Université de Montréal

# Estimation multi-robuste efficace en présence de données influentes

par

**Victoire Michal**

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

22 août 2019

# Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé

# Estimation multi-robuste efficace en présence de données influentes

présenté par

## Victoire Michal

a été évalué par un jury composé des personnes suivantes :

*Pierre Duchesne*
(président-rapporteur)

*David Haziza*
(directeur de recherche)

*Mylène Bédard*
(membre du jury)

Mémoire accepté le :
*7 août 2019*

# Sommaire

Lorsque des enquêtes sont effectuées, il est commun de faire face à de la non-réponse de la part des individus échantillonnés. Les estimateurs non-ajustés pouvant être biaisés en présence de données manquantes, on a habituellement recours à des méthodes d'imputation pour obtenir un fichier de données complété et réduire ainsi le biais de non-réponse. De plus, les estimateurs usuels de totaux ou moyennes de la population finie sont très sensibles à la présence de données influentes dans l'échantillon. Nous proposons une version efficace en présence de valeurs influentes des estimateurs multi-robustes, c'est-à-dire des estimateurs imputés par une méthode d'imputation multi-robuste. Pour ce faire, nous définissons le biais conditionnel d'une unité échantillonnée comme mesure de son influence. Nous présenterons les résultats d'une étude par simulation afin de montrer les gains de la méthode proposée en termes de biais et d'efficacité.

Mots-clefs : Robustesse ; Imputation multi-robuste ; Biais conditionnel ; Inférence basée sur le plan de sondage ; Unités influentes ; Non-réponse.

# Summary

Item nonresponse is a common issue in surveys. Because unadjusted estimators may be biased in the presence of nonresponse, it is common practice to impute the missing values, leading to the creation of a completed data file, in order to reduce the nonresponse bias. Moreover, the commonly used estimators of population totals/means are very unstable in the presence of influential units. We develop an efficient version, in the presence of influential units, of multiply robust estimators, which are estimators obtained after a multiply robust imputation method. To do so, we define the conditional bias of a sample unit as its measure of influence. We will present the results of a simulation study to show the benefits of the proposed method in terms of bias and efficiency.

Keywords: Robustness; Multiply robust imputation; Conditional bias; Design-based inference; Influential units; Item nonresponse.

# Table des matières

# Liste des tableaux

# Table des figures

# Remerciements

Il serait impensable de ne pas rendre hommage à mon superviseur, David Haziza. Il a su tisser un véritable « cocon » d'encouragements, tout en subtilité, que seul un œil (très) averti pouvait déceler ! Un grand merci d'avoir pu me permettre « d'admirer les astuces mathématiques », de repérer mes « regards hagards » et de combler toutes mes « incultures », sans oublier tous les conseils pour m'apprendre à écrire scientifiquement...

J'aimerais ensuite remercier mon fan club, dont ma maman, Mong, est la présidente et Nanou, le vice-président. Cette horde de supporters, composée de deux personnes, me soutient à la fois financièrement, mais surtout émotionnellement. Combien de fois ai-je appelé, paniquée d'avoir raté ma vie à cause d'un examen ? !

Je ne peux oublier Marc, mon admirateur officiel : j'espère que dans 10 ans, quand je relirai ce mémoire, tu seras à côté de moi, parce que sinon... malaise !

Je termine par un petit paragraphe pour mes amis. À des fins d'anonymat, je tairai leur nom et n'utiliserai que leur surnom. Quand j'ai besoin de France, je peux compter sur Mallau, Elo, Choup's et Kik ! Sinon, je me tourne vers le très anonyme Alexis, en Belgique après de nombreuses excursions chez Faure (un merci tout spécial pour la lecture et les commentaires à mi-parcours), Daniil, en Espagne et Véro... à Longueuil ! Enfin, les stateux : Gab (un énorme merci pour la lecture finale !) et Vanessou avec qui j'ai grandi, beaucoup trop ri et potiné, et Isa, la petite nouvelle, avec qui les potins vont également bon train.

# Introduction

Des enquêtes, aussi appelées sondages, sont régulièrement effectuées afin de collecter des informations sur une population finie ou sur un sous-ensemble (échantillon) de celle-ci. Dans la majorité des enquêtes, les données recueillies portent sur plusieurs variables d'intérêt et l'objectif est d'estimer des paramètres de population finie tels que des totaux, des moyennes ou encore des quantiles. Dans la base de données, une colonne de poids de sondage est fournie. Les estimateurs des paramètres souhaités pourront donc être calculés au moyen des données recueillies et des poids. Cependant, il est quasiment certain que l'on fera face à un problème de données manquantes. Il s'agit alors d'un problème de non-réponse. Deux types de non-réponse sont à distinguer : la non-réponse totale et la non-réponse partielle. La première désigne le cas où aucune réponse n'est fournie tandis que la seconde est caractérisée par le fait que seules certaines variables sont manquantes. La non-réponse totale survient lorsque, par exemple, l'unité échantillonnée n'a pas souhaité répondre au sondage ou encore s'il était impossible de la contacter. La non-réponse partielle survient, par exemple, dans un contexte de variable sensible telle que le revenu ou encore si la réponse fournie à une variable donnée n'est pas cohérente avec la réponse à d'autres variables de l'enquête, comme si une personne indiquait être mariée alors qu'âgée de 9 ans. Dans un tel cas, l'une des variables sera mise à manquante.

Le tableau 0.1, représentant un ensemble de données usuel, est séparé en trois parties : la réponse totale, la non-réponse partielle et la non-réponse totale. Le cas d'un ensemble de données complet, soit avec seulement des répondants, est traité dans la section 1.1 suivante. Généralement, la non-réponse totale est traitée par des méthodes de repondération : les individus pour lesquels aucune information n'est disponible sont retirés du fichier et les poids échantillonnaux des unités restantes sont recalculés afin de compenser l'élimination

des non-répondants du fichier. Le traitement de la non-réponse totale ne sera pas abordé dans ce mémoire.

| | $y_1$ | $y_2$ | $y_3$ | $\dots$ | $y_p$ | |
|---|---|---|---|---|---|---|
| 1 | ✓ | ✓ | ✓ | $\dots$ | ✓ | Réponse totale |
| 2 | ✓ | ✓ | ✓ | $\dots$ | ✓ | |
| $\vdots$ | ✓ | X | X | $\dots$ | ✓ | Non-réponse partielle |
| $\vdots$ | X | ✓ | X | $\dots$ | X | |
| $\vdots$ | X | X | X | $\dots$ | X | Non-réponse totale |
| $n$ | X | X | X | $\dots$ | X | |

**Tableau 0.1.** Un ensemble de données typique

Nous nous focalisons sur le traitement de la non-réponse partielle, qui est habituellement traitée par des méthodes d'imputation. Autrement dit, lorsqu'une donnée est manquante pour un individu, elle est remplacée par une valeur plausible obtenue au moyen d'une information auxiliaire disponible pour tous les individus de l'échantillon. Cette information auxiliaire, notée $\mathbf{v}_i$, est un vecteur de variables mesurées au cours de l'enquête et est complètement observé pour toute unité $i$ dans l'échantillon. La base de donnée complétée est exhibée dans le tableau 0.2, où $y_{ij}^*$ désigne la valeur imputée pour la variable $y_j$ manquante à l'unité $i$.

| | $y_1$ | $y_2$ | $y_3$ | $\dots$ | $y_p$ | |
|---|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | $y_{13}$ | $\dots$ | $y_{1p}$ | Réponse totale |
| 2 | $y_{21}$ | $y_{22}$ | $y_{23}$ | $\dots$ | $y_{2p}$ | |
| $\vdots$ | | | | | | |
| $i$ | $y_{i1}$ | $y_{i2}^*$ | $y_{i3}^*$ | $\dots$ | $y_{ip}$ | Non-réponse partielle |
| $\vdots$ | | | | | | |
| $n$ | $y_{n1}^*$ | $y_{n2}$ | $y_{n3}^*$ | $\dots$ | $y_{np}^*$ | |

**Tableau 0.2.** Un ensemble de données complété

Habituellement, les données manquantes sont remplacées au moyen d'une imputation simple où un seul modèle d'imputation est fixé. Plusieurs méthodes d'imputation existent, nous les présentons dans le chapitre 1. Dans le chapitre 2, nous mettons l'accent sur l'imputation multi-robuste, où une valeur imputée est fondée sur plusieurs modèles d'imputation et/ou de non-réponse.

D'autre part, les estimateurs usuels sont sensibles à la présence de données influentes dans la population finie étudiée. Dans le chapitre 1, nous définirons le biais conditionnel comme mesure de l'influence d'une unité puis nous définirons un estimateur efficace en présence de valeurs influentes lorsque l'échantillon ne présente pas de non-réponse. Dans le chapitre 2, nous définirons un estimateur imputé efficace en présence de valeurs influentes lorsqu'un seul modèle d'imputation est fixé. Nous proposerons enfin un estimateur imputé multi-robuste efficace en présence de données influentes. Ces estimateurs proposés seront le sujet d'études par simulation présentées dans le chapitre 3.

# Chapitre 1

---

# Introduction à la théorie de l'échantillonnage et à l'imputation

## 1.1. Théorie des sondages en présence de données complètes

Considérons une population finie $U = \{1, \ldots, i, \ldots, N\}$ de taille $N$ dont on veut estimer le total

$$t_y = \sum_{i \in U} y_i,$$

où $y$ désigne une variable d'intérêt. Pour ce faire, un échantillon $S$ de taille $n$ est sélectionné selon un certain plan d'échantillonnage $p$, tel que $p(S)$ désigne la probabilité que l'échantillon $S$ soit tiré. Soit la variable aléatoire indicatrice de sélection dans l'échantillon $I_i$, c'est-à-dire $I_i = 1$ si $i \in S$ et $I_i = 0$, sinon. Les probabilités d'inclusion du premier et du second ordre sont respectivement définies par

$$\pi_i = \mathbb{P}(i \in S) = \mathbb{P}(I_i = 1), \text{ pour tout } i \in U,$$

et pour $i, j \in U$ :

$$\pi_{ij} = \mathbb{P}(i \in S, j \in S) = \mathbb{P}(I_i = 1, I_j = 1), \ i \neq j.$$

Lorsque $i = j$, on a $\pi_{ij} = \pi_{ii} = \pi_i$.

**Résultat 1.** *Les variables indicatrices $I_i$ satisfont :*

*(i)* $\mathbb{E}_p(I_i) = \pi_i$ ;

*(ii)* $\mathbb{V}_p(I_i) = \pi_i(1 - \pi_i)$ ;

*(iii)* $\mathbb{E}_p(I_i I_j) = \pi_{ij}$ ;

*(iv)* $\mathrm{Cov}_p(I_i, I_j) = \pi_{ij} - \pi_i \pi_j \equiv \Delta_{ij}$,

*où* $\mathbb{E}_p(\cdot)$, $\mathbb{V}_p(\cdot)$ *et* $\mathrm{Cov}_p(\cdot)$ *désignent l'espérance, la variance et la covariance par rapport au plan de sondage* $p$.

DÉMONSTRATION. Voir Särndal et al., p.36. □

L'échantillonnage aléatoire simple et sans remise (EASSR) étant le premier exemple considéré dans la théorie des sondages, nous l'utiliserons tout au long de ce mémoire. Ce plan d'échantillonnage implique que tout échantillon de taille $n$ a la même probabilité d'être sélectionné que tout autre échantillon de même taille, soit $p(S_1) = p(S_2)$, où $S_1$ et $S_2$ sont deux échantillons de même cardinalité. Puisqu'il existe $\binom{N}{n}$ échantillons de taille $n$ dans une population de taille $N$, la probabilité de sélectionner $S$ de taille $n$ est $p(S) = 1/\binom{N}{n}$. Il en découle que les probabilités d'inclusion pour un échantillonnage aléatoire simple et sans remise sont les suivantes :

$$\pi_i = \mathbb{P}(i \in S) = \sum_{\substack{S \in \mathcal{Q} \\ S \ni i}} p(S) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}, \text{ pour tout } i \in U,$$

et pour $i, j \in U$ :

$$\pi_{ij} = \mathbb{P}(i \in S,\ j \in S) = \sum_{\substack{S \in \mathcal{Q} \\ S \ni i,j}} p(S) = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)},\ i \neq j,$$

où $\mathcal{Q} = \{s : p(s) > 0\}$ désigne le support du plan d'échantillonnage.

### 1.1.1. L'estimateur de Horvitz-Thompson

L'estimateur de Horvitz-Thompson du total $t_y$ est défini par

$$\widehat{t}_{y,HT} = \sum_{i \in S} w_i y_i, \tag{1.1.1}$$

où $w_i = \pi_i^{-1}$ désigne le poids d'échantillonnage associé à l'unité $i$. L'estimateur de Horvitz-Thompson est sans biais par rapport au plan de sondage, c'est-à-dire, $\mathbb{E}_p\left(\widehat{t}_{y,HT}\right) = t_y$. Sa variance s'exprime comme

$$\mathbb{V}_p\left(\widehat{t}_{y,HT}\right) = \sum_{i \in U} \sum_{j \in U} \Delta_{ij}(w_i y_i)(w_j y_j).$$

**Proposition 1.** *Sous les conditions de régularité décrites dans l'annexe A, l'estimateur* $\widehat{t}_{y,HT}$ *est convergent, au sens où*

$$\frac{\widehat{t}_{y,HT} - t_y}{N} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

DÉMONSTRATION. Voir l'annexe A.1. □

La proposition suivante nous sera utile dans le calcul du biais conditionnel défini plus tard.

**Proposition 2.** *Supposons que l'on cherche à estimer le total* $\sum_{j\in U} a_j$, *où* $a_1, \ldots, a_N$ *sont des quantités fixes. Pourvu que* $\pi_{ij} > 0$ *pour tout* $(i,j) \in U \times U$, *un estimateur conditionnellement sans biais de* $\sum_{j\in U} a_j$ *est donné par* $\sum_{j\in S} \frac{\pi_i}{\pi_{ij}} a_j$. *Autrement dit,*

$$\mathbb{E}_p\left(\sum_{j\in S} \frac{\pi_i}{\pi_{ij}} a_j \mid I_i = 1\right) = \sum_{j\in U} a_j.$$

DÉMONSTRATION. On a

$$\begin{aligned}
\mathbb{E}_p\left(\sum_{j\in S} \frac{\pi_i}{\pi_{ij}} a_j \mid I_i = 1\right) &= E_p\left(\sum_{j\in U} \frac{\pi_i}{\pi_{ij}} a_j I_j \mid I_i = 1\right) \\
&= \sum_{j\in U} \frac{\pi_i}{\pi_{ij}} a_j \mathbb{E}_p(I_j \mid I_i = 1) \\
&= \sum_{j\in U} \frac{\pi_i}{\pi_{ij}} a_j \mathbb{P}(I_j = 1 \mid I_i = 1) \\
&= \sum_{j\in U} \frac{\pi_i}{\pi_{ij}} a_j \frac{\pi_{ij}}{\pi_i} = \sum_{j\in U} a_j.
\end{aligned}$$

□

### 1.1.2. L'estimateur GREG

En pratique, il arrive fréquemment qu'une information auxiliaire soit disponible à l'étape de l'estimation. Soit $\mathbf{z}_i = (z_{1i}, \ldots, z_{Ji})^\top$ un vecteur de $J$ variables auxiliaires observé pour $i \in S$. Nous supposons que le vecteur des totaux dans la population, $\mathbf{t_z} = (t_{z_1}, \ldots, t_{z_J})^\top$, est connu, où $t_{z_j} = \sum_{i\in U} z_{ji}$. Supposons alors qu'un modèle $\xi$ relie la variable d'intérêt $y$ au vecteur $\mathbf{z}$ selon

$$\xi : \; y_i = \mathbf{z}_i^\top \boldsymbol{\beta} + \varepsilon_i,$$

où $\boldsymbol{\beta}$ est un vecteur de taille $J$ de paramètres inconnus et les erreurs $\varepsilon_i$ sont telles que

$$\mathbb{E}_\xi(\varepsilon_i) = 0, \; \mathbb{E}_\xi(\varepsilon_i\varepsilon_j) = 0 \text{ si } i \neq j \text{ et } \mathbb{V}_\xi(\varepsilon_i) = \sigma^2 c_i,$$

avec $c_i > 0$ un coefficient connu pour l'individu $i$ et où $\mathbb{E}_\xi(\cdot)$ et $\mathbb{V}_\xi(\cdot)$ désignent l'espérance et la variance par rapport au modèle $\xi$.

