Model-assisted estimation through random forests in finite population sampling
Supplementary material

Mehdi DAGDOUG[(a)], Camelia GOGA[(a)] and David HAZIZA[(b)]
(a) Université de Bourgogne Franche-Comté,
Laboratoire de Mathématiques de Besançon, Besançon, FRANCE
(b) University of Ottawa, Department of mathematics and statistics,
Ottawa, CANADA

September 17, 2021

1 Asymptotic assumptions

Assumptions: population-based RF model-assisted estimator $\hat{t}^*_{rf}$

To establish the properties of the proposed estimators, we will consider three categories of assumptions: assumptions on the sampling design, assumptions on the survey variable and, finally, assumptions on the random forests.

(H1) We assume that there exists a positive constant C such that $\sup_{k\in U_v}|y_k| \leqslant C < \infty$.

(H2) We assume that $\lim\limits_{v\to\infty}\dfrac{n_v}{N_v} = \pi \in (0;1)$.

(H3) There exist positive constants λ and λ^* such that $\min\limits_{k\in U_v}\pi_k \geqslant \lambda > 0$, $\min\limits_{k,\ell\in U_v}\pi_{k\ell} \geqslant \lambda^* > 0$ and $\limsup\limits_{v\to\infty} n_v \max\limits_{k\neq\ell\in U_v}|\pi_{k\ell} - \pi_k\pi_\ell| < \infty$.

(C1) The number of subsampled elements N'_v is such that $\lim_{v\to\infty} N'_v/N_v \in (0;1]$.

Assumptions: sample-based RF model-assisted estimator $\hat{t}_{rf}$

(H4) We assume that there exists a positive constant $C_1 > 0$ such that $n_v \max_{k\neq\ell\in U_v}\left|\mathbb{E}_p\left\{(I_k-\pi_k)(I_\ell-\pi_\ell)|\widehat{\mathcal{P}}_S\right\}\right| \leq C_1$

(H5) The random forests based on population partitions and those based on sample partitions are such that, for all $\mathbf{x} \in \mathbb{R}^p$:

$$\mathbb{E}_p \left(\widehat{\widetilde{m}}_{rf}(\mathbf{x}) - \widetilde{m}_{rf}(\mathbf{x}) \right)^2 = o(1).$$

where $\widehat{\widetilde{m}}_{rf}(\mathbf{x})$ is given by

$$\widehat{\widetilde{m}}_{rf}(\mathbf{x}) = \sum_{\ell \in U_v} \frac{1}{B} \sum_{b=1}^{B} \frac{\psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(S)}\left(\mathbf{x}, \theta_b^{(S)}\right)}}{\widehat{\widetilde{N}}(\mathbf{x}, \theta_b^{(S)})} y_\ell$$

with $\widehat{\widetilde{N}}(\mathbf{x}, \theta_b^{(S)}) = \sum_{k \in U_v} \psi_k^{(b,S)} \mathbb{1}_{\mathbf{x}_k \in A^{(S)}\left(\mathbf{x}, \theta_b^{(S)}\right)}$ and

$$\widetilde{m}_{rf}(\mathbf{x}) = \sum_{\ell \in U_v} \frac{1}{B} \sum_{b=1}^{B} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(U)}\left(\mathbf{x}, \theta_b^{(U)}\right)}}{\widetilde{N}(\mathbf{x}, \theta_b^{(U)})} y_\ell$$

with $\widetilde{N}(\mathbf{x}, \theta_b^{(U)}) = \sum_{k \in U_v} \psi_k^{(b,U)} \mathbb{1}_{\mathbf{x}_k \in A^{(U)}\left(\mathbf{x}, \theta_b^{(U)}\right)}$.

Below, we include a graph illustrating the convergence of the difference $\widehat{\widetilde{m}}_{rf} - \widetilde{m}_{rf}$ towards 0 in L^2 where the regression function was defined as $m(X) = 2+2X_1+X_2+X_3$, with X_1, X_2 and X_3 defined as in Section 6 from the main paper. The population sizes were such that the sampling fraction was of 10%. Similar results can be obtained using other simulation parameters.

(C2) The number of subsampled elements n'_v is such that $\lim_{v\to\infty} n'_v/n_v \in (0;1]$.

Consistency of the Horvitz-Thompson variance estimator

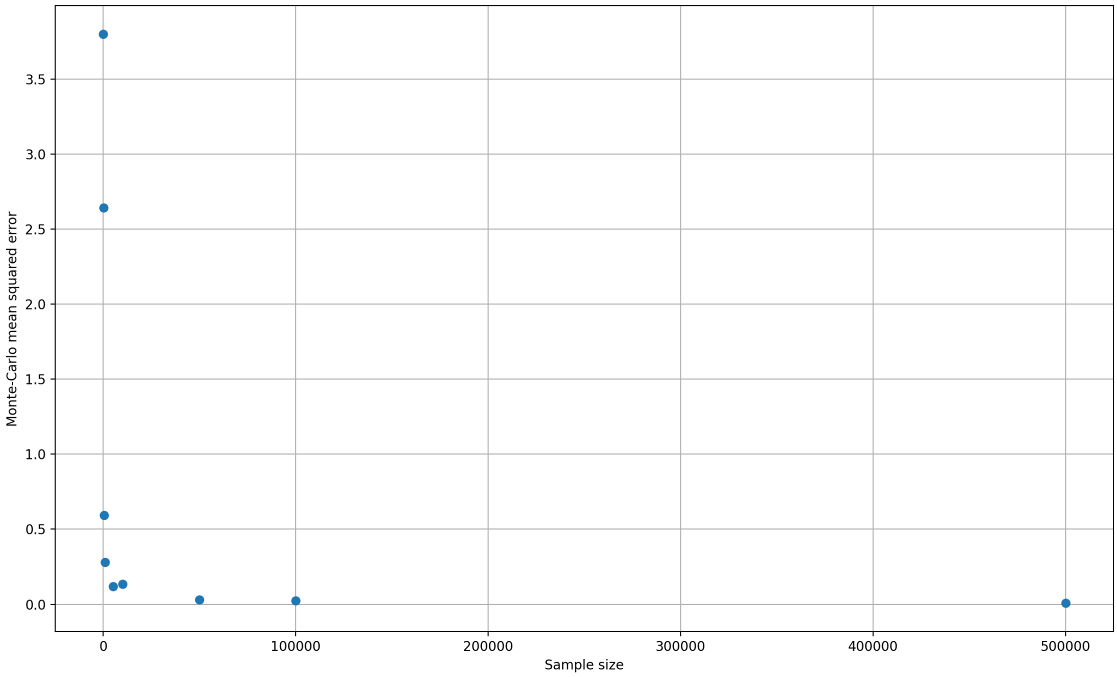
(H6) Assume that $\lim_{v\to\infty} \max_{i,j,k,\ell \in D_{4,N_v}} |\mathbb{E}_p \left\{ (I_i I_j - \pi_i \pi_j)(I_k I_\ell - \pi_k \pi_\ell) \right\}| = 0$, where D_{4,N_v} denotes the set of distinct 4-tuples from U_v.

2 Asymptotic Results

2.1 Asymptotic results of the population RF model-assisted estimator $\widehat{t}^*_{rf}$

The population RF model-assisted estimator is given by

$$\widehat{t}^*_{rf} = \sum_{k \in U_v} \widehat{m}^*_{rf}(\mathbf{x}_k) + \sum_{k \in S_v} \frac{y_k - \widehat{m}^*_{rf}(\mathbf{x}_k)}{\pi_k},$$



where $\widehat{m}^*_{rf}$ is the sample-based estimator of m by using RF built at the population level (for more details, see relation (13) from the main paper):

$$\widehat{m}^*_{rf}(\mathbf{x}_k) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \widetilde{W}^*_\ell(\mathbf{x}_k) y_\ell, \tag{1}$$

where

$$\widetilde{W}^*_\ell(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{*(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}}{\widetilde{N}^*(\mathbf{x}_k, \theta_b^{(U)})}$$

and $\widetilde{N}^*(\mathbf{x}_k, \theta_b^{(U)}) = \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{*(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}$ is the number of units falling in the terminal node $A^{*(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)$ containing $\mathbf{x}_k$. The estimator $\widehat{m}^*_{rf}(\mathbf{x}_k)$ can be written as a bagged estimator as follows:

$$\widehat{m}^*_{rf}(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \widehat{m}^{*(b)}_{tree}(\mathbf{x}_k, \theta_b^{(U)}),$$

where $\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k, \theta_b^{(U)})$ is the sample-based estimation of m based on the b-th stochastic tree:

$$\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k, \theta_b^{(U)}) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{*(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}}{\widetilde{N}^*(\mathbf{x}_k, \theta_b^{(U)})} y_\ell. \tag{2}$$

