
R_n &\equiv \left\{ \frac{1}{\sum_{w\in\mathcal{W}} \frac{1}{n}\sum_{i=1}^{n} g(V_i,w)L_{w,i}} - \frac{1}{\mathbf{E}g(X_i)} \right\} \\
&\quad \times \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}n} \sum_{i\in S_{1,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i\in S_{0,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \right\}.
\end{aligned}
$$

Now, since $\sum_{w\in\mathcal{W}} \frac{1}{n}\sum_{i=1}^{n} g(V_i,w)L_{w,i} = \mathbf{E}g(X_i) + O_P(n^{-1/2})$ and

$$
\begin{aligned}
& \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}n} \sum_{i\in S_{1,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i\in S_{0,w}} \hat{1}_{n,i} \frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)} \right\} \\
&= \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}n} \sum_{i\in S_{1,w}} \frac{g(V_i,w)Y_i}{p_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i\in S_{0,w}} \frac{g(V_i,w)Y_i}{p_0(V_i,w)} \right\} + o_P(1),
\end{aligned}
$$

we find that

$$
\begin{aligned}
R_n &= \left\{ \frac{1}{\sum_{w\in\mathcal{W}} \frac{1}{n}\sum_{i=1}^{n} g(V_i,w)L_{w,i}} - \frac{1}{\mathbf{E}g(X_i)} \right\} \\
&\quad \times \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{q_{1,w}n} \sum_{i\in S_{1,w}} \frac{g(V_i,w)Y_i}{p_1(V_i,w)} - \frac{p_{0,w}}{q_{0,w}n} \sum_{i\in S_{0,w}} \frac{g(V_i,w)Y_i}{p_0(V_i,w)} \right\} + o_P(n^{-1/2}).
\end{aligned}
$$

Observing that

$$\begin{aligned}
& \sum_{w\in\mathcal{W}}\left\{\frac{p_{1,w}}{q_{1,w}n}\sum_{i\in S_{1,w}}\frac{g(V_i,w)Y_i}{p_1(V_i,w)}-\frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\frac{g(V_i,w)Y_i}{p_0(V_i,w)}\right\} \\
= \;& \mathbf{E}g(X_i)\tau_{wate}+O_P(n^{-1/2})
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)L_{w,i}}-\frac{1}{\mathbf{E}g(X_i)} \\
= \;& -\frac{\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)L_{w,i}-\mathbf{E}g(X_i)}{\mathbf{E}g(X_i)^2}+o_P(1),
\end{aligned}$$

we can write

$$\begin{aligned}
R_n &= -\frac{\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)L_{w,i}-\mathbf{E}g(X_i)}{\mathbf{E}g(X_i)}\tau_{wate}+o_P(n^{-1/2}) \\
&= -\frac{1}{\mathbf{E}\left[g(X_i)\right]}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)L_{w,i}\tau_{wate}+\tau_{wate}+o_P(n^{-1/2}).
\end{aligned}$$

Applying Lemma A3(i) to the first sum in (32), we obtain that

$$\begin{aligned}
& \frac{1}{\mathbf{E}[g(X_i)]}\sum_{w\in\mathcal{W}}\left\{\frac{p_{1,w}}{q_{1,w}n}\sum_{i\in S_{1,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_1(V_i,w)}-\frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\hat{1}_{n,i}\frac{g(V_i,w)Y_i}{\hat{p}_0(V_i,w)}\right\}+R_n-\tau_{wate} \\
= \;& \frac{1}{\mathbf{E}[g(X_i)]}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)\left\{\frac{L_{1,w,i}\varepsilon_{1,w,i}}{p_1(V_i,w)}-\frac{L_{0,w,i}\varepsilon_{0,w,i}}{p_0(V_i,w)}\right\} \\
& +\frac{1}{\mathbf{E}[g(X_i)]}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^n g(V_i,w)(\tau(V_i,w)-\tau_{wate})L_{w,i}+o_P(n^{-1/2}).
\end{aligned}$$

By applying the Central Limit Theorem, we obtain the asymptotic distribution of $\hat{\tau}_{wate}$.

As for $\tilde{\tau}_{wate}$, observe that

$$\begin{aligned}
& \tilde{\tau}_{wate}-\tau_{wate} \qquad (33) \\
= \;& \frac{1}{\mathbf{E}g(X_i)}\sum_{w\in\mathcal{W}}\left\{\frac{p_{1,w}}{n_{1,w}}\sum_{i\in S_{1,w}}\tilde{1}_{n,i}\frac{g(V_i,w)Y_i}{\tilde{p}_1(V_i,w)}-\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{1}_{n,i}\frac{g(V_i,w)Y_i}{\tilde{p}_0(V_i,w)}\right\}+\tilde{R}_n-\tau_{wate}.
\end{aligned}$$

where $\tilde{\mathbf{E}}_d g(X_i) = \sum_{w\in\mathcal{W}} \frac{p_{d,w}}{n_{d,w}} \sum_{i\in S_{d,w}} \tilde{1}_{n,i} \frac{g(V_i,w)}{\tilde{p}_d(V_i,w)}$, $\tau_d = \mathbf{E}\left[g(X_i)\beta_d(X_i)\right]/\mathbf{E}g(X_i)$, and

$$\begin{aligned}\tilde{R}_n &\equiv \left\{\frac{1}{\tilde{\mathbf{E}}_1 g(X_i)} - \frac{1}{\mathbf{E}g(X_i)}\right\} \frac{p_{1,w}}{n_{1,w}} \sum_{i\in S_{1,w}} \tilde{1}_{n,i} \frac{g(V_i,w)Y_i}{\tilde{p}_1(V_i,w)} \\ &\quad - \left\{\frac{1}{\tilde{\mathbf{E}}_1 g(X_i)} - \frac{1}{\mathbf{E}g(X_i)}\right\} \frac{p_{0,w}}{n_{0,w}} \sum_{i\in S_{0,w}} \tilde{1}_{n,i} \frac{g(V_i,w)Y_i}{\tilde{p}_0(V_i,w)}.\end{aligned}$$

It is not hard to show that

$$\begin{aligned}\frac{p_{1,w}}{n_{1,w}} \sum_{i\in S_{1,w}} \tilde{1}_{n,i} \frac{g(V_i,w)Y_i}{\tilde{p}_1(V_i,w)} &= \tau_1 \mathbf{E}g(X_i) + O_P(n^{-1/2}) \text{ and} \\ \frac{p_{0,w}}{n_{0,w}} \sum_{i\in S_{0,w}} \tilde{1}_{n,i} \frac{g(V_i,w)Y_i}{\tilde{p}_0(V_i,w)} &= \tau_0 \mathbf{E}g(X_i) + O_P(n^{-1/2}).\end{aligned}$$

Hence

$$\begin{aligned}\tilde{R}_n &\equiv \left\{\frac{1}{\tilde{\mathbf{E}}_1 g(X_i)} - \frac{1}{\mathbf{E}g(X_i)}\right\} \tau_1 \mathbf{E}g(X_i) \\ &\quad - \left\{\frac{1}{\tilde{\mathbf{E}}_0 g(X_i)} - \frac{1}{\mathbf{E}g(X_i)}\right\} \tau_0 \mathbf{E}g(X_i) + o_P(n^{-1/2}) \\ &= -\frac{\tilde{\mathbf{E}}_1 g(X_i) - \mathbf{E}g(X_i)}{\mathbf{E}g(X_i)}\tau_1 + \frac{\tilde{\mathbf{E}}_0 g(X_i) - \mathbf{E}g(X_i)}{\mathbf{E}g(X_i)}\tau_0 + o_P(n^{-1/2}) \\ &= -\frac{\tilde{\mathbf{E}}_1 g(X_i)\tau_1 - \tilde{\mathbf{E}}_0 g(X_i)\tau_0}{\mathbf{E}g(X_i)} + \tau_{wate} + o_P(n^{-1/2}).\end{aligned}$$