Nous cherchons à estimer le total $t_y$ dans la population. Le total de la variable d'intérêt dans la population peut s'écrire comme

$$t_y = \sum_{i\in U} y_i = \sum_{i\in U} (\mathbf{z}_i^\top \boldsymbol{\beta} + \varepsilon_i) = \sum_{i\in U} \mathbf{z}_i^\top \boldsymbol{\beta} + \sum_{i\in U} \varepsilon_i = \mathbf{t}_\mathbf{z}^\top \boldsymbol{\beta} + \sum_{i\in U} \varepsilon_i.$$

L'estimateur Generalized REGression (GREG) de $t_y$ sera alors de la forme

$$\widehat{t}_{y,GREG} = \mathbf{t}_\mathbf{z}^\top \widehat{\mathbf{B}} + \sum_{i\in S} w_i e_i = \widehat{t}_{y,HT} + (\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT})^\top \widehat{\mathbf{B}}, \tag{1.1.2}$$

où $e_i = y_i - \mathbf{z}_i^\top \widehat{\mathbf{B}}$ est le résidu échantillonnal propre à l'individu $i$ et où $\widehat{\mathbf{B}}$ est donné par

$$\widehat{\mathbf{B}} = \left( \sum_{i\in S} w_i \mathbf{z}_i c_i^{-1} \mathbf{z}_i^\top \right)^{-1} \sum_{i\in S} w_i \mathbf{z}_i c_i^{-1} y_i.$$

**Proposition 3.** *L'erreur due à l'échantillonnage de $\widehat{t}_{y,GREG}$ peut être approximée comme suit :*

$$\frac{\widehat{t}_{y,GREG} - t_y}{N} = \frac{1}{N} \left( \sum_{i\in S} w_i E_i - \sum_{i\in U} E_i \right) + O_P\left(\frac{1}{n}\right), \tag{1.1.3}$$

*où $E_i = y_i - \mathbf{z}_i^\top \mathbf{B}$ avec*

$$\mathbf{B} = \left( \sum_{i\in U} \mathbf{z}_i c_i^{-1} \mathbf{z}_i^\top \right)^{-1} \sum_{i\in U} \mathbf{z}_i c_i^{-1} y_i.$$

Démonstration. En faisant apparaître $\mathbf{B}$ dans l'estimateur GREG (1.1.2), on a

$$\begin{aligned}
\widehat{t}_{y,GREG} - t_y &= \widehat{t}_{y,HT} + (\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT})^\top \widehat{\mathbf{B}} - t_y \\
&= \widehat{t}_{y,HT} + (\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT})^\top (\widehat{\mathbf{B}} - \mathbf{B}) + (\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT})^\top \mathbf{B} - t_y \\
&= \widehat{t}_{y,HT} + (\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT})^\top \mathbf{B} - t_y + O_P\left(\frac{N}{n}\right) \\
&= \sum_{i\in S} w_i E_i - \sum_{i\in U} E_i + O_P\left(\frac{N}{n}\right),
\end{aligned}$$

en remarquant que $\mathbf{t}_\mathbf{z} - \widehat{\mathbf{t}}_{\mathbf{z},HT} = O_P\left(\frac{N}{\sqrt{n}}\right)$ et $\widehat{\mathbf{B}} - \mathbf{B} = O_P\left(\frac{1}{\sqrt{n}}\right)$. □

## 1.2. L'efficacité en présence de valeurs influentes

### 1.2.1. Le biais conditionnel

Afin de quantifier l'influence d'une unité échantillonnée, nous utiliserons le concept de biais conditionnel comme Beaumont et al. (2013). Soit $\theta$ un paramètre de population finie et $\widehat{\theta}$ un estimateur de $\theta$. Le biais conditionnel de l'unité $i$ échantillonnée est défini par

$$B_{1i} = \mathbb{E}_p\left(\widehat{\theta} \mid I_i = 1\right) - \theta.$$

Supposons, par exemple, que l'on s'intéresse au total $t_y$ et que l'estimateur de Horvitz-Thompson, $\widehat{t}_{y,HT}$, est utilisé. Dans ce cas, on peut montrer que

$$B_{1i} = \mathbb{E}_p\left(\widehat{t}_{y,HT} \mid I_i = 1\right) - t_y = \sum_{j \in U} \frac{\Delta_{ij}}{\pi_i \pi_j} y_j. \tag{1.2.1}$$

Le biais conditionnel de l'unité échantillonnée $i$ est donc la moyenne de l'erreur d'échantillonnage $\widehat{t}_{y,HT} - t_y$ calculée sur tous les échantillons contenant $i$. Le biais conditionnel d'un individu dans l'échantillon est donc fonction de la variable d'intérêt $y$. Ainsi, l'unité $i$ peut être influente pour une certaine variable d'intérêt mais pourrait ne pas l'être pour une autre variable d'intérêt. Cette mesure d'influence dépend également du plan de sondage via les probabilités d'inclusion du premier et du second ordre $\pi_i$ et $\pi_{ij}$. Lorsque $\pi_i = 1$, le biais conditionnel vaut 0 et l'unité $i$ n'est pas influente : puisque toujours échantillonnée, elle ne contribue pas à la variabilité de l'estimateur $\widehat{t}_{y,HT}$. La variance de $\widehat{t}_{y,HT}$ peut d'ailleurs être exprimée en fonction du biais conditionnel :

$$\mathbb{V}_p(\widehat{t}_{y,HT}) = \sum_{i \in U} B_{1i} y_i.$$

Le biais conditionnel $B_{1i}$ étant inconnu, il devra être estimé. Puisque

$$B_{1i} = \sum_{j \in U} a_j,$$

où $a_j = \dfrac{\Delta_{ij}}{\pi_i \pi_j} y_j$, nous utilisons la Proposition 2 afin de déterminer un estimateur conditionnellement sans biais de $B_{1i}$ :

$$\widehat{B}_{1i} = \sum_{j \in S} \frac{\Delta_{ij}}{\pi_j \pi_{ij}} y_j. \tag{1.2.2}$$

Dans le cas de l'échantillonnage aléatoire simple sans remise, les expressions (1.2.1) et (1.2.2) se simplifient pour donner

$$B_{i1} = \left(\frac{N}{n} - 1\right) \frac{N}{N-1}(y_i - \overline{Y}) \quad \text{et} \quad \widehat{B}_{1i} = \left(\frac{N}{n} - 1\right) \frac{n}{n-1}(y_i - \overline{y}),$$

où $\overline{Y} = N^{-1} \sum_{j \in U} y_j$ et $\overline{y} = n^{-1} \sum_{j \in S} y_j$.

En plus de dépendre de la variable d'intéret et du plan de sondage, le biais conditionnel dépend également de l'estimateur de $t_y$. En effet, considérons l'estimateur GREG (1.1.2). En ignorant les termes d'ordre inférieur en (1.1.3), on a :

$$\begin{aligned} B_{1i}^{GREG} = \mathbb{E}_p\left(\widehat{t}_{y,GREG} \mid I_i = 1\right) - t_y &\simeq \mathbb{E}_p\left(\sum_{j \in S} w_j E_j - \sum_{j \in U} E_j \mid I_i = 1\right) \\ &= \sum_{j \in U} \frac{\Delta_{ij}}{\pi_i \pi_j} E_j. \end{aligned}$$

Ainsi, pour un même paramètre, une même unité n'aura pas la même influence sur l'estimateur de Horvitz-Thomspon et sur l'estimateur GREG.

Il est à noter que, dans une population finie, les unités non-échantillonnées peuvent également être influentes. Le biais conditionnel $B_{0i}$ d'une unité non-échantillonnée est défini par :

$$B_{0i} = \mathbb{E}_p\left(\widehat{t}_{y,HT} \mid I_i = 0\right) - t_y = -\sum_{j \in U} \frac{\Delta_{ij}}{\pi_j(1 - \pi_i)} y_j = \frac{-\pi_i}{1 - \pi_i} B_{1i}.$$

Le biais conditionnel de l'unité non-échantillonnée $i$ est donc la moyenne de l'erreur d'échantillonnage, $\widehat{t}_{y,HT} - t_y$, calculée sur tous les échantillons ne contenant pas l'unité $i$. Cependant, à l'étape de l'estimation, rien ne peut être fait quant à l'influence des unités non-échantillonnées, car leur valeur de la variable $y$ n'est pas observée.

Une unité est influente si sa présence ou son absence dans l'échantillon a un impact significatif sur l'estimation. Supposons par exemple que l'on cherche à estimer le total des revenus d'une population dont un individu à très haut revenu fait partie. S'il est sélectionné dans l'échantillon, l'estimateur du total tendra à surestimer le vrai total ; au contraire, s'il n'est pas échantillonné, un risque est couru de sous-estimer le total des revenus de la population.

### 1.2.2. Version efficace de l'estimateur de Horvitz-Thompson

L'objectif de cette section est de rendre l'estimateur $\widehat{t}_{y,HT}$ en (1.1.1) efficace en présence de valeurs influentes. Cela signifie que nous cherchons à limiter l'impact des valeurs influentes

sur $\widehat{t}_{y,HT}$. Définissons $\widehat{t}^R_{y,HT}$, l'estimateur de Horvitz-Thompson efficace tel que

$$\widehat{t}^R_{y,HT} = \widehat{t}_{y,HT} - \Delta_c,$$

où $\Delta_c$ est une variable aléatoire indépendante de $i$ et qui dépend d'un seuil $c$. On cherche alors la valeur de $\Delta_c$ qui limite l'impact des données influentes sur l'estimateur $\widehat{t}^R_{y,HT}$. Soit $B^R_{1i}$ le biais conditionnel d'une unité échantillonnée sur l'estimateur efficace. On a

$$B^R_{1i} = \mathbb{E}_p\left(\widehat{t}^R_{y,HT} \mid I_i = 1\right) - t_y = B_{1i} - \Delta_c.$$

Ainsi, on cherche la valeur de $\Delta_c$ qui minimise l'influence maximale des unités sur $\widehat{t}^R_{y,HT}$, i.e. le $\Delta_c$ qui minimise :

$$\max\{|\widehat{B}^R_{1i}|;\ i \in S\},$$

où $\widehat{B}^R_{1i}$ désigne l'estimateur de $B^R_{1i}$ tel que

$$\widehat{B}^R_{1i} = \widehat{B}_{1i} - \Delta_c.$$

Beaumont et al. (2013) montrent que cette minimisation est atteinte par

$$\Delta_{c,opt} = \frac{\widehat{B}_{\min} + \widehat{B}_{\max}}{2}, \tag{1.2.3}$$

où $\widehat{B}_{\min}$ et $\widehat{B}_{\max}$ désignent respectivement les minimum et maximum des biais conditionnels estimés de l'estimateur de Horvitz-Thompson en (1.2.2).

Pour mieux comprendre (1.2.3), considérons le cas de $n$ réels ordonnés $x_{(1)}, \ldots, x_{(n)}$. La valeur $h$ qui minimise

$$\max\left\{|x_i - h|,\ i = 1, \ldots, n\right\},$$

est donnée par le mi-point (*midrange*)

$$h = \frac{x_{(1)} + x_{(n)}}{2}.$$

Ainsi, la version efficace de l'estimateur de Horvitz-Thompson est donnée par

$$\widehat{t}^R_{y,HT} = \widehat{t}_{y,HT} - \frac{\widehat{B}_{\min} + \widehat{B}_{\max}}{2}.$$

Remarquons que Beaumont et al. (2013) parlent d'un estimateur « robuste » aux valeurs influentes. Cependant, dans le chapitre 2 de ce mémoire, il sera question d'imputation multi-robuste. Alors, pour éviter toute confusion dans ce mémoire, nous parlerons toujours d'un estimateur « efficace » en présence de valeurs influentes.

## 1.3. La non-réponse

La non-réponse affecte les estimations, car elle engendre un biais de non-réponse et une variabilité additionnelle, appelée variance due à la non-réponse, causée par une taille d'échantillon effective réduite.

Nous supposons que le vecteur $\mathbf{x}_i$ est observé pour tout $i \in S$ et seule la variable d'intérêt $y$ est sujette à la non-réponse. Soit $r$ la variable indicatrice de réponse telle que $r_i = 1$ si $y_i$ est observée pour l'individu $i$ et $r_i = 0$, sinon. Ainsi, l'échantillon $S$ est divisé en deux sous-ensembles : l'ensemble $S_r$ des répondants et l'ensemble $S_{nr}$ des non-répondants, de sorte que $S = S_r \cup S_{nr}$. Les expressions *Missing Completely At Random* (*MCAR*), *Missing At Random* (*MAR*) et *Not Missing At Random* (*MNAR*) sont fréquemment rencontrées dans un contexte de non-réponse. Le mécanisme de non-réponse est la distribution $\mathcal{F}(\mathbf{R} \mid \mathbf{y}_U, \mathbf{X}_U)$, où $\mathbf{R} = (r_1, \ldots, r_N)^\top$, $\mathbf{y}_U = (y_1, \ldots, y_N)^\top$ et $\mathbf{X}_U = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^\top$. Le premier moment de cette distribution est donné par

$$\mathbb{E}(\mathbf{R} \mid \mathbf{y}_U, \mathbf{X}_U) = (p_1, \ldots, p_N)^\top,$$

où $p_i = \mathbb{P}(r_i = 1 \mid y_i, \mathbf{x}_i)$.

Le mécanisme de non-réponse est dit *MCAR* lorsque $\mathcal{F}(\mathbf{R} \mid \mathbf{y}_U, \mathbf{X}_U) = \mathcal{F}(\mathbf{R})$, c'est-à-dire lorsque la distribution de $\mathbf{R}$ est indépendante de $\mathbf{y}_U$ et de $\mathbf{X}_U$. Dans ce cas, la probabilité de réponse $p_i$ est donnée par $p_i = \mathbb{P}(r_i = 1)$. Autrement dit, il n'y a aucun lien entre le fait de répondre et les variables mesurées.

Dans la plupart des cas, l'hypothèse *MCAR* est irréaliste, car trop forte. L'hypothèse *MAR* (voir Rubin, 1976), moins forte, stipule que

$$p_i = \mathbb{P}(r_i = 1 \mid y_i, \mathbf{z}_i) = \mathbb{P}(r_i = 1 \mid \mathbf{z}_i),$$

où le vecteur de variables confondantes $\mathbf{z}$ est le sous-ensemble des variables du vecteur $\mathbf{x}$ dont dépendent $y$ et $r$. Cela signifie qu'après avoir conditionné sur $\mathbf{z}$, il n'y a aucune corrélation entre le fait de répondre et la variable d'intérêt $y$. Une autre formulation de cette hypothèse stipule que

$$f(y \mid \mathbf{z}, r = 1) = f(y \mid \mathbf{z}, r = 0).$$

Autrement dit, la distribution de la variable d'intérêt $y$ est la même chez les répondants et chez les non-répondants après avoir conditionné sur les variables confondantes $\mathbf{z}$.

Finalement, le scénario *MNAR* s'impose lorsque la variable $y$ et la probabilité de réponse ne sont pas indépendantes, même après avoir conditionné sur $\mathbf{z}$. Par exemple, le scénario de non-réponse est *MNAR* lorsque la non-réponse dépend directement de la variable d'intérêt.

Considérons l'exemple suivant afin d'expliquer et de visualiser les trois mécanismes de non-réponse. Nous générons un ensemble de données de taille $N = 500$ de la manière suivante :

(i) nous créons d'abord une variable « sexe », $x_1$, telle que $x_{1i} = 1$ si l'individu $i$ est un homme et $x_{1i} = 0$, sinon ;

(ii) la variable d'intérêt « revenu », $y$, est ensuite générée selon le modèle

$$y_i = 43 + 15x_{1i} + 5\varepsilon_i,$$

où le bruit aléatoire est tel que $\varepsilon \sim \mathcal{N}(0, 1)$. Nous avons fait en sorte d'avoir un revenu moyen, en miliers de dollars, d'environ 50. Le revenu moyen des hommes est de 58 et celui des femmes, 43.

(iii) Trois mécanismes sont ensuite utilisés pour générer la non-réponse dans le jeu de données :

-*MCAR* : on pose la probabilité de réponse $p_i = 0{,}5$ pour tout $i$ ;

-*MAR* : on pose la probabilité de réponse $p_i = 0{,}3$ si l'individu $i$ est un homme et $p_i = 0{,}7$, sinon. Les hommes auront donc un taux de réponse plus faible que celui des femmes.

-*MNAR* : On génère la probabilité de réponse comme une fonction logistique du revenu :

$$p_i = 1 - \frac{\exp(-45 + 0{,}9y_i)}{1 + \exp(-45 + 0{,}9y_i)}.$$

Ainsi, la probabilité de fournir son revenu diminue avec le revenu.

Le mécanisme de réponse *MCAR* est illustré dans la figure 1.1, où les répondants sont distingués des non-répondants. Nous voyons alors que la distribution des revenus est approximativement la même chez les répondants et chez les non-répondants. Autrement dit, il n'y a aucun lien entre le fait de répondre et la variable mesurée, comme par définition du cas *MCAR*.

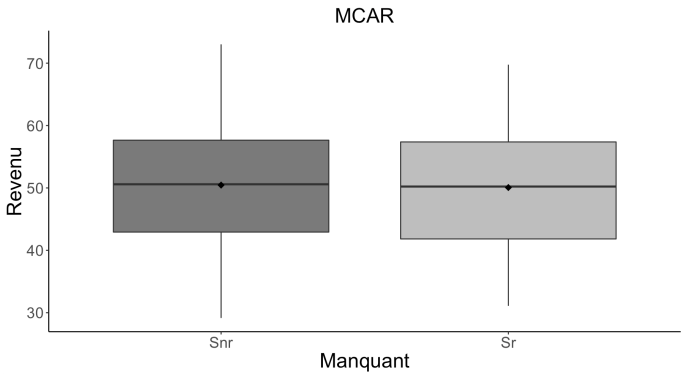



**Figure 1.1.** Illustration du cas *MCAR*

La figure 1.2 illustre le mécanisme *MAR*. Nous y représentons à nouveau le revenu pour les répondants et pour les non-répondants. Puisque nous avons défini la probabilité de réponse comme fonction du sexe, nous conditionnons de plus sur la variable confondante $z = x_1$. Nous voyons que la répartition des revenus chez les hommes répondants est approximativement la même que celle des hommes non-répondants, et donc de celle de l'échantillon initial ; de même pour les femmes répondantes – ce qui n'était pas le cas sans cette classification basée sur le sexe.

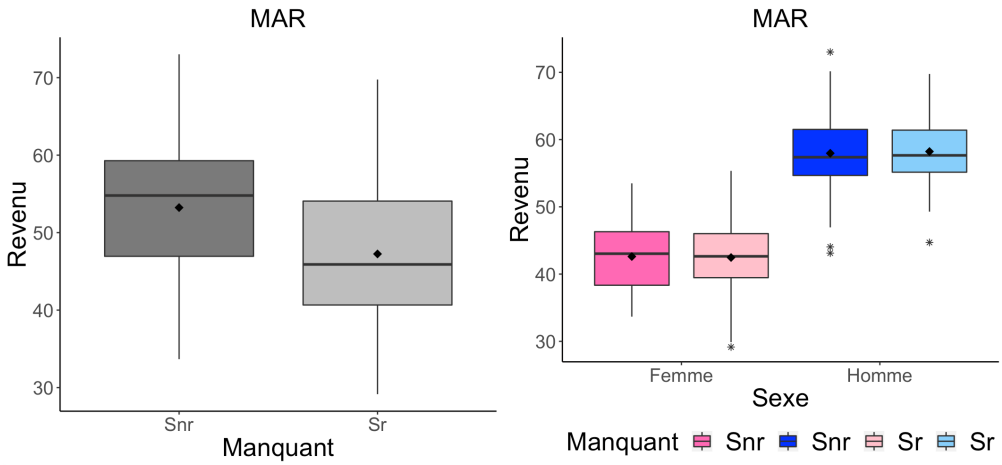



**FIGURE 1.2.** Illustration du cas *MAR*

Nous illustrons enfin le mécanisme de non-réponse *MNAR* dans la figure 1.3. À nouveau, nous distinguons les revenus des répondants de ceux des non-répondants. De plus, même après avoir conditionné sur le sexe, les distributions des revenus diffèrent. Ainsi, comme défini par le mécanisme *MNAR*, la variable d'intérêt et la probabilité de réponse ne sont pas indépendantes.