For more readability, we will use in the sequel $\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k)$ instead of $\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k, \theta^{(U)}_b)$. Consider the pseudo-generalized difference estimator:

$$\widehat{t}_{pgd} = \sum_{k \in U_v} \widetilde{m}^*_{rf}(\mathbf{x}_k) + \sum_{k \in S_v} \frac{y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k)}{\pi_k},$$

where $\widetilde{m}^*_{rf}(\mathbf{x}_k)$ is the population-based estimator of m by using RF built at the population level (for more details, see relation (12) from the main paper):

$$\widetilde{m}^*_{rf}(\mathbf{x}_k) = \sum_{\ell \in U_v} \widetilde{W}^*_{\ell}(\mathbf{x}_k) y_{\ell}.$$

The estimator $\widetilde{m}^*_{rf}$ can be written as a bagged estimator as follows:

$$\widetilde{m}^*_{rf}(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)$$

and $\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)$ is the predictor associated with unit k and based on the b-th stochastic tree:

$$\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k) = \sum_{\ell \in U_v} \frac{\psi^{(b,U)}_{\ell} \mathbb{1}_{\mathbf{x}_{\ell} \in A^{*(U)}\left(\mathbf{x}_k, \theta^{(U)}_b\right)}}{\widetilde{N}^*(\mathbf{x}_k, \theta^{(U)}_b)} y_{\ell}. \tag{3}$$

We remark that $\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k)$ is the Horvitz-Thompson estimator of $\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)$. As before, $\widetilde{m}^{*(b)}_{tree}$ depends on $\theta^{(U)}_b$ but, for more readability, we drop $\theta^{(U)}_b$ from the expression of $\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)$.

We give in the next equivalent expressions of $\widetilde{m}^{*(b)}_{tree}$ and $\widehat{m}^{*(b)}_{tree}$. Consider for that the B partitions built at the population level: $\widetilde{\mathcal{P}}^*_U = \{\widetilde{\mathcal{P}}^{*(b)}_U\}_{b=1}^B$. For a given $b = 1, \ldots, B$, the partition $\widetilde{\mathcal{P}}^{*(b)}_U$ build in the b-th stochastic tree is composed by the J^*_{bU} disjointed regions: $\widetilde{\mathcal{P}}^{*(b)}_U = \{A^{*(bU)}_j\}_{j=1}^{J^*_{bU}}$. Consider $\mathbf{z}^{*(b)}_k = (\mathbb{1}_{\mathbf{x}_k \in A^{*(bU)}_1}, \ldots, \mathbb{1}_{\mathbf{x}_k \in A^{*(bU)}_{J^*_{bU}}})^\top$ where $\mathbb{1}_{\mathbf{x}_k \in A^{*(bU)}_j} = 1$ if $\mathbf{x}_k$ belongs to the region $A^{*(bU)}_j$ and zero otherwise for all $j = 1, \ldots, J^*_{bU}$. We drop the exponent U from the expression of $\mathbf{z}^{*(b)}_k$ for more readability. Since $\widetilde{\mathcal{P}}^{*(b)}_U$ is a partition, then $\mathbf{x}_k$ belongs to only one region of the b-th tree, so the vector $\mathbf{z}^{*(b)}_k$ will contain only one non-null component. Consider for example that $\mathbf{x}_k \in A^{*(bU)}_j$, then $\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)$ is the mean of y-values of individuals ℓ for which $\mathbf{x}_{\ell} \in A^{*(bU)}_j$:

$$\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k) = \sum_{\ell \in U_v} \frac{\psi^{(b,U)}_{\ell} \mathbb{1}_{\mathbf{x}_{\ell} \in A^{*(bU)}_j}}{\widetilde{N}^{*(b)}_j} y_{\ell}, \quad \text{for} \quad \mathbf{x}_k \in A^{*(bU)}_j,$$

where $\widetilde{N}_j^{*(b)}$ is the number of units belonging to the region $A_j^{*(bU)}$:

$$\widetilde{N}_j^{*(b)} = \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{*(bU)}}, \quad j = 1, \ldots, J_{bU}^*. \tag{4}$$

Then, $\widetilde{m}_{tree}^{*(b)}(\mathbf{x}_k)$ can be written as follows:

$$\widetilde{m}_{N,rf}^{*(b)}(\mathbf{x}_k) = (\mathbf{z}_k^{*(b)})^\top \widetilde{\boldsymbol{\beta}}^{*(b)}, \quad k \in U_v \tag{5}$$

where

$$\widetilde{\boldsymbol{\beta}}^{*(b)} = \left(\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top \right)^{-1} \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} y_\ell.$$

Remark that $\widetilde{\boldsymbol{\beta}}^{*(b)}$ may be obtained as solution of the following weighted estimating equation:

$$\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (y_\ell - (\mathbf{z}_\ell^{*(b)})^\top \boldsymbol{\beta}^{*(b)}) = 0.$$

Since the regions $A_j^{*(bU)}, j = 1, \ldots, J_{bU}^*$, form a partition, then the matrix $\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top$ is diagonal with diagonal elements equal to $\widetilde{N}_j^{*(b)}$, the number of units falling in the region $A_j^{*(bU)}$ for all $j = 1, \ldots, J_{bU}^*$. By the stopping criterion, we have that all $\widetilde{N}_j^{*(b)} \geq N_{0v} > 0$ for all j, so the matrix $\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top$ is always invertible and $\widetilde{\boldsymbol{\beta}}^{*(b)}$ is well-defined.

Consider now $\widehat{m}_{tree}^{*(b)}(\mathbf{x}_k)$, the estimator of the unknown $\widetilde{m}_{tree}^{*(b)}(\mathbf{x}_k)$. Then, $\widehat{m}_{tree}^{*(b)}(\mathbf{x}_k)$ is the weighted mean of y-values for sampled individuals ℓ belonging to the same region $A_j^{*(bU)}$ as unit k :

$$\widehat{m}_{tree}^{*(b)}(\mathbf{x}_k) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{*(bU)}}}{N_{j,N}^{*(b)}} y_\ell \quad \text{for} \quad \mathbf{x}_k \in A_j^{*(bU)}$$

and we can write:

$$\widehat{m}_{tree}^{*(b)}(\mathbf{x}_k) = (\mathbf{z}_k^{*(b)})^\top \widehat{\boldsymbol{\beta}}^{*(b)}, k \in U_v \tag{6}$$

where

$$\widehat{\boldsymbol{\beta}}^{*(b)} = \left(\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top \right)^{-1} \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} y_\ell.$$

In the expression of $\widehat{\boldsymbol{\beta}}^{*(b)}$, we do not estimate the matrix $\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top$ since it is known and besides, we guarantee in this way that we will always have non-empty terminal nodes at the population level. So, $\widehat{\boldsymbol{\beta}}^{*(b)}$ will be always well-defined whatever the sample S is.

Let denote by $\alpha_k = \pi_k^{-1} I_k - 1$ for all $k \in U_v$, where I_k is the sample membership, $I_k = 1$ if $k \in S$ and zero otherwise. In order to prove the consistency of $\widehat{t}^*_{rf}$ as well as its asymptotic equivalence to the pseudo-generalized difference estimator $\widehat{t}_{pgd}$, we use the following decomposition:

$$\begin{aligned}\frac{1}{N_v}\left(\widehat{t}^*_{rf} - t_y\right) &= \frac{1}{N_v}\left(\widehat{t}_{pgd} - t_y\right) - \frac{1}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}^*_{rf}(\mathbf{x}_k) - \widetilde{m}^*_{rf}(\mathbf{x}_k))\\ &= \frac{1}{N_v}\left(\widehat{t}_{pgd} - t_y\right) - \frac{1}{B}\sum_{b=1}^{B}\left[\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\left(\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)\right)\right]. \quad (7)\end{aligned}$$

We will prove that each term form the decomposition (7) is convergent to zero. We give first several useful lemmas.

