Variable Selection for Linear Regression Imputation in Surveys

Ziming An[1], Mehdi Dagdoug[2] and David Haziza[1]

[1]*Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada*
e-mail: `zan032@uottawa.ca`*;* `dhaziza@uottawa.ca`

[2]*Department of Mathematics and Statistics, McGill University, Montréal, Canada*
e-mail: `mehdi.dagdoug@mcgill.ca`

Abstract: Survey sampling is concerned with the estimation of finite population parameters. In practice, survey data suffer from item nonresponse, which is commonly handled through imputation, i.e., replacing missing values with predicted values. As a result, the properties of the resulting imputed estimator depend critically on the properties of the prediction method used. In turn, prediction methods themselves depend on the choice of variables and tuning parameters used to fit the imputation model. In this article, we study the problem of variable selection for linear regression imputation. Although variable selection has been widely studied across many fields, primarily for identification or prediction, its role in imputation for survey data has received comparatively little attention. We introduce the notion of an optimal imputation model defined through an oracle loss function and show that, with probability tending to one, the optimal model coincides with the true model. We also examine the consequences of using misspecified models–either omitting relevant covariates or including irrelevant ones– on consistency and asymptotic variance. We then develop a complete methodological framework for constructing confidence intervals after model selection. The proposed confidence intervals are shown to be asymptotically valid and optimal among all candidate models. Simulation studies indicate that the proposed methodology performs well in finite samples.

1. Introduction

Nonresponse is a major challenge in official statistics because it affects most surveys. If missing data are ignored, the resulting estimates can be biased and inconsistent. In survey sampling, it is customary to distinguish unit from item nonresponse. The former corresponds to cases where no information is collected for a sampled unit, whereas the latter occurs when some variables are missing but others are observed. Item nonresponse is most often addressed through imputation, a procedure that replaces missing values with predicted values. Imputation can restore consistency and support valid inference, provided that the imputation model is correctly specified and that variance estimation properly reflects both sampling variability and nonresponse. In practice, this means that the imputation model matters a lot: if it is misspecified, the resulting estimates may be biased and/or inefficient.

The properties of imputed estimators have been investigated under a wide range of imputation models. A large body of work focuses on parametric imputation, such as linear regression

1

imputation (e.g., Chauvet, Deville and Haziza (2011)), including many contributions on variance estimation (Fay, 1991; Shao and Steel, 1999; Berger and Rao, 2006; Kim and Rao, 2009). Other contributions have examined nonparametric regression and related methods, including nearest neighbor (Chen and Shao, 2000, 2001; Yang and Kim, 2019), the score method (Haziza and Beaumont, 2007), predictive mean matching (Yang and Kim, 2017), and random forests (Dagdoug, Goga and Haziza, 2025).

In the context of model-assisted estimation, we highlight the work of Opsomer and Miller (2005), which provides an important reference on hyperparameter selection in survey sampling, although not in the imputation setting. In the customary i.i.d. setting, variable selection has been extensively studied with two primary objectives: identifying the best predictive model for a given task (efficiency) or recovering the true set of coefficients (consistency) (Yang, 2005). Seminal contributions include Nishii (1984), Rao and Wu (1989), Shao (1993), and Shao (1997), who established the asymptotic properties of numerous model selection criteria such as AIC (Akaike, 1970), BIC (Schwarz, 1978), and cross-validation.

In finite population sampling, the goal is to estimate finite population parameters, and the main concern is the performance of the resulting estimators, in particular their bias and efficiency. Therefore, when choosing an imputation model, the survey statistician is not primarily aiming to identify the true data-generating model, but rather to select the model that produces the most efficient imputed estimator, that is, the one with the smallest mean squared error. Although we will borrow several ideas from the i.i.d. literature, our objective is fundamentally different: we aim to optimize imputation efficiency for finite population inference, not to recover the underlying regression function. As shown later, however, these two goals are closely related in the case of linear models.

The main contributions of this paper are as follows. We introduce an oracle loss function to assess the efficiency of a candidate imputation model under survey sampling. We show that, under mild conditions, the model that minimizes this loss is asymptotically the true model, which connects model selection for imputation with model identification. We then study the effect of using a misspecified model. In particular, we give general conditions under which the resulting imputed estimators remain consistent, and we identify settings where overfitting (i.e., adding unnecessary covariates) does or does not increase the asymptotic variance. Under standard regularity conditions, we show that model selection criteria that are consistent in the i.i.d. framework remain consistent under survey sampling with missing data. Finally, when a consistent selection criterion is used, the resulting imputed estimator and its variance are asymptotically equivalent to those obtained under the true model, yielding oracle efficiency. We also show that a standard variance estimator for linear regression imputation is consistent and establish the asymptotic normality of the resulting point estimator. We propose a complete methodology for constructing confidence intervals after model selection, and show that it achieves the nominal coverage asymptotically, with asymptotically minimal width within the class of candidate models. Simulation studies under both equal and unequal probability sampling designs confirm the good performance of the proposed approach.

The remainder of the paper is organized as follows. Section 2 introduces the notation and

general framework. Section 3 defines the oracle loss function and its properties. Section 4 presents the main asymptotic results for model selection for imputation. Section 5 reports simulation results, and Section 6 concludes with final remarks and presents future works. All proofs are relegated to the Appendix.

2. Preliminaries

2.1. The setup

Consider a finite population $U := \{1, 2, ..., N\}$ of size N and a survey variable Y. We are interested in estimating the finite population mean

$$\mu := \frac{1}{N} \sum_{k \in U} y_k,$$

where y_k denotes the measurement of the survey variable Y of element $k \in U$. A sample $S \subset U$ of size n is selected according to a sampling design $\mathcal{P}$. The first and second order inclusion probabilities defined by $\pi_k := \mathbb{P}(k \in S)$ and $\pi_{kl} := \mathbb{P}(k, l \in S)$ are assumed to be strictly positive for all $k, l \in U$. The sample S is also characterized by the vector of sample selection indicators $\mathbf{I}_U := [I_1, I_2, \ldots, I_N]^\top$, where $I_k := 1$ if $k \in S$ and $I_k := 0$, otherwise.

The Horvitz-Thompson estimator of μ is defined by

$$\widehat{\mu}_\pi := \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k}. \tag{1}$$

In practice, the estimator $\widehat{\mu}_\pi$ in (1) is unfeasible when the survey variable Y suffers from nonresponse. We denote by $\mathbf{r}_U := [r_1, r_2, \ldots, r_N]^\top$ the vector of response indicators, where $r_k := 1$ if y_k is observed and $r_k := 0$, otherwise. The set of respondents and nonrespondents are denoted $S_r := \{k \in S : r_k = 1\}$ and $S_m := \{k \in S : r_k = 0\}$ with respective cardinalities n_r and n_m. We assume that p covariates $X_1, X_2, \ldots, X_p$, with measurements $\mathbf{x}_k$ are observed for every $k \in S$. The observed data are given by

$$\mathcal{D}_{imp} := \{(\mathbf{x}_k, y_k) : k \in S_r\} \cup \{\mathbf{x}_k : k \in S_m\}.$$

The vectors $[\mathbf{x}_k^\top, y_k, r_k]_{k \in U}^\top$ are assumed to be independent and identically distributed. The sampling design $\mathcal{P}$ is assumed to be non-informative, that is, $\pi_k = \mathbb{P}(I_k = 1|X_U, \mathbf{y}_U) = \mathbb{P}(I_k = 1|X_U)$, for every $k \in U$, where X_U and $\mathbf{y}_U$ denote the population design matrix and the population vector of the survey variable, respectively. The reader is referred to Pfeffermann and Sverchkov (2009) for further details. Similarly, the missing mechanism is assumed to satisfy the Missing At Random (MAR, Rubin (1976)) assumption, i.e., $\mathbb{P}(r_k = 1|\mathbf{x}_k, y_k) = \mathbb{P}(r_k = 1|\mathbf{x}_k) := p(\mathbf{x}_k)$. Finally, we assume that there exists a positive constant $\rho > 0$ such that $p(\mathbf{x}) \geq \rho$, almost surely. This is the usual positivity assumption. In this article, we restrict our attention to the customary homoscedastic linear regression model defined as

$$y_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k, \qquad k \in U, \tag{2}$$

where ϵ_k satisfies $\mathbb{E}[\epsilon_k|\mathbf{x}_k] = 0$ and $\mathbb{E}[\epsilon_k^2|\mathbf{x}_k] := \sigma^2$.

2.2. Additional notation

Probabilities, expectations and variances with respect to: (i) the distribution of the covariates' $\mathbb{P}_\mathbf{x}$ will be denoted by the subscript $\mathbf{x}$; (ii) $\mathbb{P}_{y|\mathbf{x}}$ will be denoted by the subscript m; (iii) the sampling design will be denoted by the subscript p; (iv) the nonresponse mechanism will be denoted by the subscript q. For a set of integers A, we write $\mathcal{P}(A)$ to denote its power set, where, by convention, we remove the empty set. For two functions f and g defined on the same domain D, we write $f(x) \overset{\arg\min}{\equiv} g(x)$ if $\arg\min_{x \in D} f(x) = \arg\min_{x \in D} g(x)$. Vectors like $\mathbf{x} \in \mathbb{R}^p$ will be bolded, and their j-th component will be denoted by $x^{(j)}$. We write $\mathbf{x}, y, p(\mathbf{x}), r, \epsilon$ to denote an independent copy of $(\mathbf{x}_k, y_k, p(\mathbf{x}_k), r_k, \epsilon_k)_{k \in U}$.

3. Asymptotically optimal imputation

3.1. Linear regression imputation and variable selection

Linear regression imputation replaces missing values with predictions from a linear regression model fitted using the respondent data. Let $X_r := (\mathbf{x}_k^\top)_{k \in S_r} \in \mathbb{R}^{n_r \times p}$ and $Y_r \in \mathbb{R}^{n_r}$ denote the design matrix and the vector of measurements of Y for the respondents, respectively. Provided that the matrix $X_r^\top X_r$ is positive definite, the ordinary least-squares estimator of $\boldsymbol{\beta}$ in (2) is given by

$$\widehat{\boldsymbol{\beta}} = \left(X_r^\top X_r\right)^{-1} X_r^\top Y_r. \tag{3}$$

It is not uncommon in survey sampling to fit the imputation model by weighted least squares using the design weights $w_k = \pi_k^{-1}$, for $k \in S$. However, since the sampling design is assumed to be non-informative, the unweighted least-squares fit is justified. For algebraic simplicity, we therefore focus on the unweighted estimator in (3). The same conclusions would hold for the weighted version. The linear regression imputed estimator of μ is defined as

$$\widehat{\mu}_{lr} := \frac{1}{N} \left(\sum_{k \in S_r} \frac{y_k}{\pi_k} + \sum_{k \in S_m} \frac{\mathbf{x}_k^\top \widehat{\boldsymbol{\beta}}}{\pi_k} \right).$$

In practice, some coefficients of $\boldsymbol{\beta}$ may be equal to zero. Let $\mathcal{A} \subseteq \mathcal{P}(\{1, 2, \ldots, p\})$ denote a family of candidate models. That is, each element $\alpha \in \mathcal{A}$ is a subset of $\{1, 2, ..., p\}$ which we interpret as the set of included covariates. The complement of a model α is denoted $\alpha^c := \{1, 2, \ldots, p\} \backslash \alpha$. For example, if $p = 3$, the model $\alpha = \{1, 3\}$ includes the covariates X_1 and X_3. Its complement is $\alpha^c = \{2\}$, which corresponds to the model containing only the covariate X_2.

For any $\alpha \in \mathcal{A}$, let $\mathbf{x}_{k, \alpha}$ denote the subvector of covariates for unit k corresponding to the indices in α, and let $X_{r, \alpha} := (\mathbf{x}_{k, \alpha}^\top)_{k \in S_r}$ denote the corresponding design matrix restricted to the respondents.

A model $\alpha \in \mathcal{A}$ is correct if $\|\boldsymbol{\beta}_{\alpha^c}\|_2 = 0$, i.e., α contains all covariates with non-zero

coefficients. The set of correct models is denoted C. The true model $\alpha^\star$ is defined as the support of $\boldsymbol{\beta}$, that is,

$$\alpha^\star := \left\{ j \in \{1, 2, ..., p\} \; ; \; \beta_j \neq 0 \right\}.$$

The true model $\alpha^\star$ is the smallest correct model. Under the assumed linear model, it exists and is unique. A model α is said to be wrong if it is not correct; that is, if it omits at least one covariate whose coefficient in $\boldsymbol{\beta}$ is nonzero. The set of wrong models is denoted $\mathcal{W}$.

To each $\alpha \in \mathcal{A}$, we may associate an ordinary least-squares estimator

$$\widehat{\boldsymbol{\beta}}_\alpha = \left(X_{r,\alpha}^\top X_{r,\alpha} \right)^{-1} X_{r,\alpha}^\top Y_r, \tag{4}$$

leading to the corresponding imputed estimator of μ:

$$\widehat{\mu}_\alpha := \frac{1}{N} \left(\sum_{k \in S_r} \frac{y_k}{\pi_k} + \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha}{\pi_k} \right), \qquad \alpha \in \mathcal{A}.$$

The aim of this article is to determine which estimator in

$$\Gamma_{\mathcal{A}} := \{\widehat{\mu}_\alpha \; ; \; \alpha \in \mathcal{A}\}$$

should be used, and to develop a theoretically sound methodology for conducting inference on μ as efficiently as possible.

3.2. A loss function for imputation

For an arbitrary model α, we may decompose the total error of $\widehat{\mu}_\alpha$ as

$$\widehat{\mu}_\alpha - \mu = \underbrace{(\widehat{\mu}_\alpha - \widehat{\mu}_\pi)}_{\text{imputation/nonresponse error}} + \underbrace{(\widehat{\mu}_\pi - \mu)}_{\text{sampling error}}.$$

The first term reflects the error induced by nonresponse through the imputation procedure (and therefore depends on the chosen model α), whereas the second term is the usual sampling error which does not depend on nonresponse. At the imputation stage, the aim is to reduce the imputation/nonresponse error as much as possible. To this end, we introduce the following loss function for imputation:

$$\mathcal{L}_n(\alpha) := \mathbb{E}_m \left[(\widehat{\mu}_\alpha - \widehat{\mu}_\pi)^2 \right], \qquad \alpha \in \mathcal{A}.$$

The loss $\mathcal{L}$ is a measure of the squared distance between the imputed estimator $\widehat{\mu}_\alpha$ based on model α and the complete data Horvitz-Thompson estimator. It is positive and is minimized for the model α that we would prefer to use for imputation. More formally, we define the optimal imputation model α_{opt} as as any minimizer of $\mathcal{L}_n$, that is,

$$\alpha_{\text{opt}} \in \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}} \, \mathcal{L}_n(\alpha).$$

In principle, one would like to use the imputed estimator $\widehat{\mu}_{\alpha_{\text{opt}}}$ to estimate μ. However, the loss $\mathcal{L}_n$ depends on unobserved quantities and cannot be evaluated from the available data, so α_{opt} is unknown in practice. In the remainder of the paper, we study the properties of $\widehat{\mu}_{\alpha_{\text{opt}}}$ and develop practical model selection procedures to approximate α_{opt} for inference.

Proposition 1. *Fix $\alpha \in \mathcal{A}$. Then, $\mathcal{L}_n(\alpha)$ admits the following closed-form expression:*

$$\mathcal{L}_n(\alpha) \stackrel{\arg\min}{\equiv} \left\{ \left(\sum_{k \in S_m} \frac{\mathbf{x}_k^\top}{\pi_k} - \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top}{\pi_k} A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \right) \boldsymbol{\beta} \right\}^2$$

$$+ \sigma^2 \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top}{\pi_k} \right) A_{r,\alpha}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}}{\pi_k} \right)$$

$$:= \mathcal{L}_{1,n}(\alpha) + \mathcal{L}_{2,n}(\alpha).$$

Proof. See Appendix B.1. □

Proposition 1 yields a useful decomposition of the loss into a bias term and a variance term. The component $\mathcal{L}_{1,n}(\alpha)$ corresponds to the squared model bias induced by imputing under model α, whereas $\mathcal{L}_{2,n}(\alpha)$ captures a variance contribution. For any correct model $\alpha \in C$, we have $\mathcal{L}_{1,n}(\alpha) = 0$, so that correct models contribute only through $\mathcal{L}_{2,n}(\alpha)$. Lemma 11 shows that $\mathcal{L}_{2,n}$ is a strictly increasing set function, implying that adding covariates to the model can only increase $\mathcal{L}_{2,n}$, although it may reduce $\mathcal{L}_{1,n}$. As a result, minimizing $\mathcal{L}_n(\alpha)$ over $\mathcal{A}$ amounts to a bias–variance trade-off. We do not assume uniqueness of the minimizer; any α achieving $\min_{\alpha \in \mathcal{A}} \mathcal{L}_n(\alpha)$ is considered optimal.

Proposition 2. *If $\alpha_{\mathrm{opt}} \in C$, then, almost surely*

$$\mathbb{P}_{pq}\left(\alpha_{\mathrm{opt}} = \alpha^\star \right) = 1.$$

Proof. See Appendix B.1. □

Proposition 2 states that, for any finite sample size, if the loss is minimized by a correct model (i.e., $\alpha_{\mathrm{opt}} \in C$), then the optimal model must be the true model. In particular, $\alpha_{\mathrm{opt}} = \alpha^\star$.

3.3. Asymptotic behavior of the optimal imputation model

In general, investigating the properties of α_{opt} for finite sample sizes is challenging. Nonetheless, we shall show in Section 4 that, asymptotically, $\mathcal{L}_n$ has a unique minimizer, which, under mild conditions, is also the true model.

To that aim, we consider the asymptotic framework of Isaki and Fuller (1982). Let $(U_v)_{v \in \mathbb{N}}$ be an increasing sequence of finite populations, i.e., $U_1 \subset U_2 \subset \ldots$ of respective increasing sizes $(N_v)_{v \in \mathbb{N}}$. In each population U_v, a sample S_v is drawn randomly using a sampling design $\mathcal{P}_v$ to estimate the finite population mean μ_v. All subsequent steps (i.e., imputation, variance estimation, etc.) are carried out at each $v \in \mathbb{N}$ similarly. Note that the number of covariates p remains constant as v increases. In the sequel, for simplicity of notations, we will omit the index v when no confusion arises.

The following assumptions are made about the superpopulation model.

(S1) The distribution of the covariates $\mathbb{P}_{\mathbf{x}}$ is absolutely continuous with respect to the Lebesgues measure.

(S2) The covariates $\mathbf{x}$ have bounded support, that is, there exists $C_0 > 0$ such that, almost surely, $\|\mathbf{x}\|_2 \leq C_0$.

(S3) There exists a constant M_0 such that $\mathbb{E}_m[\epsilon_1^4] \leq M_0$, almost surely.

Assumption (S1) ensures that the matrix $\sum_{k \in U} \mathbf{x}_k \mathbf{x}_k^\top$ is almost surely positive definite. Condition (S2) assumes that the covariates are almost surely bounded. This assumption is not strictly needed, but it greatly simplifies our proofs. An alternative assumption on the moments of $\|\mathbf{x}\|_2$ could be used, at the price of considerably lengthening our arguments. As such, we do not pursue this avenue further. Assumption (S3) is a moment condition that we use in various places. Note that in most of our results, we only require finite second moments. Generally, our conditions on the superpopulation may not be minimal. For brevity and simplicity of notation, we shall assume finite fourth moments directly, which will be needed for variance estimation later.

Remark 1. *Throughout the article, we work on the event $\mathcal{E}_{N_v} := \{\lambda_{min}(N_v^{-1} \sum_{k \in U_v} r_k \mathbf{x}_k \mathbf{x}_k^\top) \geqslant \gamma\}$ for some $\gamma > 0$. We show in appendix (see Lemma 2) that, for some universal constants C_1, C_2, for all large N_v, we have $\mathbb{P}_q(\mathcal{E}_{N_v}^c) \leq C_1 \exp(-C_2 N_v)$ almost surely and $\mathcal{E}_{N_v}$ is thus a high-probability event. This conditioning could be removed by considering a truncated inverse $(N_v^{-1} \sum_{k \in U_v} r_k \mathbf{x}_k \mathbf{x}_k^\top)_{trunc}^{-1} = (N_v^{-1} \sum_{k \in U_v} r_k \mathbf{x}_k \mathbf{x}_k^\top)^{-1} \mathbb{1}_{\mathcal{E}_{N_v}}$. Since the truncated and un-truncated estimators differ only with exponentially low probability, the difference between the two does not affect any of our asymptotic results. Therefore, we proceed by conditioning on $\mathcal{E}_{N_v}$ without loss of generality.*

Similarly, in addition to the conditions on the sampling design described below, we assume that $\mathcal{P}$ induces a similar high-probability event $\{X_r^\top X_r \succ 0\}$ and condition on this event. This is a natural assumption, commonly used in the literature, see e.g., Chauvet and Goga (2022).

We will refer to the following assumptions on the sequence of sampling designs $(\mathcal{P}_v)_{v \in \mathbb{N}}$.

(D1) The sampling fraction satisfies

$$\lim_{v \to \infty} \frac{n_v}{N_v} := f^\star > 0$$

almost surely.

(D2) There exists positive constants $\lambda > 0$ and $\lambda^\star > 0$ such that, for all $v \in \mathbb{N}$,

$$\min_{k \in U_v} \pi_{k,v} \geq \lambda, \qquad \text{and} \qquad \min_{k,l \in U_v} \pi_{kl,v} \geq \lambda^\star$$

almost surely.

(D3) The sampling covariances $\Delta_{kl} := \pi_{kl} - \pi_k \pi_l$, for $k, l \in U_v$, satisfy that there exists a deterministic constant $\bar{\Delta}$ such that, almost surely, for all $v \in \mathbb{N}$,

$$n_v \max_{k,l \in U_v} |\Delta_{kl}| \leq \bar{\Delta}.$$

These regularity conditions are common in the literature. Assumption (D1) states that the sample size n_v grows at the same rate as the population size N_v as v goes to infinity. Assumption (D2) requires that the first and second order inclusion probabilities remain bounded away from

zero. Finally, Assumption (D3) ensures that the sampling covariances decrease to zero at a rate of at least $O(n_v^{-1})$. For a more thorough discussion on these assumptions, we refer the reader to Breidt and Opsomer (2000).

Theorem 1. *Let $(\alpha_{opt,v})_{v \in \mathbb{N}}$ be a sequence of minimizers of $(\mathcal{L}_v)_{v \in \mathbb{N}}$. For $\alpha \in \mathcal{A}$, define*

$$
\mathbf{S}_{U_v} := \sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k) \pi_k}{N_v}
$$
$$
- \sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k) \pi_k}{N_v} \left(\sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k) \pi_k}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k) \pi_k}{N_v}. \quad (5)
$$

Assume (S1)-(S2) and (D1)-(D3) and that for $\alpha \in \mathcal{W}$,

$$
\liminf_{v \to \infty} \lambda_{min} \left(\mathbf{S}_{U_v} \right) > 0 \qquad a.s. \qquad\qquad (C_1)
$$

Then, we have

$$
\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\alpha_{opt,v} = \alpha^\star \right) = 1 \qquad a.s.
$$

Proof. See Appendix B.1.   □

Theorem 1 proves that, among all candidates in $\mathcal{A}$, the model α_{opt} minimizing the (unobservable) imputation loss $\mathcal{L}_n$ eventually is the true model $\alpha^\star$. In other words, asymptotically, the optimal set of covariates for imputation coincides with the true model $\alpha^\star$. Of course, in practice, the true model is unknown, so Theorem 1 is mainly theoretical. Its value lies in guiding the development of practical selection criteria that asymptotically identify the true model. We revisit this point in Section 4.

Assumption (C_1) ensures that the omitted covariates $\mathbf{x}_{\alpha^c}$ retain some variation once the included covariates $\mathbf{x}_\alpha$ have been accounted for. To see this, consider the simple case where $p(\mathbf{x}_k) = p$, $\pi_k = \pi$ for $k \in U$ and the intercept is included in α. Then, some algebra shows that the matrix $\mathbf{S}_{U_v}$ defined in (5) is proportional to the following empirical covariance matrix

$$
\mathbf{S}_{U_v} \propto \frac{1}{N_v} \sum_{k \in U_v} \mathbf{r}_k \mathbf{r}_k^\top
$$

with $\mathbf{r}_k := \mathbf{x}_{k,\alpha^c} - \mathbf{B}_v^\top \mathbf{x}_{k,\alpha}$ with $\mathbf{B}_v$ solving the matrix least squares problem

$$
\mathbf{B}_v = \operatorname*{argmin}_{\mathbf{B} \in \mathbb{R}^{p_\alpha \times p_\alpha^c}} \frac{1}{N_v} \sum_{k \in U_v} \|\mathbf{x}_{k,\alpha^c} - \mathbf{B}^\top \mathbf{x}_{k,\alpha}\|_2^2 = \left(\sum_{k \in U_v} \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \right)^{-1} \sum_{k \in U_v} \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top.
$$

Therefore, $\mathbf{S}_{U_v}$ represents the finite population covariance matrix of the residuals obtained by regressing $\mathbf{x}_{\alpha^c}$ on $\mathbf{x}_\alpha$. The assumption (C_1) requires that every omitted direction must preserve a strictly positive amount of variability after removing the linear effect of the included variables. Intuitively, it rules out cases in which an omitted predictor could be perfectly reconstructed from the included predictors, thereby making some incorrect models behave as if they were

correct. That is, missing covariates would have essentially no missing contribution. In our setting, assumptions (S1) and (S2) imply that $\mathbb{E}_{\mathbf{x}}[\mathbf{x}\mathbf{x}^\top]$ is positive definite, so that any covariate missing will retain some variability, even after accounting for all the other covariates. The next remark formalizes this.

Remark 2. *For an equal probability sampling design, an application of the law of large numbers gives*

$$\boldsymbol{S}_{U_v} \xrightarrow[v\to\infty]{a.s.} \boldsymbol{S}_* := \pi \left(\mathbb{E}_{\mathbf{x}}[\mathbf{x}_{\alpha^c} \mathbf{x}_{\alpha^c}^\top p(\mathbf{x})] - \mathbb{E}_{\mathbf{x}}[\mathbf{x}_{\alpha^c} \mathbf{x}_\alpha^\top p(\mathbf{x})] (\mathbb{E}_{\mathbf{x}}[\mathbf{x}_\alpha \mathbf{x}_\alpha^\top p(\mathbf{x})])^{-1} \mathbb{E}_{\mathbf{x}}[\mathbf{x}_\alpha \mathbf{x}_{\alpha^c}^\top p(\mathbf{x})] \right).$$

Now, consider the block matrix

$$\boldsymbol{M} := \pi \mathbb{E} \begin{bmatrix} p\, \mathbf{x}_\alpha \mathbf{x}_\alpha^\top & p\, \mathbf{x}_\alpha \mathbf{x}_{\alpha^c}^\top \\ p\, \mathbf{x}_{\alpha^c} \mathbf{x}_\alpha^\top & p\, \mathbf{x}_{\alpha^c} \mathbf{x}_{\alpha^c}^\top \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \boldsymbol{C} \end{bmatrix},$$

and note that $\boldsymbol{S}_* = \boldsymbol{C} - \boldsymbol{B}^\top \boldsymbol{A}^{-1} \boldsymbol{B}$ is the Schur complement of $\boldsymbol{A}$ in $\boldsymbol{M}$. Under (S1) and positivity, it can be shown that $\boldsymbol{M}$ is positive definite. This implies in turn that $\boldsymbol{S}_*$ is also positive definite, see e.g., (Horn and Johnson, 2012, p. 495). Therefore, Assumption (C_1) holds in this case. A similar reasoning would hold for Poisson sampling. More generally, Assumption (C_1) is fairly mild and is likely to hold for most sampling designs. In spirit, the assumption is in line with identifiability assumptions of the literature, see e.g., Shao (1993).

We recall that a sequence of estimators $(\widehat{\mu}_v)_{v\in\mathbb{N}}$ is consistent for μ_v if $\widehat{\mu}_v - \mu_v \xrightarrow[v\to\infty]{\mathbb{P}} 0$. Moreover, if $\mathbb{V}(\widehat{\mu}_v)$ exists for all $v \in \mathbb{N}$ and $\mathbb{V}(\widehat{\mu}_v) \xrightarrow[v\to\infty]{} \tau^2$, we call τ^2 the asymptotic variance of $\widehat{\mu}_v$ and write $\mathbb{AV}(\widehat{\mu}_v) := \lim_{v\to\infty} \mathbb{V}(\widehat{\mu}_v)$. Our next result further examines the consequences, in terms of consistency and asymptotic variance, when using an imputed estimator $\widehat{\mu}_\alpha$ of $\widehat{\mu}_v$ with $\alpha \neq \alpha^\star$.

Proposition 3. *Let $\alpha \in \mathcal{A}$ be an arbitrary model and $(\widehat{\mu}_{\alpha,v})_{v\in\mathbb{N}}$ be sequence of imputed estimators. Assume (S1)-(S2), (D1)-(D3), and assume also that the sampling design induces equal first-order inclusion probabilities ($\pi_k = \pi$ for $k \in U_v$) and equal second-order inclusion probabilities ($\pi_{kl} = \pi^\star$ for $k \neq l \in U_v$). Then, we have the following results.*

(i) *The sequence $(\widehat{\mu}_{\alpha,v})_{v\in\mathbb{N}}$ is consistent if and only if*

$$\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}_{\alpha^c}^\top \boldsymbol{\beta}_{\alpha^c} \right] = \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha^\top (1 - p(\mathbf{x})) \right] \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right]^{-1} \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha \mathbf{x}_{\alpha^c}^\top \boldsymbol{\beta}_{\alpha^c} p(\mathbf{x}) \right].$$
$$(C_2)$$

(ii) (a) *Let $\alpha \in C$ be a correct model. Assume that there exists a constant K such that the following limit exists $\lim_{v\to\infty} N_v(\pi^\star - \pi^2) = K$. Then, there exists a constant $C(\alpha^\star)$, not depending of α, and a function $M : C \to \mathbb{R}_+^*$ such that*

$$\mathbb{AV}\left(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v) \right) = C(\alpha^\star) + \frac{\sigma^2}{\pi} M(\alpha),$$

where

$$M : \alpha \mapsto \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha^\top \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha \right].$$

(b) *The function M is a non-decreasing function, that is, for $\alpha_1, \alpha_2 \in C$,*

$$\alpha_1 \subset \alpha_2 \implies M(\alpha_1) \leq M(\alpha_2),$$

with equality if and only if

$$\mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_{\alpha_2 - \alpha_1} \right]$$
$$= \mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_{\alpha_2 - \alpha_1} \mathbf{x}_{\alpha_1}^\top \right] \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha_1} \mathbf{x}_{\alpha_1}^\top \right]^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha_1} \right]. \qquad \text{(C3)}$$

Proof. See Appendix B.1.   □

The restriction on equal inclusion probabilities is essentially technical and could be removed if we considered weighted least-squares in (4) with weights $(1/\pi_k)_{k \in S_r}$. The first statement (i) provides a condition (C_2) under which the imputed estimator $\widehat{\mu}_\alpha$ is consistent. The second statement (ii) derives an expression for the asymptotic variance of the rescaled error $\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v)$ for $\alpha \in C$. By extension, the term "asymptotic variance of a model α" will refer to that of $\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v)$. The quantity $\mathbb{AV}(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v))$ decomposes into two terms, $C(\alpha^\star)$ and $M(\alpha)$. The first does not depend on α. Since $\alpha^\star$ is the smallest correct model, we have $M(\alpha^\star) \leq M(\alpha)$ for every $\alpha \in C$.