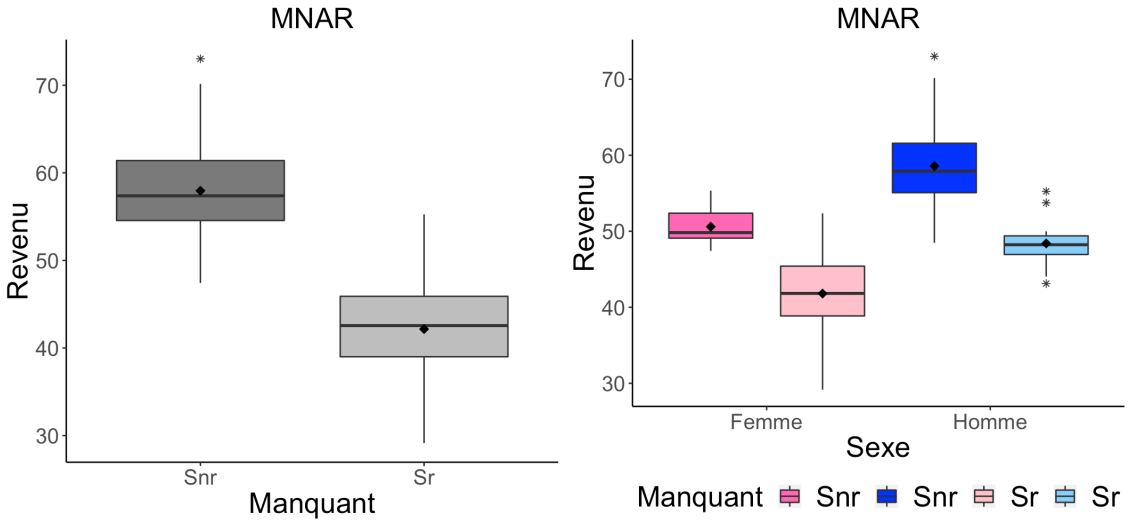



**FIGURE 1.3.** Illustration du cas *MNAR*

## 1.4. Imputation

Habituellement, l'imputation consiste à construire une valeur de remplacement à partir d'un modèle d'imputation $m$. Il existe de nombreuses méthodes d'imputation. On distingue

deux classes de méthodes : les méthodes déterministes et les méthodes aléatoires. Une méthode déterministe implique que pour un même fichier de données et une même méthode d'imputation, la même base de données complétée sera obtenue si le processus d'imputation est répété. À l'inverse, une méthode aléatoire suppose que l'imputation varie à chaque itération d'un même processus d'imputation, malgré une base de départ identique. On peut séparer les imputations paramétriques (sous-section 1.4.1) de celles non-paramétriques (sous-section 1.4.2). Pour une imputation paramétrique on postulera le modèle suivant :

$$y_i = m(\mathbf{v}_i, \boldsymbol{\beta}) + \varepsilon_i,$$

où $m(\cdot, \boldsymbol{\beta})$ est une certaine fonction, $\mathbf{v}_i$, un vecteur de variables entièrement observées pour l'individu $i$ et $\boldsymbol{\beta}$, un vecteur de paramètres inconnus. On formule les hypothèses suivantes concernant les erreurs $\varepsilon_i$ :

$$\mathbb{E}_m(\varepsilon_i|\mathbf{v}_i) = 0; \quad \mathbb{E}_m(\varepsilon_i\varepsilon_j|\mathbf{v}_i) = 0,\ i \neq j; \quad \mathbb{V}_m(\varepsilon_i|\mathbf{v}_i) = \sigma^2 c_i,$$

avec $\sigma^2$, un paramètre inconnu et $c_i > 0$, un coefficient connu pour tout $i \in S$. Notons que nous ne faisons aucune hypothèse à propos de la distribution des erreurs $\varepsilon$. La valeur imputée pour $i \in S_{nr}$ est

$$y_i^* = m(\mathbf{v}_i; \widehat{\boldsymbol{\beta}}),$$

où $\widehat{\boldsymbol{\beta}}$ est obtenu grâce aux cas complets en résolvant l'équation estimante suivante :

$$\sum_{i \in S_r} \phi_i \{y_i - m(\mathbf{v}_i, \boldsymbol{\beta})\} \frac{\partial m(\mathbf{v}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0},$$

où $\phi_i$ est un coefficient propre à l'individu $i$. Notons que ce coefficient est souvent égal à 1, pour une imputation non-pondérée, ou au poids $w_i$, pour une imputation pondérée.

Dans le cas d'une imputation aléatoire, il suffit d'ajouter un bruit aléatoire à une méthode déterministe. La valeur imputée $y_i^*$ est donnée par

$$y_i^* = m(\mathbf{v}_i; \widehat{\boldsymbol{\beta}}) + \widehat{\sigma}\sqrt{c_i} e_i^*,$$

où $\widehat{\sigma}$ est un estimateur de $\sigma$ et $e_i^*$ est le bruit aléatoire associé à l'individu $i$ tel que $\mathbb{E}_I(e_i^*) = 0$, où $\mathbb{E}_I(\cdot)$ désigne l'espérance par rapport au mécanisme d'imputation utilisé afin de générer les $e_i^*$. Soit

$$\tilde{e}_i = (\widehat{\sigma}\sqrt{c_i})^{-1} \left\{ y_i - m(\mathbf{v}_i; \widehat{\boldsymbol{\beta}}) \right\}$$

et le résidu standardisé, $e_i$, défini par

$$e_i = \tilde{e}_i - \frac{\sum_{j\in S}\phi_j r_j \tilde{e}_j}{\sum_{j\in S}\phi_j r_j}.$$

Les bruits aléatoires $e_i^*$ sont tirés aléatoirement avec remise de l'ensemble des résidus standardisés $\{e_i; i \in S_r\}$ tels que

$$e_i^* = e_i,\ i \in S_r,\ \text{avec probabilité } \frac{\phi_i}{\sum_{j\in S_r}\phi_j}.$$

Comme plus haut, le coefficient $\phi_i$ est souvent égal à 1, pour une imputation non-pondérée, ou au poids $w_i$, pour une imputation pondérée.

Soit $\tilde{y}$ la variable $y$ après imputation définie par

$$\tilde{y}_i = r_i y_i + (1 - r_i) y_i^*,\ i \in S.$$

L'estimateur du total $t_y$ après imputation est donné par

$$\widehat{t}_I = \sum_{i\in S} w_i \tilde{y}_i = \sum_{i\in S} w_i \left\{ r_i y_i + (1 - r_i) y_i^* \right\}. \tag{1.4.1}$$

### 1.4.1. Imputation par la régression linéaire

Lorsque le lien entre la variable d'intérêt, $y$, et le vecteur de variables auxiliaires, $\mathbf{v}$, est linéaire, une imputation par la régression linéaire est appropriée. Dans ce cas, on a

$$m(\mathbf{v}_i, \boldsymbol{\beta}) = \mathbf{v}_i^\top \boldsymbol{\beta} \quad \text{et} \quad y_i^* = \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}, \tag{1.4.2}$$

où $\widehat{\boldsymbol{\beta}}$ est l'estimateur des moindres carrés pondérés de $\boldsymbol{\beta}$ :

$$\widehat{\boldsymbol{\beta}} = \left( \sum_{i\in S_r} \frac{\phi_i}{c_i} \mathbf{v}_i \mathbf{v}_i^\top \right)^{-1} \sum_{i\in S_r} \frac{\phi_i}{c_i} \mathbf{v}_i y_i.$$

**Proposition 4.** *Avec une imputation par la régression linéaire, l'estimateur imputé* $\widehat{t}_I$ (1.4.1) *est sans biais pour* $t_y$*, au sens où*

$$\mathbb{E}_{mpq}\left(\widehat{t}_I - t_y\right) \equiv \mathbb{E}_m \mathbb{E}_p \mathbb{E}_q \left(\widehat{t}_I - t_y\right) = 0.$$

*Dans le cas d'une imputation par la régression linéaire aléatoire, on a également*

$$\mathbb{E}_{mpqI}(\widehat{t}_I - t_y) \equiv \mathbb{E}_m \mathbb{E}_p \mathbb{E}_q \mathbb{E}_I \left(\widehat{t}_I - t_y\right) = 0.$$

Démonstration. Voir annexe A.2. □

**Proposition 5.** *Sous certaines conditions de régularité (voir annexe A), on a*

$$\frac{\widehat{t}_I - t_y}{N} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

*Autrement dit, l'estimateur $\widehat{t}_I$ est convergent pour $t_y$.*

DÉMONSTRATION. Voir annexe A.3. □

Il existe des cas particuliers bien connus de l'imputation par la régression linéaire dont l'imputation par la moyenne, l'imputation par le ratio et l'imputation par donneur. Ces méthodes sont maintenant discutées.

1.4.1.1. *Imputation par le ratio*

Ce cas particulier de l'imputation par la régression linéaire est obtenu en posant $\mathbf{v}_i = v_i$ et $c_i = v_i$ dans (1.4.2), où $v_i$ est une variable quantitative. Cette forme d'imputation est appropriée lorsque la relation entre la variable $y$ et la variable auxiliaire $v$ est linéaire et passe par l'origine. Dans ce cas, on a

$$y_i^* = \widehat{B}_r v_i = \frac{\overline{y}_r}{\overline{v}_r} v_i,\ i \in S_{nr},$$

avec le paramètre $\widehat{B}_r = \overline{y}_r/\overline{v}_r$, où $\overline{y}_r = \left(\sum_{i\in S_r}\phi_i\right)^{-1}\sum_{i\in S_r}\phi_i y_i$ et $\overline{v}_r = \left(\sum_{i\in S_r}\phi_i\right)^{-1}\sum_{i\in S_r}\phi_i v_i$.

1.4.1.2. *Imputation par la moyenne*

Lorsque $\mathbf{v}_i = 1$ et $c_i = 1$, pour tout $i$ dans (1.4.2), on obtient

$$y_i^* = \overline{y}_r,\ i \in S_{nr},$$

où $\overline{y}_r = \left(\sum_{i\in S_r}\phi_i\right)^{-1}\sum_{i\in S_r}\phi_i y_i$ est la moyenne pondérée des répondants.

1.4.1.3. *Imputation par donneur : imputation par hot-deck aléatoire*

L'imputation par hot-deck aléatoire désigne le cas où la valeur manquante $y_i$ est remplacée par la valeur d'un donneur sélectionné au hasard parmi les cas complets, c'est-à-dire

$$y_i^* = y_j,\ j \in S_r \quad \text{avec probabilité} \quad \frac{\phi_j}{\sum_{k\in S_r}\phi_k}.$$

Ce type d'imputation peut être vu comme la version aléatoire de l'imputation par la moyenne, où $v_i = c_i = 1$ pour tout $i$. En pratique, l'utilisateur forme des voisinages d'imputation selon l'information des variables auxiliaires avant de procéder à l'imputation hot-deck aléatoire. Plusieurs méthodes peuvent être utilisées pour créer ces cellules d'imputation telles que croiser les variables auxiliaires entre elles ou encore générer des arbres de régression et de classification, voir par exemple Haziza (2009) et Chen et Haziza (2019).

### 1.4.2. Imputation non-paramétrique

#### 1.4.2.1. *Imputation par le noyau*

Lorsque la fonction $m(\cdot, \boldsymbol{\beta})$ est mal spécifiée, les estimateurs résultants peuvent être biaisés. Par exemple, lors de la sélection d'un modèle d'imputation par la régression linéaire, les interactions et les termes quadratique peuvent être omis. Une imputation non-paramétrique peut être envisagée pour se protéger contre ces possibles oublis. Le modèle devient alors :

$$y_i = m(\mathbf{v}_i) + \varepsilon_i,$$

où la fonction $m(\cdot)$ et la variance $\mathbb{V}_m(\varepsilon_i|\mathbf{v}_i)$ sont laissées libres.

Dans cette section, nous discutons de l'imputation par noyau, un cas particulier de l'imputation non-paramétrique. Considérons le cas où $v_i$ est scalaire afin de pouvoir représenter le fichier de données en un nuage de points en deux dimensions $(v_i, y_i)$. D'abord, les données sont séparées au regard des $v$, en « voisinages » de largeur $2h$, où $h$ est le paramètre de lissage arbitraire. On utilisera alors les valeurs observées du voisinage de $v_i$ afin de constituer la valeur imputée $y_i^*$. On pourra alors utiliser la moyenne pondérée des observations dans le voisinage, où les poids sont déterminés au moyen d'une fonction noyau $\mathcal{K}(\cdot, v_i)$ :

$$y_i^* = \frac{\sum_{j\in S_r} \phi_j \mathcal{K}\left(\dfrac{v_j - v_i}{h}\right) y_j}{\sum_{j\in S_r} \phi_j \mathcal{K}\left(\dfrac{v_j - v_i}{h}\right)} = \frac{\sum_{j\in S_r} \tilde{w}_j y_j}{\sum_{j\in S_r} \tilde{w}_j},\ i \in S_{nr},$$

avec $\tilde{w}_j = \phi_j \mathcal{K}\left(\dfrac{v_j - v_i}{h}\right)$. La figure 1.4 illustre le fonctionnement de l'imputation par le noyau.

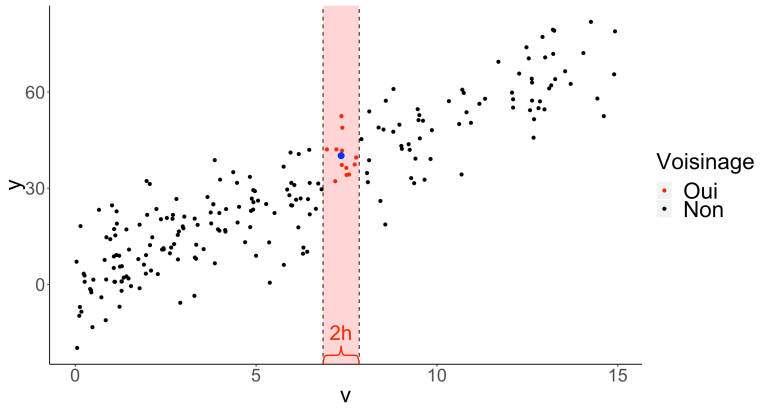



**FIGURE 1.4.** Illustration de l'imputation par le noyau

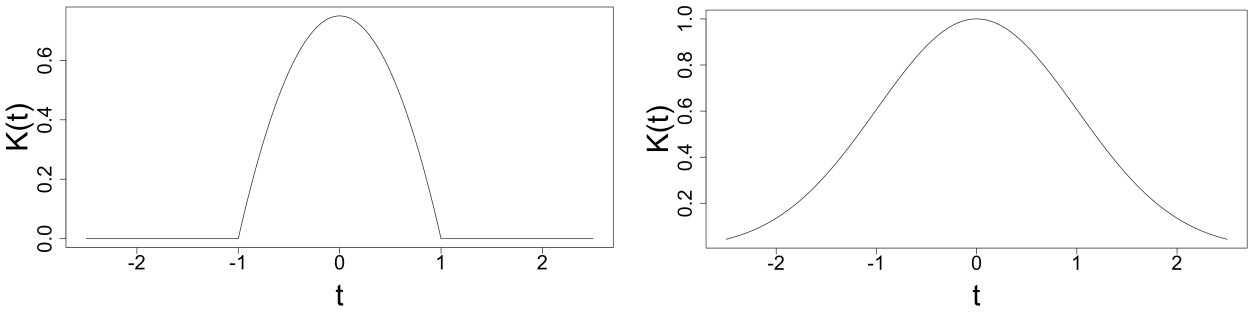
Plusieurs fonctions noyau sont conçues afin que le poids des unités diminue à mesure qu'elles s'éloignent du point focal, $v_i$. Considérons deux exemples de fonctions noyau : la fonction d'Epanechnikov (Epanechnikov, 1969)

$$\mathcal{K}(t) = (3/4)(1 - t^2)\mathbf{1}\left\{|t| \leq 1\right\},$$

qui implique qu'en dehors du voisinage prédéfini, les individus se voient assigner un poids nul, ou encore le noyau gaussien

$$\mathcal{K}(t) = \exp(-(1/2)t^2),\ t \in \mathbb{R},$$

où le poids diminue progressivement à mesure que l'on s'éloigne du point focal. Ces deux fonctions sont exhibées dans la figure 1.5 suivante.




**(a)** Noyau d'Epanechnikov **(b)** Noyau gaussien

**FIGURE 1.5.** Représentation graphique des fonctions noyau définies

Dans le but de comparer l'imputation par la régression linéaire à l'imputation par le noyau, nous avons effectué une étude par simulation. Nous avons répété le processus suivant 1000 fois :

(i) On génère 4 différentes populations finies de taille $N = 20\,000$ ;

(ii) on crée la variable auxiliaire scalaire $v$ suivant une loi uniforme : $v \sim \mathcal{U}(0,1)$ et un bruit aléatoire normal : $\varepsilon \sim \mathcal{N}(0,1)$ ;

(iii) la variable d'intérêt $y$ est générée selon les 4 modèles suivants :

- $y_{1i} = 2v_i + 0{,}5\varepsilon_i$,
- $y_{2i} = 1 + 2(v_i - 0{,}5)^2 + 0{,}5\varepsilon_i$,
- $y_{3i} = 2v_i + \exp(-200(v_i - 0{,}5)^2) + 0{,}5\varepsilon_i$,
- $y_{4i} = \exp(-8v_i) + 0{,}5\varepsilon_i$ ;

(iv) dans chaque population, on tire un échantillon de taille $n = 400$ selon un plan EASSR ;

(v) on génère ensuite dans chaque échantillon la non-réponse des individus de façon indépendante selon une loi Bernoulli de paramètre $p_i$, où

$$p_i = \frac{\exp(-1{,}5 + 5{,}7v_i)}{1 + \exp(-1{,}5 + 5{,}7v_i)}.$$

Dans chaque échantillon, le taux de réponse est d'environ 70%.

Pour l'imputation par le noyau, c'est la fonction noyau $\mathcal{K}(\cdot)$ gaussienne qui a été utilisée et plusieurs paramètres de lissage $h$ sont considérés : $h = 0{,}05$, $h = 0{,}1$ et $h = 0{,}2$.