Lemma 1. *There exists a positive constant $\tilde{c}_1$ such that:*

$$\frac{n_v}{N_v^2}\mathbb{E}_p(\widehat{t}_{pgd} - t_y)^2 \leqslant \tilde{c}_1.$$

Proof. First of all, from relation (1), $\widetilde{m}^*_{rf}(\mathbf{x}_k)$ is a weighted sum at the population level of y-values with positive weights summing to one (see proposition 2.1. from the main paper). Then, we get that $\sup_{k\in U_v}|\widetilde{m}^*_{rf}(\mathbf{x}_k)| \leqslant C$ by using also assumption (H1). We have:

$$\frac{1}{N_v}(\widehat{t}_{pgd} - t_y) = \frac{1}{N_v}\sum_{k\in U_v}\alpha_k(y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k))$$

and

$$\begin{aligned}n_v\mathbb{E}_p\left(\frac{\widehat{t}_{pgd} - t_y}{N_v}\right)^2 &= \frac{n_v}{N_v^2}\mathbb{V}_p\left(\sum_{k\in S_v}\frac{(y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k))}{\pi_k}\right)\\ &\leqslant \left(\frac{n_v}{N_v}\cdot\frac{1}{\lambda} + \frac{n_v\max_{k\neq\ell\in U_v}|\pi_{k\ell} - \pi_k\pi_\ell|}{\lambda^2}\right)\cdot\frac{2}{N_v}\sum_{k\in U_v}\left(y_k^2 + (\widetilde{m}^*_{rf}(\mathbf{x}_k))^2\right)\\ &\leqslant \tilde{c}_1\end{aligned}$$

by assumptions (H1)-(H3). ■

Lemma 2. *There exists a positive constant $\tilde{c}_2$ not depending on $b = 1, \ldots, B$, such that*

$$\mathbb{E}_p||\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)}||_2^2 \leqslant \frac{\tilde{c}_2 N_v}{N_{0v}^2} \quad \textit{for all} \quad b = 1, \ldots, B.$$

Proof. We can write

$$\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)} = \left(\sum_{\ell\in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top\right)^{-1} \left(\sum_{\ell\in S_v} \frac{1}{\pi_\ell}\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)} y_\ell - \sum_{\ell\in U_v}\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)} y_\ell\right)$$
$$= \left(\sum_{\ell\in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{*(b)} (\mathbf{z}_\ell^{*(b)})^\top\right)^{-1} \left(\sum_{\ell\in U_v} \alpha_\ell\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)} y_\ell\right) \quad (8)$$

Let denote by $\widetilde{\mathbf{T}}^{*(b)} = \sum_{\ell\in U_v}\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)}(\mathbf{z}_\ell^{*(b)})^\top$. As already mentioned before, the matrix $\widetilde{\mathbf{T}}^{*(b)}$ is diagonal with positive diagonal elements given by $\widetilde{N}_j^{*(b)}$ the number of units falling in the region $A_j^{*(bU)}$ (see relation 4) for $j = 1, \dots, J_{bU}^*$ and by the stopping criterion, we have that $\widetilde{N}_j^{*(b)} \geq N_{0v} > 0$. We obtain then

$$||(\widetilde{\mathbf{T}}^{*(b)})^{-1}||_2 = \max_{j=1,\dots,J_{bU}}\left(\frac{1}{\widetilde{N}_j^{*(b)}}\right) \leq N_{0v}^{-1}, \quad \text{for all} \quad b = 1, \dots, B \quad (9)$$

where $||\cdot||_2$ is the spectral norm matrix defined for a squared $p\times p$ matrix $\mathbf{A}$ by $||\mathbf{A}||_2 = \sup_{\mathbf{x}\in\mathbf{R}^p, ||\mathbf{x}||_2\neq 0}||\mathbf{A}\mathbf{x}||_2/||\mathbf{x}||_2$. For a symmetric and positive definite matrix $\mathbf{A}$, we have that $||\mathbf{A}||_2 = \lambda_{max}(\mathbf{A})$ where $\lambda_{max}(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$. We get, for $b = 1, \dots, B$:

$$\mathbb{E}_p||\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)}||_2^2 \leqslant \mathbb{E}_p\left[||N_v(\widetilde{\mathbf{T}}^{*(b)})^{-1}||_2^2 \cdot \left|\left|\frac{1}{N_v}\sum_{\ell\in U_v}\alpha_\ell\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)}y_\ell\right|\right|_2^2\right]$$
$$\leqslant \frac{N_v^2}{N_{0v}^2}\mathbb{E}_p\left|\left|\frac{1}{N_v}\sum_{\ell\in U_v}\alpha_\ell\psi_\ell^{(b,U)}\mathbf{z}_\ell^{*(b)}y_\ell\right|\right|_2^2 \quad (10)$$

and

$$\mathbb{E}_p\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\psi_k^{(b,U)}\mathbf{z}_k^{*(b)}y_k\right|\right|_2^2$$
$$= \frac{1}{N_v^2}\left(\sum_{k\in U_v}(\psi_k^{(b,U)})^2y_k^2||\mathbf{z}_k^{*(b)}||_2^2\mathbb{E}_p(\alpha_k^2) + \sum_{k\in U_v}\sum_{\substack{\ell\in U_v\\ \ell\neq k}}\psi_k^{(b,U)}\psi_\ell^{(b,U)}y_ky_\ell(\mathbf{z}_k^{*(b)})^\top\mathbf{z}_\ell^{*(b)}\mathbb{E}_p(\alpha_k\alpha_\ell)\right)$$
$$\leqslant \frac{1}{n_v}\left(\frac{n_v}{\lambda N_v} + \frac{n_v\max_{k,\ell\in U_v,k\neq\ell}|\pi_{k\ell}-\pi_k\pi_\ell|}{\lambda^2}\right)\left(\frac{1}{N_v}\sum_{k\in U_v}(\psi_k^{(b,U)})^2y_k^2||\mathbf{z}_k^{*(b)}||_2^2\right)$$
$$\leqslant \frac{C_0}{n_v} \quad (11)$$

by assumptions (H1)-(H3) and the fact that $||\mathbf{z}_k^{*(b)}||_2^2 = 1$ for all $k\in U_v$ and $b = 1, \dots, B$. From (10), (11) and assumption (H1), we obtain that it exists a positive constant $\tilde{c}_2$ such that

$$\mathbb{E}_p||\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)}||_2^2 \leqslant \frac{\tilde{c}_2 N_v}{N_{0v}^2}.$$

■

Result 2.1. *Consider a sequence of population RF estimators* $\{\widehat{t}^*_{rf}\}$*. Then, there exist positive constants* $\tilde{C}_1, \tilde{C}_2$ *such that*

$$\mathbb{E}_p\left|\frac{1}{N_v}\left(\widehat{t}^*_{rf} - t_y\right)\right| \leqslant \frac{\tilde{C}_1}{\sqrt{n_v}} + \frac{\tilde{C}_2}{N_{0v}}, \quad \textit{with } \xi\textit{-probability one.}$$

If $\dfrac{N_v^a}{N_{0v}} = O(1)$ *with* $1/2 \leqslant a \leqslant 1$*, then*

$$\mathbb{E}_p\left|\frac{1}{N_v}\left(\widehat{t}^*_{rf} - t_y\right)\right| \leqslant \frac{\tilde{C}}{\sqrt{n_v}}, \quad \textit{with } \xi\textit{-probability one.}$$

Proof. We get from relation (7) :

$$\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}^*_{rf} - t_y\right| \leqslant \frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{pgd} - t_y\right| + \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k))\right|.$$

Lemma 1 gives us that there exists positive constant $\tilde{C}_1$ such that

$$\frac{1}{N_v}\mathbb{E}_p\left|\widehat{t}_{pgd} - t_y\right| \leqslant \frac{\tilde{C}_1}{\sqrt{n_v}}. \tag{12}$$