The aim of Proposition 3 is therefore to highlight the consequences of using (i) too few covariates or (ii) too many. While these formulas may appear somewhat abstract, we analyze them further below to obtain more interpretable conditions.

Corollary 1. *Consider the set-up of Proposition 3. Assume that either of the following two cases holds:*

(i) *The model is correct, i.e., $\alpha \in C$.*

(ii) *Let $\alpha_{\mathrm{mis}} := \alpha^\star \cap \alpha^c$ denote the set of correct covariates, missing in α. Assume the following conditions:*

   (a) *The correct missing covariates do not influence p, i.e., for all $j \in \alpha_{\mathrm{mis}}$, we have $\left[p(\mathbf{x}) \perp\!\!\!\perp \mathbf{x}_{\alpha^c, j} \right] | \mathbf{x}_\alpha$.*

   (b) *The correct missing covariates are linearly linked to those included, that is, for all $j \in \alpha_{\mathrm{mis}}$, $\mathbb{E}[x_{\alpha^c, j} | \mathbf{x}_\alpha] = \mathbf{x}_\alpha^\top \boldsymbol{\gamma}_j$, for some $\boldsymbol{\gamma}_j \in \mathbb{R}^{p_\alpha}$.*

Then, condition (C_2) holds and thus $\widehat{\mu}_{\alpha,v} - \mu_v \xrightarrow[v \to \infty]{\mathbb{P}} 0$.

Proof. See Appendix B.1.   □

To obtain a consistent estimator, it is therefore sufficient to either: (i) include all covariates related to Y; (ii) include all covariates that are related to both Y and the probability of response, modulo condition (b). The second condition is more difficult to interpret, yet relatively weak, but essential. It states that the omitted covariates relevant to Y must be linearly related to those included. Although this may at first appear to be a technical assumption introduced to simplify the proof, it is in fact required for (C_2) to hold, as illustrated in the next example.

Example 1. *(i). Recall that a missing data mechanism is said to be Missing Completely At Random (MCAR) if $r_k \perp\!\!\!\perp y_k$. Then, condition (a) is trivially satisfied since r_k must be independent of $\mathbf{x}$. If the covariates are independent, then condition (b) reduces to requiring the intercept to be included in α. This requirement is necessary, as illustrated in the following example.*

Consider the case where $p = 2$ with $\mathbf{x}^\top = [x_1, x_2] = [1, x]$, with true model $\alpha^\star = \{1, 2\}$ and candidate $\alpha = \{2\}$, and suppose that $p(\mathbf{x}) = \bar{p}$, $0 < \bar{p} \le 1$. Then, $\alpha_{\mathrm{mis}} = \{1\}$. Condition (a) is satisfied but condition (b) is not since $\mathbb{E}[1|\mathbf{x}_\alpha] = 1$ and, because the intercept is not included, there does not exist a constant γ such that γ such that $1 = x\gamma$ almost surely, unless x is almost surely constant. Condition (b) is indeed needed since it can be shown that

$$\widehat{\mu}_{\alpha,v} - \mu_v \xrightarrow[v\to\infty]{\mathbb{P}} -\beta_0 (1 - \bar{p}) \frac{\mathbb{V}_{\mathbf{x}}(x)}{\mathbb{E}_{\mathbf{x}}[x^2]},$$

where β_0 denotes the true intercept coefficient. This asymptotic bias is non-zero whenever $\beta_0 \ne 0$, $\bar{p} \ne 1$ and $\mathbb{V}_{\mathbf{x}}(x) \ne 0$; note that if x has zero variance then there exists γ such that $1 = x\gamma$, and consistency holds.

(ii). If the intercept is included and $\mathbf{x}$ has an elliptically symmetric distribution (e.g., multivariate Gaussian), or if its components are independent, then condition (b) is always satisfied and (a) is sufficient to ensure consistency.

In the following corollary, we examine the behavior of the asymptotic variance obtained in statement (ii) of Proposition 3.

Corollary 2. *Consider the setup of Proposition 3 with $\alpha_1, \alpha_2 \in C$ such that $\alpha_1 \subset \alpha_2$. Consider the following assumptions.*

(i) Assume that:

(a) For all $j \in \alpha_2 - \alpha_1$, $\left[p(\mathbf{x}) \perp\!\!\!\perp x_{\alpha_2-\alpha_1,j} \right] | \mathbf{x}_{\alpha_1}$.

(b) For all $j \in \alpha_2 - \alpha_1$, $\mathbb{E}\left[x_{\alpha_2-\alpha_1,j} | \mathbf{x}_{\alpha_1} \right] = \mathbf{x}_{\alpha_1}^\top \boldsymbol{\gamma}_j$, for some $\boldsymbol{\gamma}_j \in \mathbb{R}^{p_{\alpha_1}}$.

Then, (C3) holds and the asymptotic variances of the models based on α_1 and α_2 are the same.

(ii) Assume that there exists a direction $\mathbf{c} \in \mathbb{R}^{\alpha_2-\alpha_1}$ such that

(a) $\mathbb{E}_{\mathbf{x}}[p(\mathbf{x})\mathbf{c}^\top \mathbf{x}_{\alpha_2-\alpha_1}\mathbf{x}_{\alpha_1}^\top] = \mathbf{0}$.

(b) $\mathbb{E}_{\mathbf{x}}[(1 - p(\mathbf{x}))\mathbf{c}^\top \mathbf{x}_{\alpha_2-\alpha_1}] \ne 0$.

Then, (C3) does not hold and α_2 has a strictly greater variance than α_1.

Proof. See Appendix B.1. □

In the above corollary, we make explicit some sufficient conditions under which one "pays a price" when adding superfluous covariates, and when one does not. Specifically, in part (i), our conditions mean that: (a) the covariates added in α_2 (compared to α_1) do not explain the nonresponse mechanism, given the covariates in α_1. This occurs for instance when

$p(\mathbf{x}) = p(\mathbf{x}_{\alpha_1})$. Condition (b) means that the addtional covariates are linearly related to those already included in α_1. Together, these conditions describe a situation in which the covariates in $\alpha_2 - \alpha_1$ are both "uninformative, in mean" and "irrelevant" for the nonresponse mechanism, once the covariates in α_1 are included. For example, consider the true model $\alpha^\star = \{1, 2\}$ with $X_1 = 1$, and suppose that X_2 and X_3 are independent. Then condition (b) is satisfied, and if $p(\mathbf{x}) = p(X_1, X_2)$, using X_1, X_2, and X_3 leads to the same asymptotic variance as using only X_1 and X_2. In part (ii), we move in the opposite direction and provide an example where adding superfluous covariates increases the variance. To better interpret these conditions, consider a new random variable $z = \mathbf{c}^\top \mathbf{x}_{\alpha_2 - \alpha_1}$, representing a direction in the space spanned by the added covariates $\alpha_2 - \alpha_1$. The first condition means that z is a new direction, not captured by the covariates in α_1, among respondents, while the second requires that there be a nonzero nonresponse signal along this direction.

4. A methodology for asymptotically optimal variable selection

Theorem 1 establishes that the true model is optimal for imputation under the loss function $\mathcal{L}$. In practice, however, the true model is unknown, and the corresponding imputed estimator $\widehat{\mu}_{\alpha^\star}$ serves only as an oracle benchmark. In this section, we develop a practical procedure that attains the same asymptotic efficiency as this oracle imputed estimator and yields asymptotically valid confidence intervals.

4.1. Asymptotic equivalence with the oracle

Model selection for identification (i.e., recovering the true non-zero coefficients of a linear regression model) has been extensively studied for i.i.d. data; we refer the reader to Shao (1993, 1997) or Rao et al. (2001) for a textbook discussion. We call a model selection criterion any (possibly data dependent) measurable map $C_n : \mathcal{A} \to \mathbb{R}$ used to select a model $\alpha \in \mathcal{A}$, that is, to choose $\widehat{\alpha}_{C_n} \in \mathcal{A}$ satisfying

$$\widehat{\alpha}_{C_n} \in \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}}\, C_n(\alpha).$$

Common model selection criteria for linear models include the Akaike Information Criterion (AIC, Akaike (1970)), the Bayesian Information Criterion (BIC, Schwarz (1978)), cross-validation, among many others. On i.i.d. data, a model selection criterion C_n is said to be consistent if

$$\lim_{n \to \infty} \mathbb{P}_m\left(\widehat{\alpha}_{C_n} = \alpha^\star\right) = 1$$

almost surely. That is, as the sample size n diverges to infinity, the model selection tends to only select the true model. Not all model selection criteria are consistent; for instance, under appropriate conditions, the BIC criterion is consistent, while leave-one-out cross-validation is not. The next lemma formalizes that, under the MAR and non-informativeness assumptions, if C_n is consistent on i.i.d. data, then it is also consistent with survey data.

Lemma 1. *Let C_{U_v} be a model selection criterion fitted on $\{(\mathbf{x}_k, y_k) : k \in U_v\}$ and C_{S_r} be the corresponding algorithm fitted on $\{(\mathbf{x}_k, y_k) : k \in S_{r,v}\}$. Denote by $\widetilde{\alpha}_{U_v}$ and $\widehat{\alpha}_{S_{r,v}}$ their*

respective minimizers. Assume (S1)-(S2), (D1)-(D3).
If

$$\lim_{v \to \infty} \mathbb{P}_m \left(\widetilde{\alpha}_{U_v} = \alpha^\star \right) = 1, \qquad a.s.$$

then

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\widehat{\alpha}_{S_r,v} = \alpha^\star \right) = 1, \qquad a.s.$$

Proof. See Appendix B.1. □

Lemma 1 formalizes the intuition that model-selection consistency transfers from the population to the sample of respondents, provided that the sampling design and the nonresponse mechanism do not shift the conditional distribution $\mathbb{P}_m$. Although this result is quite natural, it plays an important role in the developments that follow. A natural extension of Lemma 1 would be to consider informative sampling designs with appropriately weighted model selection criteria. Such criteria have been suggested, for instance, in Lumley and Scott (2015); Wieczorek, Guerin and McMahon (2022); Iparragirre et al. (2023); however, to the best of our knowledge, their consistency has not yet been formally established. This question is beyond the scope of this article and will be relegated to future work.

The first step in our methodology is to use a consistent model selection criterion to select a model $\widehat{\alpha}$. Then, to use the imputed estimator $\widehat{\mu}_{\widehat{\alpha}}$ based on the selected set of covariates $\widehat{\alpha}$. This will asymptotically lead to an optimal estimator among all possible models.

Theorem 2. *Let $(\widehat{\alpha}_v)_{v \in \mathbb{N}}$ be a sequence of models selected by a consistent model selection procedure and $(\widehat{\mu}_{\widehat{\alpha},v})_{v \in \mathbb{N}}$ be the corresponding sequence of imputed estimators. Assume (S1)-(S2), (D1)-(D3). Then,*

$$\sqrt{n_v} \left(\widehat{\mu}_{\widehat{\alpha},v} - \mu_v \right) = \sqrt{n_v} \left(\widehat{\mu}_{\alpha^\star,v} - \mu_v \right) + o_{\mathbb{P}}(1).$$

Proof. See Appendix B.1. □

The above states that, once a model selection procedure asymptotically identifies the true model $\alpha^\star$ with probability one, the asymptotic distribution of the feasible estimator $\widehat{\mu}_{\widehat{\alpha}}$ is the same as that of the oracle estimator. In other words, we can proceed with inference as if the true model were known a priori. This legitimizes the use of standard model-selection tools, such as BIC, for survey imputation and shows that the bias or extra variability introduced by using a data-driven model $\widehat{\alpha}$ disappears asymptotically. The key idea of the proof is that, on the event $\widehat{\alpha}_v = \alpha^\star$, we have $\widehat{\mu}_{\widehat{\alpha}} = \widehat{\mu}_{\alpha^\star}$. Because this event has asymptotic probability 1, what happens on the complement does not matter.

4.2. Consistent variance estimation

The second step of the methodology is to perform "classical" variance estimation, based on the model $\widehat{\alpha}$ selected by a consistent model selection procedure. By classical, we mean using the same variance estimators as traditionally used, but with model $\widehat{\alpha}$ instead of all available

covariates. In this article, we focus on the reverse approach (Fay, 1991; Shao and Steel, 1999), although a similar approach with the method of Särndal (Särndal, 1992) could also be used. For additional details on variance estimation with the reverse approach, we refer the reader to Kim and Rao (2009) and Haziza and Vallée (2020). The next theorem establishes that, for any $\alpha \in C$, there exists a first-order asymptotically equivalent linear (in the sampling indicators) estimator of $\widehat{\mu}_\alpha$. More specifically, let

$$\widetilde{\mu}_{\alpha,v} := \frac{1}{N_v} \sum_{k \in S_v} \frac{\eta_{k,\alpha}}{\pi_k},$$

denote the linearized imputed estimator with $\eta_{k,\alpha} = \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha + r_k(1 + \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha})\epsilon_k$ and

$$\mathbf{c}_{\alpha,v} = \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{N_v}. \tag{6}$$

The result below closely parallels Theorem 1 of Kim and Rao (2009); however, their proof does not apply directly to our setting because we use an unweighted least-squares estimator $\widehat{\boldsymbol{\beta}}$, whereas they rely on a weighted estimator. For completeness, we extend their theorem to unweighted regression imputation.

Theorem 3. *Let $\alpha \in C$ be a correct model and consider sequences $(\widehat{\mu}_{\alpha,v})_{v \in \mathbb{N}}$ and $(\widetilde{\mu}_{\alpha,v})_{v \in \mathbb{N}}$. Under assumptions (S1)-(S3) and (D1)-(D3), we have*

$$\sqrt{n_v}\left(\widehat{\mu}_{\alpha,v} - \widetilde{\mu}_{\alpha,v}\right) = o_\mathbb{P}(1). \tag{7}$$

Proof. See Appendix B.1. □

To perform variance estimation, consider the following decomposition of $V_T(\alpha)$, the variance of $\widetilde{\mu}_\alpha - \mu$:

$$
\begin{aligned}
V_T(\alpha) = \mathbb{V}_{mpq}(\widetilde{\mu}_\alpha - \mu) &= \mathbb{E}_q\left[\mathbb{E}_m\left[\mathbb{V}_p(\widetilde{\mu}_\alpha - \mu)\right]\right] + \mathbb{E}_q\left[\mathbb{V}_m\left(\mathbb{E}_p[\widetilde{\mu}_\alpha - \mu]\right)\right] \\
&\quad + \mathbb{V}_q\left(\mathbb{E}_m\left[\mathbb{E}_p[\widetilde{\mu}_\alpha - \mu]\right]\right) \\
&:= V_1(\alpha) + V_2(\alpha) + V_3(\alpha).
\end{aligned}
$$

Noting that for $\alpha \in C$,

$$\mathbb{E}_{mp}[\widetilde{\mu}_\alpha - \mu] = 0,$$

it follows that

$$\mathbb{V}_{mpq}(\widetilde{\mu}_\alpha - \mu) = V_1(\alpha) + V_2(\alpha).$$

We start by estimating $\boldsymbol{\beta}_\alpha$ by $\widehat{\boldsymbol{\beta}}_\alpha$ and $\boldsymbol{c}_\alpha$ with

$$\widehat{\mathbf{c}}_\alpha = \left(\sum_{k \in S} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N} \right)^{-1} \sum_{k \in S} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{N \pi_k}. \tag{8}$$

The estimator $\widehat{V}_1(\alpha)$ of $V_1(\alpha)$ is defined as

$$\widehat{V}_1(\alpha) = \frac{1}{N^2} \sum_{k \in S} \sum_{l \in S} \frac{\Delta_{kl}}{\pi_{kl}} \frac{\widehat{\eta}_k}{\pi_k} \frac{\widehat{\eta}_l}{\pi_l},$$

with

$$\widehat{\eta}_{k,\alpha} = \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha + r_k \left(1 + \pi_k \widehat{\mathbf{c}}_\alpha^\top \mathbf{x}_{k,\alpha}\right) \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right), \qquad k \in S. \tag{9}$$

To estimate $V_2(\alpha)$, we start by estimating σ^2 with

$$\widehat{\sigma}_\alpha^2 = \sum_{k \in S} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right)^2}{n_r - p_\alpha}. \tag{10}$$

We then estimate $V_2(\alpha)$ by

$$\widehat{V}_2(\alpha) = \widehat{\sigma}_\alpha^2 \sum_{k \in S} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_\alpha^\top \mathbf{x}_{k,\alpha}\right)^2}{N^2 \pi_k}.$$

Finally, for $\alpha \in C$, an estimator of the total variance, $\mathbb{V}_{mpq}(\widehat{\mu}_\alpha - \mu)$, is given by $\widehat{V}_T(\alpha) = \widehat{V}_1(\alpha) + \widehat{V}_2(\alpha)$. This is, for a fixed $\alpha \in C$, the customary variance estimator obtained via the reverse approach; see, e.g., Kim and Rao (2009); Haziza and Vallée (2020) for additional details. Although widely used, the consistency of the variance estimator $\widehat{V}_T$ has, to our knowledge, not been rigorously established. The next result provides a formal justification. To this end, we impose the following additional assumption on the sampling design:

(D4) Let D_{4,N_v} be the set of the distinct of 4-tuples from U_v. The sampling design satisfies

$$\lim_{v \to \infty} \max_{(i,j,k,l) \in D_{4,N_v}} |(I_i I_j - \pi_{ij})(I_k I_l - \pi_{kl})| = 0,$$

almost surely.

This assumption is commonly used to establish the consistency of the Horvitz-Thompson variance estimator, see, e.g., Breidt and Opsomer (2000).

Theorem 4. *Let $\alpha \in C$ and consider a sequence of variance estimators $\{\widehat{V}_{T,v}(\alpha)\}_{v \in \mathbb{N}}$. Assume (S1)-(S3) and (D1)-(D4), and that there exists $c_{1,\alpha}$ and $c_{2,\alpha}$ such that*

$$N_v \mathbb{V}_p \left[(\tilde{\mu}_{\alpha,v} - \mu_v)\right] \xrightarrow[v \to \infty]{\mathbb{P}} c_{1,\alpha}, \qquad N_v \mathbb{V}_m \left(\mathbb{E}_p \left[\tilde{\mu}_{\alpha,v} - \mu_v\right]\right) \xrightarrow[v \to \infty]{\mathbb{P}} c_{2,\alpha}.$$

Then, we have

$$n_v \left|\widehat{V}_{T,v}(\alpha) - V_{T,v}(\alpha)\right| = o_{\mathbb{P}}(1).$$

Proof. See Appendix B.1. □

We propose estimating the variance based on the model $\widehat{\alpha}$ selected by a consistent model selection criterion. The next result establishes the validity of this approach.

Theorem 5. *Let $\{\widehat{\alpha}_v\}_{v \in \mathbb{N}}$ be a sequence of models selected by a consistent model selection procedure, and $\{\widehat{V}_T(\widehat{\alpha}_v)\}_{v \in \mathbb{N}}$ be the corresponding sequence of variance estimators. Assume (S1)-(S3), (D1)-(D4), and that there exists a constant $c > 0$ such that, almost surely,*

$$\lim_{v \to \infty} n_v V_{T,v}(\alpha^\star) = c.$$

Then, the estimator $\widehat{V}_{T,v}(\widehat{\alpha}_v)$ is consistent, that is,

$$\frac{\widehat{V}_{T,v}(\widehat{\alpha}_v)}{V_{T,v}(\alpha^\star)} \xrightarrow[v\to\infty]{\mathbb{P}} 1.$$

Proof. See Appendix B.1. □

Theorem 5 shows that the variance estimator computed under the selected model $\widehat{\alpha}_v$ is asymptotically equivalent to the variance that would be obtained under the true model $\alpha^\star$. In other words, using a consistent model selection procedure does not affect first-order variance estimation: we may estimate the variance as if the true model were known.

4.3. Asymptotically valid and optimal confidence intervals

The final step of the proposed methodology is to derive asymptotically valid confidence intervals. To this end, we first establish the asymptotic distribution of the imputed estimator based on a consistent model selection procedure.

Theorem 6. *Consider a sequence $(\widehat{\alpha}_v)_{v\in\mathbb{N}}$ of consistent models and let $(\widehat{\mu}_{\widehat{\alpha}})_{v\in\mathbb{N}}$ and $(\widehat{V}_{T,v}(\alpha))_{v\in\mathbb{N}}$ be the corresponding point and variance estimators, respectively. Let $\mathcal{B}_v :=\sigma\left((r_k, \mathbf{x}_k, \epsilon_k)_{k\in U_v}\right)$. Assume (S1)-(S3), (D1)-(D4) and the following conditions.*

(i) The oracle estimator $\widetilde{\mu}_{\alpha^\star,v}$ satisfies a design central limit theorem, that is,

$$\mathbb{V}_p\left(\widetilde{\mu}_{\alpha^\star,v}\right)^{-1/2}\left(\frac{1}{N_v}\sum_{k\in U_v}\left(\frac{I_k}{\pi_k}-1\right)\eta_{k,\alpha^\star}\right)\Bigg|\,\mathcal{B}_v \xrightarrow[v\to\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

(ii) There exists a constant $c_3 > 0$ such that

$$N_v\mathbb{V}_p\left(\widetilde{\mu}_{\alpha^\star,v}\right) \xrightarrow[v\to\infty]{\mathbb{P}} c_3.$$

(iii) The first-order inclusion probabilities $(\pi_k)_{k\in U_v}$ are such that there exist constant positive definite matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ such that

$$\frac{1}{N_v}\sum_{k\in U_v} r_k\pi_k\mathbf{x}_{k,\alpha^\star}\mathbf{x}_{k,\alpha^\star}^\top \xrightarrow[v\to\infty]{\mathbb{P}} \mathbf{C}_1, \qquad \frac{1}{N_v}\sum_{k\in U_v} r_k\pi_k^2\mathbf{x}_{k,\alpha^\star}\mathbf{x}_{k,\alpha^\star}^\top \xrightarrow[v\to\infty]{\mathbb{P}} \mathbf{C}_2.$$

Then,

$$\frac{\widehat{\mu}_{\widehat{\alpha},v}-\mu_v}{\sqrt{\widehat{V}_{T,v}(\widehat{\alpha}_v)}} \xrightarrow[v\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Proof. See Appendix B.1. □

The regularity conditions (i) - (iii) in Theorem 6 are standard and fairly weak. Condition (i) means that the Horvitz-Thompson estimator with values $(\eta_{k,\alpha^\star})_{k\in U_v}$ is asymptotically normal. The asymptotic normality of Horvitz-Thompson estimators has been established for some commonly used designs: Hájek (1960) for simple random sampling without replacement,

Hájek (1964) for conditional Poisson sampling, Bickel and Freedman (1984) for stratified sampling, and Krewski and Rao (1981) for probability-proportional-to-size cluster sampling with replacement. For instance, in case of simple random sampling without replacement, is is enough to show that, (see, e.g. Thompson (1997), page 59) a.s., $\mathbb{E}_{mpq}[\eta_{k,\alpha}^{2+\delta}] < \infty$ for some $\delta > 0$. In our setting, this condition can be verified to hold. The two other conditions can also be shown to hold in simple random sampling and other common sampling designs.

The proposed methodology is summarized in Algorithm 1. The central idea is that when the imputation model is selected using a consistent model selection criterion, the effect of model uncertainty becomes asymptotically negligible. More precisely, the selected model coincides with the true model with probability tending to one, so that standard inference procedures applied conditionally on the selected model remain asymptotically valid. Consequently, point and variance estimation may be carried out as if the selected model were known in advance. This greatly simplifies inference after model selection: asymptotically valid confidence intervals can be obtained without the need for post-selection corrections or more elaborate resampling schemes, while retaining full efficiency.

Combining our previous results yields the following property for the confidence intervals obtained by Algorithm 1.

Algorithm 1 Variable selection and inference for linear regression imputation

---

Input: The observed data $\{(\mathbf{x}_k, y_k, \pi_k) : k \in S_r\} \cup \{(\mathbf{x}_k, \pi_k) : k \in S_m\}$, a family of candidates $\mathcal{A}$, and a consistent model selection criterion C_n.

Step 1. Model selection. Select the covariate set

$$\widehat{\alpha} = \arg \min_{\alpha \in \mathcal{A}} C_n(\alpha),$$

where C_n is a consistent selection criterion.

Step 2. Point estimation. Compute the imputed estimator $\widehat{\mu}_{\widehat{\alpha}}$ using the model $\widehat{\alpha}$ selected in Step 1.

Step 3. Variance estimation. Estimate the total variance $\widehat{V}_T(\widehat{\alpha})$ using the variance estimators corresponding to the model $\widehat{\alpha}$.

Step 4. Confidence interval. Construct the $(1 - 2\alpha)$ interval

$$\mathrm{CI}_{1-\alpha}(\widehat{\mu}_{\widehat{\alpha}}) = \left[\widehat{\mu}_{\widehat{\alpha}} - z_{1-\alpha/2} \sqrt{\widehat{V}_T(\widehat{\alpha})} \, , \ \ \widehat{\mu}_{\widehat{\alpha}} + z_{1-\alpha/2} \sqrt{\widehat{V}_T(\widehat{\alpha})} \right],$$

where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$-quantile of $\mathcal{N}(0, 1)$.

Output: The $100(1 - \alpha)\%$ confidence interval $\mathrm{CI}_{1-\alpha}(\widehat{\mu}_{\widehat{\alpha}})$ for the finite population mean μ.

Corollary 3. *Consider a sequence of confidence intervals* $(\mathrm{CI}_{1-\alpha}(\widehat{\mu}_{\widehat{\alpha}}))_{v \in \mathbb{N}}$ *obtained via Algorithm 1. Assume the conditions of Theorem 6. Then,*

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\mu_v \in \mathrm{CI}_{1-\alpha}(\widehat{\mu}_{\widehat{\alpha}_v}) \right) = 1 - \alpha$$

almost surely. Moreover, the confidence intervals are asymptotically optimal, attaining the minimum width achievable under any competing model.

The asymptotic coverage follows immediately from Theorem 6. The optimality follows from the fact that $\widehat{\mu}_{\widehat{\alpha}}$ has the same asymptotic variance as $\widehat{\mu}_{\alpha^\star}$, which by Proposition 3 is the lowest since M is non-decreasing.

5. Simulation Studies

In this section, we present the results of simulation studies evaluating the efficiency of the proposed methodology. We first examine the behavior of the loss function $\mathcal{L}_n(\alpha)$ across different model specifications, the efficiency of point estimators based on various model selection criteria, and then assess the performance of their variance estimators, including their ability to achieve the desired asymptotic coverage for the confidence intervals obtained from Algorithm 1.

In our simulation experiments, we adopted a finite-population framework. Specifically, for each scenario, we generated multiple finite populations of large size N from the assumed superpopulation model. From each generated population, a sample of size n was selected according to the specified sampling design. Item nonresponse was then generated within each selected sample according to the prescribed response mechanism, after which the proposed imputation procedure was applied. This entire process—population generation, sampling, nonresponse generation, and imputation—was repeated 20,000 times to evaluate the finite-sample performance of the proposed methodology.

5.1. Simulation set-up

We generated finite populations of sizes $N = (1000, 2000, 5000)$ and $p = 20$ independent covariates from a Gamma distribution with shape parameter 5 and scale parameter 2. The survey variable was generated according to

$$y_k = \mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k, \qquad k \in U,$$

with $\boldsymbol{\beta} = [10, 9, 9, 8, 8, 7, \mathbf{0}_{14}^\top]^\top$ and $\epsilon_k \sim \mathcal{N}(0, 3600)$ for $k \in U$. The first six covariates, therefore, correspond to signal variables, while the remaining fourteen covariates are noise variables with no effect on the outcome.

We considered three sample sizes $n = (100, 200, 500)$ under the following designs:

(i) Simple random sampling without replacement.
(ii) Stratified sampling: the population was first sorted by $-(3x_1 + 2x_2 + 4x_3 + 5x_4)$. It was then partitioned into $H = 4$ strata defined sequentially from the ordered list, containing respectively 50%, 25%, 20%, and 5% of the population units. In each stratum, the sample S_h was selected by simple random sampling without replacement of size n_h based on x_2-optimal allocation.

These sample sizes were chosen to correspond to the population sizes $N = (1000, 2000, 5000)$, yielding a constant sampling fraction of 10%. In other words, both the population size and

the sample size increase at the same rate across scenarios while maintaining a fixed sampling fraction.

We choose an embedded collection of models $\mathcal{A}$ consisting of the following 20 models

$$\underbrace{\{1\}}_{:=\alpha_1} \subseteq \underbrace{\{1,2\}}_{:=\alpha_2} \subseteq \cdots \subseteq \underbrace{\{1,2,\ldots,20\}}_{:=\alpha_{20}}.$$

Although we do not write it explicitly, note that the intercept is included in every model.

Response indicators were generated with

$$\text{logit}(p(\mathbf{x}_k)) = 0.1 \times \left(-70 + \mathbf{x}_k^\top \boldsymbol{\zeta}\right), \qquad k \in U,$$

with $\boldsymbol{\zeta} = \left[1,1,1,1,0,0,1,1,1,\mathbf{0}_{11}^\top\right]^\top$. This led to a response rate of approximately 50%.

5.2. Behavior of $\mathcal{L}_n(\alpha)$ with different models

We first investigated the behavior of $\mathcal{L}_n(\alpha)$ for $\alpha \in \mathcal{A}$. Here, we present only the results for the case $N = 5000$ and $n = 500$. The results obtained for the other combinations of N and n were very similar and are therefore not reported here for brevity. Given that the loss function $\mathcal{L}_n(\alpha)$ is unknown in practice, we estimated it by a Monte-Carlo approximation with $B = 20,\ 000$ iterations.