However, observe that

$$\begin{aligned}&-\frac{\tilde{\mathbf{E}}_1 g(X_i)\tau_1 - \tilde{\mathbf{E}}_0 g(X_i)\tau_0}{\mathbf{E}g(X_i)} \\ &= -\frac{1}{\mathbf{E}\left[g(X_i)\right]} \sum_{w\in\mathcal{W}} \left\{ \frac{p_{1,w}}{n_{1,w}} \sum_{i\in S_{1,w}} \tilde{1}_{n,i} \frac{g(V_i,w)\tau_1}{\tilde{p}_1(V_i,w)} - \frac{p_{0,w}}{n_{0,w}} \sum_{i\in S_{0,w}} \tilde{1}_{n,i} \frac{g(V_i,w)\tau_0}{\tilde{p}_0(V_i,w)} \right\}.\end{aligned} \tag{34}$$

By replacing $Y_i 1\{(D_i,W_i) = (1,w)\}$ by $\tau_1 1\{(D_i,W_i) = (1,w)\}$ and $Y_i 1\{(D_i,W_i) = (0,w)\}$ by $\tau_0 1\{(D_i,W_i) = (0,w)\}$ in Lemma A3(ii) and noting that $\tau_{wate} = \tau_1 - \tau_0$, we find that the last term in (34) is equal to

$$\begin{aligned}&-\frac{\tau_{wate}}{\mathbf{E}g(X_i)} \sum_{w\in\mathcal{W}} \frac{1}{n} \sum_{i=1}^n \left\{g(V_i,w) - \mathbf{E}_{1,w}\left[g(V_i,w)\right]\right\} L_{1,w,i} \\ &-\frac{\tau_{wate}}{\mathbf{E}g(X_i)} \sum_{w\in\mathcal{W}} \frac{1}{n} \sum_{i=1}^n \left\{g(V_i,w) - \mathbf{E}_{0,w}\left[g(V_i,w)\right]\right\} L_{0,w,i} \\ &-\frac{\tau_1 - \tau_0}{\mathbf{E}g(X_i)} \sum_{w\in\mathcal{W}} \left\{\mathbf{E}_{1,w}[g(V_i,w)]p_{1,w} + \mathbf{E}_{0,w}[g(V_i,w)]p_{0,w}\right\} + o_P(n^{-1/2}).\end{aligned}$$

Observe that for the last term,

$$\frac{\tau_1 - \tau_0}{\mathbf{E}g(X_i)} \sum_{w \in \mathcal{W}} \{\mathbf{E}_{1,w}[g(V_i, w)]p_{1,w} + \mathbf{E}_{0,w}[g(V_i, w)]p_{0,w}\} = \tau_{wate}.$$

Therefore, by applying Lemma A3(ii) to the leading term of (33), we conclude that $\tilde{\tau}_{wate} - \tau_{wate}$ is asymptotically equivalent to (up to $o_P(n^{-1/2})$)

$$\begin{aligned}
&\frac{1}{\mathbf{E}g(X_i)} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{g(V_i, w) L_{1,w,i} \varepsilon_{1,w,i}}{p_1(V_i, w)} - \frac{1}{n} \sum_{i=1}^{n} \frac{g(V_i, w) L_{0,w,i} \varepsilon_{0,w,i}}{p_{0,w}(V_i, w)} \right\} \\
&+ \frac{1}{\mathbf{E}g(X_i)} \sum_{w \in \mathcal{W}} \frac{1}{n} \sum_{i=1}^{n} (\zeta_1(V_i, w) L_{1,w,i} + \zeta_0(V_i, w) L_{0,w,i}) \\
&+ \frac{1}{\mathbf{E}g(X_i)} \sum_{w \in \mathcal{W}} \{\mathbf{E}_{1,w}[g(V_i, w)\tau(V_i, w)]p_{1,w} + \mathbf{E}_{0,w}[g(V_i, w)\tau(V_i, w)]p_{0,w}\} - \tau_{wate}.
\end{aligned}$$

The second to the last term is actually τ_{wate} cancelling the last τ_{wate}. The wanted result follows from the Central Limit Theorem.

Proof of Theorem 4 : (i) We first consider $\tilde{\tau}_{atet}$. Let $\mathbf{E}_1[\beta_0(X_i)] = \mathbf{E}[\beta_0(X_i)|D_i = 1]$. Note that

$$\tilde{\tau}_{atet} - \tau_{atet} = \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{p_{1,w}}{n_{1,w}} \sum_{i \in S_{1,w}} Y_i - \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \frac{\tilde{p}_1(V_i, w) Y_i}{\tilde{p}_0(V_i, w)} \right\} + \bar{R}_n - \tau_{atet}, \tag{35}$$

where

$$\bar{R}_n \equiv \left\{ \frac{1}{p_1} - \frac{1}{\sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \tilde{p}_1(V_i, w) / \tilde{p}_0(V_i, w)} \right\} \sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \frac{\tilde{p}_1(V_i, w) Y_i}{\tilde{p}_0(V_i, w)}.$$

Note that $\tau_{atet} = \mathbf{E}_1[\beta_1(X_i)] - \mathbf{E}_1[\beta_0(X_i)]$,

$$\begin{aligned}
\sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \frac{\tilde{p}_1(V_i, w)}{\tilde{p}_0(V_i, w)} &= \sum_{w \in \mathcal{W}} p_{0,w} \mathbf{E}\left[\frac{p_1(V_i, w)}{p_0(V_i, w)} | (D_i, W_i) = (0, w) \right] + O_P(n^{-1/2}) \\
&= \mathbf{E}\left[\frac{p_1(X_i)}{p_0(X_i)} | D_i = 0 \right] p_0 + O_P(n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{w \in \mathcal{W}} \frac{p_{0,w}}{n_{0,w}} \sum_{i \in S_{0,w}} \tilde{1}_{n,i} \frac{\tilde{p}_1(V_i, w) Y_i}{\tilde{p}_0(V_i, w)} &= \sum_{w \in \mathcal{W}} p_{0,w} \mathbf{E}\left[\frac{p_1(V_i, w) Y_i}{p_0(V_i, w)} | (D_i, W_i) = (0, w) \right] + O_P(n^{-1/2}) \\
&= \sum_{w \in \mathcal{W}} p_{0,w} \mathbf{E}\left[\frac{p_1(V_i, w) \beta_0(V_i, w)}{p_0(V_i, w)} | (D_i, W_i) = (0, w) \right] + O_P(n^{-1/2}) \\
&= \mathbf{E}\left[\frac{p_1(X_i) \beta_0(X_i)}{p_0(X_i)} | D_i = 0 \right] p_0 + O_P(n^{-1/2}).
\end{aligned}$$

However, we can simplify

$$\mathbf{E}\left[\frac{p_1(X_i)}{p_0(X_i)}|D_i=0\right]p_0=\mathbf{E}\left[\frac{p_1(X_i)(1-D_i)}{p_0(X_i)}\right]=\mathbf{E}\left[p_1(X_i)\right]=p_1$$

and

$$\begin{aligned}\mathbf{E}\left[\frac{p_1(X_i)\beta_0(X_i)}{p_0(X_i)}|D_i=0\right]p_0 &= \mathbf{E}\left[\frac{p_1(X_i)\beta_0(X_i)(1-D_i)}{p_0(X_i)}\right]=\mathbf{E}\left[p_1(X_i)\beta_0(X_i)\right]\\ &= \mathbf{E}\left[\beta_0(X_i)D_i\right]=\mathbf{E}\left[\beta_0(X_i)|D_i=1\right]p_1=\mathbf{E}_1\left[\beta_0(X_i)\right]p_1.\end{aligned}$$

Hence we can write

$$\begin{aligned}\bar{R}_n &= \frac{1}{p_1}\left\{\sum_{w\in\mathcal{W}}\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{1}_{n,i}\frac{\tilde{p}_1(V_i,w)}{\tilde{p}_0(V_i,w)}-p_1\right\}\mathbf{E}_1\left[\beta_0(X_i)\right]+o_P(n^{-1/2})\\ &= \frac{1}{p_1}\left\{\sum_{w\in\mathcal{W}}\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{1}_{n,i}\frac{\tilde{p}_1(V_i,w)}{\tilde{p}_0(V_i,w)}\mathbf{E}_1\left[\beta_0(X_i)\right]-p_1\mathbf{E}_1\left[\beta_1(X_i)\right]\right\}+\tau_{atet}+o_P(n^{-1/2}).\end{aligned}$$