Now, by using relations (5) and (6), we can then write for any $b = 1, \ldots, B$:

$$\sum_{k\in U_v}\alpha_k(\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k)) = \sum_{k\in U_v}\alpha_k(\mathbf{z}_k^{*(b)})^\top(\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)})$$

and

$$\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k))\right| \leqslant \left(\mathbb{E}_p\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\mathbf{z}_k^{*(b)}\right|\right|_2^2\right)^{1/2}\left(\mathbb{E}_p||\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)}||_2^2\right)^{1/2} \tag{13}$$

and

$$\begin{aligned}
\mathbb{E}_p\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\mathbf{z}_k^{*(b)}\right|\right|_2^2 &= \frac{1}{N_v^2}\left(\sum_{k\in U_v}\mathbb{E}_p(\alpha_k^2)||\mathbf{z}_k^{*(b)}||_2^2 + \sum_{k\in U_v}\sum_{\substack{\ell\in U_v\\ \ell\neq k}}\mathbb{E}_p(\alpha_k\alpha_\ell)(\mathbf{z}_k^{*(b)})^\top\mathbf{z}_\ell^{*(b)}\right)\\
&\leqslant \frac{1}{n_v}\left(\frac{n_v}{\lambda N_v} + \frac{n_v\max_{k\neq\ell\in U_v}|\pi_{k\ell} - \pi_k\pi_\ell|}{\lambda^2}\right)\cdot\frac{1}{N_v}\sum_{k\in U_v}||\mathbf{z}_k^{*(b)}||_2^2\\
&\leqslant \frac{C_2}{n_v}
\end{aligned} \tag{14}$$

by assumptions (H1)-(H3) and the fact that $||\mathbf{z}_k^{*(b)}||_2^2 = 1$ for all $k \in U_v$ and $b = 1, \ldots, B$. Then, from relations (13), (14) and lemma 2, we get that there exists a positive constant $\tilde{C}_2$ such that, for any $b = 1, \ldots, B$, we have:

$$\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{*(b)}_{rf}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{N,rf}(\mathbf{x}_k))\right| \leqslant \sqrt{\frac{C_2}{n_v}\frac{\tilde{c}_2 N_v}{N_{0v}^2}} \leqslant \frac{\tilde{C}_2}{N_{0v}} \tag{15}$$

by using also the assumption (H1). The result follows then from relations (12) and (15). ∎

Result 2.2. *Consider a sequence of RF estimators* $\{\widehat{t}^*_{rf}\}$. *If* $\dfrac{N_v^a}{N_{0v}} = O(1)$ *with* $1/2 < a \leqslant 1$, *then*

$$\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}^*_{rf} - t_y\right) = \frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd} - t_y\right) + o_{\mathbb{P}}(1).$$

Proof. We get from relation (7) and lemmas (1) and the proof of result 2.1 (relation 15) that

$$\begin{aligned}
\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}^*_{rf} - t_y\right) &= \frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd} - t_y\right) + \frac{1}{B}\sum_{b=1}^{B}\left[\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k))\right].\\
&= \frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd} - t_y\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{\sqrt{n_v}}{N_{0v}}\right)\\
&= \frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd} - t_y\right) + o_{\mathbb{P}}(1)
\end{aligned}$$

provided that $\dfrac{N_v^a}{N_{0v}} = O(1)$ with $1/2 < a \leqslant 1$. ∎

Result 2.3. *Consider a sequence of population RF estimators* $\{\widehat{t}^*_{rf}\}$. *Assume that* $\dfrac{N_v^a}{N_{0v}} = O(1)$ *with* $1/2 < a \leqslant 1$, *then the variance estimator* $\widehat{\mathbb{V}}_{rf}(\widehat{t}^*_{rf})$ *is design-consistent for the asymptotic variance* $\mathbb{AV}_p\left(\widehat{t}^*_{rf}\right)$. *That is,*

$$\lim_{v\to\infty}\mathbb{E}_p\left(\frac{n_v}{N_v^2}\left|\widehat{\mathbb{V}}_{rf}(\widehat{t}^*_{rf}) - \mathbb{AV}_p(\widehat{t}^*_{rf})\right|\right) = 0.$$

Proof Consider the following decomposition

$$\begin{aligned}
& n_v\left(\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right) - \mathbb{AV}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right)\right)\\
= \;& n_v\left(\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right) - \widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}_{pgd}\right)\right) + n_v\left(\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}_{pgd}\right) - \mathbb{AV}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right)\right)
\end{aligned}$$

where $\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}_{pgd}\right)$ is the pseudo-type variance estimator of $\mathbb{V}_p\left(N_v^{-1}\widehat{t}_{pgd}\right) = \mathbb{AV}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right)$ given by

$$\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}_{pgd}\right) = \frac{1}{N_v^2}\sum_{k\in U_v}\sum_{\ell\in U_v}\frac{\pi_{k\ell} - \pi_k\pi_\ell}{\pi_{k\ell}}\frac{y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k)}{\pi_k}\frac{y_\ell - \widetilde{m}^*_{rf}(\mathbf{x}_\ell)}{\pi_\ell}I_kI_\ell.$$

Now, to prove that the consistency of the first term from right of (16), we use the same decomposition as in Goga and Ruiz-Gazen (2014). Denote $\tilde{e}_k = y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k)$, $\hat{e}_k = y_k - \widehat{m}^*_{rf}(\mathbf{x}_k)$ and $c_{k\ell} = \dfrac{\pi_{k\ell} - \pi_k\pi_\ell}{\pi_{k\ell}\pi_k\pi_\ell}I_kI_\ell$. Then,

$$n_v(\widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}^*_{rf}\right) - \widehat{\mathbb{V}}_p\left(N_v^{-1}\widehat{t}_{pgd}\right)) = \frac{n_v}{N_v^2}\sum_{k\in U_v}\sum_{\ell\in U_v}c_{k\ell}\left(\hat{e}_k\hat{e}_\ell - \tilde{e}_k\tilde{e}_\ell\right)$$

$$= \frac{n_v}{N_v^2} \sum_{k \in U_v} \sum_{\ell \in U_v} c_{k\ell} \left[(\hat{e}_k - \tilde{e}_k)(\hat{e}_\ell - \tilde{e}_\ell) + \tilde{e}_k(\hat{e}_\ell - \tilde{e}_\ell) + \tilde{e}_\ell(\hat{e}_k - \tilde{e}_k) \right]$$
$$= A_1 + A_2 + A_3.$$

For all $k \in U_v$, $\hat{e}_k - \tilde{e}_k = \widetilde{m}^*_{rf}(\mathbf{x}_k) - \widehat{m}^*_{rf}(\mathbf{x}_k)$ and thus,

$$\mathbb{E}_p|A_1| \leq \left(\frac{n_v}{\lambda^2 N_v} + \frac{n_v \max_{k \neq \ell \in U_v} |\pi_{k\ell} - \pi_k \pi_\ell|}{\lambda^* \lambda^2} \right) \frac{1}{N_v} \sum_{k \in U_v} \mathbb{E}_p (\hat{e}_k - \tilde{e}_k)^2,$$

by assumptions (H2)-(H3). Therefore, it suffices to show that, for all $k \in U_v$, one has $\mathbb{E}_p(\hat{e}_k - \tilde{e}_k)^2 = o(1)$ uniformly in k, which we show next. We have

$$\mathbb{E}_p(\widetilde{m}^*_{rf}(\mathbf{x}_k) - \widehat{m}^*_{rf}(\mathbf{x}_k))^2 \leqslant \frac{1}{B} \sum_{b=1}^{B} \mathbb{E}_p(\widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widehat{m}^{*(b)}_{tree}(\mathbf{x}_k))^2.$$

We can write by using relations (5) and (6):

$$\widehat{m}^{*(b)}_{tree}(\mathbf{x}_k) - \widetilde{m}^{*(b)}_{tree}(\mathbf{x}_k) \quad = \quad (\mathbf{z}^{*(b)}_k)^\top (\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)})$$

and then, by using lemma (2),

$$\begin{aligned} \mathbb{E}_p(\widetilde{m}^*_{rf}(\mathbf{x}_k) - \widehat{m}^*_{rf}(\mathbf{x}_k))^2 \quad &\leqslant \quad \frac{1}{B} \sum_{b=1}^{B} \mathbb{E}_p \left(||\mathbf{z}^{*(b)}_k||_2^2 ||\widehat{\boldsymbol{\beta}}^{*(b)} - \widetilde{\boldsymbol{\beta}}^{*(b)}||_2^2 \right) \\ &\leqslant \quad \frac{\tilde{c}_2 N_v}{N_{0v}^2} \end{aligned}$$

quantity going to zero provided that $\dfrac{N_v^a}{N_{0v}} = O(1)$ with $1/2 < a \leqslant 1$.