As a measure of bias of a point estimator, we used the Monte-Carlo Relative Bias (RB) defined by

$$\text{RB}(\widehat{\mu}) = 100\% \times \frac{1}{B} \sum_{b=1}^{B} \frac{\widehat{\mu}^{(b)} - \mu^{(b)}}{\mu^{(b)}},$$

where $\widehat{\mu}^{(b)}$ denotes an arbitrary estimator and $\mu^{(b)}$ denotes the finite population mean at iteration b, respectively. As a measure of efficiency, we computed the Monte-Carlo Relative Efficiency (RE) defined by

$$\text{RE}(\widehat{\mu}) = 100\% \times \frac{\sum_{b=1}^{B}(\widehat{\mu}^{(b)} - \mu^{(b)})^2}{\sum_{b=1}^{B}(\widehat{\mu}_\pi^{(b)} - \mu^{(b)})^2},$$

where $\widehat{\mu}_\pi^{(b)}$ denotes the HT estimator in the b-th replication. The results of the simulation are presented in Table 1. We start by observing that the rankings based on our proposed loss $\mathcal{L}_n$ and on the relative efficiency (RE) were perfectly aligned: under both sampling designs, the 20 models were ranked in exactly the same order. This confirms that the loss $\mathcal{L}_n$ behaves as expected and that the model minimizing $\mathcal{L}_n$, which we refer to as the optimal imputation model, is indeed the best choice in practice.

Next, recall that the models labeled α_1–α_5 are misspecified, as they fail to include at least one of the important predictors among $X_2, \ldots, X_6$. The variables X_1 to X_4 are also correlated with the missingness mechanism. Consequently, since these variables are associated with the

Table 1
Loss $\mathcal{L}_n(\widehat{\mu}_\alpha)$, relative biases and relative efficiencies across models for sample size $n = 500$.

Model	$\mathcal{L}_n(\alpha)$	RB	RE	$\mathcal{L}(\alpha)$	RB	RE
		SRSWOR			Stratified	
α_1	491.3	4.3	2267.1	587.5	4.7	3940.9
α_2	235.5	2.9	1139.4	294.8	3.2	2023.6
α_3	63.6	1.3	381.3	89.2	1.7	677.3
α_4	15.5	0.1	166.6	16.3	0.0	202.5
α_5	12.1	0.0	152.2	12.7	0.0	180.1
α_6	9.7	0.0	142.2	10.0	0.0	162.6
α_7	10.3	0.0	144.5	10.5	0.0	165.9
α_8	11.0	0.0	147.7	11.3	0.0	170.7
α_9	11.7	0.0	150.7	11.9	0.0	174.9
α_{10}	11.8	0.0	151.0	12.0	0.0	175.3
α_{11}	11.9	0.0	151.1	12.0	0.0	175.5
α_{12}	11.9	0.0	151.3	12.1	0.0	175.9
α_{13}	11.9	0.0	151.5	12.1	0.0	176.2
α_{14}	12.0	0.0	151.7	12.2	0.0	176.6
α_{15}	12.0	0.0	151.9	12.2	0.0	177.0
α_{16}	12.1	0.0	152.3	12.3	0.0	177.4
α_{17}	12.1	0.0	152.5	12.4	0.0	177.7
α_{18}	12.2	0.0	152.6	12.4	0.0	178.1
α_{19}	12.3	0.0	153.0	12.4	0.0	178.5
α_{20}	12.3	0.0	153.2	12.5	0.0	178.8

response mechanism, omitting them from the imputation model may induce a bias that does not vanish asymptotically, leading to an inconsistent estimator, as indicated by Corollary 1. In contrast, once X_1, X_2, X_3, and X_4 are included in the imputation model, no asymptotic bias is expected. To illustrate this, consider for example model α_4, for which $\alpha_{\text{mis}} = \{5, 6\}$. Since all covariates are independent, Condition (b) of Corollary 1 is satisfied: X_5 and X_6 do not explain $p(\mathbf{x})$ conditional on $\mathbf{x}_{\alpha_4}$. Moreover,

$$\mathbb{E}[X_5 \mid X_1, X_2, X_3, X_4] = \mathbb{E}[X_5] \quad \text{and} \quad \mathbb{E}[X_6 \mid X_1, X_2, X_3, X_4] = \mathbb{E}[X_6]$$

are constants. Because the intercept is included in the model, both conditions of Corollary 1 hold for α_4. The simulations clearly reflect this behavior: models α_1 to α_3 were biased, and exhibited poor efficiency with values of RE ranging from 381.3 to 2267.1 for simple random sampling without replacement. As soon as X_4 was added to α_3, the bias vanished. However, the model α_4 was not the most efficient. Indeed, the true model, α_6, was the most efficient, thereby illustrating Theorem 1.

The model α_6 is the true model. As expected, it exhibited negligible bias, the smallest $\mathcal{L}_n(\alpha)$, and the highest efficiency. This is consistent with Theorem 1, which suggests that the true model should also be the optimal model for imputation, as observed here. Recall that the models $\alpha_7 - \alpha_{20}$ belong to the set of correct models, and thus their biases are also negligible. Moreover, $\mathcal{L}_n(\alpha_j) \geq \mathcal{L}_n(\alpha_6)$ for $j = 7, \ldots, 20$, as explained by Proposition 2. While $X_7 - X_9$ explains $p(\mathbf{x})$, including these variables in the imputation model leads to an appreciable increase in $\mathcal{L}_n(\alpha)$ and in the relative efficiency of the resulting estimator. Furthermore, Proposition 3 implies that the asymptotic variance of the imputed estimator based on $\alpha_7 - \alpha_9$ is larger than

that based on α_6, resulting in lower efficiency for these models. In contrast, $X_{10} - X_{20}$ do not explain $p(\mathbf{x})$, conditional on $\mathbf{x}_{\alpha_9}$. Moreover, $\mathbb{E}_{\mathbf{x}}[X_j|\mathbf{x}_{\alpha_9}] = \mathbb{E}_{\mathbf{x}}[X_j]$ for $j = 10, \ldots, 20$, which are constants. Because the intercept is included in the model, both conditions of Corollary 2 hold for α_9. Hence, the efficiency of the imputed estimators based on $\alpha_{10} - \alpha_{20}$ is not substantially different from that obtained under α_9.

5.3. *Point estimation with model selection criteria*

We now study the behavior of imputed estimators based on models selected by commonly used model selection criteria such as AIC, BIC, and *K*-folds cross-validation with $k = 5$.

To further investigate the model selection capabilities of each of the criteria, we computed the Monte-Carlo identification probability defined by

$$\mathbb{P}_{MC}\left(\widehat{\alpha} = \{\alpha^{\star}\}\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\widehat{\alpha}^{(b)} = \alpha^{\star}\right),$$

where $\widehat{\alpha}$ denotes a model selection criterion. We present the results in Table 2.

Table 2

Performance of imputation under different model selection criteria across different sample sizes. Note: we used the notation $C^{-} = C - \alpha^{\star}$ to denote the set of overfitted models, i.e., models that contain all true covariates, and additional superfluous ones.

Sample size	Criteria	RB	RE	$\mathcal{W}$	$\alpha^{\star}$	C^{-}	RB	RE	$\mathcal{W}$	$\alpha^{\star}$	C^{-}
				SRSWOR					Stratified		
	AIC	0.0	164.1	1.6	49.9	48.5	0.0	193.8	1.5	52.2	46.3
$n = 100$	BIC	0.0	151.5	7.5	82.0	10.5	0.0	176.3	7.0	83.0	10.0
	Cross-validation	0.0	157.7	3.9	37.5	58.6	0.0	185.7	3.5	37.5	59.0
	True model	0.0	147.9	-	-	-	0.0	170.2	-	-	-
	AIC	0.0	150.5	0.0	63.9	36.1	0.0	170.5	0.0	63.7	36.3
$n = 200$	BIC	0.0	145.0	0.5	94.5	5.0	0.0	163.6	0.3	94.9	4.8
	Cross-Validation	0.0	150.8	0.2	39.2	60.6	0.0	170.6	0.2	39.1	60.7
	True model	0.0	143.6	-	-	-	0.0	162.3	-	-	-
	AIC	0.0	146.8	0.0	69.2	30.7	0.0	168.9	0.0	68.5	31.5
$n = 500$	BIC	0.0	142.7	0.0	97.6	2.4	0.0	163.3	0.0	97.6	2.4
	Cross-Validation	0.0	146.9	0.0	39.3	60.7	0.0	170.1	0.0	38.5	60.5
	True model	0.0	142.2	-	-	-	0.0	162.6	-	-	-

Across all sample sizes, the imputed estimators based on the AIC, BIC, and cross-validation criteria showed negligible bias for both sampling designs. This is explained by the fact that these three model selection criteria are known to satisfy $\mathbb{P}(\widehat{\alpha}_v \in \mathcal{W}) \to 0$ as $v \to \infty$ at the population level (Zhang, 1993; Shao, 1997), and hence also at the sample level by an application of Lemma 1. This behavior was confirmed in our simulations, where the probability of selecting a wrong model converged to 0 for all three criteria. Consequently, the imputed estimators based on these selected models are consistent. However, AIC and cross-validation are not consistent model selection procedures and exhibited overfitting probabilities of 30.7%

and 60.7%, respectively, for $n = 500$ in our simulations. Since the imputed estimators based on these models are consistent, the lower efficiency indicates a larger variance. For $n = 500$, the imputed estimators based on AIC and cross-validation were less efficient, with RE values of 146.8% and 146.9%, respectively, compared with 142.2% for the true model. As explained in part ii) of Proposition 3, the true model achieves the smallest asymptotic variance, whereas the variance of an imputed estimator based on an overfitted model is necessarily larger. In contrast, BIC remained consistent with the probability of selecting the true model reaching 97.6% for $n = 500$. As a result, the point estimator based on BIC was the most efficient in all cases. Moreover, as established in Theorem 3, the imputed estimator based on BIC is asymptotically equivalent to the oracle imputation estimator based on the true model.

5.4. Variance estimation and confidence intervals

We now turn to the problem of variance estimation and confidence intervals, as per the procedure described in Algorithm 1. More specifically, we were interested in the relative bias of $\widehat{V}_T(\widehat{\alpha})$ as an estimator of $V_T(\widehat{\alpha})$, with $\widehat{\alpha}$ denoting a model selected via a consistent model selection procedure; here, the BIC criterion was adopted. We computed

$$\text{RB}(\widehat{V}_T(\widehat{\alpha})) = 100\% \times \frac{1}{B} \sum_{b=1}^{B} \frac{\widehat{V}_T^{(b)}(\widehat{\alpha}) - \mathbb{V}_{MC}(\widehat{\mu}_{\widehat{\alpha}})}{\mathbb{V}_{MC}(\widehat{\mu}_{\widehat{\alpha}})},$$

with

$$\mathbb{V}_{MC}(\widehat{\mu}_{\widehat{\alpha}}) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{\mu}_{\widehat{\alpha}}^{(b)} - \frac{1}{m} \sum_{m=1}^{m} \widehat{\mu}_{\widehat{\alpha}}^{(b)} \right)^2$$

denoting the Monte-Carlo variance of $\widehat{\mu}_{\widehat{\alpha}}$. We also were interested in verifying empirically that $\mathbb{P}(\mu_v \in \text{CI}_{1-\alpha}(\widehat{\mu}_{\widehat{\alpha}})) \approx 1 - \alpha$, as should be asymptotically by Corollary 3. We computed

$$\text{CP}(\widehat{\mu}_{\widehat{\alpha}}) = 100\% \times \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\left(\mu^{(b)} \in \text{CI}_{1-\alpha}^{(b)}(\widehat{\mu}_{\widehat{\alpha}}^{(b)}) \right),$$

where $\text{CI}_{1-\alpha}^{(b)}(\widehat{\mu}_{\widehat{\alpha}}^{(b)})$ denotes the output of Algorithm 1 at iteration b. The results are presented in Table 3.

Table 3

Relative biases of variance estimators and coverage probabilities evolution as n increases with various sampling fractions.

	SRSWOR						Stratified					
	5%		10%		20%		5%		10%		20%	
	RB	CP	RB	CP	RB	CP	RB	CP	RB	CP	RB	CP
$n = 200$	-2.9	94.5	-4.9	94.2	-8.0	93.8	-4.9	94.1	-5.5	94.1	-10.6	93.5
$n = 500$	-3.1	94.6	-3.0	94.6	-5.8	94.2	-4.0	94.5	-4.0	94.3	-8.2	93.8
$n = 1000$	-3.2	94.5	-1.9	94.8	-4.6	94.3	-1.8	94.7	-3.2	94.7	-4.5	94.3

For smaller sample sizes, the relative biases were slightly larger as the sampling fraction

increased. Consequently, the coverage probabilities were slightly below the nominal 95% level, particularly when $n = 200$. However, for a fixed sampling fraction, as both the sample and population sizes increased, the negative biases diminished and eventually became negligible. This pattern was consistent across both sampling designs.

These results confirmed the asymptotic validity of the proposed variance estimation procedure: as the sample size increased, the bias of the variance estimator vanished, and the empirical coverage probabilities converged to their nominal levels. Hence, Algorithm 1 produced reliable confidence intervals in large samples under both simple random sampling without replacement and stratified sampling.

6. Final Remarks

This paper develops a theoretical framework for model selection in the context of imputation under survey sampling. We introduced an oracle loss function that quantifies the efficiency of an imputation model and showed that its minimizer asymptotically coincides with the true model. Under standard regularity conditions and a non-informative sampling design, model selection procedures that are consistent in the i.i.d. setting remain consistent when applied to survey data. In particular, the BIC criterion asymptotically identifies the true imputation model.

Based on this framework, we established the asymptotic properties of the imputed estimator and its variance when the imputation model is selected using a consistent criterion. The resulting estimators are asymptotically equivalent to those obtained under the true model, therefore achieving oracle efficiency. These results provide a theoretical justification for the use of standard model-selection tools in imputation problems involving survey data.

Simulation results support the theoretical findings. The loss function $\mathcal{L}_n$ discriminates effectively between models, with bias vanishing once all relevant predictors are included. The BIC criterion consistently identifies the true model, whereas AIC and cross-validation tend to favor overly complex specifications. The proposed variance estimator performs well in finite samples, exhibiting negligible bias and empirical coverage close to the nominal level.

The proposed framework establishes a rigorous connection between model selection and imputation in survey data. It shows that, under appropriate conditions, selecting an imputation model with a consistent criterion yields asymptotically valid inference. It would be of interest to extend these results to broader classes of imputation models or to settings with informative sampling designs. A first step in this direction could be to leverage the favorable model-selection properties of cross-validation in nonparametric regression (Yang, 2007). Another related promising avenue would be to investigate the use of aggregation (Nemirovski, 2000; Bunea, Tsybakov and Wegkamp, 2007) to combine several models. This was tested empirically for the treatment of unit nonresponse with good results in Larbi et al. (2025), but a theoretical investigation of the topic is currently lacking.

Appendix A: Additional notation

The Euclidean vector norm is denoted $\|\cdot\|_2$. The operator and Frobenius norms of a matrix are denoted by $\|\cdot\|_{op}$ and $\|\cdot\|_F$, respectively. The largest and smallest eigenvalues of a symmetric matrix S are denoted $\lambda_{min}(S)$ and $\lambda_{max}(S)$, respectively. We write $\mathrm{Tr}(S)$ to denote the trace of a square matrix S. We use $S \succeq 0$ (resp, $S \succ 0$) to denote that the symmetric matrix S is a positive semi-definite (resp, positive definite) and write $S_1 \succeq S_2$ to mean that $S_1 - S_2 \succeq 0$. The identity matrix of $\mathbb{R}^{n \times n}$ is denoted I_n.

Here, $z_n = o_{\mathbb{P}}(1)$ indicates that for any $\epsilon > 0$, we have

$$\lim_{v \to \infty} \mathbb{P}_{mpq}\left(|z_n| > \epsilon\right) = 0$$

almost surely. We may omit the phrase "almost surely" or "a.s." in some intermediate steps of the proof. To express that $A \cup B$ is a disjoint union, we use $A \uplus B$. We use $x_n \sim_p z_n$ to say that $x_n - z_n = o_{\mathbb{P}}(1)$. To say that two random sequences $(x_n)_{n \in \mathbb{N}}$ and $(z_n)_{n \in \mathbb{N}}$ are of the same order in probability, we write $x_n \asymp_{\mathbb{P}} z_n$.

The matrix $X_r^\top X_r$ is denoted by $A_r := X_r^\top X_r$. Similarly, for $\alpha \in \mathcal{A}$, we write $A_{r,\alpha} := X_{r,\alpha}^\top X_{r,\alpha}$.

Appendix B: Proof of main results

B.1. Invertibility with high-probability

Lemma 2. *Assume (S1)-(S2), (D1)-(D2). Let*

$$A_v^w := \frac{1}{N_v} \sum_{k \in U_v} w_k r_k \mathbf{x}_k \mathbf{x}_k^\top,$$

for a set of $\sigma(X)$-measurable weights $(w_k)_{k \in U_v}$ satisfying that, there exists $v > 0$ such that, for all $v \in \mathbb{N}$, $w_k \geqslant v$, for all $k \in U_v$. Then, there exists constants $C_1, C_2, C_3 \in \mathbb{R}_+^*$ and an index $v_0 := v_0\left((\mathbf{x}_k)_{k \in U}\right) \in \mathbb{N}$ such that, for all $v \geqslant v_0$, almost surely,

$$\mathbb{P}_q\left(\lambda_{\min}(A_v) < C_3\right) \leq C_1 \exp\left(-C_2 N_v\right). \tag{11}$$

Proof. Since $w_k \geqslant v$ uniformly, we have

$$A_v^w \succeq v A_v, \quad \text{with} \quad A_v := \frac{1}{N} \sum_{k \in U_v} r_k \mathbf{x}_k \mathbf{x}_k^\top.$$

In particular, this ensures

$$\lambda_{\min}(A_v^w) \geq v \lambda_{\min}(A_v),$$

so we shall prove the inequality for A_v instead. Note that

$$\mathbb{E}_q\left[A_v\right] = \frac{1}{N} \sum_{k \in U_v} p(\mathbf{x}_k) \mathbf{x}_k \mathbf{x}_k^\top := A_v^p \xrightarrow[v \to \infty]{a.s.} \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}) \mathbf{x} \mathbf{x}^\top\right] := A^{(\infty)}.$$

In our setting, $\boldsymbol{A}^{(\infty)} \succ \boldsymbol{0}$ so that $\sigma^{(\infty)} := \lambda_{\min}(\boldsymbol{A}^{(\infty)}) > 0$. Since the convergence to $\boldsymbol{A}^{(\infty)}$ is almost sure, there exists an event E such that $\mathbb{P}_{\mathbf{x}}(E) = 1$, on which for all $\omega \in E$, there exists $v_0(\omega)$ such that for all $v \geq v_0$,

$$\lambda_{\min}(\boldsymbol{A}_v) \geq \frac{\sigma^{(\infty)}}{2}.$$

We now fix $\omega \in E$, and $v \geq v_0$. Let

$$\boldsymbol{Z}_k := N_v^{-1} r_k \mathbf{x}_k \mathbf{x}_k^\top, \qquad k \in U_v,$$

and note that $(\boldsymbol{Z}_k)_{k \in U_v}$ are (i) positive semi-definite, (ii) of the operator norm uniformly bounded $\|\boldsymbol{Z}_k\|_{op} \leq C_0^2/N$ with C_0 satisfying $\|\mathbf{x}_k\|_2 \leq C_0$. In particular,

$$\mu_{\min} := \lambda_{\min}\left(\mathbb{E}_q\left[\sum_{k \in U_v} \boldsymbol{Z}_k\right]\right) = \lambda_{\min}\left(\boldsymbol{A}_v^p\right) \geq \frac{\sigma^{(\infty)}}{2}.$$

An application of Matrix Chernoff's inequality (see, e.g., Theorem 5.1.1 in Tropp et al. (2015)) gives

$$\mathbb{P}_q\left(\lambda_{\min}\left(\sum_{k \in U_v} \boldsymbol{Z}_k\right) \leq \frac{\mu_{\min}}{2}\right) \leq p\left(\sqrt{\frac{2}{e}}\right)^{\frac{\mu_{\min}N_v}{C_0^2}} \leq p\left(\sqrt{\frac{2}{e}}\right)^{\frac{\sigma^{(\infty)}N_v}{2C_0^2}}.$$

Noting that $2/e < 1$, we write

$$p\left(\sqrt{\frac{2}{e}}\right)^{\frac{\sigma^{(\infty)}N_v}{2C_0^2}} = C_1 \exp\left(-C_2 N_v\right), \quad \text{with} \qquad C_1 := p, \quad C_2 := -\frac{\sigma^{(\infty)}}{2C_0^2}\ln\left(\sqrt{\frac{2}{e}}\right) > 0.$$

Overall, since

$$\left\{\lambda_{\min}\left(\sum_{k \in U_v} \boldsymbol{Z}_k\right) \leq \frac{\sigma^{(\infty)}}{4}\right\} \subset \left\{\lambda_{\min}\left(\sum_{k \in U_v} \boldsymbol{Z}_k\right) \leq \frac{\mu_{\min}}{2}\right\},$$

we get that, almost surely, for all $v \geqslant v_0$,

$$\mathbb{P}_q\left(\lambda_{\min}\left(\sum_{k \in U_v} \boldsymbol{Z}_k\right) \leq \frac{\sigma^{(\infty)}}{4}\right) \leq C_1 \exp\left(-C_2 N_v\right)$$

which also translates to its weighted version. □

Proof of Proposition 1.

In the remainder of the proof, we fix an arbitrary $\alpha \in \mathcal{A}$. Observe that

$$\mathcal{L}_n(\alpha) = \mathbb{E}_m\left[\sum_{k \in S_m} \frac{\left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right)^2}{\pi_k^2}\right] + \mathbb{E}_m\left[\sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha}{\pi_k} \frac{y_l - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha}{\pi_l}\right]$$

$$:= A_n(\alpha) + B_n(\alpha).$$

We compute each term separately.

Computation of $A_n(\alpha)$. By expanding $y_k = \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta} + \epsilon_k$ for $k \in S_m$, we obtain

$$
A_n(\alpha) = \mathbb{E}_m \left[\sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)^2}{\pi_k^2} \right] + \mathbb{E}_m \left[\sum_{k \in S_m} \frac{\epsilon_k^2}{\pi_k^2} \right] + 2\mathbb{E}_m \left[\sum_{k \in S_m} \frac{\epsilon_k}{\pi_k^2} \left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right) \right]
$$

$$
= \mathbb{E}_m \left[\sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)^2}{\pi_k^2} \right] + \sum_{k \in S_m} \frac{\sigma^2}{\pi_k^2}.
$$

The first term can be rewritten as

$$
\mathbb{E}_m \left[\sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)^2}{\pi_k^2} \right]
$$

$$
= \mathbb{E}_m \left[\sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top Y_r \right)^2}{\pi_k^2} \right]
$$

$$
= \mathbb{E}_m \left[\sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta} \right)^2}{\pi_k^2} \right]
$$

$$
+ \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top \mathbb{E}_m \left[\boldsymbol{\epsilon}_r \boldsymbol{\epsilon}_r^\top \right] X_{r,\alpha} A_{r,\alpha}^{-1} \mathbf{x}_{k,\alpha}}{\pi_k^2}
$$

$$
= \sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta} \right)^2}{\pi_k^2} + \sigma^2 \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} \mathbf{x}_{k,\alpha}}{\pi_k^2}.
$$

Overall,

$$
A_n(\alpha) = \sum_{k \in S_m} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta} \right)^2}{\pi_k^2} + \sigma^2 \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} \mathbf{x}_{k,\alpha}}{\pi_k^2} + \sum_{k \in S_m} \frac{1}{\pi_k^2} \right). \quad (12)
$$

Computation of $B_n(\alpha)$. Note that

$$
\sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right) \left(y_l - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)
$$

$$
= \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \left(\mathbf{x}_k^\top \boldsymbol{\beta} + \epsilon_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} + \epsilon_l - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)
$$

$$
= \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \epsilon_k \epsilon_l + \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \epsilon_k \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right) + \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \epsilon_l \left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right)
$$

$$
+ \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha \right). \quad (13)
$$

It follows from the above that the first three terms of (13) have mean zero. Consider the following decomposition of the last term:

$$
\sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right)}{\pi_k \quad \pi_l}
$$

$$
= \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top (X_r \boldsymbol{\beta} + \boldsymbol{\epsilon}_r)\right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top (X_r \boldsymbol{\beta} + \boldsymbol{\epsilon}_r)\right)}{\pi_k \quad \pi_l}
$$

$$
= \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta}\right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta}\right)}{\pi_k \quad \pi_l}
$$

$$
+ 2 \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha} \boldsymbol{\epsilon}_r}{\pi_k} \frac{\left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta}\right)}{\pi_l}
$$

$$
+ \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha} \boldsymbol{\epsilon}_r}{\pi_k} \frac{\boldsymbol{\epsilon}_r^\top X_{r,\alpha}^\top A_{r,\alpha}^{-1} \mathbf{x}_{l,\alpha}}{\pi_l}.
$$

It follows that

$$
\mathbb{E}_m \left[\sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha}{\pi_k} \frac{y_l - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha}{\pi_l} \right]
$$

$$
= \mathbb{E}_m \left[\sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha\right)}{\pi_k \quad \pi_l} \right]
$$

$$
= \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\left(\mathbf{x}_k^\top \boldsymbol{\beta} - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta}\right) \left(\mathbf{x}_l^\top \boldsymbol{\beta} - \mathbf{x}_{l,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \boldsymbol{\beta}\right)}{\pi_k \quad \pi_l}
$$

$$
+ \sigma^2 \sum_{k \in S_m} \sum_{\substack{l \in S_m \\ l \neq k}} \frac{\mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} \mathbf{x}_{l,\alpha}}{\pi_k \pi_l}. \tag{14}
$$

Combining the closed-form formulas of $A_n(\alpha)$ in (12) and $B_n(\alpha)$ in (14), we get

$$
\mathcal{L}_n(\alpha) = \left\{ \sum_{k \in S_m} \pi_k^{-1} \left(\mathbf{x}_k^\top - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r\right) \boldsymbol{\beta} \right\}^2
$$

$$
+ \sigma^2 \left\{ \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top}{\pi_k} \right) A_{r,\alpha}^{-1} \left(\sum_{l \in S_m} \frac{\mathbf{x}_{l,\alpha}}{\pi_l} \right) + \sum_{k \in S_m} \frac{1}{\pi_k^2} \right\}.
$$

Proof of Proposition 2.

Let $\alpha_1, \alpha_2 \in C$ such that $\alpha_1 \subset \alpha_2$. Recall that, for any $\alpha \in C$, $\mathcal{L}_{1,n}(\alpha) = 0$. Thus,

$$\mathcal{L}_n(\alpha_1) < \mathcal{L}_n(\alpha_2) \qquad \Leftrightarrow \qquad \mathcal{L}_{2,n}(\alpha_1) < \mathcal{L}_{2,n}(\alpha_2).$$

Since $\alpha^\star$ is the smallest correct model, the result follows using Lemma 11.

Proof of Theorem 1.

We will show that $\lim_{v \to \infty} \mathbb{P}_{pq}(\alpha_{opt,v} \neq \alpha^\star) = 0$. We may write the event $\{\alpha_{opt,v} \neq \alpha^\star\}$ as a disjoint union

$$\{\alpha_{opt,v} \neq \alpha^\star\} = \{\alpha_{opt,v} \in C, \alpha_{opt,v} \neq \alpha^\star\} \biguplus \{\alpha_{opt,v} \in \mathcal{W}\},$$

so that

$$\mathbb{P}_{pq}(\alpha_{opt,v} \neq \alpha^\star) = \mathbb{P}_{pq}(\alpha_{opt,v} \in C, \alpha_{opt,v} \neq \alpha^\star) + \mathbb{P}_{pq}(\alpha_{opt,v} \in \mathcal{W}) := p_{1,v} + p_{2,v}.$$

Using Proposition 2, we have directly that $p_{1,v} = 0$ for all $v \in \mathbb{N}$.