Dans chaque échantillon, nous avons calculé les estimateurs de $t_y$ suivants : l'estimateur de Horvitz-Thompson (calculé avant la non-réponse, notre valeur de référence), l'estimateur après imputation par la régression linéaire et l'estimateur après imputation par le noyau. Pour mesurer le biais, on calcule le biais relatif Monte-Carlo :

$$BR_{MC}(\widehat{t_y}) = \frac{\mathbb{E}_{MC}(\widehat{t_y}) - t_y}{t_y} \times 100, \tag{1.4.3}$$

où l'espérance Monte-Carlo est telle que $\mathbb{E}_{MC}(\widehat{t_y}) = (1/R)\sum_{k=1}^{R} \widehat{t}_y^{(k)}$, $R$ est le nombre d'itérations (ici, $R = 1000$) et $\widehat{t}_y^{(k)}$ est l'estimateur du total obtenu dans l'échantillon $k$. Finalement,

pour comparer l'efficacité des estimateurs à celui de Horvitz-Thompson, nous calculons l'efficacité relative, $ER$, soit le ratio entre l'erreur quadratique moyenne Monte-Carlo de l'estimateur après imputation et celle de l'estimateur de Horvitz-Thompson :

$$ER = \frac{EQM_{MC}(\widehat{t}_y)}{EQM_{MC}(\widehat{t}_{y,HT})}, \tag{1.4.4}$$

où

$$EQM_{MC}(\widehat{t}_y) = \frac{1}{R}\sum_{k=1}^{R}\left(\widehat{t}_y^{(k)} - t_y\right)^2. \tag{1.4.5}$$

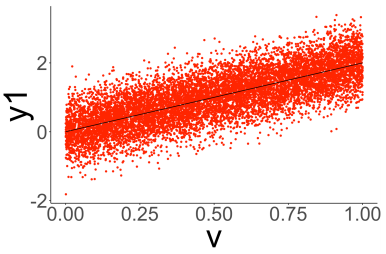
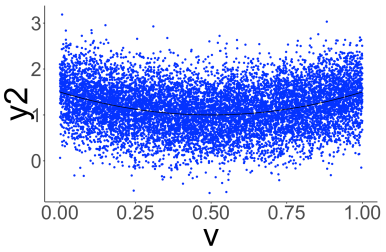
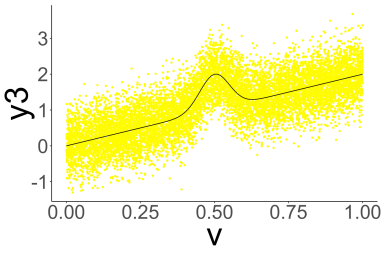
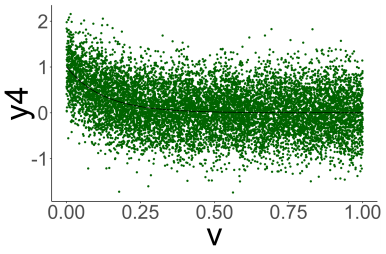
| | Estimateur | $h$ | $BR_{MC}$ (%) | $ER$ | |
|---|---|---|---|---|---|
| $y_1$ | Horvitz-Thompson | | $-0,03$ | 1,00 |  |
| | Régression linéaire | | $-0,01$ | 1,27 | |
| | | 0,05 | $-0,06$ | 1,44 | |
| | Noyau | 0,1 | $-0,00$ | 1,31 | |
| | | 0,2 | $-0,01$ | 1,29 | |
| $y_2$ | Horvitz-Thompson | | 0,04 | 1,00 |  |
| | Régression linéaire | | $-3,13$ | 4,16 | |
| | | 0,05 | 0,21 | 1,88 | |
| | Noyau | 0,1 | 0,34 | 1,89 | |
| | | 0,2 | 0,06 | 1,69 | |
| $y_3$ | Horvitz-Thompson | | $-0,08$ | $1v00$ |  |
| | Régression linéaire | | 3,33 | 2,53 | |
| | | 0,05 | $-0,04$ | 1,34 | |
| | Noyau | 0,1 | 0,08 | 1,35 | |
| | | 0,2 | $-0,71$ | 1,43 | |
| $y_4$ | Horvitz-Thompson | | $-0,30$ | 1,00 |  |
| | Régression linéaire | | $-23,26$ | 3,26 | |
| | | 0,05 | 3,47 | 1,96 | |
| | Noyau | 0,1 | 3,14 | 1,80 | |
| | | 0,2 | $-2,99$ | 1,85 | |

**Tableau 1.1.** Comparaison des estimateurs après imputation par la régression linéaire et imputation par le noyau

Pour le modèle linéaire, correspondant à la variable $y_1$, aucun des estimateurs n'est biaisé ; mais dès que le modèle n'est pas linéaire, ce qui correspond aux variables $y_2$, $y_3$, $y_4$, l'estimateur imputé par la régression linéaire devient biaisé. Par exemple, pour la variable $y_4$, le biais relatif en valeur absolue s'élève à environ 23% alors que l'estimateur non-paramétrique exhibe un biais négligeable. Considérons ensuite l'efficacité relative des estimateurs après imputation. Une valeur de $ER$ proche de 1 signifie que l'estimateur imputé est aussi performant

que l'estimateur Horvitz-Thompson et plus $ER$ est petite, plus l'estimateur est efficace en termes de $EQM$. Pour le modèle linéaire, c'est l'estimateur après imputation par la régression linéaire qui est le plus efficace avec une valeur de $ER$ aussi basse que $1,27$. Au contraire, quand le modèle n'est pas linéaire, c'est toujours l'estimateur imputé par la régression linéaire qui est le moins efficace. En effet, les valeurs de $ER$ pour l'estimateur imputé par le noyau sont toujours inférieures à 2 tandis que celles de l'estimateur imputé linéairement sont toujours supérieures, s'élevant jusqu'à 4 dans le cas de la variable $y_2$.

1.4.2.2. *Imputation par donneur : imputation par le plus proche voisin*

Un autre exemple d'imputation non-paramétrique est celle du plus proche voisin. Dans ce cas, la variable $y_i$ manquante est remplacée par $y_j$, la valeur de l'individu $j$, où $\mathbf{v}_j$ correspond au « plus proche » vecteur de variables auxiliaires de $\mathbf{v}_i$ parmi les cas complets. Pour ce faire, on introduit une notion de distance $D(\cdot,\cdot)$ et on aura, pour $\mathbf{v}$ est un vecteur de variables auxiliaires de dimension $G$ :

$$y_i^* = y_j \quad \text{si} \quad D(\mathbf{v}_j, \mathbf{v}_i) \leq D(\mathbf{v}_k, \mathbf{v}_i), \text{ pour tout } k \in S_r.$$

On peut considérer une fonction de distance générale

$$D(\mathbf{v}_j, \mathbf{v}_i) = \left( \sum_{k=1}^{G} a_k |v_{kj} - v_{ki}|^b \right)^{1/b},$$

qui dépend du poids $a_k$ de la variable $v_k$ et de $b \geq 1$. Par exemple, si $b = 1$, la distance $D$ correspond à la norme $L_1$ et si $b = 2$, la distance $D$ correspond à la norme $L_2$.

L'imputation par le plus proche voisin est une méthode d'imputation par donneur, au sens où les valeurs manquantes sont remplacées par des observations d'autres répondants. Deux avantages de ce type d'imputation sont à souligner. D'abord, lorsque la variable d'intérêt est une variable catégorielle, les valeurs imputées seront aussi catégorielles et plausibles, puisque observées parmi les cas complets. Ensuite, lorsque pour une seule unité plusieurs variables sont manquantes, il est aisé d'imputer toutes ces valeurs d'un coup avec un unique donneur.

# Chapitre 2

---

# Imputation multi-robuste efficace en présence de valeurs influentes

## 2.1. Introduction

Le risque d'une imputation fondée sur un unique modèle d'imputation est que ce dernier soit mal spécifié. Dans un tel cas, les valeurs imputées peuvent conduire à un estimateur imputé biaisé. Pour pallier cela, il est possible d'utiliser une imputation multi-robuste, où une valeur imputée est construite au moyen de plusieurs modèles d'imputation et/ou modèles de non-réponse. L'estimateur imputé est dit multi-robuste s'il est convergent lorsque tous les modèles utilisés sont mal-spécifiés, sauf un, voir Han et Wang (2013), Chan et Yam (2014), Han (2014a et 2014b) et Chen et Haziza (2017 et 2019). Les méthodes d'imputation multi-robuste sont intéressantes puisqu'elles protègent contre la mauvaise spécification des modèles. La multi-robustesse peut être vue comme une extension de la double robustesse où seuls un modèle d'imputation et un modèle de non-réponse sont définis ; voir, par exemple, Haziza et Rao (2006) et Kim et Haziza (2014).

Nous discutons de certaines situations pour lesquelles l'imputation multi-robuste peut s'avérer intéressante.

(i) Supposons que l'on souhaite estimer le total de la variable d'intérêt $y$ qui est sujette à la non-réponse et que l'on observe complètement un vecteur de variables auxiliaires $\mathbf{v}$. Postuler un modèle concernant la relation entre $y$ et $\mathbf{v}$ qui l'emporterait sur tous les autres modèles possibles peut ne pas être facile. En effet, même avec les plus récentes méthodes de sélection de modèle, telles que le LASSO, plusieurs valeurs du paramètre

de régularisation peuvent mener à différents modèles sélectionnés, d'où l'avantage de l'imputation multi-robuste qui prend en compte de multiples modèles.

(ii) Dans un contexte d'échantillonnage, nous pourrions envisager certains modèles tenant compte des caractéristiques du plan de sondage (poids de sondage, stratification, ...) et d'autres, non.

(iii) Supposons que la variable d'intérêt soit binaire. Lorsqu'elle est sujette à la non-réponse, habituellement, les valeurs manquantes sont imputées par une régression logistique. Cependant, si la fonction de lien n'est pas correctement spécifiée, l'estimateur imputé peut être biaisé. Une imputation non-paramétrique (sous-section 1.4.2) peut alors être envisagée, mais si les variables auxiliaires sont nombreuses, on se trouve face au « fléau de la dimensionalité » où l'estimation des paramètres devient difficile, voire impossible.

(iv) Il est possible qu'aucun des modèles définis ne soit correctement spécifié, mais Han (2014b) et Chen et Haziza (2017) montrent qu'avec des méthodes d'imputation multi-robuste, les estimateurs, même s'ils ne sont pas convergents, tendent à être performants numériquement. Cela n'est pas toujours le cas lors d'une imputation fondée sur un seul modèle d'imputation.

Comme l'estimateur de Horvitz-Thompson, un estimateur imputé selon un modèle d'imputation simple ou selon une méthode d'imputation multi-robuste est sensible à la présence de valeurs influentes dans la population et est volatile. L'objectif de ce chapitre est de créer un estimateur multi-robuste efficace en présence de valeurs influentes. Dans le chapitre 1, nous avons défini le biais conditionnel (1.2.1) d'une unité échantillonnée afin d'en quantifier l'influence sur l'estimateur de Horvitz-Thompson (Beaumont et al., 2013). Nous définirons ici le biais conditionnel d'une unité échantillonnée pour un estimateur imputé simple, d'abord, puis multi-robuste.

Dongmo Jiongo (2015) a développé des estimateurs efficaces en présence d'unités influentes dans le cadre d'une imputation doublement robuste. Comme celui qui sera proposé dans la section 2.4.1, Dongmo Jiongo (2015) a proposé un estimateur basé sur le biais conditionnel. Il a utilisé deux définitions du biais conditionnel : l'une pour laquelle l'espérance est évaluée par rapport à la distribution conjointe induite par le modèle d'imputation, le plan de sondage et le mécanisme de non-réponse et l'autre pour laquelle l'espérance est évaluée

par rapport à la distribution conjointe induite par le plan de sondage et le mécanisme de non-réponse. Dans ce chapitre, nous proposons des procédures d'imputation qui incluent les procédures d'imputation doublement robustes comme cas particulier. De plus, nous utiliserons une définition différente du biais conditionnel de celle utilisée par Dongmo Jiongo (2015). Dans notre cas, le biais conditionnel est évalué par rapport au plan de sondage seulement.

## 2.2. L'imputation multi-robuste

Nous travaillons avec l'imputation multi-robuste telle que présentée par Chen et Haziza (2019). Nous commençons par définir les deux classes de modèles qui sont utilisés pour construire une valeur imputée multi-robuste :

(i) la classe $\mathcal{C}_1$ des $J$ modèles de non-réponse envisagés pour estimer la probabilité de réponse $p_i$ telle que

$$\mathcal{C}_1 = \{p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j) : j = 1, \ldots, J\},$$

où $p^j(\cdot, \boldsymbol{\alpha}^j)$ est la $j$-ème fonction définie pour décrire le mécanisme de non-réponse, dépendante de $\boldsymbol{\alpha}^j$, un vecteur de paramètres inconnus, et où $\mathbf{v}_i^j$ est le vecteur de variables auxiliaires, associées au $j$-ème modèle de non-réponse, entièrement observé pour tout $i \in S$ ;

(ii) la classe $\mathcal{C}_2$ des $L$ modèles d'imputation spécifiés pour prédire $y_i$ telle que

$$\mathcal{C}_2 = \{m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell) : \ell = 1, \ldots, L\},$$

où $m^\ell(\cdot, \boldsymbol{\beta}^\ell)$ est la $\ell$-ème fonction définie pour décrire la relation entre la variable d'intérêt $y$ et les variables auxiliaires, dépendante du vecteur de paramètres inconnus $\boldsymbol{\beta}^\ell$ et où $\mathbf{v}_i^\ell$ est le vecteur de variables auxiliaires associées au $\ell$-ème modèle d'imputation.

En tout, nous avons $J + L$ modèles spécifiés pour construire une valeur imputée multi-robuste. L'imputation est donc dite multi-robuste lorsque l'estimateur imputé est convergent si au moins l'un des $J + L$ modèles est correctement spécifié. Lorsque $J = 0$ et $L = 1$, nous retrouvons le cas d'une imputation fondée sur un unique modèle d'imputation (voir chapitre 1.4). La double robustesse est un autre cas particulier de la multi-robustesse, avec un modèle de non-réponse, $J = 1$, et un modèle d'imputation, $L = 1$.

Dans un contexte d'imputation multi-robuste, trois étapes sont nécessaires à la création d'une valeur imputée $y_i^*,\ i \in S_{nr}$ :

(1) On commence par ajuster les $J$ modèles de non-réponse $p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j),\ j = 1, \ldots, J$, ce qui conduit aux estimateurs $\widehat{\boldsymbol{\alpha}}^j,\ j = 1, \ldots, J$, qui peuvent être obtenus comme solution des équations estimantes :

$$S_{\boldsymbol{\alpha}}^j(\boldsymbol{\alpha}^j) = \sum_{i \in S} \phi_i \frac{r_i - p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)}{p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)\{1 - p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)\}} \frac{\partial p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)}{\partial \boldsymbol{\alpha}^j} = \mathbf{0},$$

où $\phi_i$ est un coefficient propre à l'individu $i$. Rappelons que ce coefficient est souvent égal à 1, pour une imputation non-pondérée, ou au poids $w_i$, pour une imputation pondérée. De même, on ajuste les $L$ modèles d'imputation $m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell),\ \ell = 1, \ldots, L$ sur les répondants pour estimer les paramètres $\boldsymbol{\beta}^\ell,\ \ell = 1, \ldots, L$ en résolvant les équations estimantes suivantes :

$$S_{\boldsymbol{\beta}}^\ell(\boldsymbol{\beta}^\ell) = \sum_{i \in S_r} \phi_i \{y_i - m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell)\} \frac{\partial m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell)}{\partial \boldsymbol{\beta}^\ell} = \mathbf{0}.$$

(2) Pour tout $i \in S$ (répondants et non-répondants), deux vecteurs de tailles $J$ et $L$ sont ainsi créés :

$$\widehat{\mathbf{U}}_{pi} = \left(p^1(\mathbf{v}_i^1, \widehat{\boldsymbol{\alpha}}^1), \ldots, p^J(\mathbf{v}_i^J, \widehat{\boldsymbol{\alpha}}^J)\right)^\top \text{ et } \widehat{\mathbf{U}}_{mi} = \left(m^1(\mathbf{v}_i^1, \widehat{\boldsymbol{\beta}}^1), \ldots, m^L(\mathbf{v}_i^L, \widehat{\boldsymbol{\beta}}^L)\right)^\top.$$

La deuxième étape consiste à comprimer chacun des vecteurs $\widehat{U}_{pi}$ et $\widehat{U}_{mi}$ en un seul score, soit résumer l'information des $J$ probabilités de réponse estimées en un seul score $\widehat{p}_i$ pour tout $i \in S$ et résumer l'information contenue dans les $L$ valeurs prédites en une seule valeur $\widehat{m}_i$ pour tout $i \in S$.

Afin de condenser l'information contenue dans le vecteur $\widehat{\mathbf{U}}_{pi}$, on ajuste un modèle de régression linéaire pondérée entre l'indicatrice de réponse $r$ comme variable dépendante et le vecteur des probabilités estimées $\widehat{\mathbf{U}}_{pi}$ comme prédicteur. On obtient alors le paramètre estimé $\widehat{\boldsymbol{\eta}}_p$ de dimension $J$ de ce modèle linéaire :

$$\widehat{\boldsymbol{\eta}}_p = \left(\sum_{i \in S} w_i \widehat{\mathbf{U}}_{pi} \widehat{\mathbf{U}}_{pi}^\top\right)^{-1} \sum_{i \in S} w_i \widehat{\mathbf{U}}_{pi} r_i.$$

De même, pour condenser l'information contenue dans le vecteur $\widehat{\mathbf{U}}_{mi}$, on ajuste un deuxième modèle de régression linéaire pondérée, au moyen des unités répondantes, entre les $y_i$ comme variable dépendante et le vecteur des valeurs prédites, $\widehat{\mathbf{U}}_{mi}$, comme

prédicteur. On obtient le paramètre estimé $\widehat{\boldsymbol{\eta}}_m$ de dimension $L$ de ce modèle linéaire :

$$\widehat{\boldsymbol{\eta}}_m = \left( \sum_{i \in S_r} w_i \widehat{\mathbf{U}}_{mi} \widehat{\mathbf{U}}_{mi}^\top \right)^{-1} \sum_{i \in S_r} w_i \widehat{\mathbf{U}}_{mi} y_i.$$

Pour chaque $i \in S$, on effectue une prédiction pour chacun des deux modèles, ce qui conduit aux scores standardisés :

$$\widehat{p}_i = \widehat{\mathbf{U}}_{pi}^\top \frac{\widehat{\boldsymbol{\eta}}_p^2}{\widehat{\boldsymbol{\eta}}_p^\top \widehat{\boldsymbol{\eta}}_p} \quad \text{et} \quad \widehat{m}_i = \widehat{\mathbf{U}}_{mi}^\top \frac{\widehat{\boldsymbol{\eta}}_m^2}{\widehat{\boldsymbol{\eta}}_m^\top \widehat{\boldsymbol{\eta}}_m}.$$

Ici, si $\mathbf{a} = (a_1, \ldots, a_h)^\top$ est un vecteur de dimension $h$, $\mathbf{a}^2$ désigne le vecteur $(a_1^2, \ldots, a_h^2)^\top$.