Plugging this into (35) and defining $\tilde{\varepsilon}_{d,i}=Y_{di}-\mathbf{E}\left[\beta_d(X_i)|D_i=1\right]$, we write

$$\begin{aligned}\tilde{\tau}_{atet}-\tau_{atet} &= \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{1,w}}{n_{1,w}}\sum_{i\in S_{1,w}}\tilde{\varepsilon}_{1,i}-\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{1}_{n,i}\frac{\tilde{p}_1(V_i,w)\tilde{\varepsilon}_{0,i}}{\tilde{p}_0(V_i,w)}\right\}+o_P(n^{-1/2}) \qquad (36)\\ &= \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{1,w}}{n_{1,w}}\sum_{i\in S_{1,w}}\tilde{\varepsilon}_{1,i}-\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}}{p_0(V_i,w)}\right\}\\ &\quad -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{\varepsilon}_{0,i}\left\{\tilde{1}_{n,i}\frac{\tilde{p}_1(V_i,w)}{\tilde{p}_0(V_i,w)}-\hat{1}_{n,i}\frac{\hat{p}_1(V_i,w)}{\hat{p}_0(V_i,w)}\right\}\\ &\quad -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{\varepsilon}_{0,i}\left(\hat{1}_{n,i}\frac{\hat{p}_1(V_i,w)}{\hat{p}_0(V_i,w)}-\frac{p_1(V_i,w)}{p_0(V_i,w)}\right)\right\}+o_P(n^{-1/2})\\ &= B_n-C_n-D_n+o_P(n^{-1/2}),\ \text{say.}\end{aligned}$$

We consider D_n first. Write it as

$$
\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{0,w}}{n_{0,w}}\sum_{i\in S_{0,w}}\tilde{\varepsilon}_{0,i}\left(\frac{\hat{p}_1(V_i,w)p_0(V_i,w)-p_1(V_i,w)\hat{p}_0(V_i,w)}{p_0^2(V_i,w)}\right)\right\}+o_P(n^{-1/2}) \\
= \; & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\tilde{\varepsilon}_{0,i}\left(\frac{\hat{p}_1(V_i,w)-p_1(V_i,w)}{p_0(V_i,w)}\right)\right\} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{p_{0,w}}{q_{0,w}n}\sum_{i\in S_{0,w}}\tilde{\varepsilon}_{0,i}\left(\frac{p_1(V_i,w)\{p_0(V_i,w)-\hat{p}_0(V_i,w)\}}{p_0^2(V_i,w)}\right)\right\}+o_P(n^{-1/2}) \\
= \; & D_{1n}+D_{2n}+o_P(n^{-1/2}), \text{ say.}
\end{aligned}
$$

Apply Lemma A2(i) to write D_{1n} as (up to $o_P(n^{-1/2})$)

$$
\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]\mathcal{J}_{1,w,i}}{p_0(V_i,w)\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]}\right\} \\
& -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]p_1(V_i,w)\mathcal{J}_{w,i}}{p_0(V_i,w)\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]}\right\}.
\end{aligned}
$$

Defining $\Delta_{d,w,i}\equiv\beta_d(V_i,w)-\mathbf{E}\left[\beta_d(X_i)|D_i=1\right]$, we write the last difference as

$$
\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_{0,w,i}\mathcal{J}_{1,w,i}\right\}-\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}p_1(V_i,w)\Delta_{0,w,i}\mathcal{J}_{w,i}\right\},
$$

because similarly as in (28),

$$
\begin{aligned}
\frac{\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]}{\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]} &= p_0(V_i,w)\{\beta_0(V_i,w)-\mathbf{E}_1\left[\beta_0(X_i)\right]\} \\
&= p_0(V_i,w)\Delta_{0,w,i} \text{ and} \\
\frac{\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]p_1(V_i,w)}{p_0(V_i,w)\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]} &= \{\beta_0(V_i,w)-\mathbf{E}_1\left[\beta_0(X_i)\right]\}p_1(V_i,w) \\
&= p_1(V_i,w)\Delta_{0,w,i}.
\end{aligned}
$$

Applying Lemma A2(i), we write D_{2n} as (up to $o_P(n^{-1/2})$)

$$
\begin{aligned}
& -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]}{p_0^2(V_i,w)\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]}\mathcal{J}_{0,w,i} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\mathbf{E}_Q\left[\tilde{\varepsilon}_{0,i}L_{0,w,i}|V_i,W_i=w\right]}{p_0(V_i,w)\mathbf{E}_Q\left[L_{w,i}|V_i,W_i=w\right]}\mathcal{J}_{w,i} \\
= \; & -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}\mathcal{J}_{0,w,i}\right\}+\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}p_1(V_i,w)\Delta_{0,w,i}\mathcal{J}_{w,i}\right\}.
\end{aligned}
$$

Therefore, $D_{1n}+D_{2n}$ is equal to

$$\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \Delta_{0,w,i}\mathcal{J}_{1,w,i}\right\} - \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n p_1(V_i,w)\Delta_{0,w,i}\mathcal{J}_{w,i}\right\} \\
& -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}\mathcal{J}_{0,w,i}\right\} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n p_1(V_i,w)\Delta_{0,w,i}\mathcal{J}_{w,i}\right\} + o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \Delta_{0,w,i}\mathcal{J}_{1,w,i}\right\} - \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}\mathcal{J}_{0,w,i}\right\} + o_P(n^{-1/2}).
\end{aligned}$$

As for the last difference, recall the definition $\mathcal{J}_{d,w,i} \equiv L_{d,w,i} - \mathbf{E}_Q\left[L_{d,w,i}|V_i, W_i = w\right]$ and write it as

$$\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \Delta_{0,w,i}L_{1,w,i}\right\} - \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}L_{0,w,i}\right\} \\
& -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \Delta_{0,w,i}\mathbf{E}_Q\left[L_{1,w,i}|V_i, W_i = w\right]\right\} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}\mathbf{E}_Q\left[L_{0,w,i}|V_i, W_i = w\right]\right\}.
\end{aligned}$$

Note that from (29),

$$\begin{aligned}
& \mathbf{E}_Q\left[L_{1,w,i}|V_i, W_i = w\right] - \frac{p_1(V_i,w)}{p_0(V_i,w)}\mathbf{E}_Q\left[L_{0,w,i}|V_i, W_i = w\right] \qquad (37) \\
= & \frac{p_{1,w}}{q_{1,w}}q_1(V_i,w) - \frac{p_{0,w}}{q_{0,w}}\frac{p_1(V_i,w)q_0(V_i,w)}{p_0(V_i,w)} \\
= & p_1(V_i,w)\frac{f_P(V_i,w)}{f_Q(V_i,w)} - \frac{p_{0,w}}{q_{0,w}}\frac{q_0(V_i,w)}{p_0(V_i,w)} + \frac{p_{0,w}}{q_{0,w}}q_0(V_i,w) \\
= & p_1(V_i,w)\frac{f_P(V_i,w)}{f_Q(V_i,w)} - \frac{f_P(V_i,w)}{f_Q(V_i,w)} + p_0(V_i,w)\frac{f_P(V_i,w)}{f_Q(V_i,w)} = 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
D_n &= D_{1n}+D_{2n} \\
&= \frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \Delta_{0,w,i}L_{1,w,i}\right\} \\
&\quad -\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^n \frac{p_1(V_i,w)\Delta_{0,w,i}}{p_0(V_i,w)}L_{0,w,i}\right\} + o_P(n^{-1/2}).
\end{aligned}$$