Let $\alpha_F := \{1, \ldots, p\} \in C$ be the set of all covariates. For $p_{2,v}$, consider the following decomposition

$$
\begin{aligned}
p_{2,v} &= \mathbb{P}_{pq}\left(\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}_n(\alpha^\star)\right) \\
&= \mathbb{P}_{pq}\left(\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}_n(\alpha^\star), \mathcal{L}_n(\alpha^\star) \leq \mathcal{L}_n(\alpha_F)\right) \\
&\quad + \mathbb{P}_{pq}\left(\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}_n(\alpha^\star), \mathcal{L}_n(\alpha^\star) > \mathcal{L}_n(\alpha_F)\right) \\
&\leq \mathbb{P}_{pq}\left(\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}(\alpha_F)\right) + \mathbb{P}_{pq}\left(\mathcal{L}_n(\alpha_F) < \mathcal{L}_n(\alpha^\star)\right) := q_{1,v} + q_{2,v}.
\end{aligned}
$$

Convergence of $q_{1,v}$. On the one hand, for $\alpha_{opt,v} \in \mathcal{W}$, the loss $\mathcal{L}_{1,n}(\alpha_{opt,v})$ satisfies the following identity

$$
\begin{aligned}
\mathcal{L}_{1,n}(\alpha_{opt,v}) &= \left\{ \sum_{k \in S_{m,v}} \pi_k^{-1} \left(\mathbf{x}_k^\top - \mathbf{x}_{k,\alpha}^\top A_{r,\alpha}^{-1} X_{r,\alpha}^\top X_r \right) \boldsymbol{\beta} \right\}^2 \\
&= \left\{ \sum_{k \in S_{m,v}} \pi_k^{-1} \Big(\mathbf{x}_{k,\alpha_{opt,v}}^\top \boldsymbol{\beta}_{\alpha_{opt,v}} + \mathbf{x}_{k,\alpha_{opt,v}^c}^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \right. \hspace{2cm} (15) \\
&\quad \underbrace{- \mathbf{x}_{k,\alpha_{opt,v}}^\top A_{r,\alpha_{opt,v}}^{-1} X_{r,\alpha_{opt,v}}^\top X_{r,\alpha_{opt,v}} \boldsymbol{\beta}_{\alpha_{opt,v}}}_{= \mathbf{x}_{k,\alpha_{opt,v}}^\top \boldsymbol{\beta}_{\alpha_{opt,v}}} \\
&\quad \left. - \mathbf{x}_{k,\alpha_{opt,v}}^\top A_{r,\alpha_{opt,v}}^{-1} X_{r,\alpha_{opt,v}}^\top X_{r,\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c} \Big) \right\}^2 \\
&= \left\{ \sum_{k \in S_m} \pi_k^{-1} \left(\mathbf{x}_{k,\alpha_{opt,v}^c}^\top - \mathbf{x}_{k,\alpha_{opt,v}}^\top A_{r,\alpha_{opt,v}}^{-1} X_{r,\alpha_{opt,v}}^\top X_{r,\alpha_{opt,v}^c} \right) \boldsymbol{\beta}_{\alpha_{opt,v}^c} \right\}^2 .
\end{aligned}
$$

$$(16)$$

On the other hand, since $\alpha_F \in C$, we have $\mathcal{L}_{1,n}(\alpha_F) = 0$. Thus, $\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}_n(\alpha_F)$ is equivalent to

$$
\begin{aligned}
\mathcal{L}_n(\alpha_{opt,v}) < \mathcal{L}_n(\alpha_F) \quad &\Leftrightarrow \quad \mathcal{L}_{1,n}(\alpha_{opt,v}) + \mathcal{L}_{2,n}(\alpha_{opt,v}) < \mathcal{L}_{1,n}(\alpha_F) + \mathcal{L}_{2,n}(\alpha_F) \\
&\Leftrightarrow \quad \mathcal{L}_{1,n}(\alpha_{opt,v}) < \mathcal{L}_{2,n}(\alpha_F) - \mathcal{L}_{2,n}(\alpha_{opt,v}). \quad (17)
\end{aligned}
$$

Combining (15) and (17) leads to

$$
\begin{aligned}
&\mathcal{L}_{1,n}(\alpha_{opt,v}) < \mathcal{L}_{2,n}(\alpha_F) - \mathcal{L}_{2,n}(\alpha_{opt,v}) \\
&\Leftrightarrow \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}^c}^\top}{N_v \pi_k} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \mathbf{x}_{r,\alpha_{opt,v}^c} \right) \\
&\quad \times \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}^c}^\top}{N_v \pi_k} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \mathbf{x}_{r,\alpha_{opt,v}^c} \right)^\top \\
&\quad < \sigma^2 \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_F}^\top}{N_v \pi_k} \right) A_{r,\alpha_F}^{-1} \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_F}}{N_v \pi_k} \right) \\
&\quad - \sigma^2 \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} \right) A_{r,\alpha_{opt,v}}^{-1} \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}}{N_v \pi_k} \right).
\end{aligned}
$$

Using equality (42), by setting $\alpha_1 := \alpha_{opt,v}$, $\alpha_2 := \alpha_F$, $\alpha_{opt,v}^c := \alpha_{opt,v}^c$, we obtain

$$
\begin{aligned}
&\sigma^2 \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_F}^\top}{N_v \pi_k} \right) A_{r,\alpha_F}^{-1} \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_F}}{N_v \pi_k} \right) - \sigma^2 \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} \right) A_{r,\alpha_{opt,v}}^{-1} \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}}{N_v \pi_k} \right) \\
&= \sigma^2 \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}^c}^\top}{N_v \pi_k} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \mathbf{x}_{r,\alpha_{opt,v}^c} \right) \\
&\quad \times S_v^{-1} \left(\sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}^c}^\top}{N_v \pi_k} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \mathbf{x}_{r,\alpha_{opt,v}^c} \right)^\top,
\end{aligned}
$$

where

$$
S_v = \mathbf{x}_{r,\alpha_{opt,v}^c}^\top \left(I_{n_{r,v}} - \mathbf{x}_{r,\alpha_{opt,v}} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \right) \mathbf{x}_{r,\alpha_{opt,v}^c}
$$

is a positive and definite matrix. Let

$$
\mathbf{z}_v^\top = \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}^c}^\top}{N_v \pi_k} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha_{opt,v}}^\top}{N_v \pi_k} A_{r,\alpha_{opt,v}}^{-1} \mathbf{x}_{r,\alpha_{opt,v}}^\top \mathbf{x}_{r,\alpha_{opt,v}^c}.
$$

Then $\mathcal{L}_{1,n}(\alpha_{opt,v}) < \mathcal{L}_{2,n}(\alpha_F) - \mathcal{L}_{2,n}(\alpha_{opt,v})$ reduces to

$$
\mathcal{L}_{1,n}(\alpha_{opt,v}) < \mathcal{L}_{2,n}(\alpha_F) - \mathcal{L}_{2,n}(\alpha_{opt,v}) \quad \Leftrightarrow \quad \mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v.
$$

Consider the following decomposition of $q_{1,v}$,

$$
\lim_{v \to \infty} q_{1,v} = \lim_{v \to \infty} \mathbb{P}_{pq} \left(\mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v, \mathbf{z}_v \neq \mathbf{0} \right)
$$

$$+ \lim_{v \to \infty} \mathbb{P}_{pq} \left(\mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v, \mathbf{z}_v = \mathbf{0} \right)$$

If $\mathbf{z}_v = \mathbf{0}$, then the event $\{\mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v\}$ reduces to $\{0 < 0\} = \emptyset$. Therefore,

$$\mathbb{P}_{pq} \left(\mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v, \mathbf{z}_v = \mathbf{0} \right) = 0, \quad \text{for all } v \in \mathbb{N}.$$

Recall that for two quadratic forms $\mathbf{z}^\top A \mathbf{z}$ and $\mathbf{z}^\top B \mathbf{z}$, if for $\mathbf{z} \neq \mathbf{0}$, $\mathbf{z}^\top A \mathbf{z} < \mathbf{z}^\top B \mathbf{z}$, then $\lambda_{max}(A - B) < 0$. Recall also that from Courant-Fisher's Theorem, for real-valued symmetric matrices A and B, we have $\lambda_{max}(A) + \lambda_{min}(B) \leq \lambda_{max}(A + B) \leq \lambda_{max}(A) + \lambda_{max}(B)$. Therefore, combining the two statements above, we obtain

$$\lim_{v \to \infty} \mathbb{P}_{pq} \left(\mathbf{z}_v^\top \boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top \mathbf{z}_v < \sigma^2 \mathbf{z}_v^\top S_v^{-1} \mathbf{z}_v, \mathbf{z}_v \neq \mathbf{0} \right)$$

$$= \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{max}(\boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top - \sigma^2 S_v^{-1}) < 0 \right)$$

$$\leq \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{max}(\boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top) + \sigma^2(\lambda_{min}(-S_v^{-1})) < 0 \right)$$

$$= \lim_{v \to \infty} \left(\lambda_{max}(\boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top) - \sigma^2(\lambda_{max}(S_v^{-1})) < 0 \right)$$

$$= \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{max}(\boldsymbol{\beta}_{\alpha_{opt,v}^c} \boldsymbol{\beta}_{\alpha_{opt,v}^c}^\top) - \sigma^2(\lambda_{min}(S_v))^{-1} < 0 \right)$$

$$= \lim_{v \to \infty} \mathbb{P}_{pq} \left(\|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2 - N_v^{-1} \sigma^2 (\lambda_{min}(N_v^{-1} S_v))^{-1} < 0 \right)$$

$$= \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(N_v^{-1} S_v) < \frac{\sigma^2}{N_v \|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2} \right).$$

Using Lemma 4, there exists a constant $K_0 > 0$ such that

$$\lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(N_v^{-1} S_v) < \frac{\sigma^2}{N_v \|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2} \right)$$

$$= \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(N_v^{-1} S_v) < \frac{\sigma^2}{N_v \|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2}, \lambda_{min}(N_v^{-1} S_v) \geq K_0 \right)$$

$$+ \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(N_v^{-1} S_v) < \frac{\sigma^2}{N_v \|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2}, \lambda_{min}(N_v^{-1} S_v) < K_0 \right)$$

$$\leq \lim_{v \to \infty} \mathbb{P}_{pq} \left(K_0 \leq \frac{\sigma^2}{N_v \|\boldsymbol{\beta}_{\alpha_{opt,v}^c}\|_2^2} \right) + \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(N_v^{-1} S_v) < K_0 \right) = 0.$$

Therefore, it follows that $\lim_{v \to \infty} q_{1,v} = 0$ almost surely.

Convergence of $q_{2,v}$. It remains to show that, almost surely,

$$\lim_{v \to \infty} q_{2,v} = \lim_{v \to \infty} \mathbb{P}_{pq} \left(\mathcal{L}_n(\alpha_F) < \mathcal{L}_n(\alpha^\star) \right) = 0.$$

However, since $\alpha^\star \subset \alpha_F \in \mathcal{C}$, the event $\mathcal{L}_n(\alpha_F) < \mathcal{L}_n(\alpha^\star)$ implies $\mathcal{L}_{2,n}(\alpha_F) < \mathcal{L}_{2,n}(\alpha^\star)$, which by Proposition 2 is a negligible event. It follows that $q_{2,v} = 0$ for all $v \in \mathbb{N}$. Therefore, we conclude that $\lim_{v \to \infty} p_{2,v} = 0$ which then implies

$$\lim_{v \to \infty} \mathbb{P}_{pq} \left(\alpha_{opt,v} \neq \alpha^\star \right) = 0$$

almost surely. Finally, applying the Lebesgue convergence theorem gives

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\alpha_{opt,v} = \alpha^\star \right) = 1$$

almost surely.

Proof of Proposition 3.

Proof of statement (i).

For $\alpha \in \mathcal{A}$, consider the decomposition of $\widehat{\mu}_{\alpha,v} - \mu_v$

$$\widehat{\mu}_{\alpha,v} - \mu_v = \underbrace{\widehat{\mu}_{\alpha,v} - \widehat{\mu}_{\pi,v}}_{(a)} + \underbrace{\widehat{\mu}_{\pi,v} - \mu_v}_{(b)} .$$

As $\pi_k = \pi$ for all $k \in U_v$, we have

$$\widehat{\boldsymbol{\beta}}_{\alpha,v} = A_{r,\alpha}^{-1} X_{r,\alpha}^\top Y_r = A_{r,\alpha}^{-1} \sum_{k \in S_{r,v}} \mathbf{x}_{k,\alpha} y_k = \left(\frac{A_{r,\alpha}}{N_v \pi} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v \pi} .$$

By expanding the term (a), we obtain

$$\widehat{\mu}_{\alpha,v} - \widehat{\mu}_{\pi,v} = \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta} \right) + \epsilon_k}{N_v \pi}$$

$$= \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top}{N_v \pi} \left(\frac{A_{r,\alpha}}{N_v \pi} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v \pi} - \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{N_v \pi} + \sum_{k \in S_{r,v}} \frac{\epsilon_k}{N_v \pi}$$

$$= \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top}{N_v \pi} \left(\frac{A_{r,\alpha}}{N_v \pi} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_k^\top \boldsymbol{\beta} + \mathbf{x}_{k,\alpha} \epsilon_k}{N_v \pi}$$

$$- \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{N_v \pi} + \sum_{k \in S_{r,v}} \frac{\epsilon_k}{N_v \pi} .$$

The law of large numbers gives

$$\sum_{k \in S_{m,v}} \frac{\epsilon_k}{N_v \pi} \xrightarrow[v \to \infty]{\mathbb{P}} 0, \qquad \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{N_v \pi} \xrightarrow[v \to \infty]{\mathbb{P}} \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha^\top \boldsymbol{\beta}_\alpha \right],$$

$$\left(\frac{A_{r,\alpha}}{N_v \pi} \right)^{-1} \xrightarrow[v \to \infty]{\mathbb{P}} \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1},$$

$$\sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_k^\top \boldsymbol{\beta} + \mathbf{x}_{k,\alpha} \epsilon_k}{N_v \pi} \xrightarrow[v \to \infty]{\mathbb{P}} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}^\top \boldsymbol{\beta} \right].$$

Using the continuous mapping theorem, we obtain

$$\widehat{\mu}_{\alpha,v} - \mu_v \xrightarrow[v \to \infty]{\mathbb{P}} \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha^\top \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}^\top \boldsymbol{\beta} \right]$$

$$- \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \mathbf{x}^\top \boldsymbol{\beta} \right]$$

Recall that $\widehat{\mu}_{\alpha,v} - \mu_v$ is consistent if $\widehat{\mu}_{\alpha,v} - \mu_v \xrightarrow[v \to \infty]{\mathbb{P}} 0$. Equivalently,

$$\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \, \mathbf{x}_\alpha^\top \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}^\top \boldsymbol{\beta} \right] - \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x})) \, \mathbf{x}^\top \boldsymbol{\beta} \right] = 0.$$

To obtain (C_2), one can use the decomposition $\mathbf{x}^\top \boldsymbol{\beta} = \mathbf{x}_\alpha^\top \boldsymbol{\beta}_\alpha + \mathbf{x}_{\alpha^c}^\top \boldsymbol{\beta}_\alpha^c$ on the above equality, from which the result follows after some algebra.

Proof of statement (ii), (a).

By Theorem 3, for, $\alpha \in C$, we have

$$\widehat{\mu}_{\alpha,v} - \mu_v = \widehat{\mu}_{\alpha,v} - \widetilde{\mu}_{\alpha,v} + o_{\mathbb{P}} \left(n_v^{-1/2} \right).$$

Consequently, $\mathbb{AV} \left(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v) \right) = \mathbb{AV} \left(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v) \right)$. To obtain the asymptotic variance, we define

$$\mathbb{AV} \left(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v) \right) = \lim_{v \to \infty} \left(\mathbb{E}_{mq} \left[\mathbb{V}_p \left(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v) \right) \right] \right.$$
$$+ \mathbb{E}_q \left[\mathbb{V}_m \left(\mathbb{E}_p [\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)] \right) \right]$$
$$\left. + \mathbb{V}_q \left(\mathbb{E}_{mp} [\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)] \right) \right).$$

Recalling that $\mathbb{V}_q(\mathbb{E}_{mp}[\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)]) = 0$, only the first two terms remain and we write

$$\mathbb{AV} \left(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v) \right) = \lim_{v \to \infty} \left(\mathbb{E}_{mq} \left[\mathbb{V}_p \left(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v) \right) \right] \right.$$
$$\left. + \mathbb{E}_q \left[\mathbb{V}_m \left(\mathbb{E}_p [\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)] \right) \right] \right)$$
$$:= \mathrm{AV}_1 + \mathrm{AV}_2,$$

Asymptotic behavior of AV_1. *Main idea.* If there exists $c_{1,\alpha}$ such that

$$\mathbb{V}_p(\sqrt{N_v} \left(\widetilde{\mu}_{\alpha,v} - \mu_v \right)) \xrightarrow[v \to \infty]{\mathbb{P}} c_{1,\alpha},$$

and if $\mathbb{V}_p(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v))$ is uniformly integrable, then, almost surely, we would have

$$\lim_{v \to \infty} \mathbb{E}_{mq} \left[\left| \mathbb{V}_p(\sqrt{N_v} \left(\widetilde{\mu}_{\alpha,v} - \mu_v \right)) - c_{1,\alpha} \right| \right] = 0.$$

Furthermore, Jensen inequality would imply $\lim_{v \to \infty} |\mathbb{E}_{mq}[\mathbb{V}_p(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)) - c_{1,\alpha}]| = 0$ almost surely. We shall follow the above architecture.

Uniform integrability. The uniform integrability of $\mathbb{V}_p(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v))$ follows by Lemma 23 in which we prove that $\mathbb{E}_{mq}[(\mathbb{V}_p(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)))^2]$ is bounded.

Limit determination. We now proceed to find $c_{1,\alpha}$ such that $\mathbb{V}_p(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)) \xrightarrow[v\to\infty]{\mathbb{P}} c_{1,\alpha}$. To that aim, write

$$
\begin{aligned}
\mathbb{V}_p\left(\sqrt{N_v}\left(\widetilde{\mu}_{\alpha,v} - \mu_v\right)\right) &= \frac{1}{N_v} \sum_{k\in U_v} \sum_{l\in U_v} \frac{\eta_{k,\alpha}}{\pi} \frac{\eta_{l,\alpha}}{\pi} \Delta \\
&= \frac{1}{N_v} \sum_{k\in U_v} \sum_{l\in U_v} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi} \frac{\mathbf{x}_{l,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi} \Delta \\
&\quad + \frac{2}{N_v} \sum_{k\in U_v} \sum_{l\in U_v} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi} \frac{r_l(1 + \pi\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha})\epsilon_l}{\pi} \Delta \\
&\quad + \frac{1}{N_v} \sum_{k\in U_v} \sum_{l\in U_v} \frac{r_k(1 + \pi\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha})\epsilon_k}{\pi} \frac{r_l(1 + \pi\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha})\epsilon_l}{\pi} \Delta \\
&:= A_v + B_v + C_v.
\end{aligned}
$$

Next, we investigate the asymptotic behavior A_v, B_v, and C_v separately.

Behavior of A_v. Since $\alpha \in C$, we have $\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha = \mathbf{x}_{k,\alpha^\star}^\top \boldsymbol{\beta}_{\alpha^\star}$ and thus A_v does not depend on α. As a result, there exists a constant $C_1(\alpha^\star)$ such that $A_v - C_1(\alpha^\star) = o_{\mathbb{P}}(1)$.

Behavior of B_v. Write

$$
\begin{aligned}
B_v &= 2\frac{1-\pi}{\pi} \sum_{k\in U_v} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha r_k \left(1 + \pi\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right) \epsilon_k}{N_v} \\
&\quad + \frac{2}{N_v} \frac{1-\pi}{\pi} \sum_{k\in U_v} \sum_{\substack{l\in U_v \\ l\neq k}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi} \frac{r_l \left(1 + \pi\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha}\right) \epsilon_l}{\pi} \Delta \\
&:= B_{1,v} + B_{2,v}.
\end{aligned}
$$

By substituting $\pi\mathbf{c}_{\alpha,v}$, $B_{1,v}$ gives

$$
\begin{aligned}
B_{1,v} &= \frac{1-\pi}{\pi} \sum_{k\in U_v} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha r_k \epsilon_k}{N_v} \\
&\quad + \frac{1-\pi}{\pi} \sum_{k\in U_v} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha r_k}{N_v} \mathbf{x}_{k,\alpha}^\top \epsilon_k \left(\sum_{k\in U_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1} \sum_{k\in U_v} \frac{(1 - r_k) \mathbf{x}_{k,\alpha}}{N_v}.
\end{aligned}
$$

By the law of large numbers, we obtain

$$
\frac{1}{N_v} \sum_{k\in U_v} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha r_k \epsilon_k \xrightarrow[v\to\infty]{\mathbb{P}} \mathbb{E}_{\mathbf{x}pq}\left[\mathbf{x}_\alpha^\top \boldsymbol{\beta}_\alpha p(\mathbf{x})\epsilon\right] = 0,
$$

$$
\frac{1}{N_v} \sum_{k\in U_v} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha r_k \mathbf{x}_{k,\alpha}^\top \epsilon_k \xrightarrow[v\to\infty]{\mathbb{P}} \mathbb{E}_{\mathbf{x}pq}\left[\mathbf{x}_\alpha^\top \boldsymbol{\beta}_\alpha p(\mathbf{x})\mathbf{x}_\alpha\epsilon\right] = \mathbf{0}_\alpha^\top,
$$

and

$$\left(\frac{1}{N_v}\sum_{k\in U_v} r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^{\top}\right)^{-1}\frac{1}{N_v}\sum_{k\in U_v}(1-r_k)\mathbf{x}_{k,\alpha} \xrightarrow[v\to\infty]{\mathbb{P}} \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_{\alpha}\mathbf{x}_{\alpha}^{\top}\right]\right)^{-1}\mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\,\mathbf{x}_{\alpha}\right].$$

As a result, $B_{1,v} \xrightarrow[v\to\infty]{\mathbb{P}} 0$. For $B_{2,v}$, recall that $K := \lim_{v\to\infty} N_v\Delta$, we have

$$\begin{aligned}
B_{2,v} &= \frac{2}{N_v}\sum_{k\in U_v}\sum_{\substack{l\in U_v\\l\neq k}}\frac{\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}}{\pi}\frac{r_l\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{l,\alpha}\right)\epsilon_l}{\pi}\Delta \\
&= \frac{2}{\pi^2}N_v\Delta\sum_{k\in U_v}\sum_{\substack{l\in U_v\\l\neq k}}\frac{\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}}{N_v}\frac{r_l\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{l,\alpha}\right)\epsilon_l}{N_v} \\
&= \frac{2}{\pi^2}(K+o(1))\left(\sum_{k\in U_v}\sum_{l\in U_v}\frac{\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}}{N_v}\frac{r_l\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{l,\alpha}\right)\epsilon_l}{N_v}\right. \\
&\quad \left. -\frac{1}{N_v}\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}r_k\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{k,\alpha}\right)\epsilon_k}{N_v}\right) \\
&= \frac{2}{\pi^2}(K+o(1))\left(\left(\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}}{N_v}\right)\left(\sum_{k\in U_v}\frac{r_k\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{k,\alpha}\right)\epsilon_k}{N_v}\right)+o_{\mathbb{P}}\left(N_v^{-1}\right)\right) \\
&= \frac{2}{\pi^2}(K+o(1))\left(\left(\mathbb{E}_{\mathbf{x}}[\mathbf{x}_{k,\alpha}^{\top}\boldsymbol{\beta}_{\alpha}]+o_{\mathbb{P}}(1)\right)\left(0+o_{\mathbb{P}}(1)\right)+o_{\mathbb{P}}(N_v^{-1})\right) = o_{\mathbb{P}}(1),
\end{aligned}$$

since

$$\begin{aligned}
\sum_{k\in U_v}\frac{r_k\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{k,\alpha}\right)\epsilon_k}{N_v} &= \sum_{k\in U_v}\frac{r_k\epsilon_k}{N_v} \\
&\quad +\sum_{k\in U_v}\frac{r_k\epsilon_k\mathbf{x}_{k,\alpha}^{\top}}{N_v}\left(\sum_{k\in U_v}\frac{r_k\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^{\top}}{N_v}\right)^{-1}\sum_{k\in U_v}\frac{(1-r_k)\,\mathbf{x}_{k,\alpha}}{N_v} \\
&\xrightarrow[v\to\infty]{\mathbb{P}} 0+\mathbf{0}_{\alpha}^{\top}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_{\alpha}\mathbf{x}_{\alpha}^{\top}\right]\right)^{-1}\mathbb{E}_{\mathbf{x}}[(1-p(\mathbf{x}))\mathbf{x}_{\alpha}] = 0. \quad (18)
\end{aligned}$$

Finally, we have $B_v \xrightarrow[v\to\infty]{\mathbb{P}} 0$.

Behavior of C_v. Write

$$\begin{aligned}
C_v &= \frac{1-\pi}{\pi}\sum_{k\in U_v}\frac{r_k^2\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{k,\alpha}\right)^2\epsilon_k^2}{N_v} \\
&\quad +\frac{1}{N_v}\sum_{k\in U_v}\sum_{\substack{l\in U_v\\l\neq k}}\frac{r_k\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{k,\alpha}\right)\epsilon_k}{\pi}\frac{r_l\left(1+\pi\mathbf{c}_{\alpha,v}^{\top}\mathbf{x}_{l,\alpha}\right)\epsilon_l}{\pi}\Delta \\
&:= C_{1,v}+C_{2,v}.
\end{aligned}$$

Expanding $C_{1,v}$ gives

$$
\begin{aligned}
C_{1,v} &= \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right)^2 \epsilon_k^2}{N_v} \\
&= \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2}{N_v} + 2 \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2 \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}}{N_v} + \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2 \left(\pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right)^2}{N_v} \\
&= \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2}{N_v} + 2 \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2 \pi \mathbf{x}_{k,\alpha}^\top \mathbf{c}_{\alpha,v}}{N_v} \\
&\quad + \frac{1-\pi}{\pi} \left(\pi \mathbf{c}_{\alpha,v}^\top\right) \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2 \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \left(\pi \mathbf{c}_{\alpha,v}\right) \\
&= \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2}{N_v} + 2 \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2 \mathbf{x}_{k,\alpha}^\top}{N_v} \left(\sum_{k \in U_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}}{N_v}\right)^{-1} \sum_{k \in U_v} \frac{(1-r_k)\mathbf{x}_{k,\alpha}}{N_v} \\
&\quad + \frac{1-\pi}{\pi} \sum_{k \in U_v} \frac{(1-r_k)\mathbf{x}_{k,\alpha}^\top}{N_v} \left(\sum_{k \in U_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1} \left(\sum_{k \in U_v} \frac{r_k^2 \epsilon_k^2}{N_v} \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top\right) \\
&\quad \times \left(\sum_{k \in U_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1} \sum_{k \in U_v} \frac{(1-r_k)\mathbf{x}_{k,\alpha}}{N_v}.
\end{aligned}
$$

Hence, using the continuous mapping theorem, $C_{1,v}$ converges to

$$
\begin{aligned}
&\sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\right] - 2\sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha^\top\right] \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right]\right)^{-1} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha\right] \\
&\quad + \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha^\top\right] \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right]\right)^{-1} \\
&\quad\quad \times \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right] \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right]\right)^{-1} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha\right] \\
&= \sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}[p(\mathbf{x})] - 2\sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha^\top\right] \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right]\right)^{-1} \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha\right] \\
&\quad + \sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha^\top\right] \left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top\right]\right)^{-1} \mathbb{E}_{\mathbf{x}}\left[(1-p(\mathbf{x}))\mathbf{x}_\alpha\right] \\
&\overset{(*)}{=} \sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}[p(\mathbf{x})] - 2\sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})] + \sigma^2 \frac{1-\pi}{\pi} M(\alpha) \\
&= \sigma^2 \frac{1-\pi}{\pi} \mathbb{E}_{\mathbf{x}}[2 - p(\mathbf{x})] + \sigma^2 \frac{1-\pi}{\pi} M(\alpha).
\end{aligned}
$$

The equality $(*)$ is due to Lemma 12. For $C_{2,v}$, using (18), write

$$
\begin{aligned}
C_{2,v} &= \frac{1}{N_v} \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ l \neq k}} \frac{r_k \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right) \epsilon_k}{\pi} \frac{r_l \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha}\right) \epsilon_l}{\pi} \Delta \\
&= \frac{N_v \Delta}{\pi^2} \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ l \neq k}} \frac{r_k \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right) \epsilon_k}{N_v} \frac{r_l \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha}\right) \epsilon_l}{N_v}
\end{aligned}
$$

$$
= \frac{K + o(1)}{\pi^2} \left(\sum_{k \in U_v} \sum_{l \in U_v} \frac{r_k \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right) \epsilon_k}{N_v} \frac{r_l \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{l,\alpha} \right) \epsilon_l}{N_v} \right.
$$

$$
\left. - \frac{1}{N_v} \sum_{k \in U_v} \frac{r_k^2 \left(1 + \pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right) \epsilon_k^2}{N_v} \right)
$$

$$
= \frac{K + o(1)}{\pi^2} \left(\left(\sum_{k \in U_v} \frac{r_k (1 + \pi \mathbf{c}_{\alpha,v} \mathbf{x}_{k,\alpha}) \epsilon_k}{N_v} \right)^2 + o_{\mathbb{P}} \left(n_v^{-1} \right) \right)
$$

$$
= \frac{K + o(1)}{\pi^2} \left((0 + o_{\mathbb{P}}(1))^2 + o_{\mathbb{P}} \left(n_v^{-1} \right) \right) = o_{\mathbb{P}}(1).
$$

As a result, $C_{2,v} \xrightarrow[v \to \infty]{\mathbb{P}} 0$. Finally, we conclude

$$
C_v \xrightarrow[v \to \infty]{\mathbb{P}} \sigma^2 \frac{1 - \pi}{\pi} \mathbb{E}_\mathbf{x} \left[2 - p(\mathbf{x}) \right] + \sigma^2 \frac{1 - \pi}{\pi} M(\alpha).
$$

Overall,

$$
c_{1,\alpha} = C_1(\alpha^\star) + \sigma^2 \frac{1 - \pi}{\pi} \mathbb{E}_\mathbf{x} \left[2 - p(\mathbf{x}) \right] + \sigma^2 \frac{1 - \pi}{\pi} M(\alpha).
$$

Asymptotic behavior of AV_2. We proceed similarly as before.

Uniform integrability. In Lemma 23, we prove that $\mathbb{E}_{mq}[(\mathbb{V}_m(\mathbb{E}_p[\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v)]))^2]$ is almost surely bounded, from which we deduce uniform integrability.

Limit determination. Write

$$
\mathbb{V}_m \left(\mathbb{E}_p \left[\sqrt{N_v} \left(\widetilde{\mu}_{\alpha,v} - \mu_v \right) \right] \right)
$$

$$
= \sigma^2 \sum_{k \in U_v} \frac{1 - r_k + r_k (\pi \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha})^2}{N_v}
$$

$$
= \sigma^2 \sum_{k \in U_v} \frac{1 - r_k}{N_v} + \sigma^2 \mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_{k,\alpha}^\top \right] \left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1}
$$

$$
\times \sum_{k \in U_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_{k,\alpha} \right]
$$

$$
\xrightarrow[v \to \infty]{\mathbb{P}} \sigma^2 \mathbb{E}_\mathbf{x} \left[1 - p(\mathbf{x}) \right] + \sigma^2 \mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha^\top \right] \left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_{k,\alpha} \right]
$$

$$
= \sigma^2 \mathbb{E}_\mathbf{x} \left[1 - p(\mathbf{x}) \right] + \sigma^2 M(\alpha).
$$

As a result,

$$
c_{2,\alpha} = \sigma^2 \mathbb{E}_\mathbf{x}[1 - p(\mathbf{x})] + \sigma^2 M(\alpha).
$$

Putting all things together. By absorbing the terms that do not depend on α, called $C(\alpha^\star)$, we conclude

$$
\mathbb{AV} \left(\sqrt{N_v}(\widehat{\mu}_{\alpha,v} - \mu_v) \right) = \lim_{v \to \infty} \left(\mathbb{E}_{mq} \left[\mathbb{V}_p \left(\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v) \right) \right] \right)
$$

$$+ \mathbb{E}_q \left[\mathbb{V}_m \left(\mathbb{E}_p \left[\sqrt{N_v}(\widetilde{\mu}_{\alpha,v} - \mu_v) \right] \right) \right] \right)$$

$$= C_1(\alpha^\star) + \sigma^2 \frac{1 - \pi}{\pi} \mathbb{E}_{\mathbf{x}}[2 - p(\mathbf{x})] + \sigma^2 \frac{1 - \pi}{\pi} M(\alpha)$$

$$+ \sigma^2 \mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})] + \sigma^2 M(\alpha)$$

$$= C(\alpha^\star) + \frac{\sigma^2}{\pi} M(\alpha).$$

Proof of Corollary 1.

Statement (i) is obvious as $\|\boldsymbol{\beta}_{\alpha^c}\|_2 = 0$ which implies that $\boldsymbol{\beta}_{\alpha^c} = \mathbf{0}_{\alpha^c}$. For statement (ii), we need to show that under our assumption,

$$\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha^c}^\top \boldsymbol{\beta}_{\alpha^c} \right] = \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha^\top (1 - p(\mathbf{x})) \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha \mathbf{x}_{\alpha^c}^\top \boldsymbol{\beta}_{\alpha^c} p(\mathbf{x}) \right]$$

where we have suppressed the index for simplicity of notation. This is equivalent to showing that

$$\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))x_{\alpha^c,j} \right] = \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha^\top (1 - p(\mathbf{x})) \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha x_{\alpha^c,j} \right], \quad (19)$$

$j \in \alpha_{\text{mis}}$. Since, for $j \in \alpha^c \backslash \alpha_{\text{mis}}$, $\beta_{\alpha^c,j} = 0$, we can express the left side of (19) as

$$L_j := \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))x_{\alpha^c,j} \right] = \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_\alpha^\top \gamma_j \right],$$

using both the conditional independence (a) and the linear link (b). Similarly, using the same technique, we may write

$$\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha x_{\alpha^c,j} \right] = \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \gamma_j.$$

Therefore, the right side of (19) is equal to

$$R_j := \mathbb{E}_{\mathbf{x}} \left[\mathbf{x}_\alpha^\top (1 - p(\mathbf{x})) \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_\alpha x_{\alpha^c,j} \right]$$

$$= \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_\alpha^\top \right] \gamma_j = L_j,$$

$j \in \alpha_{\text{mis}}$. The result follows.