Using the same arguments, we obtain that $\mathbb{E}_p|A_2| = o(1)$ and $\mathbb{E}_p|A_3| = o(1)$. We get then

$$n_v \mathbb{E}_p |\widehat{\mathbb{V}}_p \left(N_v^{-1} \widehat{t}^*_{rf} \right) - \widehat{\mathbb{V}}_p \left(N_v^{-1} \widehat{t}_{pgd} \right) | = o(1).$$

The second term from right of (16) concerns the consistency of the estimator of the Horvitz-Thompson variance computed for the population residuals $y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k), k \in U_v$. The proof of this consistency (Breidt and Opsomer, 2000) requires assumptions only on the higher order inclusion probabilities (H6) as well as finite forth moment of $y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k)$:

$$\frac{1}{N_v} \sum_{k \in U_v} (y_k - \widetilde{m}^*_{rf}(\mathbf{x}_k))^4 \leq \frac{4}{N_v} \sum_{k \in U_v} (y_k^4 + (\widetilde{m}^*_{rf}(\mathbf{x}_k))^4) < \infty.$$

So,

$$n_v \mathbb{E}_p |\widehat{\mathbb{V}}_p \left(N_v^{-1} \widehat{t}_{pgd} \right) - \mathbb{A}\mathbb{V}_p \left(N_v^{-1} \widehat{t}^*_{rf} \right) | = o(1)$$

and the result follows.

2.2 Asymptotic results: the sample RF model-assisted estimator $\widehat{t}_{rf}$

The sample RF model-assisted estimator is given by

$$\widehat{t}_{rf} = \sum_{k \in U_v} \widehat{m}_{rf}(\mathbf{x}_k) + \sum_{k \in S_v} \frac{y_k - \widehat{m}_{rf}(\mathbf{x}_k)}{\pi_k},$$

where $\widehat{m}_{rf}$ is the estimator of m built at the sample level and by using RF based on partition built at the sample level (for more details, see relation (17) from the main paper):

$$\widehat{m}_{rf}(\mathbf{x}_k) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \widehat{W}_\ell(\mathbf{x}_k) y_\ell,$$

where

$$\widehat{W}_\ell(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \frac{\psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(S)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(S)}\right)}}{\widehat{N}(\mathbf{x}_k, \theta_b^{(S)})}$$

and $\widehat{N}(\mathbf{x}_k, \theta_b^{(S)}) = \sum_{\ell \in S_v} \pi_\ell^{-1} \psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(S)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(S)}\right)}$ is the estimated number of units falling in the terminal node $A^{(S)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(S)}\right)$ containing $\mathbf{x}_k$. As in Section 2.1, the estimator $\widehat{m}_{rf}(\mathbf{x}_k)$ can be written as a bagged estimator of m as follows:

$$\widehat{m}_{rf}(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \widehat{m}_{tree}^{(b)}(\mathbf{x}_k)$$

and $\widehat{m}_{tree}^{(b)}(\mathbf{x}_k)$ is the estimation of m based on the b-th stochastic tree:

$$\widehat{m}_{tree}^{(b)}(\mathbf{x}_k) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \frac{\psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(S)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(S)}\right)}}{\widehat{N}(\mathbf{x}_k, \theta_b^{(S)})} y_\ell \tag{16}$$

As in Section 2.1, for more readability, we note in the sequel $\widehat{m}_{tree}^{(b)}(\mathbf{x}_k)$ instead of $\widehat{m}_{tree}^{(b)}(\mathbf{x}_k, \theta_b^{(S)})$. Consider the pseudo-generalized difference estimator:

$$\hat{t}_{pgd} = \sum_{k \in U_v} \widetilde{m}_{rf}(\mathbf{x}_k) + \sum_{k \in S_v} \frac{y_k - \widetilde{m}_{rf}(\mathbf{x}_k)}{\pi_k}$$

where $\widetilde{m}_{rf}$ is the estimation of m built at the population level by using RF based on partition built also at the population level (relation (9) from the main paper):

$$\widetilde{m}_{rf}(\mathbf{x}_k) = \sum_{\ell \in U_v} \widetilde{W}_\ell(\mathbf{x}_k) y_\ell,$$

where

$$\widetilde{W}_\ell(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}}{\widetilde{N}(\mathbf{x}_k, \theta_b^{(U)})}$$

with $\widetilde{N}(\mathbf{x}_k, \theta_b^{(U)}) = \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}$ is the number of units falling in the terminal node $A^{(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)$ containing $\mathbf{x}_k$. The estimator $\widehat{m}_{rf}$ can be also written as a bagged estimator as follows:

$$\widetilde{m}_{rf}(\mathbf{x}_k) = \frac{1}{B} \sum_{b=1}^{B} \widetilde{m}_{tree}^{(b)}(\mathbf{x}_k)$$

and

$$\widetilde{m}_{tree}^{(b)}(\mathbf{x}_k) = \sum_{\ell \in U_v} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A^{(U)}\left(\mathbf{x}_{\boldsymbol{k}}, \theta_b^{(U)}\right)}}{\widetilde{N}(\mathbf{x}_k, \theta_b^{(U)})} y_\ell.$$

As in the previous section, we will write $\widetilde{m}_{rf}$ and $\widehat{m}_{rf}$ in equivalent forms. Consider for that the B partitions build at the population level $\widetilde{\mathcal{P}}_U = \{\widetilde{\mathcal{P}}_U^{(b)}\}_{b=1}^B$. For a given $b = 1, \ldots, B$, the partition $\widetilde{\mathcal{P}}_U^{(b)}$ is composed by the disjointed regions $\widetilde{\mathcal{P}}_U^{(b)} = \{A_j^{(bU)}\}_{j=1}^{J_{bU}}$. Consider $\mathbf{z}_k^{(b)} = (\mathbb{1}_{\mathbf{x}_k \in A_1^{(bU)}}, \ldots, \mathbb{1}_{\mathbf{x}_k \in A_{J_{bU}}^{(bU)}})^\top$ where $\mathbb{1}_{\mathbf{x}_k \in A_j^{(bU)}} = 1$ if $\mathbf{x}_k$ belongs to the region $A_j^{(bU)}$ and zero otherwise for all $j = 1, \ldots, J_{bU}$. Since $\widetilde{\mathcal{P}}_U^{(b)}$ is a partition, then $\mathbf{x}_k$ belongs to only one region at the b-th step. Suppose for example that $\mathbf{x}_k \in A_j^{(bU)}$, then $\widetilde{m}_{tree}^{(b)}(\mathbf{x}_k)$ is the mean of y-values for individuals ℓ for which $\mathbf{x}_\ell \in A_j^{(bU)}$:

$$\widetilde{m}_{tree}^{(b)}(\mathbf{x}_k) = \sum_{\ell \in U_v} \frac{\psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{(bU)}}}{\widetilde{N}_j^{(b)}} y_\ell, \quad \text{for} \quad \mathbf{x}_k \in A_j^{(bU)},$$

where $\widetilde{N}_j^{(b)}$ is the number of units belonging to the region $A_j^{(bU)}$:

$$\widetilde{N}_j^{(b)} = \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{(bU)}}, \quad j = 1, \ldots, J_{bU}. \tag{17}$$

Then, $\widetilde{m}_{tree}^{(b)}(\mathbf{x}_k)$ can be written as a regression-type estimator with $\mathbf{z}_k^{(b)}$ as explanatory variables:

$$\widetilde{m}_{tree}^{(b)}(\mathbf{x}_k) = (\mathbf{z}_k^{(b)})^\top \widetilde{\boldsymbol{\beta}}^{(b)}, \quad k \in U_v \tag{18}$$

where

$$\widetilde{\boldsymbol{\beta}}^{(b)} = \left(\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{(b)} (\mathbf{z}_\ell^{(b)})^\top \right)^{-1} \sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{(b)} y_\ell.$$

Based on the same arguments as in Section 2.1, the matrix $\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{(b)} (\mathbf{z}_\ell^{(b)})^\top$ is diagonal with diagonal elements equal to $\widetilde{N}_j^{(b)}, j = 1, \ldots, J_{bU}$. By the stopping criterion, we have that all $\widetilde{N}_j^{(b)} \geq N_0 > 0$, so the matrix $\sum_{\ell \in U_v} \psi_\ell^{(b,U)} \mathbf{z}_\ell^{(b)} (\mathbf{z}_\ell^{(b)})^\top$ is invertible and $\widetilde{\boldsymbol{\beta}}^{(b)}$ is well-defined.