Now, we turn to C_n (in (36)) which we write as

$$
\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\left\{\tilde{1}_{n,i}\frac{\tilde{p}_1(V_i,w)}{\tilde{p}_0(V_i,w)}-\hat{1}_{n,i}\frac{\hat{p}_1(V_i,w)}{\hat{p}_0(V_i,w)}\right\}+o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\tilde{p}_1(V_i,w)\hat{p}_0(V_i,w)-\hat{p}_1(V_i,w)\tilde{p}_0(V_i,w)}{p_0^2(V_i,w)}\right\}+o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\tilde{p}_1(V_i,w)\{\hat{p}_0(V_i,w)-\tilde{p}_0(V_i,w)\}}{p_0^2(V_i,w)}\right\} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\{\tilde{p}_1(V_i,w)-\hat{p}_1(V_i,w)\}\tilde{p}_0(V_i,w)}{p_0^2(V_i,w)}\right\}+o_P(n^{-1/2}). \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\tilde{p}_1(V_i,w)\{\hat{p}_0(V_i,w)-\tilde{p}_0(V_i,w)\}}{p_0^2(V_i,w)}\right\} \\
& +\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\{\hat{p}_0(V_i,w)-\tilde{p}_0(V_i,w)\}\tilde{p}_0(V_i,w)}{p_0^2(V_i,w)}\right\}+o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{0,i}L_{0,w,i}\hat{1}_{n,i}\left\{\frac{\hat{p}_0(V_i,w)-\tilde{p}_0(V_i,w)\}}{p_0^2(V_i,w)}\right\}+o_P(n^{-1/2}).
\end{aligned}
$$

As for the last term, we apply Lemma A2(ii) to write it as

$$
\begin{aligned}
& \frac{1}{p_1}\sum_{w\in\mathcal{W}}\mathbf{E}_Q\left[\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}L_{0,w,i}}{p_0(V_i,w)}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}-\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\sum_{w\in\mathcal{W}}\mathbf{E}\left[p_1(V_i,w)\Delta_{0,w,i}\right]\left(\frac{\hat{q}_{0,w}-q_{0,w}}{q_{0,w}}-\frac{\hat{q}_{1,w}-q_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2})
\end{aligned}
$$

because

$$
\begin{aligned}
\mathbf{E}_Q\left[\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}L_{0,w,i}}{p_0(V_i,w)}\right] &= \frac{p_{0,w}}{q_{0,w}}\mathbf{E}_Q\left[\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}1\{(D_i,W_i)=(0,w)\}}{p_0(V_i,w)}\right] \\
&= \mathbf{E}\left[\frac{p_1(V_i,w)}{p_0(V_i,w)}\tilde{\varepsilon}_{0,i}1\{(D_i,W_i)=(0,w)\}\right]=\mathbf{E}\left[p_1(V_i,w)\tilde{\varepsilon}_{0,i}\right] \\
&= \mathbf{E}\left[p_1(V_i,w)\Delta_{0,w,i}\right].
\end{aligned}
$$

Now, let us turn to B_n (in (36)) which we write as

$$
\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}L_{1,w,i}-\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}L_{0,w,i}}{p_0(V_i,w)}\right\}+E_n,
$$

where

$$
E_n\equiv\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}(\hat{L}_{1,w,i}-L_{1,w,i})-\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}(\hat{L}_{0,w,i}-L_{0,w,i})}{p_0(V_i,w)}\right\}.
$$

Now, we focus on E_n. Observe that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}(\hat{L}_{1,w,i}-L_{1,w,i}) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}p_{1,w}\left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}^2}\right)1\{(D_i,W_i)=(1,w)\}+o_P(n^{-1/2})\\ &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}L_{1,w,i}\left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2})\\ &= \mathbf{E}_Q\left[\tilde{\varepsilon}_{1,i}L_{1,w,i}\right]\left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2}).\end{aligned}$$

As for the last expectation,

$$\begin{aligned}\mathbf{E}_Q\left[\tilde{\varepsilon}_{1,i}L_{1,w,i}\right] &= \frac{p_{1,w}}{q_{1,w}}\mathbf{E}_Q\left[\tilde{\varepsilon}_{1,i}1\{(D_i,W_i)=(1,w)\}\right]=\mathbf{E}\left[\tilde{\varepsilon}_{1,i}1\{(D_i,W_i)=(1,w)\}\right]\\ &= \mathbf{E}\left[p_1(V_i,w)\left(\beta_1(V_i,w)-\mathbf{E}[\beta_1(X_i)|D_i=1]\right)\right]=\mathbf{E}\left[p_1(V_i,w)\Delta_{1,w,i}\right].\end{aligned}$$

Hence

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_{1,i}(\hat{L}_{1,w,i}-L_{1,w,i})=\mathbf{E}\left[p_1(V_i,w)\Delta_{1,w,i}\right]\left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)+o_P(n^{-1/2}).$$

Also,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}(\hat{L}_{0,w,i}-L_{0,w,i})}{p_0(V_i,w)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)\tilde{\varepsilon}_{0,i}}{p_0(V_i,w)}\left(\frac{q_{0,w}-\hat{q}_{0,w}}{q_{0,w}}\right)L_{0,w,i}\\ &= \mathbf{E}\left[p_1(V_i,w)\Delta_{0,w,i}\right]\left(\frac{q_{0,w}-\hat{q}_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2}).\end{aligned}$$

Therefore, we write E_n as

$$\begin{aligned}&\frac{1}{p_1}\sum_{w\in\mathcal{W}}\mathbf{E}\left[p_1(V_i,w)\Delta_{1,w,i}\right]\left(\frac{q_{1,w}-\hat{q}_{1,w}}{q_{1,w}}\right)\\ &-\frac{1}{p_1}\sum_{w\in\mathcal{W}}\mathbf{E}\left[p_1(V_i,w)\Delta_{0,w,i}\right]\left(\frac{q_{0,w}-\hat{q}_{0,w}}{q_{0,w}}\right)+o_P(n^{-1/2}).\end{aligned}$$

Now, let us collect all the results for B_n, C_n, and D_n and plug these into (36) to deduce that

$$\begin{aligned}
&\tilde{\tau}_{atet} - \tau_{atet} \\
= \quad & \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \tilde{\varepsilon}_{1,i} L_{1,w,i} - \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w) \tilde{\varepsilon}_{0,i} L_{0,w,i}}{p_0(V_i, w)} \right\} \\
& + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \mathbf{E}\left[p_1(V_i, w) \Delta_{1,w,i} \right] \left(\frac{q_{1,w} - \hat{q}_{1,w}}{q_{1,w}} \right) - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \mathbf{E}\left[p_1(V_i, w) \Delta_{0,w,i} \right] \left(\frac{q_{0,w} - \hat{q}_{0,w}}{q_{0,w}} \right) \\
& - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \mathbf{E}\left[p_1(V_i, w) \Delta_{0,w,i} \right] \left(\frac{\hat{q}_{0,w} - q_{0,w}}{q_{0,w}} - \frac{\hat{q}_{1,w} - q_{1,w}}{q_{1,w}} \right) \\
& - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Delta_{0,w,i} L_{1,w,i} \right\} + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w) \Delta_{0,w,i}}{p_0(V_i, w)} L_{0,w,i} \right\} + o_P(n^{-1/2})
\end{aligned}$$

By consolidating the second, third, and fourth terms, we rewrite

$$\begin{aligned}
\tilde{\tau}_{atet} - \tau_{atet} \quad = \quad & \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \tilde{\varepsilon}_{1,i} L_{1,w,i} - \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w) \tilde{\varepsilon}_{0,i} L_{0,w,i}}{p_0(V_i, w)} \right\} \\
& + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \mathbf{E}\left[p_1(V_i, w) \left(\tau(V_i, w) - \mathbf{E}[\tau(X_i) | D_i = 1] \right) \right] \left(\frac{q_{1,w} - \hat{q}_{1,w}}{q_{1,w}} \right) \\
& - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Delta_{0,w,i} L_{1,w,i} \right\} + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w) \Delta_{0,w,i}}{p_0(V_i, w)} L_{0,w,i} \right\} + o_P(n^{-1/2}).
\end{aligned}$$