Proof of Corollary 2.

Proof of (i).

From Proposition 3, the equality holds when, for $j \in \alpha_2 - \alpha_1$, we need to show

$$\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_1}^\top \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1} \mathbf{x}_{\alpha_1}^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}}[p(\mathbf{x})\mathbf{x}_{\alpha_1} x_{\alpha_2 - \alpha_1,j}] = \mathbb{E}_{\mathbf{x}}[(1 - p(\mathbf{x}))x_{\alpha_2 - \alpha_1,j}].$$
$$(20)$$

Recall that if Y and Z are conditionally independent given X, then $\mathbb{E}[YZ|X] = \mathbb{E}[Y|X]\mathbb{E}[Z|X]$. Next, we need to prove that the left-hand side and the right-hand side of (20) are equal. The left-hand side of (20) gives

$$L_j = \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_1}^\top \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1} \mathbf{x}_{\alpha_1}^\top \right] \right)^{-1} \mathbb{E}_{\mathbf{x}}[p(\mathbf{x})\mathbf{x}_{\alpha_1} x_{\alpha_2 - \alpha_1,j}]$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} [\mathbb{E}_{\mathbf{x}}[p(\mathbf{x})\mathbf{x}_{\alpha_1} x_{\alpha_2-\alpha_1,j}|\mathbf{x}_{\alpha_1}]] \\
&= \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} [\mathbf{x}_{\alpha_1}\mathbb{E}_{\mathbf{x}}[p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbb{E}_{\mathbf{x}}[x_{\alpha_2-\alpha_1,j}|\mathbf{x}_{\alpha_1}]] \\
&= \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[\mathbb{E}_{\mathbf{x}}[p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top}\boldsymbol{\gamma}_j \right] \\
&= \mathbb{E}_{\mathbf{x}} \left[\mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbf{x}_{\alpha_1}^{\top} \right] \boldsymbol{\gamma}_j \\
&= \mathbb{E}_{\mathbf{x}} \left[\mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbf{x}_{\alpha_1}^{\top}\boldsymbol{\gamma}_j \right]
\end{aligned}
$$

Similarly, the right-hand side of (20) gives

$$
\begin{aligned}
R_j &= \mathbb{E}_{\mathbf{x}}[(1 - p(\mathbf{x}))x_{\alpha_2-\alpha_1,j}] = \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{\mathbf{x}}[(1 - p(\mathbf{x}))x_{\alpha_2-\alpha_1,j}|\mathbf{x}_{\alpha}]] \\
&= \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbb{E}_{\mathbf{x}}[x_{\alpha_2-\alpha_1,j}|\mathbf{x}_{\alpha_1}]] \\
&= \mathbb{E}_{\mathbf{x}} \left[\mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})|\mathbf{x}_{\alpha_1}]\mathbf{x}_{\alpha_1}^{\top}\boldsymbol{\gamma}_j \right].
\end{aligned}
$$

As a result, both sides are equal under our assumptions.

Proof of (ii).

Let $\mathbf{c}$ be a direction satisfying (a) and (b); we need to show that

$$
\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_2-\alpha_1} \right] \neq \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_2-\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1} \right].
$$

We take the Euclidean inner product of the above with $\mathbf{c}$. The left side gives

$$
L = \mathbf{c}^{\top}\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_2-\alpha_1} \right] = \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{c}^{\top}\mathbf{x}_{\alpha_2-\alpha_1} \right] \neq 0
$$

by (b). On the right hand side,

$$
\begin{aligned}
R_j &= \mathbf{c}^{\top}\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha_2-\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1} \right] \\
&= \mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{c}^{\top}\mathbf{x}_{\alpha_2-\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \left(\mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] \right)^{-1} \mathbb{E}_{\mathbf{x}} \left[p(\mathbf{x})\mathbf{x}_{\alpha_1} \right] \\
&= 0
\end{aligned}
$$

since, by (a),

$$
\mathbb{E}_{\mathbf{x}} \left[(1 - p(\mathbf{x}))\mathbf{c}^{\top}\mathbf{x}_{\alpha_2-\alpha_1}\mathbf{x}_{\alpha_1}^{\top} \right] = \mathbf{0}.
$$

Proof of Lemma 1.

Assuming that

$$
\lim_{v\to\infty} \mathbb{P}_m \left(\widetilde{\alpha}_{U_v} = \alpha^{\star} \right) = 1 \tag{21}
$$

almost surely, we wish to prove that

$$
\lim_{v\to\infty} \mathbb{P}_{mpq} \left(\widehat{\alpha}_{S_{r,v}} \neq \alpha^{\star} \right) = \lim_{v\to\infty} \mathbb{P}_{mpq} \left(C_{S_{r,v}} \left(\widehat{\alpha}_{S_{r,v}} \right) < C_{S_{r,v}} \left(\alpha^{\star} \right) \right) = 0
$$

almost surely. With a slight abuse of notation, we write $\mathbb{P}_{\cdot|\cdot}$ to denote a generic conditional distribution, whose precise specification may vary but will be clear from the context. Using the missing at random assumption and non-informativeness of the sampling design, we get

$$
\mathbb{P}_m \left(C_{S_{r,v}} \left(\widehat{\alpha}_{S_{r,v}} \right) < C_{S_{r,v}} \left(\alpha^{\star} \right) \right) = \int_{\mathbb{R}^{n_{r,v}}} \mathbb{1}(C_{S_{r,v}} \left(\widehat{\alpha}_{S_{r,v}} \right) < C_{S_{r,v}} \left(\alpha^{\star} \right)) \prod_{k\in S_{r,v}} \mathbb{P}_{y_k|\mathbf{x}_k} \left(\mathrm{d}y_k \right)
$$

$$\xrightarrow{\nu \to \infty} 0,$$

where the limit follows by (21) and Lemma 13. Since

$$\mathbb{P}_{mpq}\left(C_{S_{r,\nu}}\left(\widehat{\alpha}_{S_{r,\nu}}\right) < C_{S_{r,\nu}}\left(\alpha^\star\right)\right) = \mathbb{E}_{pq}\left[\mathbb{P}_m\left(C_{S_{r,\nu}}\left(\widehat{\alpha}_{S_{r,\nu}}\right) < C_{S_{r,\nu}}\left(\alpha^\star\right)\right)\right],$$

an application of Lebesgue dominated convergence gives the result.

Proof of Theorem 2.

Write

$$\sqrt{n_\nu}\left(\widehat{\mu}_{\widehat{\alpha},\nu} - \mu_\nu\right) = \sqrt{n_\nu}\left(\widehat{\mu}_{\alpha^\star,\nu} - \mu_\nu\right) + \sqrt{n_\nu}\left(\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}\right).$$

We will show that the second term vanishes in probability. Observe that

$$\left\{\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\right\} = \left(\left\{\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\right\} \cap \{\widehat{\alpha} = \alpha^\star\}\right)$$
$$\biguplus\left(\left\{\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\right\} \cap \{\widehat{\alpha} \neq \alpha^\star\}\right).$$

Moreover, on the event $\{\widehat{\alpha} = \alpha\}$, we have $\widehat{\mu}_{\widehat{\alpha},\nu} = \widehat{\mu}_{\alpha^\star,\nu}$ so that

$$\left\{\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\right\} \cap \{\widehat{\alpha} = \alpha^\star\} = \emptyset.$$

Thus,

$$\mathbb{P}_{mpq}\left(\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\right) = \mathbb{P}_{mpq}\left(\sqrt{n_\nu}|\widehat{\mu}_{\widehat{\alpha},\nu} - \widehat{\mu}_{\alpha^\star,\nu}| > \epsilon\mid \widehat{\alpha} \neq \alpha^\star\right)\mathbb{P}_{mpq}\left(\widehat{\alpha} \neq \alpha^\star\right)$$
$$\leq \mathbb{P}_{mpq}\left(\widehat{\alpha} \neq \alpha^\star\right)$$

converges to 0 by assumption.

Proof of Theorem 3.

We adapt the proof of Kim and Rao (2009) to the unweighted case. Let

$$\widehat{U}\left(\boldsymbol{\beta}_\alpha\right) = \sum_{k \in S_\nu} \frac{r_k \mathbf{x}_{k,\alpha}\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)}{N_\nu}. \tag{22}$$

Note that $\widehat{\boldsymbol{\beta}}_{\alpha,\nu}$ given by (4) is the solution of $\widehat{U}(\boldsymbol{\beta}_\alpha) = \mathbf{0}_\alpha$. Let

$$\widehat{\mu}_\nu(\boldsymbol{\beta}_\alpha) = \frac{1}{N_\nu}\left(\sum_{k \in S_{r,\nu}} \frac{y_k}{\pi_k} + \sum_{k \in S_{m,\nu}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}\right),$$

which is the function of $\boldsymbol{\beta}_\alpha$. We further define

$$\widetilde{\mu}_\nu\left(\boldsymbol{\gamma}, \boldsymbol{\beta}_\alpha\right) = \widehat{\mu}_\nu\left(\boldsymbol{\beta}_\alpha\right) + \boldsymbol{\gamma}^\top \widehat{U}\left(\boldsymbol{\beta}_\alpha\right),$$

which is seen as a function of $\boldsymbol{\beta}_\alpha$ and $\boldsymbol{\gamma}$. Our goal is to find a particular choice of $\boldsymbol{\gamma}$, called $\boldsymbol{\gamma}^\star$ such that $\widetilde{\mu}_\nu(\boldsymbol{\gamma}^\star, \widehat{\boldsymbol{\beta}}_{\alpha,\nu}) - \widetilde{\mu}_\nu(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}_{\alpha,\nu}) = o_\mathbb{P}(n_\nu^{-1/2})$. If so, the effect of estimating $\boldsymbol{\beta}_\alpha$ can be ignored by choosing $\boldsymbol{\gamma} = \boldsymbol{\gamma}^\star$. By noting $\widetilde{\mu}_\nu(\boldsymbol{\gamma}, \widehat{\boldsymbol{\beta}}_{\alpha,\nu}) = \widehat{\mu}_{\alpha,\nu}$ for every p_α-dimensional vector $\boldsymbol{\gamma}$, we have $\widehat{\mu}_{\alpha,\nu} - \widetilde{\mu}_\nu(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}_{\alpha,\nu}) = o_\mathbb{P}(n_\nu^{-1/2})$. To find $\boldsymbol{\gamma}^\star$, we use the theory of Randles (1982), having

proved $\sqrt{N_v}(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha) = O_{\mathbb{P}}(1)$ in Lemma 3, then $\widetilde{\mu}_v(\boldsymbol{\gamma}^\star, \widehat{\boldsymbol{\beta}}_{\alpha,v}) - \widetilde{\mu}_v(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}_\alpha) = o_{\mathbb{P}}(n_v^{-1/2})$
holds if

$$\mathbb{E}_{mp}\left[\left.\frac{\partial \widetilde{\mu}_v(\boldsymbol{\gamma}, \boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\beta}_\alpha}\right|_{\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha}\right] = \mathbb{E}_{mp}\left[\left.\frac{\partial \widehat{\mu}_v(\boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\beta}_\alpha}\right|_{\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha}\right] - \boldsymbol{\gamma}^\top \mathbb{E}_{mp}\left[\left.\frac{\partial \widehat{U}(\boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\beta}_\alpha}\right|_{\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha}\right] = 0.$$

The solution of $\boldsymbol{\gamma}^\star$ is given by

$$\boldsymbol{\gamma}^\star = -\left(\mathbb{E}_{mp}\left[\left.\frac{\partial \widehat{U}(\boldsymbol{\beta}_\alpha)^\top}{\partial \boldsymbol{\beta}_\alpha}\right|_{\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha}\right]\right)^{-1} \mathbb{E}_{mp}\left[\left.\frac{\partial \widehat{\mu}_v(\boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\beta}_\alpha}\right|_{\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_\alpha}\right]$$

$$= \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1} \sum_{k \in U_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{N_v} = \mathbf{c}_{\alpha,v}.$$

Next we will show that $\widetilde{\mu}_v(\boldsymbol{\gamma}^\star, \boldsymbol{\beta}_\alpha) = \widetilde{\mu}_{\alpha,v}$. Write $\widetilde{\mu}_v(\boldsymbol{\gamma}, \boldsymbol{\beta}_\alpha)$ as

$$\widetilde{\mu}_v(\boldsymbol{\gamma}, \boldsymbol{\beta}_\alpha) = \widehat{\mu}_v(\boldsymbol{\beta}_\alpha) + \boldsymbol{\gamma}^\top \widehat{U}(\boldsymbol{\beta}_\alpha)$$

$$= \frac{1}{N_v}\left(\sum_{k \in S_{r,v}} \frac{y_k}{\pi_k} + \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k} + \boldsymbol{\gamma}^\top \sum_{k \in S_{r,v}} \frac{\pi_k\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)\mathbf{x}_{k,\alpha}}{\pi_k}\right)$$

$$= \frac{1}{N_v}\left(\sum_{k \in S_{r,v}} \frac{y_k + \pi_k \boldsymbol{\gamma}^\top\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)\mathbf{x}_{k,\alpha}}{\pi_k} + \sum_{k \in S_{m,v}} \frac{\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}\right)$$

$$= \frac{1}{N_v}\sum_{k \in S_v} \frac{r_k y_k + r_k \pi_k \boldsymbol{\gamma}^\top \mathbf{x}_{k,\alpha}\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right) + (1 - r_k)\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}$$

$$= \frac{1}{N_v}\sum_{k \in S_v} \frac{r_k\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right) + r_k \pi_k \boldsymbol{\gamma}^\top \mathbf{x}_{k,\alpha}\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right) + \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}$$

$$= \frac{1}{N_v}\sum_{k \in S_v} \frac{r_k\left(1 + \pi_k \boldsymbol{\gamma}^\top \mathbf{x}_{k,\alpha}\right)\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right) + \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}.$$

Recall that

$$\widetilde{\mu}_{\alpha,v} = \frac{1}{N_v}\sum_{k \in S_v} \frac{r_k\left(1 + \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha}\right)\left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right) + \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha}{\pi_k}.$$

Taking $\boldsymbol{\gamma}^\star = \mathbf{c}_{\alpha,v}$ leads to $\widetilde{\mu}_v(\mathbf{c}_{\alpha,v}, \boldsymbol{\beta}_\alpha) = \widetilde{\mu}_{\alpha,v}$, we finally conclude $\widehat{\mu}_{\alpha,v} - \widetilde{\mu}_{\alpha,v} = o_{\mathbb{P}}(n_v^{-1/2})$.

Proof of Theorem 4.

Write

$$n_v\left|\widehat{V}_{T,v}(\alpha) - V_{T,v}(\alpha)\right| \le n_v\left|\widehat{V}_{1,v}(\alpha) - V_{1,v}(\alpha)\right| + n_v\left|\widehat{V}_{2,v}(\alpha) - V_{2,v}(\alpha)\right| := (a) + (b).$$

Consistency of (a). We further decompose $n_v |\widehat{V}_1(\alpha) - V_1(\alpha)|$ as

$$n_v \left| \widehat{V}_{1,v}(\alpha) - V_{1,v}(\alpha) \right| \le n_v \left| \widehat{V}_{1,v}(\alpha) - \widetilde{V}_{1,v}(\alpha) \right| + n_v \left| \widetilde{V}_{1,v}(\alpha) - \bar{V}_{1,v}(\alpha) \right|$$
$$+ n_v \left| \bar{V}_{1,v}(\alpha) - V_{1,v}(\alpha) \right|$$
$$:= A_v + B_v + C_v,$$

where

$$\widetilde{V}_{1,v}(\alpha) = \frac{1}{N_v^2} \sum_{k \in S_v} \sum_{l \in S_v} \frac{\eta_{k,\alpha}}{\pi_k} \frac{\eta_{l,\alpha}}{\pi_l} \frac{\Delta_{kl}}{\pi_{kl}},$$

$$\bar{V}_{1,v}(\alpha) = \mathbb{V}_p \left(\tilde{\mu}_{\alpha,v} - \mu_v \right) = \frac{1}{N_v^2} \sum_{k \in U_v} \sum_{l \in U_v} \frac{\eta_{k,\alpha}}{\pi_k} \frac{\eta_{l,\alpha}}{\pi_l} \Delta_{kl}, \tag{23}$$

Consistency of A_v. Let

$$c_{kl} = \frac{\Delta_{kl} I_k I_l}{\pi_k \pi_l \pi_{kl}}, \qquad k, l \in U_v.$$

Using the equality, for arbitrary $k, l \in U_v$,

$$\widehat{\eta}_{k,\alpha} \widehat{\eta}_{l,\alpha} - \eta_{k,\alpha} \eta_{l,\alpha} = \left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right) \left(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha} \right) + \eta_{k,\alpha} \left(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha} \right) + \eta_{l,\alpha} \left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right),$$

A_v can be decomposed as

$$A_v \le \frac{n_v}{N_v^2} \sum_{k \in U_v} \sum_{l \in U_v} \left| c_{kl} \left(\widehat{\eta}_{k,\alpha} \widehat{\eta}_{l,\alpha} - \eta_{k,\alpha} \eta_{l,\alpha} \right) \right|$$
$$\le \frac{n_v}{N_v^2} \sum_{k \in U_v} \sum_{l \in U_v} \left| c_{kl} \left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right) \left(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha} \right) \right|$$
$$+ \frac{n_v}{N_v^2} \sum_{k \in U_v} \sum_{l \in U_v} \left| c_{kl} \left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right) \eta_{l,\alpha} \right| + \frac{n_v}{N_v^2} \sum_{k \in U_v} \sum_{l \in U_v} \left| c_{kl} \eta_{k,\alpha} \left(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha} \right) \right|$$
$$:= A_{1,v} + A_{2,v} + A_{3,v}.$$

Next, we will treat $A_{1,v}$ and $A_{2,v}$, and $A_{3,v}$ separately. For $A_{1,v}$, since Lemma 8 shows

$$\sum_{k \in U_v} \frac{\left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right)^2}{N_v} = o_{\mathbb{P}}(1),$$

we have

$$A_{1,v} \le \frac{n_v}{N_v \lambda^2} \sum_{k \in U_v} \frac{\left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right)^2}{N_v}$$
$$+ \frac{n_v}{N_v \lambda^2 \lambda^\star} \max_{k \ne l \in U_v} |\Delta_{kl}| \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ k \ne l}} \frac{|(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha})|}{N_v}$$
$$\le \frac{n_v}{N_v \lambda^2} \sum_{k \in U_v} \frac{\left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right)^2}{N_v}$$

$$+ \frac{n_v}{N_v \lambda^2 \lambda^\star} \max_{k \neq l \in U_v} |\Delta_{kl}| \sum_{k \in U_v} \sum_{l \in U_v} \frac{|(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})(\widehat{\eta}_{l,\alpha} - \eta_{l,\alpha})|}{N_v}$$

$$\leq \left(\frac{n_v}{N_v \lambda^2} + \frac{n_v}{\lambda^2 \lambda^\star} \max_{k \neq l \in U_v} |\Delta_{kl}| \right) \sum_{k \in U_v} \frac{(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})^2}{N_v} = o_{\mathbb{P}}(1).$$

By symmetry, for $A_{2,v}$, using $\sum_{k \in U_v} \sum_{l \in U_v} a_k b_l \leq N_v \left(\sum_{k \in U_v} a_k^2 \right)^{1/2} \left(\sum_{l \in U_v} b_l \right)^{1/2}$, we have

$$A_{2,v} \leq \frac{n_v}{N_v \lambda^2} \sum_{k \in U_v} \frac{|(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})\eta_{k,\alpha}|}{N_v} + \frac{n_v}{N_v^2 \lambda^2 \lambda^\star} \max_{k \neq l \in U_v} |\Delta_{kl}| \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ k \neq l}} |(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})\eta_{l,\alpha}|$$

$$\leq \frac{n_v}{N_v \lambda^2} \sum_{k \in U_v} \frac{|(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})\eta_{k,\alpha}|}{N_v} + \frac{n_v}{N_v^2 \lambda^2 \lambda^\star} \max_{k \neq l \in U_v} |\Delta_{kl}| \sum_{k \in U_v} \sum_{l \in U_v} |(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})\eta_{l,\alpha}|$$

$$\leq \left(\frac{n_v}{N_v \lambda^2} + \frac{n_v}{\lambda^2 \lambda^\star} \max_{k \neq l \in U_v} |\Delta_{kl}| \right) \sqrt{\sum_{k \in U_v} \frac{\eta_{k,\alpha}^2}{N_v}} \sqrt{\sum_{k \in U_v} \frac{(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})^2}{N_v}} = o_{\mathbb{P}}(1).$$

As a result, we have $A_v = o_{\mathbb{P}}(1)$. Furthermore, we have $n_v |\widehat{V}_{1,v}(\alpha) - V_{1,v}(\alpha)| = o_{\mathbb{P}}(1)$.

Consistency of B_v. Write

$$\mathbb{E}_p \left[\left(n_v (\widetilde{V}_{1,v}(\alpha) - \bar{V}_{1,v}(\alpha)) \right)^2 \right]$$

$$\leq \frac{2n_v^2}{N_v^4} \sum_{k \in U_v} \sum_{l \in U_v} \frac{1 - \pi_k}{\pi_k} \frac{1 - \pi_l}{\pi_l} \frac{\eta_{k,\alpha}^2}{\pi_k} \frac{\eta_{l,\alpha}^2}{\pi_l} \Delta_{kl}$$

$$+ \frac{2n_v^2}{N_v^4} \sum_{i \in U_v} \sum_{\substack{j \in U_v \\ j \neq i}} \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ l \neq k}} \frac{\eta_{i,\alpha} \eta_{j,\alpha} \eta_{k,\alpha} \eta_{l,\alpha}}{\pi_i \pi_j \pi_k \pi_l} \mathbb{E}_p \left[\frac{(I_i I_j - \pi_{ij})(I_k I_l - \pi_{kl})}{\pi_{ij} \pi_{kl}} \right] \Delta_{ij} \Delta_{kl}$$

$$:= B_{1,v} + B_{2,v}.$$

We will use the same argument of Theorem 3 in Breidt and Opsomer (2000). For $B_{1,v}$, we have

$$B_{1,v} \leq \left(\frac{2}{N_v \lambda^3} + \frac{2n_v \max_{k \neq l \in U_v} |\Delta_{kl}|}{N_v \lambda^4} \right) \sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v}.$$

Recall from Lemma 20 gives

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} < \infty,$$

almost surely. It follows that $B_{1,v}$ converges to 0 almost surely. For $B_{2,v}$, we have

$$B_{2,v} \leq L_v + \frac{2 \left(n_v \max_{k \neq l \in U_v} |\Delta_{kl}| \right)^2}{\lambda^4 \lambda^{\star 2}}$$

$$\times \max_{(i,j,k,l) \in D_{4,N_v}} \left| \mathbb{E}_p \left[\frac{(I_i I_j - \pi_{ij})(I_k I_l - \pi_{kl})}{\pi_{ij} \pi_{kl}} \right] \right| \sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v}.$$

Here, L_v converges to 0 almost surely. It follows that $B_{2,v}$ converges to 0 almost surely. As a result, for any $\epsilon > 0$, an application of Chebyshev's inequality gives

$$\lim_{v \to \infty} \mathbb{P}_p \left(\left| n_v (\widetilde{V}_{1,v}(\alpha) - \bar{V}_{1,v}(\alpha)) \right| > \epsilon \right) = 0,$$

almost surely. It follows that

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\left| n_v (\widetilde{V}_{1,v}(\alpha) - \bar{V}_{1,v}(\alpha)) \right| > \epsilon \right) = 0,$$

almost surely. This concludes $B_v = o_{\mathbb{P}}(1)$.

Consistency of C_v. C_v can be decomposed as

$$C_v = n_v \left| \bar{V}_{1,v}(\alpha) - V_{1,v}(\alpha) \right| \leq \left| n_v \bar{V}_{1,v}(\alpha) - c^\star_{1,\alpha} \right| + \left| c^\star_{1,\alpha} - n_v V_{1,v}(\alpha) \right|,$$

where $c^\star_{1,\alpha} = f^\star c_{1,\alpha}$ with $f^\star = \lim_{v \to \infty} n_v / N_v$. On the one hand, we have

$$\left| n_v \bar{V}_{1,v}(\alpha) - c^\star_{1,\alpha} \right| = o_{\mathbb{P}}(1),$$

by assumption. The remaining part is to show

$$\lim_{v \to \infty} \left| c^\star_{1,\alpha} - n_v V_{1,v}(\alpha) \right| = \lim_{v \to \infty} \left| \mathbb{E}_{mq} \left[n_v \bar{V}_{1,v}(\alpha) - c^\star_{1,\alpha} \right] \right| = 0$$

almost surely. We aim to apply Vitali convergence theorem, that is to show that $|n_v \bar{V}_{1,v}(\alpha) - c^\star_{1,\alpha}| = o_{\mathbb{P}}(1)$ and that $n_v \bar{V}_{1,v}(\alpha)$ is uniformly integrable. From this, it will follow that $\lim_{v \to \infty} \mathbb{E}_{mq}[|n_v \bar{V}_{1,v}(\alpha) - c^\star_{1,\alpha}|] = 0$ almost surely. To show the uniform integrability of $n_v \bar{V}_{1,v}(\alpha)$, it suffices to prove that $\mathbb{E}_{mq}[(n_v \bar{V}_{1,v}(\alpha))^2]$ is almost surely bounded, which follows from Lemma 23.

Consistency of (b). The term $n_v |\hat{V}_{2,v}(\alpha) - V_{2,v}(\alpha)|$ can be decomposed as

$$n_v \left| \hat{V}_{2,v}(\alpha) - V_{2,v}(\alpha) \right| \leq n_v \left| \hat{V}_{2,v}(\alpha) - \bar{V}_{2,v}(\alpha) \right| + n_v \left| \bar{V}_{2,v}(\alpha) - V_{2,v}(\alpha) \right| := D_v + E_v,$$

where

$$\bar{V}_{2,v}(\alpha) = \mathbb{V}_m \left(\mathbb{E}_p \left[\tilde{\mu}_{\alpha,v} - \mu_v \right] \right) = \sigma^2 \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}^\top_{\alpha,v} \mathbf{x}_{k,\alpha} \right)^2}{N_v^2} \qquad (24)$$

Consistency of D_v. Combining the results $|\hat{\sigma}^2_{\alpha,v} - \sigma^2| = o_{\mathbb{P}}(1)$ in Lemma 5 and

$$\left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}^\top_{\alpha,v} \mathbf{x}_{k,\alpha} \right)^2}{\pi_k N_v} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}^\top_{\alpha,v} \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right| = o_{\mathbb{P}}(1),$$

from (40), we have

$$\left| \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}^\top_{\alpha,v} \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right| \leq 1 + \|\mathbf{c}_{\alpha,v}\|_2^2 \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} = O_{\mathbb{P}}(1).$$

It follows that

$$n_v \left| \widehat{V}_{2,v}(\alpha) - \bar{V}_{2,v}(\alpha) \right| \leq \frac{n_v}{N_v} \left| \left(\widehat{\sigma}^2_{\alpha,v} - \sigma^2 \right) \left(\sum_{k \in S_v} \frac{1 - r_k + r_k (\pi_k \widehat{\mathbf{c}}^\top_{\alpha,v} \mathbf{x}_{k,\alpha})^2}{\pi_k N_v} \right. \right.$$

$$- \sum_{k \in U_v} \frac{1 - r_k + r_k (\pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha})^2}{N_v} \Bigg) \Bigg|$$

$$+ \frac{n_v}{N_v} \sigma^2 \Bigg| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2}{\pi_k N_v}$$

$$- \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2}{N_v} \Bigg|$$

$$+ \frac{n_v}{N_v} \Bigg| \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2}{N_v} \Bigg| \left| \widehat{\sigma}_{\alpha,v}^2 - \sigma^2 \right| = o_{\mathbb{P}}(1).$$

Consistency of E_v.   The term E_v can be decomposed as

$$E_v = n_v \left| \bar{V}_{2,v}(\alpha) - V_{2,v}(\alpha) \right| \le \left| n_v \bar{V}_{2,v}(\alpha) - c_{2,\alpha}^{\star} \right| + \left| c_{2,\alpha}^{\star} - n_v V_{2,v}(\alpha) \right|,$$

where $c_{2,\alpha}^{\star} = f^{\star} c_{2,\alpha}$ with $f^{\star} = \lim_{v \to \infty} n_v / N_v$. On the one hand, we have

$$\left| n_v \bar{V}_{2,v}(\alpha) - c_{2,\alpha}^{\star} \right| = o_{\mathbb{P}}(1),$$

by assumption. The remaining part is to show

$$\lim_{v \to \infty} \left| c_{2,\alpha}^{\star} - n_v V_{2,v}(\alpha) \right| = \lim_{v \to \infty} \left| \mathbb{E}_q \left[n_v \bar{V}_{2,v}(\alpha) - c_{2,\alpha}^{\star} \right] \right| = 0$$

almost surely. We proceed again by Vitali's convergence theorem. The uniform integrability of $n_v \bar{V}_{2,v}(\alpha)$ follows from Lemma 23.

Proof of Theorem 5.

Write

$$\frac{\widehat{V}_{T,v}(\widehat{\alpha})}{V_{T,v}(\alpha^{\star})} = 1 + \frac{n_v \left(\widehat{V}_{T,v}(\widehat{\alpha}) - \widehat{V}_{T,v}(\alpha^{\star}) \right)}{n_v V_{T,v}(\alpha^{\star})} + \frac{n_v \left(\widehat{V}_{T,v}(\alpha^{\star}) - V_{T,v}(\alpha^{\star}) \right)}{n_v V_{T,v}(\alpha^{\star})} := 1 + A_v + B_v. \tag{25}$$

It is therefore enough to show that $A_v = o_{\mathbb{P}}(1)$ and $B_v = o_{\mathbb{P}}(1)$. Notice that, by assumption, satisfies $n_v \lim_{v \to \infty} V_{T,v}(\alpha^{\star}) > 0$ so that it is enough to show that the numerators of A_v and B_v converge to 0 in probability. For A_v, let $\epsilon > 0$ and write

$$\mathbb{P}_{mpq} \left(n_v \left| \widehat{V}_{T,v}(\widehat{\alpha}) - \widehat{V}_{T,v}(\alpha^{\star}) \right| > \epsilon \right) = \mathbb{P}_{mpq} \left(n_v \left| \widehat{V}_{T,v}(\widehat{\alpha}) - \widehat{V}_{T,v}(\alpha^{\star}) \right| > \epsilon , \ \widehat{\alpha} = \alpha^{\star} \right)$$

$$+ \mathbb{P}_{mpq} \left(n_v \left| \widehat{V}_{T,v}(\widehat{\alpha}) - \widehat{V}_{T,v}(\alpha^{\star}) \right| > \epsilon , \ \widehat{\alpha} \ne \alpha^{\star} \right)$$

$$\le \mathbb{P}_{mpq} \left(\widehat{\alpha} \ne \alpha^{\star} \right) \xrightarrow[v \to \infty]{} 0$$

almost surely since the model selection criterion is consistent.