Consider now the B partitions build at the sample level $\widehat{\mathcal{P}}_S = \{\widehat{\mathcal{P}}_S^{(b)}\}_{b=1}^B$. For a given $b = 1, \ldots, B$, the partition $\widehat{\mathcal{P}}_S^{(b)}$ is composed by the disjointed regions $\widehat{\mathcal{P}}_S^{(b)} = \{A_j^{(bS)}\}_{j=1}^{J_{bS}}$. Consider $\hat{\mathbf{z}}_k^{(b)} = (\mathbb{1}_{\mathbf{x}_k \in A_{1S}^{(b)}}, \ldots, \mathbb{1}_{\{\mathbf{x}_k \in A_{J_{bS}}^{(b)}\}})^\top$ where $\mathbb{1}_{\mathbf{x}_k \in A_j^{(bS)}} = 1$ if $\mathbf{x}_k$ belongs to the region $A_j^{(bS)}$ and zero otherwise for all $j = 1, \ldots, J_{bS}$. Here, the hat notation is to design the fact that the vector $\hat{\mathbf{z}}_k^{(b)}$ depends on random dummy variables $\mathbb{1}_{\mathbf{x}_k \in A_j^{(bS)}}$. Since $\{A_j^{(bS)}\}_{j=1}^{J_{bS}}$ form a partition, then $\mathbf{x}_k$ belongs to only one terminal node. Suppose for example that $\mathbf{x}_k \in A_j^{(bS)}$, then $\widehat{m}_{tree}^{(b)}(\mathbf{x}_k)$ is a Hajek-type estimator:

$$\widehat{m}_{tree}^{(b)}(\mathbf{x}_k) = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \frac{\psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{(bS)}} y_\ell}{\widehat{N}_j^{(b)}}, \quad \text{for} \quad \mathbf{x}_k \in A_j^{(bS)},$$

where $\widehat{N}_j^{(b)}$ is the estimated number of units falling in the terminal node $A_j^{(bS)}$:

$$\widehat{N}_j^{(b)} = \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A_j^{(bS)}}, \quad j = 1, \ldots, J_{bS}.$$

Then, $\widehat{m}_{tree}^{(b)}(\mathbf{x}_k)$ can be written also as a regression-type estimator with $\hat{\mathbf{z}}_k^{(b)}$ as explanatory variables:

$$\widehat{m}_{tree}^{(b)}(\mathbf{x}_k) = (\hat{\mathbf{z}}_k^{(b)})^\top \widehat{\boldsymbol{\beta}}^{(b)}, k \in U_v, \tag{19}$$

where

$$\widehat{\boldsymbol{\beta}}^{(b)} = \left(\sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} (\hat{\mathbf{z}}_\ell^{(b)})^\top \right)^{-1} \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} y_\ell.$$

As in Section 2.1, remark that $\widehat{\boldsymbol{\beta}}^{(b)}$ may be obtained as solution of the following weighted estimating equation:

$$\sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} (y_\ell - (\hat{\mathbf{z}}_\ell^{(b)})^\top \boldsymbol{\beta}^{(b)}) = 0.$$

Since $\{A_j^{(bS)}\}_{j=1}^{J_{bS}}$ is a partition, then the matrix $\sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} (\hat{\mathbf{z}}_\ell^{(b)})^\top$ is diagonal with diagonal elements equal to $\widehat{N}_j^{(b)}, j = 1, \ldots, J_{bS}$. By the stopping criterion and assumption (H3), we have that $\sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \mathbb{1}_{\mathbf{x}_\ell \in A_{jS}^{(b)}} \geq n_{0v} > 0$, so $\sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} (\hat{\mathbf{z}}_\ell^{(b)})^\top$ is always invertible is and $\widehat{\boldsymbol{\beta}}^{(b)}$ is well-defined whatever the sample S is.

We need to consider also a second pseudo-generalized difference estimator:

$$\widehat{\widehat{t}}_{pgd} = \sum_{k \in U_v} \widehat{\widehat{m}}_{rf}(\mathbf{x}_k) + \sum_{k \in S_v} \frac{y_k - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k)}{\pi_k}$$

where

$$\begin{aligned}\widehat{\widehat{m}}_{rf}(\mathbf{x}_k) &= \sum_{\ell\in U_v}\left(\frac{1}{B}\sum_{b=1}^{B}\frac{\psi_\ell^{(b,S)}\mathbb{1}_{\mathbf{x}_\ell\in A^{(S)}(\mathbf{x}_k,\theta_b^{(S)})}}{\widehat{\widehat{N}}(\mathbf{x}_k,\theta_b^{(S)})}\right)y_\ell \\ &= \frac{1}{B}\sum_{b=1}^{B}\widehat{\widehat{m}}_{tree}^{(b)}(\mathbf{x}_k)\end{aligned}$$

with $\widehat{\widehat{N}}(\mathbf{x}_k,\theta_b^{(S)}) = \sum_{\ell\in U_v}\psi_\ell^{(b,S)}\mathbb{1}_{\mathbf{x}_\ell\in A(\mathbf{x}_k,\theta_b^{(S)})}$ and

$$\widehat{\widehat{m}}_{tree}^{(b)}(\mathbf{x}_k) = \sum_{\ell\in U_v}\frac{\psi_\ell^{(b,S)}\mathbb{1}_{\mathbf{x}_\ell\in A^{(S)}(\mathbf{x}_k,\theta_b^{(S)})}y_\ell}{\widehat{\widehat{N}}(\mathbf{x}_k,\theta_b^{(S)})} = (\hat{\mathbf{z}}_k^{(b)})^\top\widehat{\widehat{\boldsymbol{\beta}}}^{(b)}, \quad k\in U_v \tag{20}$$

for

$$\widehat{\widehat{\boldsymbol{\beta}}}^{(b)} = \left(\sum_{\ell\in U_v}\psi_\ell^{(b,S)}\hat{\mathbf{z}}_\ell^{(b)}(\hat{\mathbf{z}}_\ell^{(b)})^\top\right)^{-1}\sum_{\ell\in U_v}\psi_\ell^{(b,S)}\hat{\mathbf{z}}_\ell^{(b)}y_\ell.$$

The matrix $\sum_{\ell\in U_v}\psi_\ell^{(b,S)}\hat{\mathbf{z}}_\ell^{(b)}(\hat{\mathbf{z}}_\ell^{(b)})^\top$ is also diagonal with diagonal elements equal to $\sum_{\ell\in U_v}\psi_\ell^{(b,S)}\mathbb{1}_{\mathbf{x}_\ell\in A_j^{(bS)}} \geq n_{0v} > 0, j = 1,\ldots,J_{bS}$ so $\widehat{\widehat{\boldsymbol{\beta}}}^{(b)}$ is also well-defined whatever the sample S is. In order to prove the consistency of the sample-based RF estimator $\widehat{t}_{rf}$, we use the following decomposition:

$$\frac{1}{N_v}(\hat{t}_{rf} - t_y) = \frac{1}{N_v}(\widehat{\widehat{t}}_{pgd} - t_y) - \frac{1}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k) - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k)). \tag{21}$$

We will give first several useful lemmas. The constants used in the following results may not be the same as the ones from Section 2.1 even if they are denoted in the same way for simplicity.

Lemma 3. *There exists a positive constant $\tilde{c}_1$ such that:*

$$\frac{n_v}{N_v^2}\mathbb{E}_p(\widehat{t}_{pgd} - t_y)^2 \leqslant \tilde{c}_1.$$

Proof. The proof is similar to that of lemma 1. We also have that $\sup_{k\in U_v}|\widetilde{m}_{rf}(\mathbf{x}_k)| \leqslant C$ by using assumption (H1). Further,

$$\begin{aligned} n_v\mathbb{E}_p\left(\frac{\widehat{t}_{pgd} - t_y}{N_v}\right)^2 &\leqslant \left(\frac{n_v}{N_v}\cdot\frac{1}{\lambda} + \frac{n_v\max_{k\neq\ell\in U_v}|\pi_{k\ell} - \pi_k\pi_\ell|}{\lambda^2}\right)\cdot\frac{2}{N_v}\sum_{k\in U_v}\left(y_k^2 + (\widetilde{m}_{rf}(\mathbf{x}_k))^2\right) \\ &\leqslant \tilde{c}_1\end{aligned}$$

by assumptions (H1)-(H3). ■

Lemma 4. *There exists a positive constant $\tilde{c}_2$ such that:*

$$\frac{n_v}{N_v^2}\mathbb{E}_p(\widehat{\widehat{t}}_{pgd} - t_y)^2 \leqslant \tilde{c}_2.$$