(3) Enfin, on construit la valeur imputée multi-robuste $y_i^*$ en exécutant, sur l'ensemble des répondants, une troisième régression linéaire pondérée entre la variable $y$ comme variable dépendante et le score $\widehat{m}_i$ comme prédicteur. Les poids de cette régression linéaire sont donnés par $w_i \left( \widehat{p}_i^{-1} - 1 \right)$, $i \in S_r$. On aboutit aux valeurs imputées :

$$y_i^* = \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}, \ i \in S_{nr},$$

où $\mathbf{h}_i = (1, \widehat{m}_i)^\top$ et

$$\widehat{\boldsymbol{\tau}} = \left( \sum_{i \in S_r} w_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \mathbf{h}_i \mathbf{h}_i^\top \right)^{-1} \sum_{i \in S_r} w_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \mathbf{h}_i y_i.$$

Soit $\widehat{t}_{MR}$ l'estimateur du total $t_y$ après imputation multi-robuste. Il est défini par

$$\widehat{t}_{MR} = \sum_{i \in S} w_i r_i y_i + \sum_{i \in S} w_i \left(1 - r_i\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}. \tag{2.2.1}$$

Les théorèmes 2.2.1 et 2.2.2 suivants établissent la caractéristique multi-robuste de $\widehat{t}_{MR}$.

**Théorème 2.2.1.** *Si l'un des modèles d'imputation est correctement spécifié ou que le vrai modèle est une combinaison linéaire de plusieurs modèles d'imputation, alors $\widehat{t}_{MR}/t_y$ converge en probabilité vers 1 quand $n$ et $N$ tendent vers l'infini.*

Démonstration. Voir Chen et Haziza (2017). □

**Théorème 2.2.2.** *Si l'un des modèles de non-réponse est correctement spécifié, alors $\widehat{t}_{MR}/t_y$ converge en probabilité vers 1 quand $n$ et $N$ tendent vers l'infini.*

Démonstration. Voir Chen et Haziza (2017). □

**Proposition 2.2.3.** *L'estimateur $\widehat{t}_{MR}$ peut s'écrire comme*

$$\widehat{t}_{MR} = \sum_{i\in S} w_i \frac{r_i}{\widehat{p}_i} y_i + \sum_{i\in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \boldsymbol{h}_i^\top \widehat{\boldsymbol{\tau}}.$$

Démonstration On a

$$\begin{aligned}
\widehat{t}_{MR} &= \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i \left(1 - r_i\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}} \\
&= \sum_{i\in S} w_i \left(r_i + \frac{r_i}{\widehat{p}_i} - \frac{r_i}{\widehat{p}_i}\right) y_i + \sum_{i\in S} w_i \left(1 - r_i + \frac{r_i}{\widehat{p}_i} - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}} \\
&= \sum_{i\in S} w_i \frac{r_i}{\widehat{p}_i} y_i + \sum_{i\in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}} + \sum_{i\in S} w_i \left(r_i - \frac{r_i}{\widehat{p}_i}\right) (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) \\
&= \sum_{i\in S} w_i \frac{r_i}{\widehat{p}_i} y_i + \sum_{i\in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}},
\end{aligned}$$

en remarquant que l'équation estimante

$$\sum_{i\in S_r} w_i \left(\widehat{p}_i^{-1} - 1\right) (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) = 0.$$

Bien que l'estimateur multi-robuste nous protège contre une éventuelle mauvaise spécification du modèle d'imputation, il peut s'avérer inefficace en présence de valeurs influentes.

## 2.3. Estimation efficace en présence de valeurs influentes : Imputation basée sur un seul modèle d'imputation

On commencera par traiter le cas d'une imputation basée sur un modèle de régression linéaire. L'objectif de cette section est de créer une version efficace de l'estimateur après imputation simple, $\widehat{t}_I$, donné en (1.4.1). Dans ce cas, on a

$$y_i^* = \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}_r,\ i \in S_{nr},$$

où $\widehat{\boldsymbol{\beta}}_r = \widehat{\mathbf{T}}_r^{-1} \widehat{\mathbf{t}}_r$, avec $\widehat{\mathbf{T}}_r = \sum_{i\in S} w_i r_i c_i^{-1} \mathbf{v}_i \mathbf{v}_i^\top$ et $\widehat{\mathbf{t}}_r = \sum_{i\in S} w_i r_i c_i^{-1} \mathbf{v}_i y_i$. L'estimateur du total après imputation est donc donné par

$$\widehat{t}_I = \sum_{i\in S} w_i \{r_i y_i + (1 - r_i) \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}_r\}.$$

Soit $B_{1i}^I$, le biais conditionnel d'une unité échantillonnée par rapport à $\widehat{t}_I$. Il est défini par

$$B_{1i}^I = \mathbb{E}_p \left(\widehat{t}_I - t_y \mid I_i = 1\right). \tag{2.3.1}$$

Il est virtuellement impossible de calculer ce biais conditionnel car $\widehat{t}_I$ est une fonction complexe de totaux estimés. Nous utilisons alors une série de Taylor pour linéariser $\widehat{t}_I$ et l'approximer par un estimateur de la forme

$$\widehat{t}_I \simeq \sum_{j \in S} w_j \psi_j,$$

où $\psi_j$ est la variable linéarisée pour l'unité $j$. Nous détaillons maintenant comment nous obtenons $\psi_j$. Désignons par $\boldsymbol{\beta}^\bullet$ la limite probabiliste de $\widehat{\boldsymbol{\beta}}_r$. On peut écrire :

$$\begin{aligned}
\widehat{t}_I &= \sum_{j \in S} w_j \{ r_j y_j + (1 - r_j) \mathbf{v}_j^\top \widehat{\boldsymbol{\beta}}_r \} \\
&= \sum_{j \in S} w_j \{ r_j y_j + (1 - r_j) \mathbf{v}_j^\top (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet + \boldsymbol{\beta}^\bullet) \} \\
&= \sum_{j \in S} w_j \{ r_j y_j + (1 - r_j) \mathbf{v}_j^\top \boldsymbol{\beta}^\bullet \} + \sum_{j \in S} w_j (1 - r_j) \mathbf{v}_j^\top (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet) \\
&= \sum_{j \in S} w_j \{ r_j y_j + (1 - r_j) \mathbf{v}_j^\top \boldsymbol{\beta}^\bullet \} + (\widehat{\mathbf{t}}_{\mathbf{v},HT} - \widehat{\mathbf{t}}_{\mathbf{v}_r})^\top (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet), \qquad (2.3.2)
\end{aligned}$$

où $\widehat{\mathbf{t}}_{\mathbf{v},HT} = \sum_{i \in S} w_i \mathbf{v}_i$ et $\widehat{\mathbf{t}}_{\mathbf{v}_r} = \sum_{i \in S} w_i r_i \mathbf{v}_i$. Pour obtenir une forme linéarisée de $\widehat{t}_I$, nous approximons par série de Taylor la différence $\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet$. Soit $\widehat{S}_{\boldsymbol{\beta}}$ l'équation estimante à résoudre pour estimer $\boldsymbol{\beta}$ dans le cadre d'une régression linéaire. On a

$$\widehat{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i \in S_r} w_i c_i^{-1} (y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \mathbf{v}_i = \mathbf{0}.$$

L'approximation par série de Taylor du premier ordre est alors telle que :

$$\widehat{S}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_r) = \widehat{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^\bullet) + \mathbb{E}\left( \frac{\partial \widehat{S}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta}^\bullet) \right) (\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet) + o_P(n^{-1/2}),$$

ou encore

$$\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet = \mathbb{E}^{-1}\left( \frac{\partial \widehat{S}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta}^\bullet) \right) \left( \widehat{S}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_r) - \widehat{S}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^\bullet) \right) + o_P(n^{-1/2}).$$

Or, nous avons par définition $\widehat{S}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_r) = \mathbf{0}$. Ainsi, l'expression linéarisée de $\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet$, en ignorant les termes d'ordre supérieur, est de la forme

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet &\simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{S}_{\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^\bullet)\right)\widehat{S}^\bullet_{\boldsymbol{\beta}} \\
&= -\mathbb{E}^{-1}\left(-\sum_{j\in S} w_j \frac{r_j}{c_j}\mathbf{v}_j\mathbf{v}_j^\top\right)\sum_{j\in S} w_j \frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j \\
&= \mathbb{E}^{-1}\left(\sum_{j\in S} w_j \frac{r_j}{c_j}\mathbf{v}_j\mathbf{v}_j^\top\right)\sum_{j\in S} w_j \frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j \\
&= \left(\sum_{j\in S} w_j \frac{r_j}{c_j}\mathbf{v}_j\mathbf{v}_j^\top\right)^{-1}\sum_{j\in S} w_j \frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j \\
&= \widehat{\mathbf{T}}_r^{-1}\sum_{j\in S} w_j \frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j.
\end{aligned}$$

Finalement, en remplaçant la linéarisée de $\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^\bullet$ dans (2.3.2) et en notant que

$$(\widehat{\mathbf{t}}_{\mathbf{v},HT} - \widehat{\mathbf{t}}_{\mathbf{v}_r})^\top\widehat{\mathbf{T}}_r^{-1} = (\mathbf{t}_{\mathbf{v}} - \mathbf{t}_{\mathbf{v}_r})^\top\mathbf{T}_r^{-1} + o_P(N/\sqrt{n}),$$

où $\mathbf{t}_{\mathbf{v}} = \sum_{i\in U}\mathbf{v}_i$, $\mathbf{t}_{\mathbf{v}_r} = \sum_{i\in U} r_i\mathbf{v}_i$ et $\mathbf{T}_r = \sum_{i\in U} r_i c_i^{-1}\mathbf{v}_i\mathbf{v}_i^\top$, on obtient

$$\begin{aligned}
\widehat{t}_I &\simeq \sum_{j\in S} w_j\{r_j y_j + (1-r_j)\mathbf{v}_j^\top\boldsymbol{\beta}^\bullet\} + (\widehat{\mathbf{t}}_{\mathbf{v},HT} - \widehat{\mathbf{t}}_{\mathbf{v}_r})^\top\widehat{\mathbf{T}}_r^{-1}\sum_{j\in S} w_j\frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j \\
&\simeq \sum_{j\in S} w_j\{r_j y_j + (1-r_j)\mathbf{v}_j^\top\boldsymbol{\beta}^\bullet\} + (\mathbf{t}_{\mathbf{v}} - \mathbf{t}_{\mathbf{v}_r})^\top\mathbf{T}_r^{-1}\sum_{j\in S} w_j\frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j \\
&= \sum_{j\in S} w_j\{r_j y_j + (1-r_j)\mathbf{v}_j^\top\boldsymbol{\beta}^\bullet + (\mathbf{t}_{\mathbf{v}} - \mathbf{t}_{\mathbf{v}_r})^\top\mathbf{T}_r^{-1}\frac{r_j}{c_j}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j\} \\
&= \sum_{j\in S} w_j\psi_j,
\end{aligned}$$

avec

$$\psi_j = r_j y_j + (1-r_j)\mathbf{v}_j^\top\boldsymbol{\beta}^\bullet + (\mathbf{t}_{\mathbf{v}} - \mathbf{t}_{\mathbf{v}_r})^\top\mathbf{T}_r^{-1}r_j c_j^{-1}(y_j - \mathbf{v}_j^\top\boldsymbol{\beta}^\bullet)\mathbf{v}_j.$$

Nous obtenons finalement l'approximation du biais conditionnel suivante :

$$\begin{aligned} B_{1i}^I &= \mathbb{E}_p\left(\widehat{t}_I - t_y \mid I_i = 1\right) \\ &\simeq \mathbb{E}_p\left(\sum_{j\in S} w_j\psi_j + \sum_{j\in U}\psi_j - \sum_{j\in U}\psi_j - t_y \mid I_i = 1\right) \\ &= \mathbb{E}_p\left(\sum_{j\in S} w_j\psi_j - \sum_{j\in U}\psi_j \mid I_i = 1\right) + \sum_{j\in U}\psi_j - t_y \\ &= \sum_{j\in U}\frac{\Delta_{ij}}{\pi_i\pi_j}\psi_j + \sum_{j\in U}\psi_j - t_y. \end{aligned} \tag{2.3.3}$$

Puisque $\sum_{j\in U}\psi_j - t_y$ ne dépend pas de $i$, cet ajout est le même pour tous les biais conditionnels des unités échantillonnées.

Un estimateur de $B_{1i}^I$ en (2.3.3) est donné par

$$\widehat{B}_{1i}^I = \sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}\pi_j}\widehat{\psi}_j, \tag{2.3.4}$$

où

$$\widehat{\psi}_j = r_j y_j + (1-r_j)\mathbf{v}_j^\top\widehat{\boldsymbol{\beta}}_r + (\widehat{\mathbf{t}}_{\mathbf{v},HT} - \widehat{\mathbf{t}}_{\mathbf{v}_r})^\top\widehat{\mathbf{T}}_r^{-1} r_j c_j^{-1}(y_j - \mathbf{v}_j^\top\widehat{\boldsymbol{\beta}}_r)\mathbf{v}_j.$$

Notons que l'estimateur $\widehat{B}_{1i}^I$ en (2.3.4) ne tient pas compte du terme $\sum_{j\in U}\psi_j - t_y$ en (2.3.3) car un estimateur de $t_y$ est donné par $\widehat{t}_I$ alors qu'un estimateur de $\sum_{j\in U}\psi_j$ est donné par $\sum_{j\in S} w_j\psi_j$. Or, $\widehat{t}_I \simeq \sum_{j\in S} w_j\psi_j$ et donc $\widehat{t}_I - \sum_{j\in S} w_j\psi_j \simeq 0$.

Soit $\widehat{t}_I^R$ l'estimateur du total après imputation efficace en présence de données influentes. Le terme $\widehat{t}_I^R$ est obtenu par la même méthode que celle présentée dans la section 1.2.2, soit

$$\widehat{t}_I^R = \widehat{t}_I - \frac{\widehat{B}_{\min}^I + \widehat{B}_{\max}^I}{2}, \tag{2.3.5}$$

avec $\widehat{B}_{\min}^I$ et $\widehat{B}_{\max}^I$ les extrema de $\widehat{B}_{1i}^I$ en (2.3.4).

### 2.3.1. Décomposition du biais conditionnel

Nous pouvons réécrire $B_{1i}^I$ en (2.3.3) comme

$$B_{1i}^I = \sum_{j\in U}\frac{\Delta_{ij}}{\pi_i\pi_j}y_j + \sum_{j\in U}\frac{\Delta_{ij}}{\pi_i\pi_j}z_j + \left(\sum_{j\in U}\psi_j - t_y\right), \tag{2.3.6}$$

où

$$z_j = \left[r_j \left\{1 + (\mathbf{t}_{\mathbf{v}} - \mathbf{t}_{\mathbf{v}_r})^\top \mathbf{T}_r^{-1} c_j^{-1} \mathbf{v}_j\right\} - 1\right] (y_j - \mathbf{v}_j^\top \boldsymbol{\beta}^\bullet).$$

Dans le cas d'un plan d'échantillonage EASSR d'une population de taille $N$ et d'un échantillon de taille $n$, en ignorant le terme $\sum_{j\in U} \psi_j - t_y$, le biais conditionnel en (2.3.6) peut être écrit comme

$$B_{1i}^I = \left(\frac{N}{n} - 1\right) \frac{N}{N-1} \left\{(y_i - \overline{Y}) + (z_i - \overline{Z})\right\},$$

où $\overline{Z} = N^{-1} \sum_{j\in U} z_j$. Nous étudions la forme de ce biais conditionnel dans le cas de l'imputation par le ratio (voir section 1.4.1.1), où

$$y_i^* = \frac{\overline{y}_r}{\overline{v}_r} v_i, \ i \in S_{nr}.$$

Dans ce cas, on a

$$z_i = \left(\frac{r_i}{P_r} \frac{\overline{V}}{\overline{V}_r} - 1\right) (y_i - B_r v_i), \tag{2.3.7}$$

où $\overline{V} = N^{-1} \sum_{j\in U} v_j$, $B_r = \overline{Y}_r / \overline{V}_r$ et $P_r = N_r/N$ avec $N_r = \sum_{j\in U} r_j$, $\overline{Y}_r = N_r^{-1} \sum_{j\in U} r_j y_j$ et $\overline{V}_r = N_r^{-1} \sum_{j\in U} r_j v_j$.

Nous commençons par noter que lorsque $r_i = 1$ pour tout $i$ (cas de réponse complète), on a $z_i = 0$ pour tout $i$ et le biais conditionnel $B_{1i}^I$ se simplifie pour donner

$$B_{1i}^I = \left(\frac{N}{n} - 1\right) \frac{N}{N-1} (y_i - \overline{Y}).$$

Si l'unité $i$ est répondante ($r_i = 1$), alors $z_i$ aura tendance à être grande si le résidu $y_i - B_r v_i$ est grand et que le taux de réponse $P_r$ est petit. Pour une unité non-répondante ($r_i = 0$), $z_i = -(y_i - B_r v_i)$ et son influence est grande si son résidu est grand.