For B_v, given that $\alpha^{\star} \in C$, it follows directly that $B_v = o_{\mathbb{P}}(1)$ by Theorem 4.

Proof of Theorem 6.

To prove the asymptotic normality of $\widehat{\mu}_{\alpha^\star,v} - \mu_v$, we verify the conditions of Theorem 2 of Chen and Rao (2007). We decompose

$$\widehat{\mu}_{\alpha^\star,v} - \mu_v = \underbrace{\frac{1}{N_v} \sum_{k \in U_v} v_k \eta_{k,\alpha^\star}}_{:=U_v} + \underbrace{\frac{1}{N_v} \sum_{k \in U_v} (\eta_{k,\alpha} - y_k)}_{:=V_v}, \tag{26}$$

with $v_k := I_k/\pi_k - 1$ for $k \in U_v$. Define $\mathcal{B}_v := \sigma\left((\mathbf{x}_k, r_k, y_k)_{k \in U_v}\right)$. We need to verify the three conditions of Theorem 2 of Chen and Rao (2007), which we label as (i), (ii), and (iii), respectively.

Verification of (i).　We wish to show that V_v is asymptotically normal and $\mathcal{B}_v$-measurable. Measurability follows immediately by noting, for an arbitrary $k \in U_v$,

$$\eta_{k,\alpha} - y_k = \epsilon_k (r_k + r_k \pi_k \mathbf{c}_{\alpha^\star,v}^\top \mathbf{x}_{k,\alpha^\star} - 1) := \epsilon_k w_k$$

with $\mathbf{c}_{\alpha^\star,v}$ depending only on the covariates and response indicators. Moreover, note that $\mathbb{E}_m[\eta_{k,\alpha} - y_k] = 0$. We will start by establishing a conditional central limit theorem via the conditional Lyapunov condition. More precisely, we need to show that there exists $\delta > 0$ such that

$$L_v := \frac{1}{\left[\sum_{k \in U_v} \mathbb{V}_m \left(\frac{\eta_k - y_k}{N_v}\right)\right]^{1+\delta/2}} \sum_{k \in U_v} \mathbb{E}_m \left[\left(\frac{\eta_k - y_k}{N_v}\right)^{2+\delta}\right] \xrightarrow[v \to \infty]{\mathbb{P}} 0.$$

Given our assumptions, it is convenient to show it for $\delta = 2$. We start by noting that

$$\sum_{k \in U_v} \mathbb{V}_m \left(\frac{\eta_k - y_k}{N_v}\right) = \frac{\sigma^2}{N_v^2} \sum_{k \in U_v} \left((1 - r_k) + r_k(\pi_k \mathbf{c}_{\alpha^\star,v}^\top \mathbf{x}_{k,\alpha^\star})^2\right) \geq \frac{\sigma^2}{N_v^2} N_{m,v} \asymp_{\mathbb{P}} \frac{1}{N_v}.$$

Moreover, note that $w_k^4 = 1$ if $r_k = 0$ and $w_k^4 = (\pi_k \mathbf{c}_{\alpha^\star,v}^\top \mathbf{x}_{k,\alpha^\star})^4$ otherwise. Thus, using $(a+b)^4 \leq 8(a^4 + b^4)$, Cauchy-Schwartz inequality and (S3), we get

$$\sum_{k \in U_v} \mathbb{E}_m \left[\left(\frac{\eta_k - y_k}{N_v}\right)^4\right] \leqslant \frac{8M_0}{N_v^4} \sum_{k \in U_v} \left(1 + \|\mathbf{c}_{\alpha^\star,v}\|_2^4 \|\mathbf{x}_{k,\alpha^\star}\|_2^4\right) = O_{\mathbb{P}}\left(\frac{1}{N_v^3}\right),$$

by using (40) and (S2). This shows that $L_v = O_{\mathbb{P}}\left(N_v^{-1}\right)$ and thus the Lyapunov condition holds. Therefore, (i) with

$$\sigma_{1v}^2 := \mathbb{V}_m \left(\sum_{k \in U_v} \frac{\eta_{k,\alpha^\star} - y_k}{N_v}\right)$$

holds. Conditional Gaussianity follows. Moreover, since the asymptotic distribution does not depend on the conditioning, a dominated convergence argument can be used to extend it to an unconditional central limit theorem.

Verification of (ii). We have

$$\mathbb{E}\left[U_v|\mathcal{B}_v\right] = \frac{1}{N_v}\sum_{k\in U_v}\mathbb{E}\left[\nu_k|\mathcal{B}_v\right]\eta_{k,\alpha^\star} = 0,$$

$$\mathbb{V}\left(U_v|\mathcal{B}_v\right) = \frac{1}{N_v^2}\sum_{k\in U_v}\sum_{l\in U_v}\Delta_{kl}\frac{\eta_{k,\alpha^\star}}{\pi_k}\frac{\eta_{l,\alpha^\star}}{\pi_l} := \sigma_{2v}^2,$$

since, for all $k \in U_v$, η_k is $\mathcal{B}_v$-measurable. By assumption, $\sigma_{2v}^{-1}U_v|\mathcal{B}_v \xrightarrow[v\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$, which can be equivalently stated as

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\sigma_{2v}^{-1}U_v \leqslant t|\mathcal{B}_n\right) - \Phi(t)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

where Φ denotes cumulative distribution function of a standard normal random variable. Assumption (ii) is verified.

Verification of (iii). We have

$$N_v\mathbb{V}_m\left(\sum_{k\in U_v}\frac{\eta_{k,\alpha^\star} - y_k}{N_v}\right) = \frac{\sigma^2}{N_v}\sum_{k\in U_v}(1 - r_k) + \sigma^2\mathbf{c}_{\alpha^\star,v}^\top\frac{1}{N_v}\sum_{k\in U_v}r_k\pi_k^2\mathbf{x}_{k,\alpha^\star}\mathbf{x}_{k,\alpha^\star}^\top\mathbf{c}_{\alpha^\star,v}$$

$$\xrightarrow[v\to\infty]{\mathbb{P}} \sigma^2\left\{\mathbb{E}\left[1 - p(\mathbf{x}_{\alpha^\star})\right] + \overline{\mathbf{c}}_{\alpha^\star}^\top C_2\overline{\mathbf{c}}_{\alpha^\star}\right\}$$

where

$$\overline{\mathbf{c}}_{\alpha^\star} = C_1^{-1}\mathbb{E}_{\mathbf{x}}\left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha^\star}\right].$$

Thus, we have

$$\frac{\sigma_{1v}^2}{\sigma_{2v}^2} \xrightarrow[v\to\infty]{\mathbb{P}} \frac{\sigma^2\left\{\mathbb{E}\left[1 - p(\mathbf{x}_{\alpha^\star})\right] + \overline{\mathbf{c}}_{\alpha^\star}^\top C_2\overline{\mathbf{c}}_{\alpha^\star}\right\}}{c_3}.$$

Therefore, (iii) follows.

Putting pieces together. Since the three conditions of Chen and Rao (2007) are satisfied, it follows that $\widehat{\mu}_{\alpha^\star,v} - \mu_v$ is asymptotically normal. Moreover, by Theorem 3, the asymptotic distribution of $\widehat{\mu}_{\widehat{\alpha},v} - \mu_v$ is the same as that of $\widehat{\mu}_{\alpha^\star,v} - \mu_v$, and is thus asymptotically normal as well. An application of Theorem 5 proves the unit variance.

B.2. Consistency of estimators

B.2.1. Consistency of single estimators

Lemma 3. *Let $\alpha \in C$ and $\left(\widehat{\boldsymbol{\beta}}_{\alpha,v}\right)_{v\in\mathbb{N}}$ be a sequence of least-squares estimators given by (4). Assume (S1)-(S3) and (D1)-(D3). Then,*

$$\left\|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}}\right).$$

Proof. Let

$$\widetilde{\boldsymbol{\beta}}_{\alpha,v} = \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_k y_k}{N_v}.$$

We decompose $\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha$ as

$$\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha = \underbrace{\widehat{\boldsymbol{\beta}}_{\alpha,v} - \widetilde{\boldsymbol{\beta}}_{\alpha,v}}_{(a)} + \underbrace{\widetilde{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha}_{(b)}.$$

We need to show that both terms (a) and (b) converge to 0 with rate $O_\mathbb{P}(n_v^{-1/2})$.

Convergence of (a). The triangle inequality gives

$$\begin{aligned}
\left\| \widehat{\boldsymbol{\beta}}_{\alpha,v} - \widetilde{\boldsymbol{\beta}}_{\alpha,v} \right\|_2 &= \left\| \left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v} \right. \hspace{3cm} (27) \\
&\quad \left. - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 \\
&= \left\| \left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} \left(\sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right) \right. \\
&\quad \left. + \left(\left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right) \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 \\
&\leq \left\| \left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} \left(\sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right) \right\|_2 \\
&\quad + \left\| \left(\left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right) \right. \hspace{2cm} (28) \\
&\quad \left. \times \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 \\
&:= A_{1,v} + A_{2,v}
\end{aligned}$$

Next, we consider $A_{1,v}$ and $A_{2,v}$ separately.

Treatment of $A_{1,v}$. For $A_{1,v}$, using the [Schwarz matrix inequality], we obtain

$$A_{1,v} \leq \left\| \left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2$$

We will apply Lemma 9 to obtain

$$\left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} y_k}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 = O_\mathbb{P}\left(\frac{1}{\sqrt{N_v}} \right). \hspace{2cm} (29)$$

Setting $w_k := r_k$, $\boldsymbol{T}_k := \mathbf{x}_{k,\alpha} y_k$, as $\{r_k\}_{k \in U_v}$ is i.i.d. random variables with $\mathbb{E}_q[r_k^2] = p_k \le 1$ almost surely, it remains to verify

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\mathbf{x}_{k,\alpha} y_k\|_2^2\right]}{N_v} < \infty$$

almost surely. Recall from (S2) that the covariates have bounded support, from which we get

$$\sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\mathbf{x}_{k,\alpha} y_k\|_2^2\right]}{N_v} = \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2 \mathbb{E}_m\left[y_k^2\right]}{N_v} = \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2 \mathbb{E}_m\left[\left(\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha + \epsilon_k\right)^2\right]}{N_v}$$

$$\le C_0^2 \sum_{k \in U_v} \frac{2\sigma^2 + 2\left(\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2}{N_v} \le 2C_0^2\sigma^2 + 2C_0^4\|\boldsymbol{\beta}_\alpha\|_2^2 < \infty.$$

It follows that

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha} y_k\|_2^2}{N_v} < \infty,$$

almost surely, from which (29) follows. On the other hand, Lemma 18 shows

$$\left\|\left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v}\right)^{-1}\right\|_{op} = O_{\mathbb{P}}(1).$$

As a result, we obtain $A_{1,v} = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of $A_{2,v}$. The term $A_{2,v}$ can be decomposed as

$$A_{2,v} \le \left\|\left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v}\right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op} \left\|\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} y_k}{N_v}\right\|_2.$$

On the one hand, using Lemma 9 with $w_k := r_k$ and $\boldsymbol{T}_k := \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top$. Since

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top\|_F^2\right]}{N_v} = \limsup_{v \to \infty} \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v} \le C_0^4 \tag{30}$$

almost surely, it follows that

$$\left\|\frac{\boldsymbol{A}_{r,\alpha}}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}}\right). \tag{31}$$

On the other hand, recall that Lemma 15 shows

$$\left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op} = O_{\mathbb{P}}(1).$$

An application of Lemma 14 gives

$$\left\|\left(\frac{\boldsymbol{A}_{r,\alpha}}{N_v}\right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}}\right).$$

Meanwhile,

$$\left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 \leq \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2 |y_k|}{N_v} \leq C_0 \sum_{k \in U_v} \frac{|y_k|}{N_v},$$

which implies

$$\limsup_{v \to \infty} \left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} y_k}{N_v} \right\|_2 < \infty.$$

It follows that $A_{2,v} = O_{\mathbb{P}}(n_v^{-1/2})$ and, therefore,

$$\left\| \widehat{\boldsymbol{\beta}}_{\alpha,v} - \widetilde{\boldsymbol{\beta}}_{\alpha,v} \right\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Convergence of (b). Recall from Lemma 16, we have

$$\limsup_{v \to \infty} \left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \right\|_{op} < \infty$$

almost surely. An application of Chebyshev's inequality shows that for any $\epsilon > 0$, we have

$$\mathbb{P}_{mpq}\left(\left\| \widetilde{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right\|_2^2 > \epsilon \right) \tag{32}$$

$$\leq \frac{1}{\epsilon^2} \mathbb{E}_{mpq}\left[\left(\widetilde{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha \right)^\top \left(\widetilde{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha \right) \right]$$

$$= \frac{1}{\epsilon^2} \mathbb{E}_m \left[\sum_{k \in U_v} \frac{\epsilon_k \mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k) \pi_k}{N_v} \left(\sum_{k \in U_v} \frac{\pi_k p(\mathbf{x}_k) \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \sum_{k \in U_v} \frac{\epsilon_k \mathbf{x}_{k,\alpha} p(\mathbf{x}_k) \pi_k}{N_v} \right]$$

$$\overset{(*)}{=} \frac{\sigma^2}{\epsilon^2} \sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k) \pi_k}{N_v} \left(\sum_{k \in U_v} \frac{\pi_k p(\mathbf{x}_k) \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \frac{\mathbf{x}_{k,\alpha} p(\mathbf{x}_k) \pi_k}{N_v}$$

$$\overset{(**)}{\leq} \frac{\sigma^2}{N_v \epsilon^2} \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} \left\| \left(\sum_{k \in U_v} \frac{\pi_k p(\mathbf{x}_k) \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \right\|_{op}$$

$$\leq \frac{\sigma^2 C_0^2}{N_v \epsilon^2} \left\| \left(\sum_{k \in U_v} \frac{\pi_k p(\mathbf{x}_k) \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \right\|_{op}. \tag{33}$$

Here, $(*)$ is due to the fact that $\mathbb{E}_m[\epsilon_k \epsilon_l] = 0$ for $k \neq l \in U_v$ and $(**)$ from the inequality $\mathbf{z}^\top A \mathbf{z} \leq \|\mathbf{z}\|_2^2 \|A\|_{op}$. The last line of (32) converges to 0 almost surely. Putting all things together, we obtain

$$\left\| \widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

$\square$

Lemma 4. *Assume (S1)-(S2) and (D1)-(D3). If (C_1) holds, then there exists a constant K_0 such that*

$$\lim_{v \to \infty} \mathbb{P}_{mpq}\left(\lambda_{min}(N_v^{-1} \mathbf{S}_v) \geq K_0 \right) = 1$$

almost surely.

Proof. **Step 1.** Prove that

$$\|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op} = O_{\mathbb{P}}(n_v^{-1/2}).$$

Write

$$\|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op}$$

$$= \left\| \frac{1}{N_v}\mathbf{A}_{r,\alpha^c} \right.$$

$$- \sum_{k\in S_{r,v}} \frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top}{N_v}\left(\frac{\mathbf{A}_{r,\alpha}}{N_v}\right)^{-1}\sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top}{N_v} - \left(\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k)\pi_k}{N_v}\right.$$

$$\left.\left. - \sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k)\pi_k}{N_v}\left(\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k)\pi_k}{N_v}\right)^{-1}\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k)\pi_k}{N_v}\right)\right\|_{op}$$

$$\leq \left\| \frac{1}{N_v}\mathbf{A}_{r,\alpha^c} - \sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k)\pi_k}{N_v} \right\|_{op}$$

$$+ \left\| \sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top}{N_v}\left(\frac{\mathbf{A}_{r,\alpha}}{N_v}\right)^{-1}\sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top}{N_v} \right.$$

$$\left. - \sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k)\pi_k}{N_v}\left(\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k)\pi_k}{N_v}\right)^{-1}\sum_{k\in U_v}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top p(\mathbf{x}_k)\pi_k}{N_v} \right\|_{op}$$

$$:= A_v + B_v.$$

Recall that (31) shows $A_v = O_{\mathbb{P}}(n_v^{-1/2})$. For B_v, we decompose

$$B_v \leq \left\| \sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top}{N_v}\left(\left(\frac{\mathbf{A}_{r,\alpha}}{N_v}\right)^{-1}\sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top}{N_v}\right.\right.$$

$$\left.\left. - \left(\sum_{k\in U_v}\frac{p(\mathbf{x}_k)\pi_k\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\sum_{k\in U_v}\frac{p(\mathbf{x}_k)\pi_k\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top}{N_v}\right)\right\|_{op}$$

$$+ \left\| \left(\sum_{k\in S_{r,v}}\frac{\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k\in U_v}\frac{p(\mathbf{x}_k)\pi_k\mathbf{x}_{k,\alpha^c}\mathbf{x}_{k,\alpha}^\top}{N_v}\right) \right.$$

$$\left. \times \left(\sum_{k\in U_v}\frac{p(\mathbf{x}_k)\pi_k\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\sum_{k\in U_v}\frac{p(\mathbf{x}_k)\pi_k\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op}$$

$$:= B_{1,v} + B_{2,v}.$$

We need to prove $B_{1,v}$ and $B_{2,v}$ converges to 0 with $O_{\mathbb{P}}(n_v^{-1/2})$.

Convergence of $B_{1,v}$. We have

$$B_{1,v} \leq \left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op}$$

$$\times \left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right.$$

$$\left. - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_2 .$$

Using (S2), we get

$$\left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} \leq \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2 \|\mathbf{x}_{k,\alpha^c}\|_2}{N_v} \leq C_0^2,$$

almost surely. Next, using the Schwarz matrix inequality, we decompose

$$\left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op}$$

$$\leq \left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} \left(\sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right) \right\|_{op}$$

$$+ \left\| \left(\left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right) \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op}$$

$$\leq \left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op}$$

$$+ \left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op} . \quad (34)$$

Next, we evaluate every component of (34). Lemma 15 and Lemma 18 show that

$$\left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1), \qquad \left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1),$$

respectively. We proceed to apply Lemma 9 with $w_k := r_k$ and $\boldsymbol{T}_k = \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top$. Since

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m \left[\|\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top\|_F^2 \right]}{N_v} = \limsup_{v \to \infty} \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2 \|\mathbf{x}_{k,\alpha^c}\|_2^2}{N_v} \leq C_0^4$$

almost surely, it follows that

$$\left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).$$

Recall again from (31) that

$$\left\| \frac{\mathbf{A}_{r,\alpha}}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

so that, using Lemma 14 gives

$$\left\| \left(\frac{\mathbf{A}_{r,\alpha}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Finally, by noting

$$\left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op} \leq \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \|\mathbf{x}_{k,\alpha}\|_2 \|\mathbf{x}_{k,\alpha^c}\|}{N_v} \leq C_0^2$$

almost surely, we conclude that (34) is $O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of $B_{2,v}$. Using the Schwarz matrix inequality, we have

$$
\begin{aligned}
B_{2,v} &= \left\| \left(\sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \right. \\
&\quad \left. \times \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op} \\
&\leq \left\| \sum_{k \in S_{r,v}} \frac{\mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha^c} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} \\
&\quad \times \left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha^c}^\top}{N_v} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).
\end{aligned}
$$

Putting all things together, we obtain

$$\|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Step 2. On the one hand, Step 1 gives for any $\epsilon > 0$,

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op} \leq \epsilon \right) = 1$$

almost surely. On the other hand, by Weyl's inequality,

$$\left| \lambda_{min}(N_v^{-1}\mathbf{S}_v) - \lambda_{min}(\mathbf{S}_{U_v}) \right| \leq \|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op},$$

from which it follows that

$$\lim_{v \to \infty} \mathbb{P}_{pq} \left(\left| \lambda_{min}(N_v^{-1}\mathbf{S}_v) - \lambda_{min}(\mathbf{S}_{U_v}) \right| \leq \epsilon \right) = 1.$$

As a result, by setting $K_0 = C_0 - \epsilon$, we have

$$
\lim_{v \to \infty} \mathbb{P}_{pq} \left(\left| \lambda_{min}(N_v^{-1}\mathbf{S}_v) - \lambda_{min}(\mathbf{S}_{U_v}) \right| \leq \epsilon \right)
$$

$$
= \lim_{v \to \infty} \mathbb{P}_{pq} \left(-\epsilon \leq \lambda_{min}(N_v^{-1}\mathbf{S}_v) - \lambda_{min}(\mathbf{S}_{U_v}) \leq \epsilon \right)
$$

$$
\leq \lim_{v \to \infty} \mathbb{P}_{pq} \left(\lambda_{min}(\mathbf{S}_{U_v}) - \epsilon \leq \lambda_{min}(N_v^{-1}\mathbf{S}_v) \right)
$$

$$
\leq \lim_{v \to \infty} \mathbb{P}_{pq} \left(C_0 - \epsilon \leq \lambda_{min}(N_v^{-1}\mathbf{S}_v) \right)
$$

$$
= \lim_{v \to \infty} \mathbb{P}_{pq} \left(K_0 \leq \lambda_{min}(N_v^{-1}\mathbf{S}_v) \right) = 1
$$

almost surely. Furthermore, using the Lebesgue convergence theorem, we obtain

$$
\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\|N_v^{-1}\mathbf{S}_v - \mathbf{S}_{U_v}\|_{op} \leq \epsilon \right) = 1
$$

almost surely.

$\square$

Lemma 5. *Let $\alpha \in C$ and $(\widehat{\sigma}_{\alpha,v}^2)_{v \in \mathbb{N}}$ be the sequence of the estimator defined in* (10). *Assume (S1)-(S3) and (D1)-(D3). We have*

$$
|\widehat{\sigma}_{\alpha,v}^2 - \sigma^2| = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).
$$

Proof. Let

$$
\widehat{S}_v^2 = \frac{1}{n_{r,v}} \sum_{k \in S_{r,v}} \left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha \right)^2.
$$

Then, $|\sigma_{\alpha,v}^2 - \sigma^2|$ can be decomposed as

$$
\left| \widehat{\sigma}_{\alpha,v}^2 - \sigma^2 \right| \leq \frac{n_{r,v}}{n_{r,v} - p_\alpha} \left| \widehat{S}_v^2 - \sigma^2 \right| + \frac{p_\alpha}{n_{r,v} - p_\alpha} \sigma^2.
$$

Since $\sigma^2 p_\alpha / (n_{r,v} - p_\alpha) = o_{\mathbb{P}}(n_v^{-1/2})$, the remaining part is to show $|\widehat{S}_v^2 - \sigma| = O_{\mathbb{P}}(n_v^{-1/2})$. Note that $|\widehat{S}_v^2 - \sigma^2|$ can be further decomposed as

$$
|\widehat{S}_v^2 - \sigma^2| \leq \underbrace{|\widehat{S}_v^2 - \widehat{S}_r^2|}_{(a)} + \underbrace{|\widehat{S}_r^2 - \widetilde{S}_v^2|}_{(b)} + \underbrace{|\widetilde{S}_v^2 - \sigma^2|}_{(c)},
$$

where

$$
\widehat{S}_r^2 = \sum_{k \in S_v} \frac{r_k \epsilon_k^2}{N_v} \left(\frac{n_{r,v}}{N_v} \right)^{-1}
$$

and

$$
\widetilde{S}_v^2 = \sum_{k \in S_v} \frac{p(\mathbf{x}_k)\pi_k \epsilon_k^2}{N_v} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k}{N_v} \right)^{-1}.
$$

We need to show the terms (a), (b), and (c) converge to 0 with rate $O_{\mathbb{P}}(n_v^{-1/2})$.

Convergence of (a). $|\widehat{S}_v^2 - \widehat{S}_r^2|$ can be expressed as

$$\left|\widehat{S}_v^2 - \widehat{S}_r^2\right| = \left|\frac{N_v}{n_{r,v}} \sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v}\right)^2}{N_v} - \frac{N_v}{n_{r,v}} \sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2}{N_v}\right|$$

$$= \left|\left(\frac{n_{r,v}}{N_v}\right)^{-1}\right| \underbrace{\left|\sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v}\right)^2}{N_v} - \sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2}{N_v}\right|}_{:=A_v}.$$

Since Lemma 13 shows

$$\lim_{v \to \infty} \mathbb{P}_{mpq}\left(\frac{n_{r,v}}{n_v} \geq \xi\right) = 1$$

we deduce $(n_{r,v}/N_v)^{-1} = O_\mathbb{P}(1)$. Using

$$\left(\mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v}\right)^2 - \left(\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2 = \left(\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2 + 2\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right),$$

we decompose A_v as

$$\left|\sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v}\right)^2}{N_v} - \sum_{k \in S_v} \frac{r_k \left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2}{N_v}\right|$$

$$\leq \sum_{k \in U_v} \frac{\left|(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v})^2 - (y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha)^2\right|}{N_v}$$

$$= \sum_{k \in U_v} \frac{\left|-2\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right) + \left(\mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v}\right)^2 - \left(\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha\right)^2\right|}{N_v}$$

$$= \sum_{k \in U_v} \frac{\left|-2\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right) + \left(\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2 + 2\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right|}{N_v}$$

$$\leq \frac{2}{N_v} \sum_{k \in U_v} |y_k| \left|\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right|$$

$$+ \frac{1}{N_v} \sum_{k \in U_v} \left(\mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2 + \frac{2}{N_v} \sum_{k \in U_v} \left|\mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right|$$

$$:= A_{1,v} + A_{2,v} + A_{3,v}.$$

By the law of the large numbers, we have

$$\sum_{k \in U_v} \frac{y_k^2}{N_v} \xrightarrow[v \to \infty]{a.s.} \mathbb{E}_{\mathbf{x}m}[y_1^2] = \mathbb{E}_{\mathbf{x}m}\left[\left(\mathbf{x}_{1,\alpha}^\top \boldsymbol{\beta}_\alpha + \epsilon_1\right)^2\right]$$

$$\leq \mathbb{E}_{\mathbf{x}m} \left[2 \left(\mathbf{x}_{1,\alpha}^{\top} \boldsymbol{\beta}_\alpha \right) + \epsilon_1^2 \right] = 2\sigma^2 + 2C_0^2 \|\boldsymbol{\beta}_\alpha\|_2^2.$$

For $A_{1,v}$, using Cauchy-Schwarz inequality, we have

$$A_{1,v} \leq 2 \sqrt{\sum_{k \in U_v} \frac{y_k^2}{N_v}} \sqrt{\sum_{k \in U_v} \frac{\left(\mathbf{x}_{k,\alpha}^{\top} \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) \right)^2}{N_v}}$$

$$\leq 2 \|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2 \sqrt{\sum_{k \in U_v} \frac{y_k^2}{N_v}} \sqrt{\sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v}}$$

$$\leq 2C_0 \|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2 \sqrt{\sum_{k \in U_v} \frac{y_k^2}{N_v}} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right),$$

which also implies $A_{2,v} = O_{\mathbb{P}}(n_v^{-1/2})$. For $A_{3,v}$, using Cauchy-Schwarz inequality, we have

$$A_{3,v} \leq 2 \sqrt{\sum_{k \in U_v} \frac{\left(\mathbf{x}_{k,\alpha}^{\top} \boldsymbol{\beta}_\alpha \right)^2}{N_v}} \sqrt{\sum_{k \in U_v} \frac{\left(\mathbf{x}_{k,\alpha}^{\top} \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) \right)^2}{N_v}}$$

$$\leq 2C_0^2 \|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2 \|\boldsymbol{\beta}_\alpha\|_2 = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).$$

Finally, we conclude $|\widehat{S}_v - \widehat{S}_r| = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of term (b). Term (b) can be decomposed as

$$\left| \widehat{S}_r^2 - \widetilde{S}_v^2 \right| \leq \left| \left(\frac{n_{r,v}}{N_v} \right)^{-1} \left(\sum_{k \in S_v} \frac{r_k \epsilon_k^2}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \epsilon_k^2}{N_v} \right) \right|$$

$$+ \left| \left(\left(\frac{n_{r,v}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k}{N_v} \right)^{-1} \right) \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \epsilon_k^2}{N_v} \right|$$

$$:= B_{1,v} + B_{2,v}.$$

For $B_{1,v}$, we apply Lemma 9 with $\boldsymbol{T}_k := \epsilon_k^2$ and $w_k := r_k$. Since

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m \left[\epsilon_k^4 \right]}{N_v} \leq M_0$$

almost surely, it follows that

$$\left| \sum_{k \in S_v} \frac{r_k \epsilon_k^2}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \epsilon_k^2}{N_v} \right| = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).$$

Recall from Lemma 13 that $\left(n_{r,v} / N_v \right)^{-1} = O_{\mathbb{P}}(1)$, so that $B_{1,v} = O_{\mathbb{P}}(n_v^{-1/2})$. For $B_{2,v}$, we have

$$\left| \left(\frac{n_{r,v}}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k}{N_v} \right)^{-1} \right| = \left| \left(\frac{n_{r,v}}{N_v} \right)^{-1} \right| \left| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k}{N_v} \right)^{-1} \right| \left| \frac{n_{r,v}}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k}{N_v} \right|.$$

Applying Lemma 9, set $T_k := 1$ and $w_k := r_k$ for $k \in U_v$, we have

$$\left| \frac{n_{r,v}}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k}{N_v} \right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

By (D2) and positivity,

$$\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k}{N_v} \right)^{-1} \le (\rho\lambda)^{-1}$$

so that $B_{2,v} = O_{\mathbb{P}}(n_v^{-1/2})$ and $|\widehat{S}_r^2 - \widetilde{S}_v^2| = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of term (c). By Chebyshev's inequality

$$\begin{aligned}
\mathbb{P}_{mpq}\left(\left| \widetilde{S}_v^2 - \sigma^2 \right| > \epsilon \right) &\le \frac{1}{\epsilon^2} \mathbb{E}_{mpq}\left[\left(\widetilde{S}_v^2 - \sigma^2 \right)^2 \right] \\
&\le \frac{1}{N_v} \sum_{k \in U_v} \frac{p(\mathbf{x}_k)^2 \pi_k^2 \mathbb{E}_m[\epsilon_k^4]}{N_v} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k}{N_v} \right)^{-2} \\
&\le \frac{1}{N_v} \sum_{k \in U_v} \frac{M_0}{N_v} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k}{N_v} \right)^{-2} \\
&\le \frac{M_0}{N_v \rho^2 \lambda^2}.
\end{aligned} \tag{35}$$