Proof. Using (20), we get that $\widehat{\widehat{m}}_{rf}(\mathbf{x}_k)$ can be written as a weighted sum of y-values with positive weights summing to unity, so $\sup_{k\in U_v} |\widehat{\widehat{m}}_{rf}(\mathbf{x}_k)| \leqslant C$ by using also assumption (H1). Now,

$$\widehat{\widehat{t}}_{pgd} - t_y = \sum_{k\in U_v} \alpha_k (y_k - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k))$$

and

$$\begin{aligned}
\frac{n_v}{N_v^2}\mathbb{E}_p(\widehat{\widehat{t}}_{pgd} - t_y) &= \frac{n_v}{N_v^2}\sum_{k\in U_v}\mathbb{E}_p\left[\alpha_k^2(y_k - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k))^2\right] \\
&\quad + \frac{n_v}{N_v^2}\sum_{k\in U_v}\sum_{\ell\neq k,\ell\in U_v}\mathbb{E}_p\left[(y_k - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k))(y_\ell - \widehat{\widehat{m}}_{rf}(\mathbf{x}_\ell))\mathbb{E}_p(\alpha_k\alpha_\ell|\hat{\mathcal{P}}_S)\right] \\
&\leqslant \frac{2n_vC^2}{\lambda N_v} + \frac{n_v}{N_v^2}\sum_{k\in U_v}\sum_{\ell\neq k,\ell\in U_v}\mathbb{E}_p\left[|y_k - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k)||y_\ell - \widehat{\widehat{m}}_{rf}(\mathbf{x}_\ell)| \max_{\ell\neq k\in U_v}|\mathbb{E}_p(\alpha_k\alpha_\ell|\hat{\mathcal{P}}_S)|\right] \\
&\leqslant \tilde{c}_2,
\end{aligned}$$

by assumptions (H2) and (H4). ■

Lemma 5. *There exists a positive constant $\tilde{c}_3$ not depending on b such that:*

$$\mathbb{E}_p\left|\left|\widehat{\boldsymbol{\beta}}^{(b)} - \widehat{\widehat{\boldsymbol{\beta}}}^{(b)}\right|\right|_2^2 \leqslant \frac{\tilde{c}_3 n_v}{n_{0v}^2},$$

for all $b = 1, \ldots, B$.

Proof. Let denote by $\widehat{\mathbf{T}}^{(b)} = \sum_{\ell\in S_v}\frac{1}{\pi_\ell}\psi_\ell^{(b,S)}\hat{\mathbf{z}}_\ell^{(b)}(\hat{\mathbf{z}}_\ell^{(b)})^\top$. As already mentioned, the $J_{bS} \times J_{bS}$ dimensional matrix $\widehat{\mathbf{T}}^{(b)}$ is diagonal with diagonal elements given by $\widehat{N}_j^{(b)} = \sum_{\ell\in S_v}\frac{1}{\pi_\ell}\psi_\ell^{(b,S)}\mathbb{1}_{\mathbf{x}_\ell\in A_{jS}^{(b)}}$ the weighted somme of units falling in the region $A_{jS}^{(b)}$ for $j = 1, \ldots, J_{bS}$ and by the stopping criterion, we have that $\widehat{N}_j^{(b)} \geq n_{0v} > 0$. The matrix $\widehat{\mathbf{T}}^{(b)}$ is then always invertible with

$$||(\widehat{\mathbf{T}}^{(b)})^{-1}||_2 \leq n_{0v}^{-1} \quad \text{for all} \quad b = 1, \ldots B. \tag{22}$$

Now, write

$$\widehat{\boldsymbol{\beta}}^{(b)} - \widehat{\widehat{\boldsymbol{\beta}}}^{(b)} = (\widehat{\mathbf{T}}^{(b)})^{-1}\left(\sum_{\ell\in S_v}\frac{1}{\pi_\ell}\psi_\ell^{(b,S)}\hat{\mathbf{z}}_\ell^{(b)}y_\ell - \widehat{\mathbf{T}}^{(b)}\widehat{\widehat{\boldsymbol{\beta}}}^{(b)}\right)$$

$$
\begin{aligned}
&= (\widehat{\mathbf{T}}^{(b)})^{-1} \sum_{\ell \in S_v} \frac{1}{\pi_\ell} \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} \left(y_\ell - \widehat{\widehat{m}}_{tree}^{(b)}(\mathbf{x}_\ell) \right) \\
&= (\widehat{\mathbf{T}}^{(b)})^{-1} \sum_{\ell \in U_v} \alpha_\ell \widehat{E}_\ell^{(b)} \qquad (23)
\end{aligned}
$$

where $\widehat{E}_\ell^{(b)} = \psi_\ell^{(b,S)} \hat{\mathbf{z}}_\ell^{(b)} (y_\ell - \widehat{\widehat{m}}_{tree}^{(b)}(\mathbf{x}_\ell))$ with $\sum_{\ell \in U_v} \widehat{E}_\ell^{(b)} = 0$. We have that $||\hat{\mathbf{z}}_\ell^{(b)}||_2 = 1$ and $\sup_{\ell \in U_v} |\widehat{\widehat{m}}_{tree}^{(b)}(\mathbf{x}_\ell))| \leq C$ for all $\ell \in U_v$ and $b = 1, \ldots, B$, then:

$$
||\widehat{E}_\ell^{(b)}||_2^2 \leqslant 2C^2.
$$

Following the same lines as in lemma 4, we get that it exists a positive constant $\tilde{C}_0$ not depending on b such that

$$
\frac{1}{N_v^2} \mathbb{E}_p \left|\left| \sum_{\ell \in U_v} \alpha_\ell \widehat{E}_\ell^{(b)} \right|\right|_2^2 \leqslant \frac{\tilde{C}_0}{n_v}, \quad \text{for all} \quad b = 1, \ldots B. \qquad (24)
$$

We obtain then from relations (22) and (23) that:

$$
\begin{aligned}
\mathbb{E}_p \left|\left| \widehat{\boldsymbol{\beta}}^{(b)} - \widehat{\widehat{\boldsymbol{\beta}}}^{(b)} \right|\right|_2^2 &\leqslant \mathbb{E}_p \left(N_v^2 ||(\widehat{\mathbf{T}}^{(b)})^{-1}||_2^2 \frac{1}{N_v^2} \left|\left| \sum_{\ell \in U_v} \alpha_\ell \widehat{E}_\ell^{(b)} \right|\right|_2^2 \right) \\
&\leqslant \frac{N_v^2}{n_{0v}^2} \frac{1}{N_v^2} \mathbb{E}_p \left|\left| \sum_{\ell \in U_v} \alpha_\ell \widehat{E}_\ell^{(b)} \right|\right|_2^2 \\
&\leqslant \frac{N_v^2}{n_{0v}^2} \frac{\tilde{C}_0}{n_v} \\
&\leqslant \frac{\tilde{c}_3 n_v}{n_{0v}^2} \qquad (25)
\end{aligned}
$$

by assumption (H2). ■

Result 2.4. *Consider a sequence of sample RF estimators* $\{\widehat{t}_{rf}\}$. *Then, there exist positive constants* $\tilde{C}_1, \tilde{C}_2$ *such that*

$$
\frac{1}{N_v} \mathbb{E}_p |\hat{t}_{rf} - t_y| \leqslant \frac{\tilde{C}_1}{\sqrt{n_v}} + \frac{\tilde{C}_2}{n_{0v}}.
$$

If $\frac{n_v^u}{n_{0v}} = O(1)$ *with* $1/2 \leqslant u \leqslant 1$, *then*

$$
\mathbb{E}_p \left| \frac{1}{N_v} \left(\widehat{t}_{rf} - t_y \right) \right| \leqslant \frac{\tilde{C}}{\sqrt{n_v}}, \quad \textit{with } \xi\textit{-probability one.}
$$

Proof. We use the decomposition given in relation (21):

$$
\frac{1}{N_v} (\hat{t}_{rf} - t_y) = \frac{1}{N_v} (\widehat{\widehat{t}}_{pgd} - t_y) - \frac{1}{N_v} \sum_{k \in U_v} \alpha_k (\widehat{m}_{rf}(\mathbf{x}_k) - \widehat{\widehat{m}}_{rf}(\mathbf{x}_k)).
$$