By writing $\tilde{\varepsilon}_{d,i} = Y_{di} - \beta_d(X_i) + \beta_d(X_i) - \mathbf{E}_1\left[\beta_1(X_i)\right]$ and splitting the sums, we rewrite

$$\begin{aligned}
&\tilde{\tau}_{atet} - \tau_{atet} \\
= \quad & \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_{1i} - \beta_1(X_i)) L_{1,w,i} - \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w)(Y_{0i} - \beta_0(X_i)) L_{0,w,i}}{p_0(V_i, w)} \right\} \\
& + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\beta_1(X_i) - \mathbf{E}_1\left[\beta_1(X_i)\right]) L_{1,w,i} \right\} \\
& - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w)(\beta_0(X_i) - \mathbf{E}_1\left[\beta_0(X_i)\right]) L_{0,w,i}}{p_0(V_i, w)} \right\} \\
& + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \mathbf{E}\left[p_1(V_i, w) \left(\tau(V_i, w) - \mathbf{E}[\tau(X_i) | D_i = 1] \right) \right] \left(\frac{q_{1,w} - \hat{q}_{1,w}}{q_{1,w}} \right) \\
& - \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \{\beta_0(V_i, w) - \mathbf{E}_1\left[\beta_0(X_i)\right]\} L_{1,w,i} \right\} \\
& + \frac{1}{p_1} \sum_{w \in \mathcal{W}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{p_1(V_i, w)\{\beta_0(V_i, w) - \mathbf{E}_1\left[\beta_0(X_i)\right]\}}{p_0(V_i, w)} L_{0,w,i} \right\} + o_P(n^{-1/2}).
\end{aligned}$$

Note that

$$\mathbf{E}\left[p_1(V_i,w)\left(\tau(V_i,w)-\mathbf{E}[\tau(X_i)|D_i=1]\right)\right]=\mathbf{E}_{1,w}\left[\tau(V_i,w)-\mathbf{E}[\tau(X_i)|D_i=1]\right]p_{1,w}.$$

Using this and noting that $\mathbf{E}\left[\tau(X_i)|D_i=1\right]=\tau_{atet}$ and cancelling out some terms, we rewrite

$$\begin{aligned}
&\tilde{\tau}_{atet}-\tau_{atet}\\
=\;&\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}(Y_{1i}-\beta_1(X_i))L_{1,w,i}-\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)(Y_{0i}-\beta_0(X_i))L_{0,w,i}}{p_0(V_i,w)}\right\}\\
&+\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}(\tau(V_i,w)-\tau_{atet})L_{1,w,i}\\
&+\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{p_{1,w}}{nq_{1,w}}\mathbf{E}_{1,w}\left[\tau(V_i,w)-\tau_{atet}\right](q_{1,w}-\hat{q}_{1,w})+o_P(n^{-1/2})\\
=\;&\frac{1}{p_1}\sum_{w\in\mathcal{W}}\left\{\frac{1}{n}\sum_{i=1}^{n}(Y_{1i}-\beta_1(X_i))L_{1,w,i}-\frac{1}{n}\sum_{i=1}^{n}\frac{p_1(V_i,w)(Y_{0i}-\beta_0(X_i))L_{0,w,i}}{p_0(V_i,w)}\right\}\\
&+\frac{1}{p_1}\sum_{w\in\mathcal{W}}\frac{1}{n}\sum_{i=1}^{n}(\tau(V_i,w)-\tau_{atet}-\mathbf{E}_{d,w}\left[\tau(V_i,w)-\tau_{atet}\right])L_{1,w,i}+o_P(n^{-1/2}).
\end{aligned}$$

The last equality follows because

$$\begin{aligned}
\frac{1}{p_1}\sum_{w\in\mathcal{W}}p_{1,w}\mathbf{E}_{1,w}\left[\tau(V_i,w)-\tau_{atet}\right] &= \frac{1}{p_1}\mathbf{E}\left[\left(\tau(X_i)-\tau_{atet}\right)1\{D_i=1\}\right]\\
&= \mathbf{E}\left[\tau(X_i)-\tau_{atet}|D_i=1\right]=0.
\end{aligned}$$

Hence the wanted result follows by the Central Limit Theorem.

(ii) The case of $\tilde{\tau}_{atet,p}$ is a special case of $\tilde{\tau}_{atet}$ with $W_i=1$ for all $i=1,\cdots,n$. Hence we focus on $\hat{\tau}_{atet,p}$. We write it as

$$\hat{\tau}_{atet,p}-\tau_{atet}=\frac{1}{p_1}\left\{\frac{p_1}{q_1 n}\sum_{i\in S_1}Y_i-\frac{p_0}{q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\frac{\hat{p}_1(X_i)Y_i}{\hat{p}_0(X_i)}\right\}+R_n-\tau_{atet}+o_P(n^{-1/2}),$$

where $L_{d,i}=\sum_{d\in\mathcal{D}}1\{D_i=d\}p_d/q_d$ and

$$\begin{aligned}
R_n &\equiv \left\{\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\hat{1}_{n,i}\hat{p}_1(X_i)\left\{L_{1,i}+L_{0,i}\right\}}-\frac{1}{p_1}\right\}p_1\tau_{atet}+o_P(n^{-1/2})\\
&= \left\{p_1-\frac{1}{n}\sum_{i=1}^{n}\hat{1}_{n,i}\hat{p}_1(X_i)\left\{L_{1,i}+L_{0,i}\right\}\right\}\frac{\tau_{atet}}{p_1}+o_P(n^{-1/2}).
\end{aligned}$$

Hence we can write $\hat{\tau}_{atet,p} - \tau_{atet}$ as

$$\begin{aligned}
& \frac{1}{p_1}\left\{\frac{p_1}{q_1 n}\sum_{i\in S_1} Y_i - \frac{p_0}{q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\frac{\hat{p}_1(X_i)Y_i}{\hat{p}_0(X_i)}\right\} - \frac{\tau_{atet}}{np_1}\sum_{i=1}^{n}\hat{1}_{n,i}\hat{p}_1(X_i)\left\{L_{1,i}+L_{0,i}\right\} + o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\left\{\frac{p_1}{q_1 n}\sum_{i\in S_1}\{Y_i - \tau_{atet}\hat{1}_{n,i}\hat{p}_1(X_i)\} - \frac{p_0}{q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\hat{p}_1(X_i)\left\{\frac{Y_i}{\hat{p}_0(X_i)} + \tau_{atet}\right\}\right\} + o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\left\{\frac{p_1}{q_1 n}\sum_{i\in S_1}\{Y_i - \tau_{atet}p_1(X_i)\} - \frac{p_0}{q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\hat{p}_1(X_i)\left\{\frac{Y_i}{p_0(X_i)} + \tau_{atet}\right\}\right\} \\
& + \frac{1}{p_1}\left\{\frac{p_1\tau_{atet}}{q_1 n}\sum_{i\in S_1}\hat{1}_{n,i}\{p_1(X_i) - \hat{p}_1(X_i)\} + \frac{p_0}{q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\hat{p}_1(X_i)Y_i\left\{\frac{1}{p_0(X_i)} - \frac{1}{\hat{p}_0(X_i)}\right\}\right\} + o_P(n^{-1/2}) \\
= & \frac{1}{p_1}\left\{\frac{p_1}{q_1 n}\sum_{i\in S_1}\{Y_i - \tau_{atet}p_1(X_i)\} - \frac{p_0}{q_0 n}\sum_{i\in S_0}p_1(X_i)\left\{\frac{Y_i}{p_0(X_i)} + \tau_{atet}\right\}\right\} + F_n + G_n + o_P(n^{-1/2}),
\end{aligned}$$

where

$$\begin{aligned}
F_n & \equiv \frac{p_0}{p_1 q_0 n}\sum_{i\in S_0}\hat{1}_{n,i}\{p_1(X_i) - \hat{p}_1(X_i)\}\left\{\frac{Y_i}{p_0(X_i)} + \tau_{atet}\right\} \text{ and} \\
G_n & \equiv \frac{1}{p_1}\left\{\frac{p_1\tau_{atet}}{q_1 n}\sum_{i\in S_1}\hat{1}_{n,i}\{p_1(X_i) - \hat{p}_1(X_i)\} + \frac{p_0}{q_0 n}\sum_{i\in S_0}\frac{p_1(X_i)}{p_0(X_i)^2}Y_i\hat{1}_{n,i}\left\{\hat{p}_0(X_i) - p_0(X_i)\right\}\right\}.
\end{aligned}$$