## 2.4. Estimation efficace en présence de valeurs influentes : Imputation multi-robuste

Dans la section précédente, l'estimateur après imputation a été rendu efficace en présence de valeurs influentes. Dans cette section, nous appliquons à nouveau la même méthode à l'estimateur après imputation multi-robuste, $\widehat{t}_{MR}$ en (2.2.1). Nous considérons la forme suivante de $\widehat{t}_{MR}$ (voir proposition 2.2.3) :

$$\widehat{t}_{MR} = \sum_{i\in S} w_i \frac{r_i}{\widehat{p}_i} y_i + \sum_{i\in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}.$$

Soit $B_{1i}^{MR}$, le biais conditionnel d'une unité échantillonnée par rapport à $\widehat{t}_{MR}$. Ce biais conditionnel est défini par

$$B_{1i}^{MR} = \mathbb{E}_p\left(\widehat{t}_{MR} - t_y \mid I_i = 1\right). \tag{2.4.1}$$

Comme plus haut, cette espérance est virtuellement impossible à calculer et nous utilisons à nouveau une série de Taylor pour linéariser $\widehat{t}_{MR}$ et l'approximer par un estimateur de la forme

$$\widehat{t}_{MR} \simeq \sum_{j\in S} w_j\psi_j,$$

où $\psi_j$ est la variable linéarisée pour l'unité $j$. Nous détaillons maintenant comment nous obtenons $\psi_j$. Rappelons que $\widehat{t}_{MR}$ dépend des $J+L+3$ paramètres estimés $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\alpha}}^1, \ldots, \widehat{\boldsymbol{\alpha}}^J)^\top$, $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}^1, \ldots, \widehat{\boldsymbol{\beta}}^L)^\top$, $\widehat{\boldsymbol{\eta}}_p$, $\widehat{\boldsymbol{\eta}}_m$ et $\widehat{\boldsymbol{\tau}}$ obtenus au moyen des équations estimantes suivantes :

$$\begin{aligned}
\widehat{S}^j_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^j) &= \sum_{i\in S} w_i \frac{r_i - p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)}{p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)\{1 - p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)\}} \frac{\partial p^j(\mathbf{v}_i^j, \boldsymbol{\alpha}^j)}{\partial \boldsymbol{\alpha}^j} = \mathbf{0},\ j = 1, \ldots, J;\\
\widehat{S}^\ell_{\boldsymbol{\beta}}(\boldsymbol{\beta}^\ell) &= \sum_{i\in S_r} w_i\{y_i - m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell)\} \frac{\partial m^\ell(\mathbf{v}_i^\ell, \boldsymbol{\beta}^\ell)}{\partial \boldsymbol{\beta}^\ell} = \mathbf{0},\ \ell = 1, \ldots, L;\\
\widehat{U}_p(\boldsymbol{\alpha}, \boldsymbol{\eta}_p) &= \sum_{i\in S} w_i(r_i - \mathbf{U}_{p_i}^\top \boldsymbol{\eta}_p)\mathbf{U}_{p_i} = \mathbf{0};\\
\widehat{U}_m(\boldsymbol{\beta}, \boldsymbol{\eta}_m) &= \sum_{i\in S_r} w_i(y_i - \mathbf{U}_{m_i}^\top \boldsymbol{\eta}_m)\mathbf{U}_{m_i} = \mathbf{0};\\
\widehat{U}_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}_p, \boldsymbol{\eta}_m, \boldsymbol{\tau}) &= \sum_{i\in S_r} w_i \frac{1 - \widehat{p}_i}{\widehat{p}_i}(y_i - \mathbf{h}_i^\top \boldsymbol{\tau})\mathbf{h}_i = \mathbf{0}.
\end{aligned}$$

Désignons par $\boldsymbol{\alpha}^\bullet$, $\boldsymbol{\beta}^\bullet$, $\boldsymbol{\eta}_p^\bullet$, $\boldsymbol{\eta}_m^\bullet$ et $\boldsymbol{\tau}^\bullet$ les limites probabilistes respectives de $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\eta}}_p$, $\widehat{\boldsymbol{\eta}}_m$ et $\widehat{\boldsymbol{\tau}}$ et notons $\mathbf{h}_i^{\bullet\top} = (1, m_i^\bullet)^\top$. Puisque $\widehat{t}_{MR}$ est une fonction des $J + L + 3$ paramètres estimés, une première série de Taylor nous donne :

$$
\begin{aligned}
\widehat{t}_{MR}(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\eta}}_p,\widehat{\boldsymbol{\eta}}_m,\widehat{\boldsymbol{\tau}}) &= \sum_{i\in S} w_i \frac{r_i}{\widehat{p}_i} y_i + \sum_{i\in S} w_i \left(1-\frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}} \\
&= \sum_{i\in S} w_i \frac{r_i}{p_i^\bullet} y_i + \sum_{i\in S} w_i \left(1-\frac{r_i}{p_i^\bullet}\right) \mathbf{h}_i^{\bullet\top} \boldsymbol{\tau}^\bullet \\
&\quad + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\alpha}}\right)(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\bullet) \\
&\quad + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\eta}_p}\right)(\widehat{\boldsymbol{\eta}}_p-\boldsymbol{\eta}_p^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m-\boldsymbol{\eta}_m^\bullet) \\
&\quad + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\tau}}\right)(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}^\bullet) + o_P\left(\frac{N}{\sqrt{n}}\right), \qquad (2.4.2)
\end{aligned}
$$

où $\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\theta}} \equiv \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\theta}}(\boldsymbol{\alpha}^\bullet,\boldsymbol{\beta}^\bullet,\boldsymbol{\eta}_p^\bullet,\boldsymbol{\eta}_m^\bullet,\boldsymbol{\tau}^\bullet)$ et $\boldsymbol{\theta}$ désigne l'un des $J+L+3$ paramètres. Soient $\widehat{\mathbf{S}}_{\boldsymbol{\alpha}} = (\widehat{S}^1_{\boldsymbol{\alpha}},\ldots,\widehat{S}^J_{\boldsymbol{\alpha}})^\top$ et $\widehat{\mathbf{S}}_{\boldsymbol{\beta}} = (\widehat{S}^1_{\boldsymbol{\beta}},\ldots,\widehat{S}^J_{\boldsymbol{\beta}})^\top$. Pour alléger, nous notons $\widehat{\mathbf{f}}_{\boldsymbol{\theta}}^\bullet \equiv \widehat{\mathbf{f}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\bullet)$ et $\partial\widehat{\mathbf{f}}_{\boldsymbol{\theta}}^\bullet/\partial\boldsymbol{\theta} \equiv \partial\widehat{\mathbf{f}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\bullet)/\partial\boldsymbol{\theta}$, où $\widehat{\mathbf{f}}_{\boldsymbol{\theta}}$ désigne une équation estimante. En ignorant les termes d'ordre supérieur, les linéarisées sont de la forme :

$$
\begin{aligned}
&(i) \qquad \widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{\mathbf{S}}_{\boldsymbol{\alpha}}^\bullet}{\partial \boldsymbol{\alpha}}\right)\widehat{\mathbf{S}}_{\boldsymbol{\alpha}}^\bullet; \\
&(ii) \qquad \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{\mathbf{S}}_{\boldsymbol{\beta}}^\bullet}{\partial \boldsymbol{\beta}}\right)\widehat{\mathbf{S}}_{\boldsymbol{\beta}}^\bullet; \\
&(iii) \qquad \widehat{\boldsymbol{\eta}}_p-\boldsymbol{\eta}_p^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_p^\bullet}{\partial \boldsymbol{\eta}_p}\right)\widehat{U}_p^\bullet + \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_p^\bullet}{\partial \boldsymbol{\eta}_p}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_p^\bullet}{\partial \boldsymbol{\alpha}}\right)\mathbb{E}^{-1}\left(\frac{\partial \widehat{\mathbf{S}}_{\boldsymbol{\alpha}}^\bullet}{\partial \boldsymbol{\alpha}}\right)\widehat{\mathbf{S}}_{\boldsymbol{\alpha}}^\bullet; \\
&(iv) \quad \widehat{\boldsymbol{\eta}}_m-\boldsymbol{\eta}_m^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_m^\bullet}{\partial \boldsymbol{\eta}_m}\right)\widehat{U}_m^\bullet + \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_m^\bullet}{\partial \boldsymbol{\eta}_m}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_m^\bullet}{\partial \boldsymbol{\beta}}\right)\mathbb{E}^{-1}\left(\frac{\partial \widehat{\mathbf{S}}_{\boldsymbol{\beta}}^\bullet}{\partial \boldsymbol{\beta}}\right)\widehat{\mathbf{S}}_{\boldsymbol{\beta}}^\bullet; \\
&(v) \qquad \widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\tau}}\right)\widehat{U}_{\boldsymbol{\tau}}^\bullet - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\alpha}}\right)(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^\bullet) \\
&\qquad\qquad - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\bullet) - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\eta}_p}\right)(\widehat{\boldsymbol{\eta}}_p-\boldsymbol{\eta}_p^\bullet) \\
&\qquad\qquad - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m-\boldsymbol{\eta}_m^\bullet).
\end{aligned}
$$

Finalement, en remplaçant ces cinq linéarisées dans (2.4.2), on obtient

$$\widehat{t}_{MR} \simeq \sum_{j \in S} w_j \psi_j,$$

où

$$\begin{aligned}\psi_j = \frac{r_j}{p_j^\bullet} y_j + \left(1 - \frac{r_j}{p_j^\bullet}\right) \mathbf{h}_j^{\bullet\top} \boldsymbol{\tau}^\bullet + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\alpha}}\right)(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\bullet) \\ + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\eta}_p}\right)(\widehat{\boldsymbol{\eta}}_p - \boldsymbol{\eta}_p^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m - \boldsymbol{\eta}_m^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}_{MR}^\bullet}{\partial \boldsymbol{\tau}}\right)(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}^\bullet).\end{aligned}$$

Nous obtenons alors l'approximation du biais conditionnel en (2.4.1) suivante :

$$\begin{aligned} B_{1i}^{MR} &= \mathbb{E}_p\left(\widehat{t}_{MR} - t_y \mid I_i = 1\right) \\ &\simeq \mathbb{E}_p\left(\sum_{j \in S} w_j \psi_j + \sum_{j \in U} \psi_j - \sum_{j \in U} \psi_j - t_y \mid I_i = 1\right) \\ &= \sum_{j \in U} \frac{\Delta_{ij}}{\pi_i \pi_j} \psi_j + \sum_{j \in U} \psi_j - t_y. \end{aligned} \tag{2.4.3}$$

Un estimateur de ce biais conditionnel est donné par

$$\widehat{B}_{1i}^{MR} = \sum_{j \in S} \frac{\Delta_{ij}}{\pi_{ij} \pi_j} \widehat{\psi}_j, \tag{2.4.4}$$

avec

$$\begin{aligned}\widehat{\psi}_j = \frac{r_j}{\widehat{p}_j} y_j + \left(1 - \frac{r_j}{\widehat{p}_j}\right) \mathbf{h}_j^\top \widehat{\boldsymbol{\tau}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\alpha}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\alpha}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}} \\ + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_p}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_p} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\tau}},\end{aligned} \tag{2.4.5}$$

où les matrices $\widehat{\mathbf{A}}_{\boldsymbol{\theta}}$ désignent les estimations des différences linéarisées $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\bullet$. Toutes les dérivées nécessaires au calcul de l'estimation du biais conditionnel sont explicitées en annexe B.1. Pour alléger, nous notons $\widehat{\mathbf{f}}_{\widehat{\boldsymbol{\theta}}} \equiv \widehat{\mathbf{f}}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$ et $\partial \widehat{\mathbf{f}}_{\widehat{\boldsymbol{\theta}}} / \partial \boldsymbol{\theta} \equiv \widehat{\mathbf{f}}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) / \partial \boldsymbol{\theta}$, où $\widehat{\mathbf{f}}_{\boldsymbol{\theta}}$ désigne une équation estimante. Les matrices $\widehat{\mathbf{A}}_{\boldsymbol{\theta}}$ sont données par

$$
\begin{aligned}
&(i) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\alpha}} \simeq -\left(\frac{\partial \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\alpha}}}}{\partial \boldsymbol{\alpha}}\right)^{-1} \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\alpha}}};\\
&(ii) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\beta}} \simeq -\left(\frac{\partial \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\beta}}}}{\partial \boldsymbol{\beta}}\right)^{-1} \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\beta}}};\\
&(iii) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\eta}_p} \simeq -\left(\frac{\partial \widehat{U}_{\widehat{p}}}{\partial \boldsymbol{\eta}_p}\right)^{-1} \widehat{U}_{\widehat{p}} + \left(\frac{\partial \widehat{U}_{\widehat{p}}}{\partial \boldsymbol{\eta}_p}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{p}}}{\partial \boldsymbol{\alpha}}\right) \left(\frac{\partial \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\alpha}}}}{\partial \boldsymbol{\alpha}}\right)^{-1} \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\alpha}}}; \qquad (2.4.6)\\
&(iv) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m} \simeq -\left(\frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\eta}_m}\right)^{-1} \widehat{U}_{\widehat{m}} + \left(\frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\eta}_m}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\beta}}\right) \left(\frac{\partial \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\beta}}}}{\partial \boldsymbol{\beta}}\right)^{-1} \widehat{\mathbf{S}}_{\widehat{\boldsymbol{\beta}}};\\
&(v) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\tau}} \simeq -\left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \widehat{U}_{\widehat{\boldsymbol{\tau}}} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\alpha}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\alpha}} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}}\\
&\qquad\qquad - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_p}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_p} - \left(\frac{\partial \widehat{U}^{\triangle}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_m}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m}.
\end{aligned}
$$

Notons que l'estimateur $\widehat{B}_{1i}^{MR}$ en (2.4.4) ne tient pas compte du terme $\sum_{j\in U}\psi_j - t_y$ en (2.4.3) car un estimateur de $t_y$ est donné par $\widehat{t}_{MR}$ alors qu'un estimateur de $\sum_{j\in U}\psi_j$ est donné par $\sum_{j\in S} w_j\psi_j$. Or, $\widehat{t}_{MR} \simeq \sum_{j\in S} w_j\psi_j$ et donc $\widehat{t}_{MR} - \sum_{j\in S} w_j\psi_j \simeq 0$.

Soit $\widehat{t}_{MR}^{R}$, la version efficace en présence de valeurs influentes de l'estimateur après imputation multi-robuste. L'estimateur $\widehat{t}_{MR}^{R}$ est obtenu par la même méthode que celle présentée dans la section 1.2.2, soit

$$
\widehat{t}_{MR}^{R} = \widehat{t}_{MR} - \frac{\widehat{B}_{\min}^{MR} + \widehat{B}_{\max}^{MR}}{2}, \qquad (2.4.7)
$$

où $\widehat{B}_{\min}^{MR}$ et $\widehat{B}_{\max}^{MR}$ sont les extrema de $\widehat{B}_{1i}^{MR}$ en (2.4.4).

Dans les prochaines sous-sections nous présentons des cas particuliers de cette méthode générale : la double robustesse, où un modèle d'imputation $m$ et un modèle de non-réponse $p$ sont spécifiés ; puis l'imputation multi-robuste, où deux modèles d'imputation $m$ sont spécifiés mais aucun modèle de non-réponse.

### 2.4.1. Imputation doublement robuste

Nous traitons d'abord le cas $1m-1p$ où un seul modèle d'imputation $m(\mathbf{v}_i, \boldsymbol{\beta})$ et un seul modèle de non-réponse $p(\mathbf{v}_i, \boldsymbol{\alpha})$ sont spécifiés. On a alors

$$\widehat{p}_i = p(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}})\frac{\widehat{\eta}_p^2}{\widehat{\eta}_p^2} = p(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}) \quad \text{et} \quad \widehat{m}_i = m(\mathbf{v}_i, \widehat{\boldsymbol{\beta}})\frac{\widehat{\eta}_m^2}{\widehat{\eta}_m^2} = m(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}).$$

Dans ce cas, la linéarisation du total estimé $\widehat{t}_{MR}$ en (2.4.2) devient :

$$\widehat{t}_{MR} = \sum_{i\in S} w_i\psi_i,$$

où

$$\psi_i = \frac{r_i}{p_i^\bullet}y_i + \left(1-\frac{r_i}{p_i^\bullet}\right)\mathbf{h}_i^{\bullet\top}\boldsymbol{\tau}^\bullet + \mathbb{E}\left(\frac{\partial\widehat{t}_{MR}^\bullet}{\partial\boldsymbol{\alpha}}\right)(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^\bullet) + \mathbb{E}\left(\frac{\partial\widehat{t}_{MR}^\bullet}{\partial\boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\bullet)$$
$$+ \mathbb{E}\left(\frac{\partial\widehat{t}_{MR}^\bullet}{\partial\boldsymbol{\tau}}\right)(\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}^\bullet),$$

avec le développement de Taylor de $\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}^\bullet$ simplifié :

$$\widehat{\boldsymbol{\tau}}-\boldsymbol{\tau}^\bullet \simeq -\mathbb{E}^{-1}\left(\frac{\partial\widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial\boldsymbol{\tau}}\right)\widehat{U}_{\boldsymbol{\tau}}^\bullet - \mathbb{E}^{-1}\left(\frac{\partial\widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial\boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial\widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial\boldsymbol{\alpha}}\right)(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^\bullet)$$
$$- \mathbb{E}^{-1}\left(\frac{\partial\widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial\boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial\widehat{U}_{\boldsymbol{\tau}}^\bullet}{\partial\boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\bullet).$$

Finalement, le biais conditionnel $B_{1i}^{MR}$ s'exprime, en ignorant le terme $\sum_{j\in U}\psi_j - t_y$, comme

$$B_{1i}^{MR} = \sum_{j\in U}\frac{\Delta_{ij}}{\pi_i\pi_j}\psi_j.$$

Par exemple, une sélection d'échantillon aléatoire simple et sans remise mène au biais conditionnel de la forme

$$B_{1i}^{MR} = \left(\frac{N}{n}-1\right)\frac{N}{N-1}(\psi_i - \overline{\Psi}),$$

où $\overline{\Psi} = \sum_{j\in U}\psi_j/N$. L'estimation du biais conditionnel est telle que

$$\widehat{B}_{1i}^{MR} = \sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}\pi_j}\widehat{\psi}_j, \tag{2.4.8}$$

où

$$\widehat{\psi}_j = \frac{r_j}{\widehat{p}_j}y_j + \left(1-\frac{r_j}{\widehat{p}_j}\right)\mathbf{h}_j^\top\widehat{\boldsymbol{\tau}} + \left(\frac{\partial\widehat{t}_{MR}}{\partial\boldsymbol{\alpha}}\right)\widehat{\mathbf{A}}_{\boldsymbol{\alpha}} + \left(\frac{\partial\widehat{t}_{MR}}{\partial\boldsymbol{\beta}}\right)\widehat{\mathbf{A}}_{\boldsymbol{\beta}} + \left(\frac{\partial\widehat{t}_{MR}}{\partial\boldsymbol{\tau}}\right)\widehat{\mathbf{A}}_{\boldsymbol{\tau}},$$

où $\widehat{\mathbf{A}}_{\boldsymbol{\alpha}}$ et $\widehat{\mathbf{A}}_{\boldsymbol{\beta}}$ sont définies en (2.4.6) et où $\widehat{\mathbf{A}}_{\boldsymbol{\tau}}$ est simplifiée telle que

$$\widehat{\mathbf{A}}_{\boldsymbol{\tau}} \simeq -\left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \widehat{U}_{\widehat{\boldsymbol{\tau}}} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\alpha}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\alpha}} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}}.$$

L'estimateur efficace $\widehat{t}_{MR}^{R}$ lors de l'imputation doublement robuste est donc

$$\widehat{t}_{MR}^{R} = \widehat{t}_{MR} - \frac{\widehat{B}_{\min}^{MR} + \widehat{B}_{\max}^{MR}}{2},$$

où $\widehat{B}_{\min}^{MR}$ et $\widehat{B}_{\max}^{MR}$ sont les minimum et maximum de (2.4.8).