Now,

$$\mathbb{E}_p\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widehat{\widehat{m}}_{rf}(\mathbf{x}_k))\right| \leqslant \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{(b)}_{tree}(\mathbf{x}_k)-\widehat{\widehat{m}}^{(b)}_{tree}(\mathbf{x}_k))\right|$$

and using relations (19) and (20), we get:

$$\begin{aligned}\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{(b)}_{tree}(\mathbf{x}_k)-\widehat{\widehat{m}}^{(b)}_{tree}(\mathbf{x}_k))\right| &\leqslant \mathbb{E}_p\left(\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\hat{\mathbf{z}}^{(b)}_k\right|\right|_2\left|\left|\widehat{\boldsymbol{\beta}}^{(b)}-\widehat{\widehat{\boldsymbol{\beta}}}^{(b)}\right|\right|_2\right)\\ &\leqslant \sqrt{\mathbb{E}_p\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\hat{\mathbf{z}}^{(b)}_k\right|\right|_2^2\mathbb{E}_p\left|\left|\widehat{\boldsymbol{\beta}}^{(b)}-\widehat{\widehat{\boldsymbol{\beta}}}^{(b)}\right|\right|_2^2}.\end{aligned}$$

We have that $||\hat{\mathbf{z}}^{(b)}_k||_2=1$ for all $k\in U_v$ and $b=1,\ldots,B$. We can show then by using the same arguments as in the proof of lemma 4, that there exists a positive constant $\tilde{C}'_0$ such that

$$E_p\left|\left|\frac{1}{N_v}\sum_{k\in U_v}\alpha_k\hat{\mathbf{z}}^{(b)}_k\right|\right|_2^2 \leqslant \frac{\tilde{C}'_0}{n_v}$$

which together with lemma 5 gives us that there exists a positive constant $\tilde{C}_2$ such that

$$\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widehat{\widehat{m}}_{rf}(\mathbf{x}_k))\right| \leqslant \frac{\tilde{C}_2}{n_{0v}}. \tag{26}$$

Now,

$$\begin{aligned}\frac{1}{N_v}\mathbb{E}_p\left|\hat{t}_{rf}-t_y\right| &\leqslant \frac{1}{N_v}\mathbb{E}_p\left|\widehat{\widehat{t}}_{pgd}-t_y\right|+\frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_v}\mathbb{E}_p\left|\sum_{k\in U_v}\alpha_k(\widehat{m}^{(b)}_{tree}(\mathbf{x}_k)-\widehat{\widehat{m}}^{(b)}_{tree}(\mathbf{x}_k))\right|\\ &\leqslant \frac{\tilde{C}_1}{\sqrt{n_v}}+\frac{\tilde{C}_2}{n_{0v}}\end{aligned}$$

by using lemma 4 and relation (26). ■

Result 2.5. *Consider a sequence of RF estimators* $\{\widehat{t}_{rf}\}$. *Assume that* $\dfrac{n_v^u}{n_{0v}}=O(1)$ *with* $1/2<u\leqslant 1$. *Then,*

$$\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{rf}-t_y\right)=\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd}-t_y\right)+o_{\mathbb{P}}(1).$$

Proof. We have

$$\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{rf}-t_y\right)=\frac{\sqrt{n_v}}{N_v}\left(\widehat{t}_{pgd}-t_y\right)+\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widetilde{m}_{N,rf}(\mathbf{x}_k)). \tag{27}$$

Now,

$$
\begin{aligned}
&\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)) \\
=\ &\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k))+\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)). \quad (28)
\end{aligned}
$$

Relation (26) gives us that

$$
\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{m}_{rf}(\mathbf{x}_k)-\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k))=O_{\mathbb{P}}\left(\frac{\sqrt{n_v}}{n_{0v}}\right)=o_{\mathbb{P}}(1) \tag{29}
$$

provided that $\dfrac{n_v^u}{n_{0v}}=O(1)$ with $1/2<u\leqslant 1$. Consider now the second term from the right-side of relation (28). We have:

$$
\begin{aligned}
&\frac{n_v}{N_v^2}\mathbb{E}_p\left(\sum_{k\in U_v}\alpha_k(\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k))\right)^2 \\
\leqslant\ &\frac{n_v}{N_v^2}\frac{(1+\lambda)^2}{\lambda^2}\sum_{k\in U_v}\mathbb{E}_p\left(\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)\right)^2 \\
+\ &\frac{n_v}{N_v^2}\sum_{k\in U_v}\sum_{\ell\neq k,\ell\in U_v}\mathbb{E}_p\left[|\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)||\widehat{\widetilde{m}}_{rf}(\mathbf{x}_\ell)-\widetilde{m}_{rf}(\mathbf{x}_\ell)|\max_{\ell\neq k\in U_v}|\mathbb{E}_p(\alpha_k\alpha_\ell|\hat{\mathcal{P}}_S)|\right] \\
\leqslant\ &\left(\frac{n_v}{N_v}\frac{(1+\lambda)^2}{\lambda^2}+\frac{C_1}{\lambda^2}\right)\frac{1}{N_v}\sum_{k\in U_v}\mathbb{E}_p\left(\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)\right)^2=o(1),
\end{aligned}
$$

by assumptions (H2), (H3), (H4) and (H5). It follows then that

$$
\frac{\sqrt{n_v}}{N_v}\sum_{k\in U_v}\alpha_k(\widehat{\widetilde{m}}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k))=o_{\mathbb{P}}(1). \tag{30}
$$

Relations (27), (28), (29) and (30) give then the result. ∎

Result 2.6. *Consider a sequence of population RF estimators $\{\widehat{t}_{rf}\}$. Assume also that $\dfrac{n_v^u}{n_{0v}}=O(1)$ with $1/2<u\leqslant 1$. Then, the variance estimator $\widehat{\mathbb{V}}_{rf}(\widehat{t}_{rf})$ is asymptotically design-consistent for the asymptotic variance $\mathbb{AV}_p\left(\widehat{t}_{rf}\right)$. That is,*

$$
\lim_{v\to\infty}\mathbb{E}_p\left(\frac{n_v}{N_v^2}\left|\widehat{\mathbb{V}}_{rf}(\widehat{t}_{rf})-\mathbb{AV}_p(\widehat{t}_{rf})\right|\right)=0. \tag{31}
$$

Proof. The proof follows the same steps as those of result (2.3). We need to show that

$$
\mathbb{E}_p\left[\left(\widehat{m}_{rf}(\mathbf{x}_k)-\widetilde{m}_{rf}(\mathbf{x}_k)\right)^2\right]=o(1), \tag{32}
$$

uniformly in $k \in U_v$. We have $\widehat{m}_{rf}(\mathbf{x}_k) - \widetilde{m}_{rf}(\mathbf{x}_k) = \widehat{m}_{rf}(\mathbf{x}_k) - \widehat{\widetilde{m}}_{rf}(\mathbf{x}_k) + \widehat{\widetilde{m}}_{rf}(\mathbf{x}_k) - \widetilde{m}_{rf}(\mathbf{x}_k)$ and

$$\begin{aligned}
\mathbb{E}_p(\widehat{m}_{rf}(\mathbf{x}_k) - \widehat{\widetilde{m}}_{rf}(\mathbf{x}_k))^2 &\leqslant \frac{1}{B}\sum_{b=1}^{B} \mathbb{E}_p(\widehat{m}^{(b)}_{tree}(\mathbf{x}_k) - \widehat{\widetilde{m}}^{(b)}_{tree}(\mathbf{x}_k))^2 \\
&\leqslant \frac{1}{B}\sum_{b=1}^{B} \mathbb{E}_p\left(||\hat{\mathbf{z}}^{(b)}_k||^2_2 \left|\left|\widehat{\boldsymbol{\beta}}^{(b)} - \widehat{\widetilde{\boldsymbol{\beta}}}^{(b)}\right|\right|^2_2\right) \\
&\leqslant \frac{1}{B}\sum_{b=1}^{B} \mathbb{E}_p\left(\left|\left|\widehat{\boldsymbol{\beta}}^{(b)} - \widehat{\widetilde{\boldsymbol{\beta}}}^{(b)}\right|\right|^2_2\right) \\
&\leqslant \frac{\tilde{c}_3 n_v}{n^2_{0v}} = o(1)
\end{aligned}$$

by lemma 5 and provided that $\dfrac{n_v^u}{n_{0v}} = O(1)$ with $1/2 < u < 1$. The result (32) follows then by using also assumption (H5). ■

References

Breidt, F.-J. and Opsomer, J.-D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28(4):1023–1053.

Goga, C. and Ruiz-Gazen, A. (2014). Efficient estimation of non-linear finite population parameters by using non-parametrics. *Journal of the Royal Statistical Society, B*, 76:113–140.