By applying Lemma A2(i), we write F_n as

$$\begin{aligned}
& -\frac{1}{p_1 n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q\left[\left(Y_i/p_0(X_i) + \tau_{atet}\right)L_{0,i}|X_i\right]}{\mathbf{E}_Q\left[L_i|X_i\right]}\mathcal{J}_{1,i} \\
& + \frac{1}{p_1 n}\sum_{i=1}^{n}\frac{\mathbf{E}_Q\left[\left(Y_i/p_0(X_i) + \tau_{atet}\right)L_{0,i}|X_i\right]p_1(X_i)}{\mathbf{E}_Q\left[L_i|X_i\right]}\mathcal{J}_i + o_P(n^{-1/2}) \\
= & -\frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)\mathcal{J}_{1,i} \\
& + \frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)p_1(X_i)\mathcal{J}_i + o_P(n^{-1/2}) \\
= & -\frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)p_0(X_i)\mathcal{J}_{1,i} \\
& + \frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)p_1(X_i)\mathcal{J}_{0,i} + o_P(n^{-1/2}).
\end{aligned}$$

Again by applying Lemma A2(i), we write G_n as

$$\begin{aligned}
& -\frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)\mathcal{J}_{1,i} + \frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)^2\mathcal{J}_i \\
& +\frac{1}{p_1 n}\sum_{i=1}^{n}\frac{p_1(X_i)\beta_0(X_i)}{p_0(X_i)}\mathcal{J}_{0,i} \\
& -\frac{1}{p_1 n}\sum_{i=1}^{n} p_1(X_i)\beta_0(X_i)\mathcal{J}_i + o_P(n^{-1/2}) \\
= \quad & -\frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)p_0(X_i)\mathcal{J}_{1,i} + \frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)^2\mathcal{J}_{0,i} \\
& +\frac{1}{p_1 n}\sum_{i=1}^{n}\frac{p_1^2(X_i)\beta_0(X_i)}{p_0(X_i)}\mathcal{J}_{0,i} - \frac{1}{p_1 n}\sum_{i=1}^{n} p_1(X_i)\beta_0(X_i)\mathcal{J}_{1,i} + o_P(n^{-1/2}).
\end{aligned}$$

Collecting these results, we write

$$\begin{aligned}
\hat{\tau}_{atet,p} - \tau_{atet,p} \quad = \quad & \frac{1}{p_1}\left\{\frac{1}{n}\sum_{i=1}^{n} L_{1,i}\{Y_i - \tau_{atet}p_1(X_i)\} - \frac{1}{n}\sum_{i=1}^{n} L_{0,i}p_1(X_i)\left\{\frac{Y_i}{p_0(X_i)} + \tau_{atet}\right\}\right\} \\
& -\frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)p_0(X_i)\mathcal{J}_{1,i} \\
& +\frac{1}{p_1 n}\sum_{i=1}^{n}\left(\beta_0(X_i) + \tau_{atet}p_0(X_i)\right)p_1(X_i)\mathcal{J}_{0,i} \\
& -\frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)p_0(X_i)\mathcal{J}_{1,i} + \frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n} p_1(X_i)^2\mathcal{J}_{0,i} \\
& +\frac{1}{p_1 n}\sum_{i=1}^{n}\frac{p_1^2(X_i)\beta_0(X_i)}{p_0(X_i)}\mathcal{J}_{0,i} - \frac{1}{p_1 n}\sum_{i=1}^{n} p_1(X_i)\beta_0(X_i)\mathcal{J}_{1,i} + o_P(n^{-1/2}).
\end{aligned}$$

By rearranging the terms, we rewrite $\hat{\tau}_{atet,p} - \tau_{atet,p}$ as

$$
\begin{aligned}
&\frac{1}{p_1}\left\{\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\beta_1(X_i)\}L_{1,i}-\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{Y_i-\beta_0(X_i)\}p_1(X_i)}{p_0(X_i)}\right\}L_{0,i}\right\}\\
&+\frac{1}{p_1}\frac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\beta_1(X_i)-\tau_{atet}p_1(X_i)-(\beta_0(X_i)+\tau_{atet}p_0(X_i))\,p_0(X_i)\\ -\tau_{atet}p_1(X_i)p_0(X_i)-p_1(X_i)\beta_0(X_i)\end{array}\right\}L_{1,i}\\
&-\frac{1}{p_1}\frac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\beta_0(X_i)p_1(X_i)/p_0(X_i)+\tau_{atet}p_1(X_i)-(\beta_0(X_i)+\tau_{atet}p_0(X_i))\,p_1(X_i)\\ -\tau_{atet}p_1(X_i)^2-p_1^2(X_i)\beta_0(X_i)/p_0(X_i)\end{array}\right\}L_{0,i}\\
&+\frac{1}{p_1 n}\sum_{i=1}^{n}(\beta_0(X_i)+\tau_{atet}p_0(X_i))\,\frac{p_1}{q_1}p_0(X_i)q_1(X_i)\\
&-\frac{1}{p_1 n}\sum_{i=1}^{n}(\beta_0(X_i)+\tau_{atet}p_0(X_i))\,\frac{p_0}{q_0}p_1(X_i)q_0(X_i)\\
&+\frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n}p_1(X_i)p_0(X_i)\frac{p_1}{q_1}q_1(X_i)-\frac{\tau_{atet}}{p_1 n}\sum_{i=1}^{n}p_1(X_i)^2\frac{p_0}{q_0}q_0(X_i)\\
&-\frac{1}{p_1 n}\sum_{i=1}^{n}\frac{p_1^2(X_i)\beta_0(X_i)}{p_0(X_i)}\frac{p_0}{q_0}q_0(X_i)+\frac{1}{p_1 n}\sum_{i=1}^{n}p_1(X_i)\beta_0(X_i)\frac{p_1}{q_1}q_1(X_i)+o_P(n^{-1/2}).
\end{aligned}
$$

Or by cancelling terms out,

$$
\begin{aligned}
\hat{\tau}_{atet,p}-\tau_{atet,p} \quad = \quad &\frac{1}{p_1}\left\{\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\beta_1(X_i)\}L_{1,i}-\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{Y_i-\beta_0(X_i)\}p_1(X_i)}{p_0(X_i)}\right\}L_{0,i}\right\}\\
&+\frac{1}{p_1}\frac{1}{n}\sum_{i=1}^{n}L_{1,i}\left\{\tau(X_i)-\tau_{atet}\right\}\\
&+\frac{1}{p_1 n}\sum_{i=1}^{n}(\beta_0(X_i)+\tau_{atet}p_0(X_i))\,\frac{p_1}{q_1}q_1(X_i)\\
&-\frac{1}{p_1 n}\sum_{i=1}^{n}\left(\frac{\beta_0(X_i)p_1(X_i)}{p_0(X_i)}+\tau_{atet}p_1(X_i)\right)\frac{p_0}{q_0}q_0(X_i).
\end{aligned}
$$