À titre d'exemple, nous considérerons le cas où le modèle d'imputation et le modèle de non-réponse fixés sont de la forme

$$\widehat{p}_i = p(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}) = \frac{\exp(\mathbf{v}_i^\top \widehat{\boldsymbol{\alpha}})}{1 + \exp(\mathbf{v}_i^\top \widehat{\boldsymbol{\alpha}})} \quad \text{et} \quad \widehat{m}_i = m(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}) = \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}.$$

Toutes les dérivées propres à ce cas $1m - 1p$ se trouvent en annexe B.2 et des résultats de simulation sont présentés dans le chapitre 3, à la section 3.3.1.

### 2.4.2. Imputation multi-robuste avec 2 modèles d'imputation

Supposons que l'on n'estime pas la non-réponse et que deux modèles d'imputation soient spécifiés. On a alors

$$\widehat{m}_i^1 \equiv m^1(\mathbf{v}_i^1, \widehat{\boldsymbol{\beta}}^1) \quad \text{et} \quad \widehat{m}_i^2 \equiv m^2(\mathbf{v}_i^2, \widehat{\boldsymbol{\beta}}^2),$$

où $\mathbf{v}_i^\ell$, $\ell = 1, 2$, désigne le vecteur de variables auxiliaires complètement observées utilisé dans le modèle $\ell$. Puisque la non-réponse n'est pas estimée, l'estimateur du total $\widehat{t}_{MR}$ est de la forme

$$\widehat{t}_{MR} = \sum_{i \in S} w_i r_i y_i + \sum_{i \in S} w_i (1 - r_i) y_i^*. \tag{2.4.9}$$

Dans ce cas, la linéarisation du total estimé en (2.4.9) devient :

$$\widehat{t}_{MR} = \sum_{i \in S} w_i \psi_i,$$

où

$$\begin{aligned}\psi_i &= r_i y_i + (1-r_i)\,\mathbf{h}_i^{\bullet\top}\boldsymbol{\tau}^\bullet + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m - \boldsymbol{\eta}_m^\bullet) \\ &\quad + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\tau}}\right)(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}^\bullet) \\ &= r_i y_i + (1-r_i)\,\mathbf{h}_i^{\bullet\top}\boldsymbol{\tau}^\bullet + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\beta}^1}\right)(\widehat{\boldsymbol{\beta}}^1 - \boldsymbol{\beta}^{1\bullet}) + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\beta}^2}\right)(\widehat{\boldsymbol{\beta}}^2 - \boldsymbol{\beta}^{2\bullet}) \\ &\quad + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m - \boldsymbol{\eta}_m^\bullet) + \mathbb{E}\left(\frac{\partial \widehat{t}^\bullet_{MR}}{\partial \boldsymbol{\tau}}\right)(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}^\bullet)\end{aligned}$$

avec le développement de Taylor de $\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}^\bullet$ simplifié :

$$\begin{aligned}\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}^\bullet &\simeq -\mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\widehat{U}^\bullet_{\boldsymbol{\tau}} - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\beta}}\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\bullet) \\ &\quad - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m - \boldsymbol{\eta}_m^\bullet) \\ &= -\mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\widehat{U}^\bullet_{\boldsymbol{\tau}} - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\beta}^1}\right)(\widehat{\boldsymbol{\beta}}^1 - \boldsymbol{\beta}^{1\bullet}) \\ &\quad - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\beta}^2}\right)(\widehat{\boldsymbol{\beta}}^2 - \boldsymbol{\beta}^{2\bullet}) - \mathbb{E}^{-1}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\tau}}\right)\mathbb{E}\left(\frac{\partial \widehat{U}^\bullet_{\boldsymbol{\tau}}}{\partial \boldsymbol{\eta}_m}\right)(\widehat{\boldsymbol{\eta}}_m - \boldsymbol{\eta}_m^\bullet).\end{aligned}$$

Finalement, le biais conditionnel $B_{1i}^{MR}$ s'exprime, en ignorant le terme $\sum_{j\in U}\psi_j - t_y$, comme

$$B_{1i}^{MR} = \sum_{j\in U}\frac{\Delta_{ij}}{\pi_i\pi_j}\psi_j.$$

Un estimateur de ce biais conditionnel est de la forme

$$\widehat{B}_{1i}^{MR} = \sum_{j\in S}\frac{\Delta_{ij}}{\pi_{ij}\pi_j}\widehat{\psi}_j, \tag{2.4.10}$$

avec

$$\widehat{\psi}_j = r_j y_j + (1-r_j)\,\mathbf{h}_j^\top\widehat{\boldsymbol{\tau}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}^1}\right)\widehat{\mathbf{A}}^1_{\boldsymbol{\beta}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}^2}\right)\widehat{\mathbf{A}}^2_{\boldsymbol{\beta}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m}\right)\widehat{\mathbf{A}}_{\boldsymbol{\eta}_m} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}}\right)\widehat{\mathbf{A}}_{\boldsymbol{\tau}},$$

où $\widehat{\mathbf{A}}_{\boldsymbol{\eta}_m}$ est définie en (2.4.6) et où $\widehat{\mathbf{A}}_{\boldsymbol{\beta}}^{\ell}$, $\ell = 1, 2$ et $\widehat{\mathbf{A}}_{\boldsymbol{\tau}}$ sont telles que

$$(i) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\beta}}^{\ell} \simeq -\left(\frac{\partial \widehat{S}_{\widehat{\boldsymbol{\beta}}}^{\ell}}{\partial \boldsymbol{\beta}^{\ell}}\right)^{-1} \widehat{S}_{\widehat{\boldsymbol{\beta}}}^{\ell}, \ \ell = 1, 2;$$

$$(i) \qquad \widehat{\mathbf{A}}_{\boldsymbol{\tau}} \simeq -\left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \widehat{U}_{\widehat{\boldsymbol{\tau}}} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}^1}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}}^{1} - \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}^2}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}}^{2}$$
$$- \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}^{\bullet}}{\partial \boldsymbol{\tau}}\right)^{-1} \left(\frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_m}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m}.$$

Par exemple, un plan de sondage EASSR nous mène à nouveau à

$$B_{1i}^{MR} = \left(\frac{N}{n} - 1\right) \frac{N}{N-1} \left(\psi_i - \overline{\Psi}\right) \quad \text{et} \quad \widehat{B}_{1i}^{MR} = \left(\frac{N}{n} - 1\right) \frac{n}{n-1} \left(\widehat{\psi}_i - \overline{\widehat{\psi}}\right),$$

avec $\overline{\Psi} = \sum_{j \in U} \psi_j / N$ et $\overline{\widehat{\psi}} = \sum_{j \in S} \widehat{\psi}_j / n$.

L'estimateur efficace $\widehat{t}_{MR}^{R}$ lors de l'imputation multi-robuste avec deux modèles d'imputation est donc

$$\widehat{t}_{MR}^{R} = \widehat{t}_{MR} - \frac{\widehat{B}_{\min}^{MR} + \widehat{B}_{\max}^{MR}}{2},$$

où $\widehat{B}_{\min}^{MR}$ et $\widehat{B}_{\max}^{MR}$ sont les minimum et maximum de (2.4.10).

À titre d'exemple, nous considérerons le cas où les modèles d'imputation fixés sont les modèles linéaires

$$m^1(\mathbf{v}_i^1, \widehat{\boldsymbol{\beta}}^1) = \mathbf{v}_i^{1\top} \widehat{\boldsymbol{\beta}}^1 \quad \text{et} \quad m^2(\mathbf{v}_i^2, \widehat{\boldsymbol{\beta}}^2) = \mathbf{v}_i^{2\top} \widehat{\boldsymbol{\beta}}^2,$$

où $\mathbf{v}_i^1$ et $\mathbf{v}_i^2$ sont observés pour chaque individu échantillonné. Les dérivées relatives à ce scenario sont présentées en annexe B.3 et les résultats de simulation se trouvent dans le chapitre 3, à la section 3.3.2.

# Chapitre 3

---

# Études par simulation

Dans ce chapitre, nous effectuons des études par simulation afin d'évaluer le comportement des estimateurs proposés en termes de biais et d'erreur quadratique moyenne.

## 3.1. L'imputation multi-robuste

### 3.1.1. Population et modèles d'imputation simulés

Nous répétons 1000 fois le processus suivant. Nous commençons par générer une population finie de taille $N = 20\,000$. Pour ce faire, nous créons d'abord les variables auxiliaires $v_1, \ldots, v_{15}$ suivant une loi normale centrée réduite : $v_k \sim \mathcal{N}(0,1),\ k = 1, \ldots, 15$, puis nous générons un bruit aléatoire normal tel que $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Ensuite, la variable d'intérêt $y$ est générée selon le modèle

$$y = 2 + v_1 + \cdots + v_9 + \varepsilon.$$

La valeur de $\sigma_\varepsilon^2 = 12$ a été fixée afin que le coefficient de corrélation entre la variable d'intérêt et les variables auxiliaires $\mathbf{v}$ soit approximativement de 0,45. Enfin, dans chaque population, nous sélectionnons un échantillon de taille $n = 400$ selon un plan aléatoire simple et sans remise.

Dans chaque échantillon, la non-réponse est générée selon deux mécanismes : une probabilité de réponse $p_i$ égale pour tous les individus de l'échantillon, $p_i = 0{,}7$ (mécanisme uniforme) ; et selon des expériences indépendantes de Bernoulli avec probabilité de réponse

$p_i$ (mécanisme non-uniforme) définie par

$$p_i = \frac{\exp(1{,}1 + v_1 - 0{,}5v_2 + 0{,}25v_3 - 0{,}12v_4 + 0{,}5v_5)}{1 + \exp(1{,}1 + v_1 - 0{,}5v_2 + 0{,}25v_3 - 0{,}12v_4 + 0{,}5v_5)},$$

de manière à avoir en moyenne 70% de répondants dans chaque échantillon.

Finalement, nous imputons $y_i^*$ aux valeurs manquantes selon 6 méthodes d'imputation : une imputation avec le modèle d'imputation correctement spécifié ; 4 imputations avec les modèles d'imputation mal spécifiés $m^1$, $m^2$, $m^3$, $m^4$ suivants :

- $m^1(\mathbf{v}_i^1, \boldsymbol{\beta}^1) = v_1 + v_2 + v_3$,
- $m^2(\mathbf{v}_i^2, \boldsymbol{\beta}^2) = v_4 + v_5 + v_6$,
- $m^3(\mathbf{v}_i^3, \boldsymbol{\beta}^3) = v_7 + v_8 + v_9$,
- $m^4(\mathbf{v}_i^4, \boldsymbol{\beta}^4) = v_{10} + v_{11} + v_{12} + v_{13} + v_{14} + v_{15}$ ;

et une imputation multi-robuste basée sur les 4 modèles mal-spécifiés $m^1$, $m^2$, $m^3$ et $m^4$. Il sera intéressant de regarder comment se comporte cette imputation multi-robuste basée uniquement sur des modèles mal-spécifiés. Plusieurs estimations du total sont alors calculées : l'estimateur de Horvitz-Thompson $\widehat{t}_{y,HT}$, utilisé comme valeur de référence, l'estimateur $\widehat{t}_{MR}$ en (2.2.1), 4 estimateurs après imputation $\widehat{t}_I^{\ell},\ \ell = 1, 2, 3, 4$ en (1.4.1) basés sur les 4 modèles $m^1$, $m^2$, $m^3$ et $m^4$, respectivement et l'estimateur après imputation $\widehat{t}_I^{vrai}$ basé sur le modèle correctement spécifié.

En plus de calculer les divers estimateurs considérés, nous avons représenté graphiquement, à la figure 3.1, les méthodes d'imputation utilisées. Pour ce faire, nous avons sélectionné les 30 premiers non-répondants du fichier de données généré complété et tracé les valeurs imputées selon les 6 méthodes, en plus d'indiquer la véritable valeur de $y$ pour chaque individu.

### 3.1.2. Résultats

Dans le tableau 3.1, nous comparons l'estimateur après imputation multi-robuste, $\widehat{t}_{MR}$, à l'estimateur de Horvitz-Thompson, $\widehat{t}_{y,HT}$, et aux 5 estimateurs après imputation basée sur un seul modèle en calculant leur biais relatif Monte-Carlo $BR_{MC}$ défini en (1.4.3). Rappelons qu'il est tel que :

$$BR_{MC}(\widehat{t}_y) = \frac{\mathbb{E}_{MC}(\widehat{t}_y) - t_y}{t_y} \times 100.$$

Nous calculons également l'efficacité relative, $ER$, des 6 estimateurs imputés à l'estimateur $\widehat{t}_{y,HT}$ définie en (1.4.4) et dont on rappelle la forme ici :

$$ER = \frac{EQM_{MC}(\widehat{t}_y)}{EQM_{MC}(\widehat{t}_{y,HT})}.$$

| | Mécanisme de non-réponse | | | |
|---|---|---|---|---|
| | Uniforme | | Non-uniforme | |
| Estimateur | $BR_{MC}$ (%) | $ER$ | $BR_{MC}$ (%) | $ER$ |
| $\widehat{t}_{y,HT}$ | 0,63 | 1,00 | 0,63 | 1,00 |
| $\widehat{t}_I^1$ | 0,63 | 1,44 | 5,64 | 1,77 |
| $\widehat{t}_I^2$ | 0,79 | 1,43 | 9,54 | 2,14 |
| $\widehat{t}_I^3$ | 0,64 | 1,40 | 13,70 | 2,86 |
| $\widehat{t}_I^4$ | 0,69 | 1,50 | 13,74 | 2,99 |
| $\widehat{t}_I^{vrai}$ | 0,73 | 0,84 | 0,54 | 0,81 |
| $\widehat{t}_{MR}$ | 0,74 | 1,33 | 3,76 | 1,55 |

**Tableau 3.1.** Résultats des différents estimateurs après imputation considérés

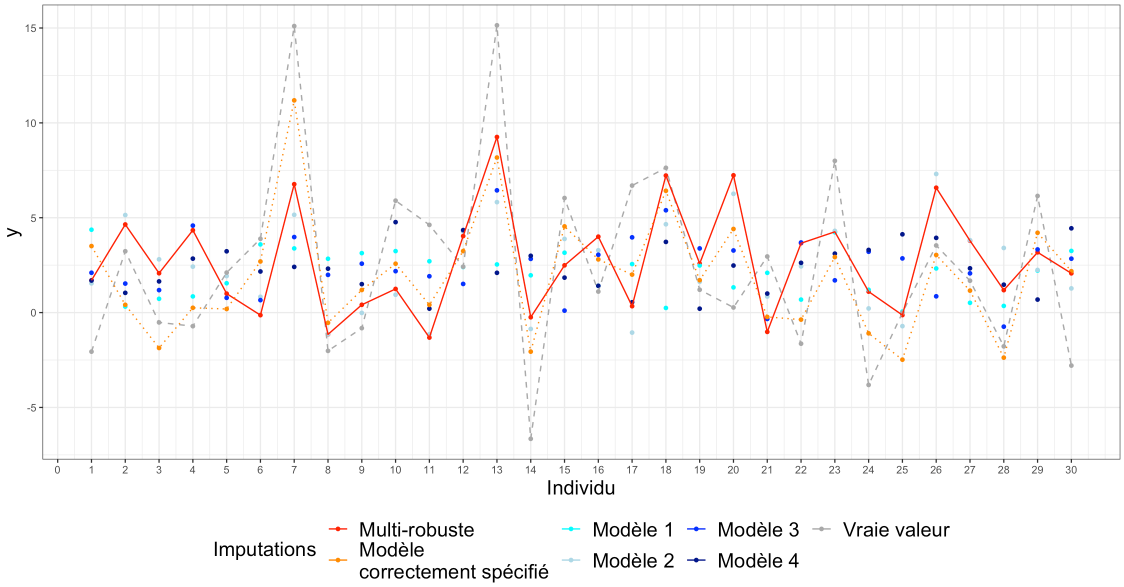



**Figure 3.1.** Représentation graphique des valeurs de $y$ générées et prédites selon les 6 modèles d'imputation

### 3.1.3. Discussion

Considérons premièrement la figure 3.1. Notons que la valeur imputée par imputation multi-robuste semble, en général, suivre la véritable valeur de $y$. De même, les valeurs obtenues par imputation simple suivent généralement la trajectoire de la vraie valeur de $y$, mais de plus loin. Cette observation est confirmée au moyen des biais relatifs du tableau 3.1. Puisque la valeur multi-robuste résulte des 4 modèles erronés, elle semble à mi-chemin entre ces données imputées et la véritable valeur. Remarquons également que la valeur imputée par le modèle correctement spécifié suit la même trajectoire que la valeur multi-robuste en étant encore plus proche de la vraie valeur.

Remarquons, dans le tableau 3.1, qu'aucun des estimateurs n'est biaisé lorsque la probabilité de réponse est la même pour tous les individus de l'échantillon, tandis que seuls les estimateurs de Horvitz-Thompson, $\widehat{t}_{y,HT}$, et celui après imputation par le modèle correctement spécifié, $\widehat{t}_I^{vrai}$, sont approximativement sans biais quand la probabilité de réponse est une fonction des variables auxiliaires $v_1, \ldots, v_5$. Cependant, tandis que les biais relatifs des estimateurs après imputation par des modèles mal spécifiés, $\widehat{t}_I^{\ell},\ \ell = 1, 2, 3, 4$, sont élevés – allant jusqu'à 13% pour les modèles incorrects $m^3$ et $m^4$ – c'est l'estimateur multi-robuste qui est le moindrement biaisé avec un biais relatif de 4%. Notons que ce faible biais est obtenu alors qu'aucun des modèles n'est correctement spécifié pour l'imputation multi-robuste, ce qui semble rejoindre Han (2014b) et Chen et Haziza (2017).