Combining the terms (a), (b), and (c), we conclude $|\widehat{S}_v^2 - \sigma^2| = O_{\mathbb{P}}(n_v^{-1/2})$ and thus $|\widehat{\sigma}_{\alpha,v}^2 - \sigma^2| = O_{\mathbb{P}}(n_v^{-1/2})$. $\qquad\square$

Lemma 6. *Let $\alpha \in C$ and $(\widehat{\mathbf{c}}_{\alpha,v})_{v \in \mathbb{N}}$ be a sequence of estimator given by* (8). *Assume (S1)-(S3) and (D1)-(D3). Then,*

$$\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Proof. Let

$$\widetilde{\mathbf{c}}_{\alpha,v} = \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v}. \tag{36}$$

Write

$$\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 = \underbrace{\|\widehat{\mathbf{c}}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2}_{(a)} + \underbrace{\|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2}_{(b)}.$$

Treatment of (a). We decompose

$$\begin{aligned}
&\|\widehat{\mathbf{c}}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2 \\
&= \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in S_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{\pi_k N_v} \right.
\end{aligned}$$

$$- \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \bigg\|_2$$

$$= \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \left(\sum_{k \in S_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{\pi_k N_v} - \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right) \right.$$

$$+ \left. \left(\left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right) \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right\|_2$$

$$\leq \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \left(\sum_{k \in S_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{\pi_k N_v} - \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right) \right\|_2$$

$$+ \left\| \left(\left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right) \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right\|_2$$

$$:= A_{1,v} + A_{2,v}.$$

For $A_{1,v}$, we have

$$A_{1,v} \leq \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in S_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{\pi_k N_v} - \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right\|_2.$$

By Lemma 17,

$$\left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1),$$

Moreover, applying Lemma 9, by setting $\boldsymbol{T}_k := \pi_k^{-1}\mathbf{x}_{k,\alpha}$, $w_k := 1 - r_k$, since

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\pi_k^{-1}\mathbf{x}_{k,\alpha}\|_F^2 \right]}{N_v} \leq \frac{C_0^2}{\lambda^2}$$

almost surely, it follows that

$$\left\| \sum_{k \in U_v} \frac{(1 - r_k)\mathbf{x}_{k,\alpha}}{\pi_k N_v} - \sum_{k \in U_v} \frac{(1 - p_k)\mathbf{x}_{k,\alpha}}{N_v} \right\|_2^2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Hence, $A_{1,v} = o_{\mathbb{P}}(1)$. For $A_{2,v}$, we obtain

$$A_{2,v} \leq \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op}$$

$$\times \left\| \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right\|_2.$$

Equation (31) gives

$$\left\| \sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Lemma 15 and Lemma 18 state

$$\left\| \left(\sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k) \pi_k}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1), \qquad \left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1)$$

respectively. Using Lemma 14, we obtain

$$\left\| \left(\sum_{k \in S_v} \frac{r_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Finally, note that

$$\left\| \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k)) \mathbf{x}_{k,\alpha}}{N_v} \right\|_2 \leq \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2}{N_v} \leq C_0$$

almost surely, we conclude that $A_{2,v} = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of (b). The term (b) can be bounded by

$$\begin{aligned}
&\|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 \\
&\leq \left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} \left\| \sum_{k \in U_v} \frac{(1 - r_k) \mathbf{x}_{k,\alpha}}{N_v} - \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k)) \mathbf{x}_{k,\alpha}}{N_v} \right\|_2 \\
&\quad + \left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} \\
&\quad \times \left\| \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k)) \mathbf{x}_{k,\alpha}}{N_v} \right\|_2 \\
&:= B_{1,v} + B_{2,v}.
\end{aligned}$$

For $B_{1,v}$, applying Lemma 10, by setting $w_k := 1 - r_k$, $\boldsymbol{T}_k := \mathbf{x}_{k,\alpha}$ and recalling that

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m \left[\|\mathbf{x}_{k,\alpha}\|_2^2 \right]}{N_v} \leq C_0^2 < \infty,$$

we have

$$\left\| \sum_{k \in U_v} \frac{(1 - r_k) \mathbf{x}_{k,\alpha}}{N_v} - \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k)) \mathbf{x}_{k,\alpha}}{N_v} \right\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).$$

Also, using Lemma 17, we obtain $B_{1,v} = O_{\mathbb{P}}(n_v^{-1/2})$. For $B_{2,v}$, applying Lemma 10, setting $w_k := r_k$, $\boldsymbol{T}_k := \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top$, (30) implies

$$\left\| \sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right). \tag{37}$$

An application of Lemma 14 gives

$$\left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} - \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).$$

As a result, $B_{2,v} = O_{\mathbb{P}}(n_v^{-1/2})$. Putting all things together, we obtain $A_v = O_{\mathbb{P}}(n_v^{-1/2})$ and $B_v = O_{\mathbb{P}}(n_v^{-1/2})$. This concludes $\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 = O_{\mathbb{P}}(n_v^{-1/2})$. □

B.2.2. *Consistency of plug-in estimators*

Lemma 7. *Let $\alpha \in C$ and $(\widehat{\mathbf{c}}_{\alpha,v})_{v \in \mathbb{N}}$ be a sequence defined in* (8). *Assume (S1)-(S3) and (D1)-(D3). We have*

$$\left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right| = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).$$

Proof. Consider the following decomposition,

$$\left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right|$$

$$\leq \left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} - \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} \right|$$

$$+ \left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right|$$

$$+ \left| \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right|$$

$$:= A_v + B_v + C_v,$$

where $\widetilde{\mathbf{c}}_{\alpha,v}$ is given by (36). Next, we will show A_v, B_v and C_v converge to 0 with rate $O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of A_v. Using

$$\left(\widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 - \left(\widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 = \left(\left(\widehat{\mathbf{c}}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \right)^2 + 2 \left(\widehat{\mathbf{c}}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha},$$

we obtain

$$A_v = \left| \sum_{k \in S_v} r_k \frac{\left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 - \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} \right|$$

$$\leq \sum_{k \in U_v} \frac{\left| \left(\widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 - \left(\widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 \right|}{N_v \lambda}$$

$$\leq \sum_{k \in U_v} \frac{\left(\left(\widehat{\mathbf{c}}_{\alpha,v}^\top - \widetilde{\mathbf{c}}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \right)^2 + 2 \left| \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \left(\widehat{\mathbf{c}}_{\alpha,v}^\top - \widetilde{\mathbf{c}}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \right|}{N_v \lambda}$$

$$:= A_{1,v} + A_{2,v}. \tag{38}$$

For $A_{1,v}$, having proved $\|\widehat{\mathbf{c}}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2 = O_{\mathbb{P}}(n_v^{-1/2})$ in Lemma 6, we have

$$\begin{aligned}
A_{1,v} &\leq \frac{1}{\lambda} \sum_{k \in U_v} \frac{\left((\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}) \mathbf{x}_{k,\alpha} \right)^2}{N_v} \\
&\leq \frac{1}{\lambda} \|\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2^2 \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} \\
&\leq \frac{C_0^2}{\lambda} \|\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2^2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).
\end{aligned}$$

For $A_{2,v}$, recall from Lemma 19, we have

$$\limsup_{v \to \infty} \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 < \infty$$

almost surely. It follows that

$$\begin{aligned}
A_{2,v} &\leq \frac{2}{\lambda} \sqrt{\sum_{k \in U_v} \frac{(\widetilde{\mathbf{c}}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha})^2}{N_v}} \sqrt{\sum_{k \in U_v} \frac{\left((\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}) \mathbf{x}_{k,\alpha} \right)^2}{N_v}} \\
&\leq \frac{2}{\lambda} \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 \|\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2 \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} \\
&\leq \frac{2C_0^2}{\lambda} \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 \|\widehat{\mathbf{c}}_{\alpha,v}^{\top} - \widetilde{\mathbf{c}}_{\alpha,v}\|_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n_v}} \right).
\end{aligned}$$

As a result, we have $A_v = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of B_v. Let $a_k = 1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v} \mathbf{x}_{k,\alpha} \right)^2$ for $k \in U_v$. Applying Lemma 9, with $T_k := a_k$ and $w_k := 1$ for $k \in U_v$, it remains to show

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m \left[a_k^2 \right]}{N_v} < \infty$$

almost surely. To this aim, write

$$\begin{aligned}
\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m \left[a_k^2 \right]}{N_v} &\leq \limsup_{v \to \infty} \frac{1}{N_v} \sum_{k \in U_v} \left(2(1 - r_k)^2 + 2r_k^2 \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2 \right) \\
&\leq \limsup_{v \to \infty} \frac{1}{N_v} \sum_{k \in U_v} \left(2 + 2\|\widetilde{\mathbf{c}}_{\alpha,v}\|_2^2 \|\mathbf{x}_{k,\alpha}\|_2^2 \right) \\
&\leq \limsup_{v \to \infty} 2 \left(1 + C_0^2 \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2^2 \right) < \infty
\end{aligned}$$

almost surely. This concludes $B_v = O_{\mathbb{P}}(n_v^{-1/2})$.

Treatment of C_v. Note that

$$\left(\mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2 - \left(\widetilde{\mathbf{c}}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha} \right)^2 = \left((\mathbf{c}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v}) \mathbf{x}_{k,\alpha} \right)^2 + 2 \left(\mathbf{c}_{\alpha,v} - \widetilde{\mathbf{c}}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \widetilde{\mathbf{c}}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha}.$$

Using the result $\|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 = O_{\mathbb{P}}(n_v^{-1/2})$ in Lemma 6 and following the same decomposition in (38), we obtain

$$
\begin{aligned}
&\left| \frac{1}{N_v} \sum_{k \in U_v} 1 - r_k + r_k \left(\pi_k \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 - \frac{1}{N_v} \sum_{k \in U_v} 1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 \right| \\
&\leq \frac{1}{N_v} \sum_{k \in U_v} \left| \left(\widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 - \left(\mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2 \right| \\
&\leq \frac{1}{N_v} \sum_{k \in U_v} \left(\left(\widetilde{\mathbf{c}}_{\alpha,v}^\top - \mathbf{c}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \right)^2 + 2 \left| \widetilde{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \left(\widetilde{\mathbf{c}}_{\alpha,v}^\top - \mathbf{c}_{\alpha,v} \right) \mathbf{x}_{k,\alpha} \right| \\
&\leq \|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 \frac{1}{N_v} \sum_{k \in U_v} \|\mathbf{x}_{k,\alpha}\|_2^2 \left(2\|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 + \|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 \right) \\
&\leq C_0^2 \|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 \left(2\|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 + \|\widetilde{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 \right) = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right). \quad (39)
\end{aligned}
$$

Overall

$$
\left| \sum_{k \in S_v} \frac{1 - r_k + r_k \left(\pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v \pi_k} - \sum_{k \in U_v} \frac{1 - r_k + r_k \left(\pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right)^2}{N_v} \right| = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n_v}} \right).
$$

$\square$

Lemma 8. *Let $\alpha \in \mathcal{A}$ and assume (D1)-(D3) and (S1)-(S3). We have*

$$
\frac{1}{N_v} \sum_{k \in U_v} \left(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha} \right)^2 = o_{\mathbb{P}}(1),
$$

where $\widehat{\eta}_{k,\alpha}$ is given by (9).

Proof. For an arbitrary $k \in U_v$, we decompose

$$
\begin{aligned}
\eta_{k,\alpha} - \widehat{\eta}_{k,\alpha} &= \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha + r_k \left(1 + \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right) \left(y_k - \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_{\alpha,v} \right) \\
&\quad - \left(\mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v} + r_k \left(1 + \pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \right) \left(y_k - \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_{\alpha,v} \right) \right) \\
&\quad + r_k \pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha - r_k \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta} \\
&= (1 - r_k) \mathbf{x}_{k,\alpha}^\top \left(\boldsymbol{\beta}_\alpha - \widehat{\boldsymbol{\beta}}_{\alpha,v} \right) - r_k \pi_k \left(\mathbf{c}_{\alpha,v} - \widehat{\mathbf{c}}_{\alpha,v} \right)^\top \mathbf{x}_{k,\alpha} y_k \\
&\quad + r_k \pi_k (\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v})^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) \\
&\quad + r_k \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) + r_k \pi_k (\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v})^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha,
\end{aligned}
$$

since

$$
\begin{aligned}
&r_k \pi_k \widehat{\mathbf{c}}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \widehat{\boldsymbol{\beta}}_\alpha - r_k \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta} \\
&= r_k \pi_k \left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v} \right)^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) \\
&\quad + r_k \pi_k \mathbf{c}_{\alpha,v}^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha \right) + r_k \pi_k \left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v} \right)^\top \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top \boldsymbol{\beta}_\alpha.
\end{aligned}
$$

On the other hand, according to Lemma 17, we obtain,

$$\|\mathbf{c}_{\alpha,v}\|_2 \leq \left\|\left(\sum_{k\in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op} \left\|\sum_{k\in U_v} \frac{(1-r_k)\mathbf{x}_{k,\alpha}}{N_v}\right\|_2 = O_\mathbb{P}(1). \qquad (40)$$

Combining the results that $\|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_{\alpha,v}\| = o_\mathbb{P}(1)$ in Lemma 3 and $\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2 = o_\mathbb{P}(1)$ in Lemma 6, an application of Cauchy-Schwarz inequality gives

$$\sum_{k\in U_v} \frac{(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})^2}{N_v}$$

$$= \sum_{k\in U_v} \frac{(\eta_{k,\alpha} - \widehat{\eta}_{k,\alpha})^2}{N_v}$$

$$= \frac{1}{N_v}\sum_{k\in U_v}\left((1-r_k)\mathbf{x}_{k,\alpha}^\top\left(\boldsymbol{\beta}_\alpha - \widehat{\boldsymbol{\beta}}_{\alpha,v}\right) - r_k\pi_k\left(\mathbf{c}_{\alpha,v} - \widehat{\mathbf{c}}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}y_k\right.$$

$$+ r_k\pi_k\left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)$$

$$\left.+ r_k\pi_k\mathbf{c}_{\alpha,v}^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right) + r_k\pi_k\left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\boldsymbol{\beta}_\alpha\right)^2$$

$$\leq \frac{5}{N_v}\sum_{k\in U_v}\left(\left(\mathbf{x}_{k,\alpha}^\top\left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2 + \left(\left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}y_k\right)^2\right.$$

$$+ \left(\left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2 + \left(\mathbf{c}_{\alpha,v}^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\left(\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\right)\right)^2$$

$$\left.+ \left(\left(\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\right)^\top\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top\boldsymbol{\beta}_\alpha\right)^2\right)$$

$$\leq 5\|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2^2\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} + 5\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2^2\sqrt{\sum_{k\in U_v}\frac{y_k^4}{N_v}}\sqrt{\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v}}$$

$$+ 5\|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2^2\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2^2\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v}$$

$$+ 5\|\widehat{\boldsymbol{\beta}}_{\alpha,v} - \boldsymbol{\beta}_\alpha\|_2^2\|\mathbf{c}_{\alpha,v}\|_2^2\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v}$$

$$+ 5\|\boldsymbol{\beta}_\alpha\|_2^2\|\widehat{\mathbf{c}}_{\alpha,v} - \mathbf{c}_{\alpha,v}\|_2^2\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v} = o_\mathbb{P}(1).$$

Finally, we conclude

$$\sum_{k\in U_v}\frac{(\widehat{\eta}_{k,\alpha} - \eta_{k,\alpha})^2}{N_v} = o_\mathbb{P}(1).$$

$\square$

B.3. Additional Lemmas

B.3.1. Technical Lemmas

Lemma 9. *Let* $(w_k)_{k \in U_v}$ *be i.i.d. random variable satisfying* $\mathbb{E}_q[w_k^2] \le b$ *for some* $b \ge 0$ *almost surely. Let* $(\widehat{\boldsymbol{Z}}_{\pi,v})_{v \in \mathbb{N}}$ *and* $(\boldsymbol{Z}_v)_{v \in \mathbb{N}}$ *be a sequence of matrices in* $\mathbb{R}^{d_1 \times d_2}$ *defined by*

$$\widehat{\boldsymbol{Z}}_{\pi,v} = \sum_{k \in S_v} \frac{w_k \boldsymbol{T}_k}{N_v},$$

and

$$\boldsymbol{Z}_v = \sum_{k \in U_v} \frac{\mathbb{E}_q[w_k] \boldsymbol{T}_k}{N_v},$$

respectively. Assume [(D1)-(D3)]. If $\{\boldsymbol{T}_k\}_{k \in U_v}$ *satisfies*

$$\limsup_{v \to \infty} \frac{1}{N_v} \sum_{k \in U_v} \mathbb{E}_m \left[\|\boldsymbol{T}_k\|_F^2 \right] < \infty$$

almost surely, then we have

$$\left\| \widehat{\boldsymbol{Z}}_{\pi,v} - \boldsymbol{Z}_v \right\|_{op} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{N_v}} \right).$$

Proof. Recall that for arbitrary matrices $\boldsymbol{B}$, we have $\|\boldsymbol{B}\|_{op} \le \|\boldsymbol{B}\|_F$. For any $\epsilon > 0$, an application of Chebyshev's inequality gives

$$\mathbb{P}_{mpq} \left(\|\widehat{\boldsymbol{Z}}_{\pi,v} - \boldsymbol{Z}_v\|_{op} > \epsilon \right) \le \frac{1}{\epsilon^2} \mathbb{E}_{mpq} \left[\|\widehat{\boldsymbol{Z}}_{\pi,v} - \boldsymbol{Z}_v\|_{op}^2 \right]$$
$$\le \frac{1}{\epsilon^2} \mathbb{E}_{mpq} \left[\|\widehat{\boldsymbol{Z}}_{\pi,v} - \boldsymbol{Z}_v\|_F^2 \right].$$

We proceed to evaluate

$$\mathbb{E}_{mpq} \left[\|\widehat{\boldsymbol{T}}_{\pi,v} - \boldsymbol{T}_v\|_F^2 \right] = \mathbb{E}_{mpq} \left[\left\| \sum_{k \in S_v} \frac{w_k \boldsymbol{T}_k}{N_v} - \sum_{k \in U_v} \frac{\mathbb{E}_q[w_k] \boldsymbol{T}_k}{N_v} \right\|_F^2 \right]$$
$$= \mathbb{E}_{mpq} \left[\text{Tr} \left(\sum_{k \in U_v} \sum_{l \in U_v} \frac{(I_k w_k - \pi_k \mathbb{E}_q[w_k])(I_l w_l - \pi_l \mathbb{E}_q[w_l]) \boldsymbol{T}_k^\top \boldsymbol{T}_l}{N_v^2} \right) \right]$$
$$= \mathbb{E}_{mpq} \left[\sum_{k \in U_v} \sum_{l \in U_v} \frac{(I_k w_k - \pi_k \mathbb{E}_q[w_k])(I_l w_l - \pi_l \mathbb{E}_q[w_l]) \text{Tr} \left(\boldsymbol{T}_k^\top \boldsymbol{T}_l \right)}{N_v^2} \right].$$
$$\tag{41}$$

When $k = l$, we have $\mathbb{E}_{pq}[(I_k w_k - \pi_k \mathbb{E}_q[w_k])(I_l w_l - \pi_l \mathbb{E}_q[w_l])] = \pi_k(1-\pi_k) \mathbb{V}_q[w_k]$. When $k \ne l \in U_v$, we have $\mathbb{E}_{pq}[(I_k w_k - \pi_k \mathbb{E}_q[w_k])(I_l w_l - \pi_l \mathbb{E}_q[w_l])] = \Delta_{kl} \mathbb{E}_q[w_k] \mathbb{E}_q[w_l]$. As a result, (41) reduces to

$$\mathbb{E}_{mpq} \left[\sum_{k \in U_v} \sum_{l \in U_v} \frac{(I_k w_k - \pi_k \mathbb{E}_q[w_k])(I_l w_l - \pi_l \mathbb{E}_q[w_l]) \text{Tr} \left(\boldsymbol{T}_k^\top \boldsymbol{T}_l \right)}{N_v^2} \right]$$

$$= \sum_{k \in U_v} \frac{\pi_k(1-\pi_k)\mathbb{V}_q(w_k)}{N_v^2} \mathbb{E}_m \left[\mathrm{Tr}\left(\boldsymbol{T}_k^\top \boldsymbol{T}_k \right) \right]$$

$$+ \sum_{k \in U_v} \sum_{\substack{l \in U_v \\ l \neq k}} \frac{\Delta_{kl}\mathbb{E}_q\left[w_k \right] \mathbb{E}_q\left[w_k \right]}{N_v^2} \mathbb{E}_m \left[\mathrm{Tr}\left(\boldsymbol{T}_k^\top \boldsymbol{T}_l \right) \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} + \max_{k \neq l \in U_v} |\Delta_{kl}| \sum_{k \in U_v} \sum_{l \in U_v} \frac{|\mathbb{E}_q\left[w_k \right] \mathbb{E}_q\left[w_k \right]|}{N_v^2} \mathbb{E}_m \left[\left| \mathrm{Tr}\left(\boldsymbol{T}_k^\top \boldsymbol{T}_l \right) \right| \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v}$$

$$+ \frac{n_v \max_{k \neq l \in U_v} |\Delta_{kl}|}{n_v} \sum_{k \in U_v} \sum_{l \in U_v} \frac{|\mathbb{E}_q\left[w_k \right] \mathbb{E}_q\left[w_k \right]|}{N_v^2} \mathbb{E}_m \left[\|\boldsymbol{T}_k\|_F \|\boldsymbol{T}_l\|_F \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} + \frac{\bar{\Delta}}{n_v} \sum_{k \in U_v} \sum_{l \in U_v} \frac{|\mathbb{E}_q\left[w_k \right] \mathbb{E}_q\left[w_k \right]|}{N_v^2} \mathbb{E}_m^{1/2} \left[\|\boldsymbol{T}_k\|_F^2 \right] \mathbb{E}_m^{1/2} \left[\|\boldsymbol{T}_l\|_F^2 \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} + \frac{\bar{\Delta}}{n_v} \sum_{k \in U_v} \frac{\mathbb{E}_q^2\left[w_k \right]}{N_v} \mathbb{E}_m \left[\|\boldsymbol{T}_k\|_F^2 \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} + \frac{\bar{\Delta}}{n_v} \sum_{k \in U_v} \frac{\mathbb{E}_q\left[w_k^2 \right]}{N_v} \mathbb{E}_m \left[\|\boldsymbol{T}_k\|_F^2 \right]$$

$$\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} + \frac{b\bar{\Delta}}{n_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v}.$$

The result follows. $\qquad \square$

Lemma 10. *Let $(w_k)_{k \in U_v}$ be i.i.d. random variable satisfying $\mathbb{E}_q[w_k^2] \leq b$ for some $b \geq 0$ almost surely. Let $(\widehat{\boldsymbol{Z}}_{p,v})_{v \in \mathbb{N}}$ and $(\boldsymbol{Z}_v)_{v \in \mathbb{N}}$ be a sequence of matrices in $\mathbb{R}^{d_1 \times d_2}$ defined by*

$$\widehat{\boldsymbol{Z}}_{p,v} = \sum_{k \in U_v} \frac{w_k \boldsymbol{T}_k}{N_v},$$

and

$$\boldsymbol{Z}_v = \sum_{k \in U_v} \frac{\mathbb{E}_q[w_k] \boldsymbol{T}_k}{N_v},$$

respectively. If $(\boldsymbol{T}_k)_{k \in U_v}$ satisfies

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2 \right]}{N_v} < \infty$$

almost surely, then we have

$$\left\| \widehat{\boldsymbol{Z}}_{p,v} - \boldsymbol{Z}_v \right\|_{op} = O_{\mathbb{P}} \left(\frac{1}{\sqrt{N_v}} \right).$$

Proof. By evaluating $\mathbb{E}_{mpq}[\|\widehat{\boldsymbol{Z}}_{p,v} - \boldsymbol{Z}_v\|_{op}^2]$, we have

$$
\begin{aligned}
\mathbb{E}_{mpq}\left[\|\widehat{\boldsymbol{Z}}_{p,v} - \boldsymbol{Z}_v\|_{op}^2\right] &\leq \mathbb{E}_{mq}\left[\left\|\sum_{k \in U_v} \frac{(w_k - \mathbb{E}_m[w_k])\,\boldsymbol{T}_k}{N_v}\right\|_F^2\right] \\
&= \mathbb{E}_{mq}\left[\sum_{k \in U_v}\sum_{l \in U_v} \frac{(w_k - \mathbb{E}_m[w_k])(w_l - \mathbb{E}_m[w_l])\,\mathrm{Tr}\left(\boldsymbol{T}_k^\top \boldsymbol{T}_l\right)}{N_v^2}\right] \\
&= \sum_{k \in U_v} \frac{\mathbb{V}_q(w_k)\mathbb{E}_m\left[\mathrm{Tr}\left(\boldsymbol{T}_k^\top \boldsymbol{T}_k\right)\right]}{N_v^2} \\
&\leq \sum_{k \in U_v} \frac{\mathbb{E}_q(w_k^2)\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2\right]}{N_v^2} \\
&\leq \frac{b}{N_v} \sum_{k \in U_v} \frac{\mathbb{E}_m\left[\|\boldsymbol{T}_k\|_F^2\right]}{N_v},
\end{aligned}
$$

from which the result follows. $\qquad\square$

Lemma 11. *For any $n \in \mathbb{N}$, the map $\mathcal{L}_{2,n} : \mathcal{P}(\{1, 2, ..., p\}) \longrightarrow \mathbb{R}_+$ defined by*

$$
\alpha \longmapsto \sigma^2 \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}^\top}{\pi_k}\right) A_{r,\alpha}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha}}{\pi_k}\right)
$$

is, almost surely, a strictly increasing set function, that is, $\alpha_1 \subset \alpha_2$ implies $\mathcal{L}_2(\alpha_1) < \mathcal{L}_2(\alpha_2)$.

Proof. Let $\alpha_1, \alpha_2 \in \mathcal{P}\{1, 2, ..., p\})$ such that $\alpha_1 \subset \alpha_2$. Then,

$$
\begin{aligned}
&\mathcal{L}_{2,n}\left(\alpha_1\right) < \mathcal{L}_{2,n}\left(\alpha_2\right) \\
&\Leftrightarrow \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top}{N\pi_k}\right) A_{r,\alpha_1}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}}{N\pi_k}\right) < \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2}^\top}{N\pi_k}\right) A_{r,\alpha_2}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2}}{N\pi_k}\right).
\end{aligned}
$$

This amounts to showing that one quadratic form is almost surely less than the other. We start by expressing A_{r,α_2}^{-1} in terms of A_{r,α_1}^{-1}. Let $A_{r,\alpha_1,\alpha_2-\alpha_1} := \mathbf{x}_{r,\alpha_1}^\top \mathbf{x}_{r,\alpha_2-\alpha_1}$ for arbitrary $\alpha_1 \subset \alpha_2$. Note that

$$
A_{r,\alpha_2} = \begin{pmatrix} A_{r,\alpha_1} & A_{r,\alpha_1,\alpha_2-\alpha_1} \\ A_{r,\alpha_2-\alpha_1,\alpha_1} & A_{r,\alpha_2-\alpha_1} \end{pmatrix}.
$$

Using a [block-matrix inversion formula](#) (see, e.g., [Horn and Johnson (2012)](#), page 25), one can show that

$$
A_{r,\alpha_2}^{-1} = \begin{pmatrix} A_{r,\alpha_1}^{-1} + A_{r,\alpha_1}^{-1} A_{r,\alpha_1,\alpha_2-\alpha_1} S^{-1} A_{r,\alpha_2-\alpha_1,\alpha_1} A_{r,\alpha_1}^{-1} & -A_{r,\alpha_1}^{-1} A_{r,\alpha_1,\alpha_2-\alpha_1} S^{-1} \\ -S^{-1} A_{r,\alpha_2-\alpha_1,\alpha_1} A_{r,\alpha_1}^{-1} & S^{-1} \end{pmatrix}.
$$

Therefore, writing

$$
\left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2}^\top}{N\pi_k}\right) = \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top}{N\pi_k}, \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k}\right),
$$

some algebra shows that

$$
\left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top}{N\pi_k}, \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k} \right) A_{r,\alpha_2}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}}{N\pi_k}, \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}}{N\pi_k} \right)
$$

$$
= \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top}{N\pi_k} \right) A_{r,\alpha_1}^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}}{N\pi_k} \right)
$$

$$
+ \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k} - \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top A_{r,\alpha_1}^{-1} A_{r,\alpha_2-\alpha_1}}{N\pi_k} \right)
$$

$$
\times S^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k} - \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top A_{r,\alpha_1}^{-1} A_{r,\alpha_2-\alpha_1}}{N\pi_k} \right)^\top. \tag{42}
$$

Hence, showing $\mathcal{L}_{2,n}(\alpha_1) < \mathcal{L}_{2,n}(\alpha_2)$ amounts to show that

$$
\left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k} - \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top A_{r,\alpha_1}^{-1} A_{r,\alpha_2-\alpha_1}}{N\pi_k} \right)
$$

$$
\times S^{-1} \left(\sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_2-\alpha_1}^\top}{N\pi_k} - \sum_{k \in S_m} \frac{\mathbf{x}_{k,\alpha_1}^\top A_{r,\alpha_1}^{-1} A_{r,\alpha_2-\alpha_1}}{N\pi_k} \right)^\top > 0 \tag{43}
$$

almost surely. To that aim, observe that S can be written as

$$
S = \mathbf{x}_{r,\alpha_2-\alpha_1}^\top \left(I_{n_r} - \mathbf{x}_{r,\alpha_1} A_{r,\alpha_1}^{-1} \mathbf{x}_{r,\alpha_1}^\top \right) \mathbf{x}_{r,\alpha_2-\alpha_1}.
$$

The matrix S is a gram matrix of linearly independent vectors, thus positive definite, which shows (43) and thus $\mathcal{L}_{2,n}(\alpha_1) < \mathcal{L}_{2,n}(\alpha_2)$ holds almost surely. □

Lemma 12. *Let $\alpha \in \mathcal{A}$ and assume that the intercept is included in α. Then,*

$$
\mathbb{E}_\mathbf{x} \left[(1 - p(\mathbf{x})) \mathbf{x}_\alpha^\top \right] \left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \right] = \mathbb{E}_\mathbf{x} \left[1 - p(\mathbf{x}) \right].
$$

Proof. Write $\mathbf{x}_\alpha^\top = \left(1, \mathbf{x}_{\alpha-1}^\top \right)$, where $\mathbf{x}_{\alpha-1}^\top$ is the covariate without the intercept. As a result,

$$
\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] = \begin{pmatrix} \mathbb{E}_\mathbf{x}[p(\mathbf{x})] & \mathbb{E}_\mathbf{x}[p(\mathbf{x}) \mathbf{x}_{\alpha-1}^\top] \\ \mathbb{E}_\mathbf{x}[p(\mathbf{x}) \mathbf{x}_{\alpha-1}] & \mathbb{E}_\mathbf{x}[p(\mathbf{x}) \mathbf{x}_{\alpha-1} \mathbf{x}_{\alpha-1}^\top] \end{pmatrix}.
$$

Using the [block matrix inverse formula](), we obtain

$$
\left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1}
$$

$$
= \begin{pmatrix} \frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} + \frac{1}{\mathbb{E}_\mathbf{x}^2[p(\mathbf{x})]} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1}^\top \right] M_{\alpha-1}^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1} \right] & -\frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} B_{\alpha-1} \\ -\frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} B_{\alpha-1}^\top & M_{\alpha-1}^{-1} \end{pmatrix},
$$

where $M_{\alpha-1} = \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1} \mathbf{x}_{\alpha-1}^\top \right] - \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1} \right] \left(\mathbb{E}_\mathbf{x}[p(\mathbf{x})] \right)^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1}^\top \right]$ is a positive and definite matrix and $B_{\alpha-1} = \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1}^\top \right] M_{\alpha-1}^{-1}$. As a result,

$$
\left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \mathbf{x}_\alpha^\top \right] \right)^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_\alpha \right]
$$

$$
= \begin{pmatrix} \frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} + \frac{1}{\mathbb{E}_\mathbf{x}^2[p(\mathbf{x})]} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1}^\top \right] M_{\alpha-1}^{-1} \mathbb{E}_\mathbf{x} \left[p(\mathbf{x}) \mathbf{x}_{\alpha-1} \right] & -\frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} B_{\alpha-1} \\ -\frac{1}{\mathbb{E}_\mathbf{x}[p(\mathbf{x})]} B_{\alpha-1}^\top & M_{\alpha-1}^{-1} \end{pmatrix}
$$

$$\times \begin{pmatrix} \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\right] \\ \mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_{\alpha-1}\right] \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ \mathbf{0}_{\alpha-1} \end{pmatrix}.$$

This leads to

$$\mathbb{E}_{\mathbf{x}}\left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha}^{\top}\right]\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_{\alpha}\mathbf{x}_{\alpha}^{\top}\right]\right)^{-1}\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x})\mathbf{x}_{\alpha}\right]$$

$$= \left(\mathbb{E}_{\mathbf{x}}\left[1 - p(\mathbf{x})\right], \mathbb{E}_{\mathbf{x}}\left[(1 - p(\mathbf{x}))\mathbf{x}_{\alpha-1}^{\top}\right]\right)\begin{pmatrix} 1 \\ \mathbf{0}_{\alpha-1} \end{pmatrix} = \mathbb{E}_{\mathbf{x}}[1 - p(\mathbf{x})].$$

$\square$

Lemma 13. *Assume (S1)-(S2). Assume also that there exists $\rho > 0$ such that $p(\mathbf{x}) > \rho$ almost surely. Then, there exists $\xi > 0$ such that*

$$\lim_{v \to \infty} \mathbb{P}_{mpq}\left(\frac{n_{r,v}}{N_v} \geq \xi\right) = 1$$

almost surely.