We rearrange the terms to write

$$\begin{aligned}
\hat{\tau}_{atet,p} - \tau_{atet} &= \frac{1}{p_1}\left\{\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \beta_1(X_i)\}L_{1,i} - \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\{Y_i - \beta_0(X_i)\}p_1(X_i)}{p_0(X_i)}\right\}L_{0,i}\right\} \\
&+\frac{1}{p_1}\frac{1}{n}\sum_{i=1}^{n}L_{1,i}\left\{\tau(X_i) - \tau_{atet}\right\} \\
&+\frac{1}{p_1 n}\sum_{i=1}^{n}\beta_0(X_i)\left(\frac{p_1}{q_1}q_1(X_i) - \frac{p_0}{q_0}\frac{p_1(X_i)}{p_0(X_i)}q_0(X_i)\right) \\
&+\frac{1}{p_1 n}\sum_{i=1}^{n}\tau_{atet}\left(p_0(X_i)\frac{p_1}{q_1}q_1(X_i) - \frac{p_0}{q_0}q_0(X_i)p_1(X_i)\right) \\
&= \frac{1}{p_1}\left\{\frac{1}{n}\sum_{i=1}^{n}L_{1,i}\{Y_i - \beta_1(X_i)\} - \frac{1}{n}\sum_{i=1}^{n}L_{0,i}\left\{\frac{\{Y_i - \beta_0(X_i)\}p_1(X_i)}{p_0(X_i)}\right\}\right\} \\
&+\frac{1}{p_1}\frac{1}{n}\sum_{i=1}^{n}L_{1,i}\left\{\tau(X_i) - \tau_{atet}\right\}.
\end{aligned}$$

The last equality follows because $p_0(X_i)q_1(X_i)p_1/q_1 = p_1(X_i)q_0(X_i)p_0/q_0$ (e.g. see (37)). The wanted result follows by the Central Limit Theorem. ■

The following lemma is used to prove Lemma A2(i) and useful for other purposes. Hence we make the notations and assumptions self-contained here. Let $(Z_i, H_i, X_i)_{i=1}^n$ be an i.i.d. sample from P, where Z_i and H_i are random variables. Let $X_i = (X_{1i}, X_{2i}) \in \mathbf{R}^{L_1+L_2}$ where X_{1i} is continuous and X_{2i} is discrete, and let $K_{ji} = K_h(X_{1j} - X_{1i})1\{X_{2j} = X_{2i}\}$, $K_h(\cdot) = K(\cdot/h)/h^{L_1}$. Let $\mathcal{X}$ be the support of X_i and $f(\cdot)$ be its density with respect to a σ-finite measure.

Assumption B1 : (i) For some $\sigma \geq 4$, $\sup_{x\in\mathcal{X}}||x_1||^{L_1}\mathbf{E}[|Z_i|^\sigma|X_i = (x_1,x_2)] < \infty$, $\mathbf{E}[|H_i|^\sigma] < \infty$, and $\mathbf{E}||X_i||^\sigma < \infty$.
(ii) $f(\cdot,x_2)$, $\mathbf{E}[Z_i|X_{1i} = \cdot, X_{2i} = x_2]f(\cdot,x_2)$ and $\mathbf{E}[H_i|X_{1i} = \cdot, X_{2i} = x_2]f(\cdot,x_2)$ are L_1+1 times continuously differentiable with bounded derivatives on $\mathbf{R}^{L_1}$ and their (L_1+1)-th derivatives are uniformly continuous.
(iii) $\mathbf{E}\left[f^{-\bar{a}}(X_{1i})\right] < \infty$ for some $\bar{a} \geq 4$.

Assumption B2 : For the kernel K and the bandwidth h, Assumption 3 holds.

Lemma B1 : *Suppose that* Assumptions B1-B2 *hold. Let* $1_{n,i}^* = \{(n-1)^{-1}\sum_{j=1,j\neq i}^n K_{ji} \geq \delta_n\}$. *Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i\left\{\mathbf{E}[Z_i|X_i] - \frac{1_{n,i}^*\sum_{j=1,j\neq i}^n Z_jK_{ji}}{\sum_{j=1,j\neq i}^n K_{ji}}\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[H_i|X_i\right]\left\{\mathbf{E}[Z_i|X_i] - Z_i\right\} + o_P(1).$$

Proof of Lemma B1 : For simplicity, we only prove the result for the case where $X_i = X_{1,i}$ so that X_i is continuous. Let $\hat{f}_j(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n K_h(X_j - x_1)$ and define $1_{n,i} = 1\{f(X_i) \geq \delta_n\}$.

Observe that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i\left\{\mathbf{E}[Z_i|X_i]-\frac{1_{n,i}^*\sum_{j=1,j\neq i}^{n}Z_jK_{ji}}{\sum_{j=1,j\neq i}^{n}K_{ji}}\right\}\\
=\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i\mathbf{E}[Z_i|X_i]\left\{1-\frac{1_{n,i}^*\hat{f}_i(X_i)}{f(X_i)}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\left\{\mathbf{E}[Z_i|X_i]\frac{\hat{f}_i(X_i)}{f(X_i)}-\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}Z_jK_{ji}}{f(X_i)}\right\}\\
&+\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\sum_{j=1,j\neq i}^{n}Z_jK_{ji}\left\{\frac{1}{f(X_i)}-\frac{1}{\hat{f}_i(X_i)}\right\}\\
=\;&A_{1n}+A_{2n}+A_{3n}\text{, say.}
\end{aligned}$$

As for A_{3n}, we write it as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]f(X_i)\left\{\frac{1}{f(X_i)}-\frac{1}{\hat{f}_i(X_i)}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\left\{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}Z_jK_{ji}-\mathbf{E}\left[Z_i|X_i\right]f(X_i)\right\}\left\{\frac{1}{f(X_i)}-\frac{1}{\hat{f}_i(X_i)}\right\}\\
=\;&B_{1n}+B_{2n}\text{, say.}
\end{aligned}$$

As for B_{1n},

$$\begin{aligned}
B_{1n}\;=\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]\left\{\frac{\hat{f}_i(X_i)-f(X_i)}{\hat{f}_i(X_i)}\right\}\\
=\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]\left\{\frac{\hat{f}_i(X_i)-f(X_i)}{f_i(X_i)}\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]\left\{\hat{f}_i(X_i)-f(X_i)\right\}\left\{\frac{1}{f(X_i)}-\frac{1}{\hat{f}_i(X_i)}\right\}\\
=\;&C_{1n}+C_{2n},\text{ say.}
\end{aligned}$$

As for C_{2n},

$$\begin{aligned}
|C_{2n}|\;\leq\;&\frac{1}{\delta_n\sqrt{n}}\sum_{i=1}^{n}\left|H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]\left\{\hat{f}_i(X_i)-f(X_i)\right\}^2\frac{1}{f(X_i)}\right|\\
\leq\;&\frac{\sup_{x\in\mathbf{R}^{L_1}}|\hat{f}(x)-f(x)|^2}{\delta_n\sqrt{n}}\sum_{i=1}^{n}\left|\frac{H_i1_{n,i}^*\mathbf{E}\left[Z_i|X_i\right]}{f(X_i)}\right|.
\end{aligned}$$

Since $\mathbf{E}\left[f(X_i)^{-4}\right]<\infty$, we find that $\mathbf{E}\left|H_i\mathbf{E}\left[Z_i|X_i\right]/f(X_i)\right|<\infty$ by Cauchy-Schwartz inequality.

By Theorem 6 of Hansen (2008), we have

$$\sup_{x \in \mathbf{R}^{L_1}} |\hat{f}(x) - f(x)| = O_P(\varepsilon_n).$$

Therefore,

$$|C_{2n}| = \sqrt{n}\delta_n^{-1} \times O_P\left(\varepsilon_n^2\right) = o_P(1).$$

We turn to C_{1n}. We write it as

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i \mathbf{E}\left[Z_i|X_i\right] \left\{ \frac{1_{n,i}^* \hat{f}_i(X_i) - f(X_i)}{f_i(X_i)} \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i \mathbf{E}\left[Z_i|X_i\right] \left(1 - 1_{n,i}^*\right) \\
= \;& D_{1n} + D_{2n}, \text{ say.}
\end{aligned}$$

Note that

$$\begin{aligned}
D_{2n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i \mathbf{E}\left[Z_i|X_i\right] 1_{n,i} \left(1 - 1_{n,i}^*\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i \mathbf{E}\left[Z_i|X_i\right] \left(1 - 1_{n,i}\right) \left(1 - 1_{n,i}^*\right) \\
&= E_{1n} + E_{2n}, \text{ say.}
\end{aligned}$$

As for E_{2n},

$$\mathbf{E}\left[|E_{2n}|\right] \leq C\sqrt{n}\mathbf{E}\left[|1 - 1_{n,i}|\right] \leq C\sqrt{n}\delta_n^{\bar{a}} \mathbf{E}\left[f_i(X_i)^{-\bar{a}}\right] \to 0,$$

so that $D_{2n} = E_{1n} + o_P(1)$. As for E_{1n}, note that

$$\begin{aligned}
\mathbf{E}\left[|E_{1n}|\right] &\leq C\sqrt{n}\mathbf{E}\left[1_{n,i}\left|1 - 1_{n,i}^*\right|\right] = C\sqrt{n}P\left\{\hat{f}_i(X_i) \leq \delta_n \text{ and } f(X_i) \geq \delta_n\right\} \\
&\leq C\sqrt{n}P\left\{\delta_n \leq f_i(X_i) \leq \delta_n + v_n\right\} + o(1),
\end{aligned}$$

where $v_n \to 0$ such that $\sqrt{n}v_n^{\bar{a}} \to 0$ and $v_n/\varepsilon_n \to \infty$. (If we take $v_n = \delta_n$, then the latter condition is satisfied by Assumption 3(ii).) The last inequality follows by Theorem 6 of Hansen (2008). The last quantity is bounded by

$$C\sqrt{n}(\delta_n + v_n)^{\bar{a}} \mathbf{E}\left[f_i(X_i)^{-\bar{a}}\right] \to 0.$$

Hence $D_{2n} = o_P(1)$. Therefore, we can write

$$B_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i \mathbf{E}\left[Z_i|X_i\right] \left\{ \frac{1_{n,i}^* \hat{f}_i(X_i) - f(X_i)}{f_i(X_i)} \right\} + o_P(1).$$

Following previous steps, it is not hard to show that $B_{2n} = o_P(1)$. Hence we conclude that

$$A_{1n} + A_{3n} = o_P(1).$$

We are left with A_{2n}. It remains to show that

$$A_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{E}\left[H_i|X_i\right] \left\{\mathbf{E}[Z_i|X_i] - Z_i\right\} + o_P(1).$$

First we write

$$\begin{aligned} A_{2n} &= \frac{1}{(n-1)\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} q_h(S_i, S_j) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{E}\left[q_h(S_i, S_j)|S_j\right] + r_n \end{aligned}$$

where $q_h(S_i, S_j) = H_i \left\{\mathbf{E}[Z_i|X_i] - Z_j\right\} K_h(X_j - X_i)/f(X_i)$ and $S_i = (X_i, Z_i, H_i)$, and

$$r_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{q_h(S_i, S_j) - \mathbf{E}\left[q_h(S_i, S_j)|S_j\right]\}.$$

Observe that

$$\begin{aligned} n^{-1}\mathbf{E}\left(q_h(S_i, S_j)^2\right) &= n^{-1}\mathbf{E}\left[H_i^2 \left\{\mathbf{E}[Z_i|X_i] - Z_j\right\}^2 \left\{K_h(X_j - X_i)\right\}^2 / f^2(X_i)\right] \\ &\leq n^{-1}\sqrt{\mathbf{E}\left[\left\{K_h(X_j - X_i)\right\}^4\right]} = O(n^{-1}h^{-2L_1}) = o(1) \end{aligned}$$

by change of variables and by Assumption B2. Therefore, by Lemma 3.1 of Powell, Stock, and Stoker (1989), $r_n = o_P(1)$. As for $\mathbf{E}\left[q_h(S_i, S_j)|S_j\right]$, we use change of variables, Taylor expansion, and deduce that

$$\mathbf{E}\left[\left|\mathbf{E}\left[q_h(S_i, S_j)|S_j\right] - \mathbf{E}[H_j|X_j] \left\{\mathbf{E}[Z_j|X_j] - Z_j\right\}\right|\right] = o(n^{-1/2}).$$

The wanted result follows from this. ■

References

[1] Ashenfelter, O. and D. Card (1985) Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics*, **67**, 648-660.

[2] Andrews, D. W. K. (1995) Nonparametric kernel estimation for semiparametric models, *Econometric Theory* 11, 560-596.

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

[4] Bickel, P. and J. Kwon (2001), "Inference for semiparametric models: some questions and an answer," *Statistica Sinica*, 11, 863-960.

[5] Breslow, N., B. McNeney, and J. A. Wellner (2003) Large sample theory for semiparametric regression models with two-phase, outcome-dependent sampling. *Annals of Statistics*, **31**, 1110-1139.

[6] Chen, X., H. Hong, and A. Tarozzi (2008) Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics*, **36**, 808-843.

[7] Cosslett, S. R., (1981a) Efficient estimation of discrete choice models. In C. F. Manski and D. McFadden, eds. *Structural Analysis of Discrete Choice Data with Economic Applications*, MIT Press, Cambridge, Massachusetts.

[8] Cosslett, S. R., (1981b) Maximum likelihood estimator for choice-based samples. *Econometrica*, **49**, 1289-1316.

[9] Dehejia, R. and S. Wahba (1998) Propensity score matching methods for nonexperimental causal studies. NBER Working Paper No. 6829.

[10] Dehejia, R. and S. Wahba (1999) Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, **94**, 1053-1062.

[11] Hahn, J. (1998) On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, **66**, 315-331.

[12] Hahn, J., K. Hirano, and D. Karlan (2008) Adaptive experimental design using the propensity score. Forthcoming in *Journal of Business, Economics, and Statistics.*

[13] Hansen, B. E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**, 726-748.

[14] Heckman, J. J., Ichimura, H. and P. Todd (1997) Matching as an econometric evaluation estimator : evidence from evaluating a job training programme. *Review of Economic Studies*, **64**, 605-654.

[15] Heckman, J. J. and P. E. Todd (2008) A note on adapting propensity score matching and selection models to choice based samples. Forthcoming in *Econometrics Journal.*

[16] Hirano, K., G. Imbens, and G. Ridder, (2003) Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, **71**, 1161-1189.

[17] Imbens, G. (1992) An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica*, **60**, 1187-1214.

[18] Imbens, G. and T. Lancaster (1996) Efficient estimation and stratified sampling. *Journal of Econometrics*, **74**, 289-318.

[19] Khan, S. and E. Tamer (2009), Irregular identification, support conditions, and inverse weight estimation. Working Paper.

[20] Lalonde, R. (1986) Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, **76**, 604-620.

[21] Manski, C. F. (1986) Semiparametric analysis of binary response from response-based samples. *Journal of Econometrics*, **31**, 31-40.

[22] Manski, C. F. and D. McFadden (1981) Alternative estimators and sample design for discrete choice analysis. In C. F. Manski and D. McFadden, eds. *Structural Analysis of Discrete Choice Data with Economic Applications*, MIT Press, Cambridge, Massachusetts.

[23] Manski, C. F. and S. Lerman (1977) The estimation of choice probabilities from choice-based samples, *Econometrica,* **45**, 1977-1988.

[24] Newey, W. K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics*, **5**, 99-135.

[25] Newey, W. K. (1994) The asymptotic variance of semiparametric estimators. *Econometrica*, **62**, 1349-1382.

[26] Powell, J., Stock, J. and T. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica*, **57**, 1403-1430.

[27] Rosenbaum, P. and D. Rubin (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika,* **70**, 41-55.

[28] Smith, J. A. and P. E. Todd (2005) Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, **125**, 305-353.

[29] Song, K. (2009) Two-step extremum estimation with estimated single-indices. PIER Workin Paper 09-012.

[30] Tripathi, G. (2008) Moment-based inference with stratified data. Forthcoming in *Econometric Theory.*

[31] van der Vaart, A. W. (1998) *Asymptotic Statistics*. Cambridge University Press. New York.

[32] van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

[33] Wooldridge, J. M. (1999) Asymptotic properties of weighted M-estimators for variable probability samples. *Econometrica*, **67**, 1385-1406.

[34] Wooldridge, J. M. (2001) Asymptotic properties of weighted M-estimators for standard stratified samples. *Econometric Theory*, **17**, 451-470