Proof. Observe that

$$\mathbb{E}_q\left[\frac{n_{r,v}}{N_v}\right] = \frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k), \qquad \text{and,} \qquad \mathbb{V}_q\left(\frac{n_{r,v}}{n_v}\right) = \frac{1}{n_v^2}\sum_{k \in S_v} p(\mathbf{x}_k)\{1 - p(\mathbf{x}_k)\} \leq \frac{1}{4n_v}.$$

Therefore, Chebyshev's inequality shows that, for all $\epsilon > 0$,

$$\mathbb{P}_q\left(\left|\frac{n_{r,v}}{n_v} - \frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k)\right| > \epsilon\right) \leq \frac{1}{4\epsilon^2 n_v} \xrightarrow[v \to \infty]{} 0$$

almost surely. On the one hand,

$$\frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k) \geq \min_{k \in S_v} p(\mathbf{x}_k) \geq \rho$$

almost surely. On the other hand,

$$\mathbb{P}_q\left(\left|\frac{n_{r,v}}{n_v} - \frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k)\right| \leq \epsilon\right) = \mathbb{P}_q\left(\frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k) - \epsilon \leq \frac{n_{r,v}}{n_v} \leq \frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k) + \epsilon\right)$$

$$\leq \mathbb{P}_q\left(\rho - \epsilon \leq \frac{n_{r,v}}{n_v}\right).$$

Using that $\min_{k \in U_v} p(\mathbf{x}_k) \geq \rho$, we set $\xi := \rho - \epsilon$ from which it follows that

$$\lim_{v \to \infty} \mathbb{P}_q\left(\xi \leq \frac{n_{r,v}}{n_v}\right) \geq \lim_{v \to \infty} \mathbb{P}_q\left(\left|\frac{n_{r,v}}{n_v} - \frac{1}{n_v}\sum_{k \in S_v} p(\mathbf{x}_k)\right| > \epsilon\right) = 1$$

almost surely. Furthermore,

$$\lim_{v \to \infty} \mathbb{P}_{mpq}\left(\frac{n_{r,v}}{N_v} \geq \xi\right) = 1$$

almost surely.

$\square$

Lemma 14. *Let $\{A_v\}_{v\in\mathbb{N}}$ and $\{B_v\}_{\mathbb{N}}$ be $p \times p$ invertible matrices such that $\|A_v^{-1}\|_{op} = O_{\mathbb{P}}(1)$ and $\|B_v^{-1}\|_{op} = O_{\mathbb{P}}(1)$. If there exists a sequence $\{u_v\}_{v\in\mathbb{N}}$ such that*

$$\|A_v - B_v\|_{op} = O_{\mathbb{P}}\left(u_v^{-1}\right).$$

Then, we have

$$\|A_v^{-1} - B_v^{-1}\|_{op} = O_{\mathbb{P}}\left(u_v^{-1}\right).$$

Proof. Recall the Schwarz matrix inequality (see Hansen (2022), B.15 in Page 981), for any squared matrices A and B, for the spectral norm, $\|AB\|_{op} \leq \|A\|_{op}\|B\|_{op}$. As a result,

$$
\begin{aligned}
\|A_v^{-1} - B_v^{-1}\|_{op} &= \|A_v^{-1}\left(I_p - A_v B_v^{-1}\right)\|_{op} \leq \|A_v^{-1}\|_{op}\|I_p - A_v B_v^{-1}\|_{op} \\
&= \|A_v^{-1}\|_{op}\|\left(B_v - A_v\right) B_v^{-1}\|_{op} \\
&\leq \|A_v^{-1}\|_{op}\|B_v^{-1}\|_{op}\|\mathbf{B}_v - A_v\|_{op} = O_{\mathbb{P}}\left(u_v^{-1}\right).
\end{aligned}
$$

This concludes the proof since $\|B_v - A_v\|_{op} = \|A_v - B_v\|_{op}$ for all $v \in \mathbb{N}$. □

B.3.2. *Tightness of various statistics of interest*

Lemma 15. *Assume (S1)-(S2) and (D1)-(D3). For $\alpha \in \mathcal{A}$, we have*

$$\left\|\left(\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op} = O_{\mathbb{P}}(1).$$

Proof. Using (D2) gives

$$\left\|\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right\|_{op} \geq \lambda \left\|\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right\|_{op},$$

On the other hand, using the law of large numbers and the continuous mapping theorem gives

$$\lambda_{min}\left(\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right) \xrightarrow[v\to\infty]{a.s.} \lambda_{min}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha}\mathbf{x}_{1,\alpha}\right]\right),$$

which is a positive-definite matrix. As a result, for any $\epsilon > 0$, we have

$$
\begin{aligned}
\lim_{v\to\infty} \mathbb{P}_{mpq}&\left(\left|\lambda_{min}\left(\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right) - \lambda_{min}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha}\mathbf{x}_{1,\alpha}\right]\right)\right| \leq \epsilon\right) \\
&\leq \lim_{v\to\infty} \mathbb{P}_{mpq}\left(\lambda_{min}\left(\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right) \geq \lambda_{min}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha}\mathbf{x}_{1,\alpha}\right]\right) - \epsilon\right) = 1.
\end{aligned}
$$

Furthermore,

$$\lim_{v\to\infty} \mathbb{P}_{mpq}\left(\lambda_{min}\left(\sum_{k\in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha}\mathbf{x}_{k,\alpha}^\top}{N_v}\right) \geq \lambda\lambda_{min}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha}\mathbf{x}_{1,\alpha}\right]\right) - \epsilon\right) = 1$$

almost surely. This concludes

$$\left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = \lambda_{min}^{-1} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) = O_{\mathbb{P}}(1).$$

$\square$

Lemma 16. *Assume (S1)-(S2) and (D1)-(D3). For $\alpha \in \mathcal{A}$, we have*

$$\limsup_{v \to \infty} \left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \right\|_{op} < \infty.$$

Proof. We have

$$\left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} \le \lambda \left\| \sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha}^\top \mathbf{x}_{k,\alpha}}{N_v} \right\|_{op},$$

Also,

$$\lambda_{min}^{-2} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \xrightarrow[v \to \infty]{a.s.} \lambda_{min}^{-2} \left(\mathbb{E}_\mathbf{x} \left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha} \mathbf{x}_{1,\alpha} \right] \right),$$

which is a positive-definite matrix. As a result,

$$\limsup_{v \to \infty} \lambda_{min}^{-2} \left(\sum_{k \in U_v} \frac{\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top p(\mathbf{x}_k)}{N_v} \right) < \infty$$

almost surely. It follows that

$$\limsup_{v \to \infty} \left\| \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-2} \right\|_{op} = \limsup_{v \to \infty} \lambda_{min}^{-2} \left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) < \infty$$

almost surely.

$\square$

Lemma 17. *Assume (S1)-(S2) and (D1)-(D3). For $\alpha \in \mathcal{A}$, we have*

$$\left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = O_{\mathbb{P}}(1).$$

Proof. Recall from Lemma 2, there exists universal C_1, C_2, and $\gamma > 0$ such that

$$\mathbb{P}_q \left(\lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \ge \gamma \right) \ge 1 - C_1 e^{-C_2 N_v}$$

almost surely for large v. As a result, we obtain

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \ge \gamma \right) = 1.$$

It follows that

$$\left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right)^{-1} \right\|_{op} = \lambda_{min}^{-1} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) = O_\mathbb{P}(1).$$

This concludes the proof.   □

Lemma 18. *Assume (S1)-(S2) and (D1)-(D3). For $\alpha \in \mathcal{A}$, we have*

$$\left\| \left(\frac{A_{r,\alpha}}{N_v} \right)^{-1} \right\|_{op} = O_\mathbb{P}(1).$$

Proof. Recall from (31) and (37), we obtain

$$\left\| \frac{A_{r,\alpha}}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_\mathbb{P} \left(\frac{1}{\sqrt{n_v}} \right)$$

and

$$\left\| \sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} - \sum_{k \in U_v} \frac{p(\mathbf{x}_k) \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_\mathbb{P} \left(\frac{1}{\sqrt{n_v}} \right),$$

respectively. An application of the triangle inequality of operator norm gives

$$\left\| \frac{A_{r,\alpha}}{N_v} - \sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op} = O_\mathbb{P} \left(\frac{1}{\sqrt{n_v}} \right).$$

It follows that

$$\left| \lambda_{min} \left(\frac{A_{r,\alpha}}{N_v} \right) - \lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \right| \leq \left\| \frac{A_{r,\alpha}}{N_v} - \sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right\|_{op}$$

$$= O_\mathbb{P} \left(\frac{1}{\sqrt{n_v}} \right).$$

As a result, for any $\epsilon > 0$

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\left| \lambda_{min} \left(\frac{A_{r,\alpha}}{N_v} \right) - \lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \right| \leq \epsilon \right)$$

$$\leq \lim_{v \to \infty} \mathbb{P}_{mpq} \left(\lambda_{min} \left(\frac{A_{r,\alpha}}{N_v} \right) \geq \lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) - \epsilon \right) = 1.$$

Recall from Lemma 17, there exists $\gamma > 0$ such that

$$\lim_{v \to \infty} \mathbb{P}_{mpq} \left(\lambda_{min} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v} \right) \geq \gamma \right) = 1.$$

Note that for any random variables X, Y and any $\gamma, \epsilon > 0$, we have

$$\mathbb{P}_{mpq} \left(Y \geq \gamma \right) \leq \mathbb{P}_{mpq} \left(X \geq \gamma - \epsilon \right) + \mathbb{P}(|Y - X| > \epsilon).$$

Thus

$$
\mathbb{P}_{mpq}\left(\lambda_{min}\left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right) \geq \gamma\right)
$$
$$
\leq \mathbb{P}_{mpq}\left(\lambda_{min}\left(\frac{A_{r,\alpha}}{N_v}\right) \geq \gamma - \epsilon\right)
$$
$$
+ \mathbb{P}_{mpq}\left(\left|\lambda_{min}\left(\frac{A_{r,\alpha}}{N_v}\right) - \lambda_{min}\left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)\right| > \epsilon\right).
$$

It follows that

$$
\lim_{v \to \infty} \mathbb{P}_{mpq}\left(\lambda_{min}\left(\frac{A_{r,\alpha}}{N_v}\right) \geq \gamma - \epsilon\right) = 1.
$$

This suffices to conclude

$$
\left\|\left(\frac{A_{r,\alpha}}{N_v}\right)^{-1}\right\|_{op} = O_{\mathbb{P}}(1).
$$

$\square$

Lemma 19. *Let $(\widetilde{\mathbf{c}}_{\alpha,v})_{v \in \mathbb{N}}$ be a sequence of estimators given by* (36). *Assume (S1)-(S3) and (D1)-(D3). We have*

$$
\limsup_{v \to \infty} \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2^4 < \infty.
$$

Proof. Recall that

$$
\|\widetilde{\mathbf{c}}_{\alpha,v}\|_2 = \left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1} \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v}\right\|_2^4
$$
$$
\leq \left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op}^4 \left\|\sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v}\right\|_2^4
$$
$$
\leq \lambda^4 \left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op}^4 \left\|\sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v}\right\|_2^4.
$$

On the one hand, by the law of large numbers and the continuous mapping theorem, we obtain

$$
\left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op}^4 = \lambda_{min}^{-4}\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)
$$
$$
\xrightarrow[v \to \infty]{a.s.} \lambda_{min}^{-4}\left(\mathbb{E}_{\mathbf{x}}\left[p(\mathbf{x}_1)\mathbf{x}_{1,\alpha} \mathbf{x}_{1,\alpha}^\top\right]\right).
$$

This implies

$$
\limsup_{v \to \infty} \left\|\left(\sum_{k \in U_v} \frac{p(\mathbf{x}_k)\mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^\top}{N_v}\right)^{-1}\right\|_{op}^4 < \infty.
$$

On the other hand,

$$\left\| \sum_{k \in U_v} \frac{(1 - p(\mathbf{x}_k))\mathbf{x}_{k,\alpha}}{N_v} \right\|_2^4 \le \sum_{k \in U_v} \frac{\|\mathbf{x}_{k,\alpha}\|_2^4}{N_v} \le C_0^4$$

almost surely. As a result, we conclude

$$\limsup_{v \to \infty} \|\widetilde{\mathbf{c}}_{\alpha,v}\|_2^4 < \infty$$

almost surely.   □

Lemma 20. *Let $\alpha \in C$. Assume (S1)-(S3) and (D1)-(D3). Then,*

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} < \infty$$

almost surely.

Proof. Using the result of (46), we obtain

$$\sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} \le \sum_{k \in U_v} \frac{8\|\mathbf{x}_{k,\alpha}\|_2^4 \|\boldsymbol{\beta}_\alpha\|_2^4 + 8(8 + 8\pi_k^2 \|\mathbf{c}_{\alpha,v}\|_2^4 \|\mathbf{x}_{k,\alpha}\|_2^4)\epsilon_k^4}{N_v}$$

$$\le 8C_0^4 \|\boldsymbol{\beta}_\alpha\|_2^4 + 64(1 + C_0^4 \|\mathbf{c}_{\alpha,v}\|_2^4) \sum_{k \in U_v} \frac{\epsilon_k^4}{N_v}.$$

Since

$$\sum_{k \in U_v} \frac{\epsilon_k^4}{N_v} \xrightarrow[v \to \infty]{a.s.} \mathbb{E}_{\mathbf{x}m} \left[\epsilon_1^4 \right] \le M_0,$$

we have

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\epsilon_k^4}{N_v} < \infty$$

almost surely. Combining with Lemma 19, we obtain

$$\limsup_{v \to \infty} \sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} < \infty$$

almost surely.   □

B.3.3. *Finite moments of various statistics of interest*

Lemma 21. *Let $\alpha \in \mathcal{A}$ and $\{\boldsymbol{c}_{\alpha,v}\}_{v \in \mathbb{N}}$ be the sequence of statistics given by (6). Assume (S1)-(S3) and (D1)-(D3). For $\alpha \in C$, there exists a constant K such that*

$$\mathbb{E}_q \left[\|\boldsymbol{c}_{\alpha,v}\|_2^4 \right] \le K$$

almost surely.

Proof. Write

$$\mathbf{c}_{\alpha,v} = \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^{\top}}{N_v} \right)^{-1} \sum_{k \in U_v} \frac{(1 - r_k) \mathbf{x}_{k,\alpha}}{N_v} := A_{\alpha,v}^{-1} \boldsymbol{b}_{\alpha,v}.$$

We have

$$\|\boldsymbol{b}_{\alpha,v}\|_2 = \left\| \frac{1}{N_v} \sum_{k \in U_v} (1 - r_k) \mathbf{x}_{k,\alpha} \right\|_2 \leq \frac{1}{N_v} \sum_{k \in U_v} (1 - r_k) \|\mathbf{x}_{k,\alpha}\|_2 \leq C_0. \tag{44}$$

Moreover, recalling that we condition on the high-probability event $\mathcal{E}_N$ defined in Remark [1], we have

$$\|A_{\alpha,v}^{-1}\|_{op} = \left\| \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^{\top}}{N_v} \right)^{-1} \right\|_{op} = \lambda_{min}^{-1} \left(\sum_{k \in U_v} \frac{r_k \pi_k \mathbf{x}_{k,\alpha} \mathbf{x}_{k,\alpha}^{\top}}{N_v} \right) \leq \frac{1}{\gamma}, \tag{45}$$

for some constant $\gamma > 0$, independent of N_v. Finally, combining (44) and (45), we have, pointwise

$$\|\mathbf{c}_{\alpha,v}\|_2^4 \leq \|A_{\alpha,v}^{-1}\|_{op}^4 \|\boldsymbol{b}_{\alpha,v}\|_2^4 \leq \frac{C_0^4}{\gamma^4}.$$

The result follows by integrating on both sides.

$\square$

Lemma 22. *Assume (S1)-(S3) and (D1)-(D3). For $\alpha \in C$, there exists a constant M such that*

$$\mathbb{E}_{mq} \left[\sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} \right] \leq M$$

almost surely.

Proof. Using $(a + b)^4 \leq 8a^4 + 8b^4$, we obtain

$$\mathbb{E}_{mq} \left[\sum_{k \in U_v} \frac{\eta_{k,\alpha}^4}{N_v} \right] = \mathbb{E}_{mq} \left[\sum_{k \in U_v} \frac{(\mathbf{x}_{k,\alpha}^{\top} \boldsymbol{\beta}_{\alpha} + r_k(1 + \pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha}) \epsilon_k)^4}{N_v} \right]$$

$$\leq \mathbb{E}_{mq} \left[\sum_{k \in U_v} \frac{8(\mathbf{x}_{k,\alpha}^{\top} \boldsymbol{\beta}_{\alpha})^4 + 8(r_k(1 + \pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha}) \epsilon_k)^4}{N_v} \right]$$

$$\leq \mathbb{E}_{mq} \left[\sum_{k \in U_v} \frac{8(\mathbf{x}_{k,\alpha}^{\top} \boldsymbol{\beta}_{\alpha})^4 + 8(1 + \pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha})^4 \epsilon_k^4}{N_v} \right]$$

$$\leq \sum_{k \in U_v} \frac{8 \left(\mathbf{x}_{k,\alpha}^{\top} \boldsymbol{\beta}_{\alpha} \right)^4 + 8\mathbb{E}_m \left[\epsilon_k^4 \right] \mathbb{E}_q \left[(1 + \pi_k \mathbf{c}_{\alpha,v}^{\top} \mathbf{x}_{k,\alpha})^4 \right]}{N_v}$$

$$\leq \sum_{k \in U_v} \frac{8\|\mathbf{x}_{k,\alpha}\|_2^4 \|\boldsymbol{\beta}_{\alpha}\|_2^4 + 8\mathbb{E}_m \left[\epsilon_k^4 \right] (8 + 8\pi_k^4 \mathbb{E}_q \left[\|\mathbf{c}_{\alpha,v}\|_2^4 \right] \|\mathbf{x}_{k,\alpha}\|_2^4)}{N_v}$$

$$\leq 8\|\boldsymbol{\beta}_{\alpha}\|_2^4 C_0^4 + 64M_0 + 64M_0 C_0^4 \mathbb{E}_q \left[\|\mathbf{c}_{\alpha,v}\|_2^4 \right]. \tag{46}$$

Using Lemma 21, we conclude

$$\mathbb{E}_{mq}\left[\sum_{k\in U_v}\frac{\eta_{k,\alpha}^4}{N_v}\right]\le M,$$

almost surely. $\square$

Lemma 23. *Assume (S1)-(S3) and (D1)-(D3). For $\alpha\in C$, there exist constants K_1 and K_2 such that*

$$\mathbb{E}_{mq}\left[\left(n_v\bar{V}_{1,v}(\alpha)\right)^2\right]\le K_1,\qquad \mathbb{E}_q\left[\left(n_v\bar{V}_{2,v}(\alpha)\right)^2\right]\le K_2$$

almost surely.

Proof. Write

$$
\begin{aligned}
n_v\left|\bar{V}_{1,v}(\alpha)\right| &= \frac{n_v}{N_v^2}\left|\sum_{k\in U_v}\sum_{l\in U_v}\frac{\eta_{k,\alpha}}{\pi_k}\frac{\eta_{l,\alpha}}{\pi_l}\Delta_{kl}\right| \\
&\le \frac{n_v}{N_v^2}\sum_{k\in U_v}\sum_{l\in U_v}\frac{|\eta_{k,\alpha}|}{\pi_k}\frac{|\eta_{l,\alpha}|}{\pi_l}|\Delta_{kl}| \\
&\le \frac{n_v}{N_v}\frac{1-\lambda}{\lambda}\sum_{k\in U_v}\frac{\eta_{k,\alpha}^2}{N_v}+\frac{n_v}{N_v}\sum_{k\in U_v}\sum_{\substack{l\in U_v\\l\ne k}}\frac{|\eta_{k,\alpha}||\eta_{l,\alpha}|}{N_v\pi_k\pi_l}|\Delta_{kl}| \\
&\le \frac{n_v}{N_v}\frac{1-\lambda}{\lambda}\sum_{k\in U_v}\frac{\eta_{k,\alpha}^2}{N_v}+\frac{n_v}{N_v\lambda^2}\max_{k\ne l\in U_v}|\Delta_{kl}|\sum_{k\in U_v}\sum_{\substack{l\in U_v\\l\ne k}}\frac{|\eta_{k,\alpha}||\eta_{l,\alpha}|}{N_v} \\
&\le \frac{n_v}{N_v}\frac{1-\lambda}{\lambda}\sum_{k\in U_v}\frac{\eta_{k,\alpha}^2}{N_v}+\frac{n_v}{\lambda^2}\max_{k\ne l\in U_v}|\Delta_{kl}|\sum_{k\in U_v}\frac{\eta_{k,\alpha}^2}{N_v}. \qquad (47)
\end{aligned}
$$

As a result, using Lemma 22 and integrating on both sides of (47) leads to

$$
\begin{aligned}
\mathbb{E}_{mq}\left[\left(n_v\bar{V}_{1,v}(\alpha)\right)^2\right] &\le \left(\frac{n_v}{N_v}\frac{1-\lambda}{\lambda}+\frac{n_v}{\lambda^2}\max_{k\ne l\in U_v}|\Delta_{kl}|\right)^2\mathbb{E}_{mq}\left[\left(\sum_{k\in U_v}\frac{\eta_{k,\alpha}^2}{N_v}\right)^2\right] \\
&\le \left(\frac{n_v}{N_v}\frac{1-\lambda}{\lambda}+\frac{\bar{\Delta}}{\lambda^2}\right)^2\mathbb{E}_{mq}\left[\sum_{k\in U_v}\frac{\eta_{k,\alpha}^4}{N_v}\right]\le K_1
\end{aligned}
$$

almost surely. Next, write

$$
\begin{aligned}
n_v\left|\bar{V}_{2,v}(\alpha)\right| &= \frac{n_v\sigma^2}{N_v}\left|\sum_{k\in U_v}\frac{1-r_k+r_k\left(\pi_k\mathbf{c}_{\alpha,v}^\top\mathbf{x}_{k,\alpha}\right)^2}{N_v}\right| \\
&\le \frac{n_v\sigma^2}{N_v}+\frac{n_v\sigma^2}{N_v}\|\mathbf{c}_{\alpha,v}\|_2^2\sum_{k\in U_v}\frac{\|\mathbf{x}_{k,\alpha}\|_2^2}{N_v} \\
&= \frac{n_v\sigma^2}{N_v}\left(1+C_0^2\|\mathbf{c}_{\alpha,v}\|_2^2\right) \qquad (48)
\end{aligned}
$$

Integrating on both sides of (48) leads to

$$\mathbb{E}_q \left[\left(n_v \bar{V}_{2,v}(\alpha) \right) \right] \le \frac{n_v^2 \sigma^4}{N_v^2} \left(2 + 2C_0^4 \mathbb{E}_q \left[\| \mathbf{c}_{\alpha,v} \|_2^4 \right] \right) \le K_2$$

almost surely using Lemma 21.

□

References

AKAIKE, H. (1970). Statistical predictor identification. *Annals of the institute of Statistical Mathematics* **22** 203–217.

BERGER, Y. G. and RAO, J. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **68** 531–547.

BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The annals of statistics* 470–482.

BREIDT, F. J. and OPSOMER, J. D. (2000). Local polynomial regression estimators in survey sampling. *Annals of statistics* 1026–1053.

BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian Regression. *The Annals of Statistics* 1674–1697.

CHAUVET, G., DEVILLE, J.-C. and HAZIZA, D. (2011). On balanced random imputation in surveys. *Biometrika* **98** 459–471.

CHAUVET, G. and GOGA, C. (2022). Asymptotic efficiency of the calibration estimator in a high-dimensional data setting. *Journal of Statistical Planning and Inference* **217** 177–187.

CHEN, J. and RAO, J. (2007). Asymptotic normality under two-phase sampling designs. *Statistica sinica* 1047–1064.

CHEN, J. and SHAO, J. (2000). Nearest neighbor imputation for survey data. *Journal of official statistics* **16** 113.

CHEN, J. and SHAO, J. (2001). Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association* **96** 260–269.

DAGDOUG, M., GOGA, C. and HAZIZA, D. (2025). Statistical inference in the presence of imputed survey data through regression trees and random forests. *Scandinavian Journal of Statistics* **52** 960–998.

FAY, R. E. (1991). *A design-based perspective on missing data variance*. US Census Bureau [custodian].

HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** 361–374.

HÁJEK, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics* **35** 1491–1523.

HANSEN, B. (2022). *Econometrics*. Princeton University Press.

HAZIZA, D. and BEAUMONT, J.-F. (2007). On the construction of imputation classes in surveys. *International Statistical Review* **75** 25–43.

HAZIZA, D. and VALLÉE, A.-A. (2020). Variance estimation procedures in the presence of singly imputed survey data: a critical review. *Japanese Journal of Statistics and Data Science* **3** 583–623.

HORN, R. A. and JOHNSON, C. R. (2012). *Matrix analysis*. Cambridge university press.

IPARRAGIRRE, A., LUMLEY, T., BARRIO, I. and AROSTEGUI, I. (2023). Variable selection with LASSO regression for complex survey data. *Stat* **12** e578.

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association* **77** 89–96.

KIM, J. K. and RAO, J. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96** 917–932.

KREWSKI, D. and RAO, J. N. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics* 1010–1019.

LARBI, K., TSANG, J., HAZIZA, D. and DAGDOUG, M. (2025). On the use of machine learning methods for the treatment of unit nonresponse in surveys. *Survey Methodology* **51** 275–303.

LUMLEY, T. and SCOTT, A. (2015). AIC and BIC for modeling with complex survey data. *Journal of Survey Statistics and Methodology* **3** 1–18.

NEMIROVSKI, A. (2000). Topics in Non-parametric Statistics, vol. 1738 of Ecole d'été de Probabilités de Saint-Flour 1998. *Lecture Notes in Mathematics. Springer, NY. MR1775640.*

NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics* 758–765.

OPSOMER, J. and MILLER, C. (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *Nonparametric Statistics* **17** 593–611.

PFEFFERMANN, D. and SVERCHKOV, M. (2009). Inference under informative sampling. In *Handbook of statistics*, **29** 455–487. Elsevier.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* 462–474.

RAO, R. and WU, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika* **76** 369–374.

RAO, C. R., WU, Y., KONISHI, S. and MUKERJEE, R. (2001). On model selection. *Lecture Notes-Monograph Series* 1–64.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.

SÄRNDAL, C.-E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. *Survey methodology* **18** 241–252.

SCHWARZ, G. (1978). Estimating the dimension of a model. *The annals of statistics* 461–464.

SHAO, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association* **88** 486–494.

SHAO, J. (1997). An asymptotic theory for linear model selection. *Statistica sinica* 221–242.

SHAO, J. and STEEL, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association* **94** 254–265.

THOMPSON, M. (1997). *Theory of sample surveys* **74**. CRC Press.

TROPP, J. A. et al. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* **8** 1–230.

WIECZOREK, J., GUERIN, C. and MCMAHON, T. (2022). K-fold cross-validation for complex sample surveys. *Stat* **11** e454.

YANG, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika* **92** 937–950.

YANG, Y. (2007). Consistency of cross validation for comparing regression procedures. *Annals of Statistics* **35** 2450–2473.

YANG, S. and KIM, J. K. (2017). Predictive mean matching imputation in survey sampling. *arXiv preprint arXiv:1703.10256.*

YANG, S. and KIM, J. K. (2019). Nearest neighbor imputation for general parameter estimation in survey sampling. In *The econometrics of complex survey data: Theory and applications* 209–234. Emerald Publishing Limited.

ZHANG, P. (1993). Model selection via multifold cross validation. *The annals of statistics* **21** 299–313.