Concernant les efficacités relatives à l'estimateur de Horvitz-Thompson, une valeur de $ER$ inférieure à 1 implique que l'estimateur étudié est plus efficace que l'estimateur de Horvitz-Thompson calculé sur le fichier de données complet, tandis que plus la valeur de $ER$ s'éloigne de 1, moins l'estimateur est efficace par rapport à celui de Horvitz-Thompson. Seul l'estimateur après imputation par le bon modèle, $\widehat{t}_I^{vrai}$, obtient de meilleurs résultats en termes d'erreur quadratique moyenne que $\widehat{t}_{y,HT}$, avec une valeur de $ER$ d'environ 0,80. Les 5 autres estimateurs étudiés ont une $ER$ supérieure à 1. Cependant, dans les deux configurations de probabilités de réponse envisagées, celui qui a les meilleures valeurs de $ER$, c'est-à-dire les plus petites, est l'estimateur multi-robuste, $\widehat{t}_{MR}$. En effet, pour le mécanisme uniforme, on note une $ER$ de 1,33 et pour le mécanisme non-uniforme, on a $ER = 1{,}55$ et ce, malgré une mauvaise spécification de tous les modèles le composant.

## 3.2. Estimation efficace en présence de valeurs influentes : Imputation basée sur un seul modèle d'imputation

Nous étudions dans cette section l'efficacité de l'estimateur proposé, $\widehat{t}_I^R$ en (2.3.5) en présence de valeurs influentes. Pour ce faire, nous le comparons à l'estimateur de base, $\widehat{t}_I$ en (1.4.1).

### 3.2.1. Populations, échantillons et modèles simulés

Pour juger de l'efficacité de $\widehat{t}_I^R$ par rapport à $\widehat{t}_I$, nous générons 10 000 répétitions provenant de 10 populations finies différentes de taille $N = 5000$. Une variable auxiliaire, $v_{1i}$, a d'abord été générée de la manière suivante : $v_{1i} \sim \mathcal{U}(0; 5)$. Ensuite, la variable d'intérêt a été générée telle que

$$y_i|\mathbf{v}_i \sim \mathcal{D}(\mu_i; \nu_i),$$

selon 4 distributions $\mathcal{D}$ : normale, Gamma, lognormale et Pareto. Nous avons fait en sorte (voir annexe C.1) que les moments des distributions soient les mêmes pour les 4 distributions :

$$\mu_i = \beta_0 + \beta_1 v_{1i} + \beta_2 v_{1i}^2 \quad \text{et} \quad \nu_i = \sigma^2.$$

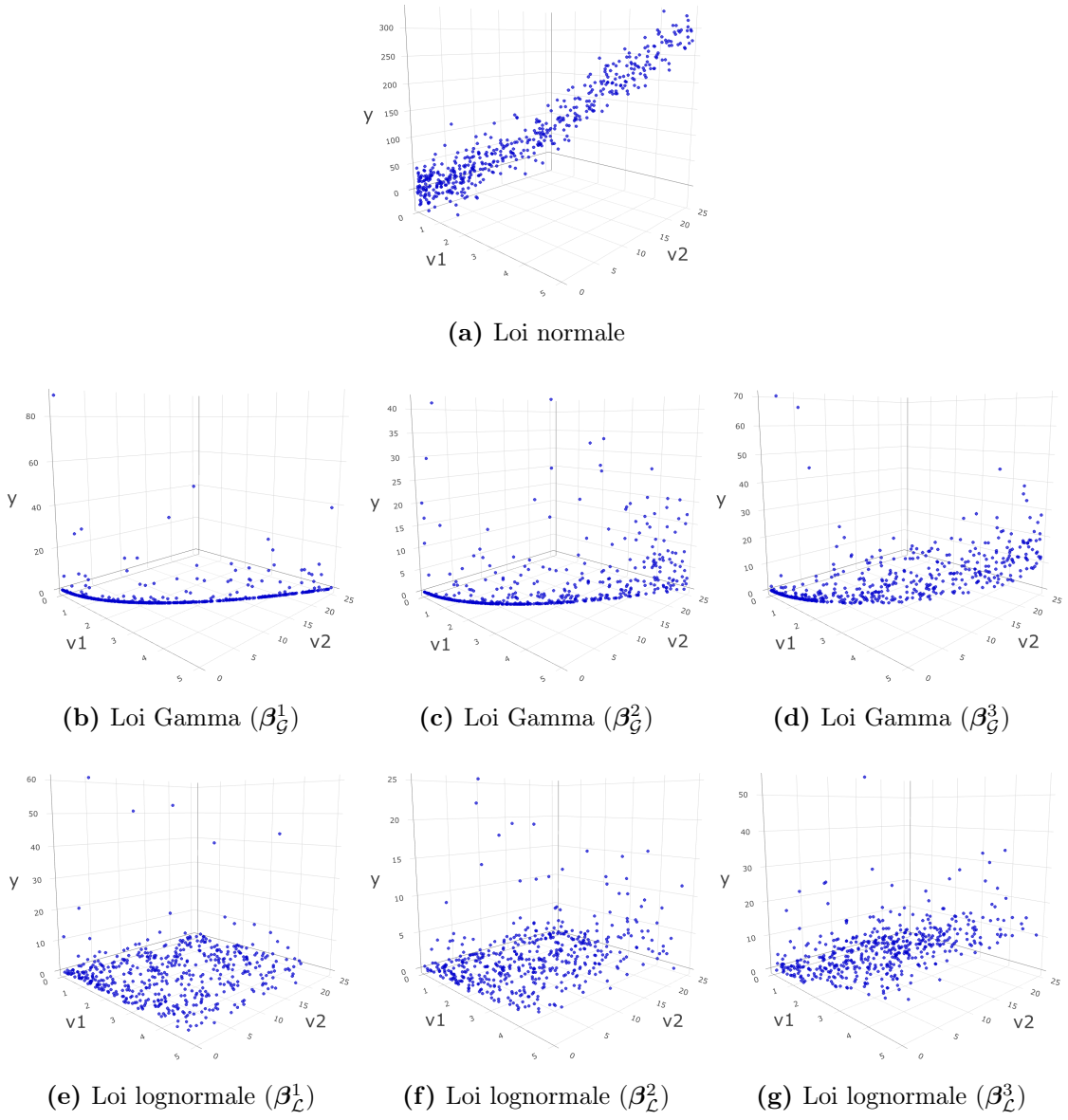
Plusieurs paramètres $\boldsymbol{\beta}$ et $\sigma^2$ ont été considérés afin de jouer sur la forme de la distribution de la variable réponse $y$ et sur la proportion de valeurs influentes. Pour la loi normale, seuls $\boldsymbol{\beta}_{\mathcal{N}} = (10; 10; 10)^\top$ et $\sigma^2 = 500$ ont été utilisés. Pour chacune des lois Gamma, lognormale et Pareto, trois combinaisons de paramètres ont été choisies, permettant d'avoir, pour chacune, de moins en moins de valeurs $y$ aberrantes. Pour la loi Gamma, nous avons fixé $\sigma^2 = 50$, $\boldsymbol{\beta}_{\mathcal{G}}^1 = (1; 0{,}05; 0{,}05)^\top$, $\boldsymbol{\beta}_{\mathcal{G}}^2 = (1; 0{,}2; 0{,}2)^\top$ et $\boldsymbol{\beta}_{\mathcal{G}}^3 = (1; 1; 0{,}4)^\top$. Pour la loi lognormale, nous avons choisi $\sigma^2 = 30$, $\boldsymbol{\beta}_{\mathcal{L}}^1 = (1; 0{,}2; 0{,}1)^\top$, $\boldsymbol{\beta}_{\mathcal{L}}^2 = (1; 0{,}3; 0{,}2)^\top$ et $\boldsymbol{\beta}_{\mathcal{L}}^3 = (1; 2{,}3; 0{,}2)^\top$. Enfin, pour la loi Pareto, $\sigma^2 = 20$, $\boldsymbol{\beta}_{\mathcal{P}}^1 = (1; 0{,}1; 0{,}1)^\top$, $\boldsymbol{\beta}_{\mathcal{P}}^2 = (1; 0{,}2; 0{,}2)^\top$ et $\boldsymbol{\beta}_{\mathcal{P}}^3 = (1; 1{,}5; 0, ,5)^\top$. Les 10 distributions sont illustrées par les figures 3.2 et 3.3.

Dans chacune de ces 10 populations, nous sélectionnons deux échantillons de tailles $n = 50$ et $n = 100$ selon le plan aléatoire simple et sans remise. Dans chaque échantillon, nous générons la non-réponse selon des expériences de Bernoulli avec probabilité de réponse

$$p_i = \frac{\exp(1{,}5 - 1{,}5 v_{1i} + 0{,}4 v_{1i}^2)}{1 + \exp(1{,}5 - 1{,}5 v_{1i} + 0{,}4 v_{1i}^2)},$$

ce qui garantit d'avoir en moyenne 70% de répondants.

La non-réponse générée, nous remplaçons les valeurs manquantes avec une méthode d'imputation basée sur le modèle de régression linéaire correctement spécifié. Ensuite, l'estimateur du total $\widehat{t_I}$ est calculé. Pour comparer cet estimateur à l'estimateur proposé, $\widehat{t_I^R}$, nous calculons d'abord l'influence estimée de chaque unité dans l'échantillon avec le biais conditionnel, $\widehat{B}_{1i}^I$ en (2.3.1). Nous pouvons ensuite estimer le total avec l'estimateur efficace, $\widehat{t_I^R}$.

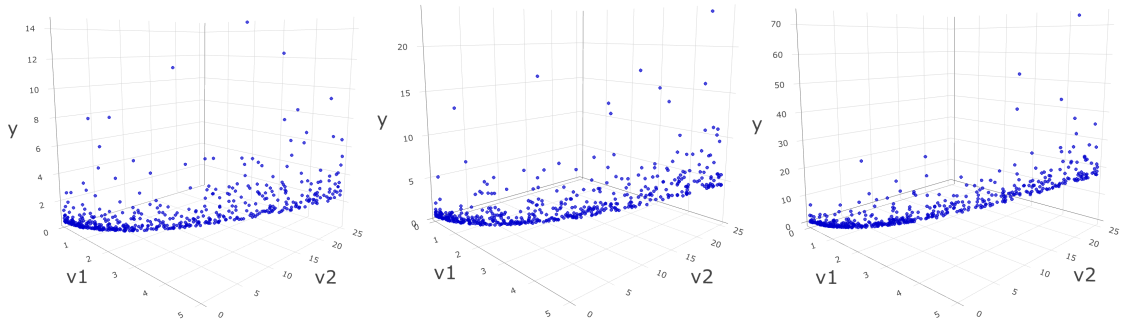



**(a)** Loi normale

**(b)** Loi Gamma ($\boldsymbol{\beta}_{\mathcal{G}}^1$) **(c)** Loi Gamma ($\boldsymbol{\beta}_{\mathcal{G}}^2$) **(d)** Loi Gamma ($\boldsymbol{\beta}_{\mathcal{G}}^3$)

**(e)** Loi lognormale ($\boldsymbol{\beta}_{\mathcal{L}}^1$) **(f)** Loi lognormale ($\boldsymbol{\beta}_{\mathcal{L}}^2$) **(g)** Loi lognormale ($\boldsymbol{\beta}_{\mathcal{L}}^3$)

**FIGURE 3.2.** Illustrations des distributions de $y$ pour les lois normale, Gamma et lognormale




**(a)** Loi Pareto ($\boldsymbol{\beta}^1_{\mathcal{P}}$) **(b)** Loi Pareto ($\boldsymbol{\beta}^2_{\mathcal{P}}$) **(c)** Loi Pareto ($\boldsymbol{\beta}^3_{\mathcal{P}}$)

**FIGURE 3.3.** Illustrations des distributions de $y$ pour la loi Pareto

### 3.2.2. Résulats

Pour comparer les deux estimateurs, nous calculons leur biais relatif Monte-Carlo (1.4.3) et l'efficacité relative en pourcentage de l'estimateur proposé par rapport à l'original, $ER$, définie par le ratio des erreurs quadratiques moyennes Monte-Carlo, $EQM_{MC}$ en (1.4.5), tel que :

$$ER = \frac{EQM_{MC}(\widehat{t}_I^R)}{EQM_{MC}(\widehat{t}_I)} \times 100.$$

| Distribution | $\boldsymbol{\beta}$ | $n$ | $BR_{MC}(\widehat{t}_I)$ | $BR_{MC}(\widehat{t}_I^R)$ | $ER$ |
|---|---|---|---|---|---|
| Normale | $\boldsymbol{\beta}_{\mathcal{N}}$ | 50 | 0,22 | −0,14 | 103 |
| | | 100 | 0,09 | −0,07 | 101 |
| Gamma | $\boldsymbol{\beta}^1_{\mathcal{G}}$ | 50 | −0,33 | −22,87 | 67 |
| | | 100 | −0,36 | −17,36 | 76 |
| Gamma | $\boldsymbol{\beta}^2_{\mathcal{G}}$ | 50 | 0,19 | −10,08 | 82 |
| | | 100 | 0,26 | −7,10 | 86 |
| Gamma | $\boldsymbol{\beta}^3_{\mathcal{G}}$ | 50 | −0,04 | −3,66 | 94 |
| | | 100 | −0,09 | −2,57 | 96 |

**Tableau 3.2.** Résultats de la comparaison entre $\widehat{t}_I^R$ et $\widehat{t}_I$ pour les distributions normale et Gamma

| Distribution | $\boldsymbol{\beta}$ | $n$ | $BR_{MC}(\widehat{t}_I)$ | $BR_{MC}(\widehat{t}_I^R)$ | $ER$ |
|---|---|---|---|---|---|
| Lognormale | $\boldsymbol{\beta}_{\mathcal{L}}^1$ | 50 | −0,32 | −9,68 | 67 |
| | | 100 | −0,23 | −7,18 | 72 |
| Lognormale | $\boldsymbol{\beta}_{\mathcal{L}}^2$ | 50 | 0,19 | −6,16 | 75 |
| | | 100 | 0,12 | −4,51 | 80 |
| Lognormale | $\boldsymbol{\beta}_{\mathcal{L}}^3$ | 50 | 0,10 | −1,81 | 94 |
| | | 100 | 0,07 | −1,28 | 94 |
| Pareto | $\boldsymbol{\beta}_{\mathcal{P}}^1$ | 50 | −0,09 | −4,69 | 57 |
| | | 100 | 0,18 | −3,43 | 57 |
| Pareto | $\boldsymbol{\beta}_{\mathcal{P}}^2$ | 50 | 0,02 | −4,00 | 59 |
| | | 100 | 0,04 | −2,95 | 70 |
| Pareto | $\boldsymbol{\beta}_{\mathcal{P}}^3$ | 50 | −0,28 | −1,98 | 92 |
| | | 100 | 0,01 | −1,25 | 91 |

**Tableau 3.3.** Résultats de la comparaison entre $\widehat{t}_I^R$ et $\widehat{t}_I$ pour les distributions lognormale et Pareto

### 3.2.3. Discussion

Pour la population normale, symétrique, qui ne présente aucune valeur potentiellement influente, les deux estimateurs étudiés sont sans biais, peu importe la taille échantillonnale. Cependant, l'efficacité relative de l'estimateur proposé est supérieure à 100% – seuil en-dessous duquel l'estimateur $\widehat{t}_I^R$ est plus efficace en termes de $EQM$ – ce qui signifie qu'il y a une très légère perte d'efficacité : on a une efficacité relative d'environ 102%.

Lorsque la distribution de la variable d'intérêt est asymétrique et propice aux données influentes, comme c'est le cas des distributions Gamma, lognormale et Pareto, les conclusions sont différentes. Notons que l'estimateur du total après imputation simple, $\widehat{t}_I$, est toujours approximativement sans biais mais nous introduisons, sans surprise, un biais avec l'estimateur proposé, $\widehat{t}_I^R$. C'est pour la loi Gamma que les biais relatifs sont les plus élevés en valeur absolue : entre 2% et 23%. En fait, plus les valeurs de $ER$ sont basses, donc

plus $\widehat{t}_I^R$ est efficace en termes de $EQM$, plus le biais introduit est élevé : le biais relatif de 23% en valeur absolue correspond à une efficacité relative de 67% et $BR_{MC} = -2{,}57\%$ correspond à $ER = 96\%$. Ainsi, pour la loi Gamma, l'efficacité relative de $\widehat{t}_I^R$ varie de 67% à 96%. Puisque $ER$ est toujours inférieure à 100%, cela signifie que $\widehat{t}_I^R$ est plus efficace que $\widehat{t}_I$. Remarquons également que lorsque la taille échantillonnale augmente, les valeurs de $ER$ tendent vers 100%, et l'efficacité de $\widehat{t}_I^R$ se rapproche de celle de $\widehat{t}_I$. Cela s'explique par le fait que plus la taille de l'échantillon augmente, moins il y a de données potentiellement influentes.

Les résultats pour les distributions lognormale et Pareto sont sensiblement les mêmes que ceux de la distribution Gamma. Les biais introduits sont moins élevés : de $-1{,}28\%$ à $-9{,}68\%$ pour la loi lognormale et de $-1{,}25\%$ à $-4{,}69\%$ pour la loi Pareto. Dans le cas de la loi lognormale, l'estimateur efficace présente des valeurs de $ER$ semblables à celles de la distribution Gamma : on note des valeurs de $ER$ entre 67% et 94%. Dans le cas de la loi Pareto, $\widehat{t}_{MR}^R$ semble encore plus efficace, avec des valeurs de $ER$ allant aussi bas que 57%. Cela implique que l'erreur quadratique moyenne Monte-Carlo de l'estimation est quasiment divisée par deux lorsque c'est l'estimateur $\widehat{t}_I^R$ qui est utilisé.

## 3.3. Estimation efficace en présence de valeurs influentes : Imputation multi-robuste

Cette section est consacrée à l'efficacité en présence de valeurs influentes de l'estimateur $\widehat{t}_{MR}^R$ en (2.4.7). Nous le comparons donc à l'estimateur multi-robuste, $\widehat{t}_{MR}$ en (2.2.1), dans deux cadres : celui de la double robustesse (section 3.3.1) et celui de la multi-robustesse avec deux modèles d'imputation (section 3.3.2).

### 3.3.1. Imputation doublement robuste

#### 3.3.1.1. *Populations, échantillons et modèles d'imputation simulés*

Pour comparer $\widehat{t}_{MR}^R$ et $\widehat{t}_{MR}$, nous avons généré 10 000 réplicats des mêmes populations que lors de l'étude par simulations décrite à la section 3.2. Ces populations sont présentées dans la sous-section 3.2.1. De même, les échantillons et la non-réponse ont été sélectionnés et générés comme à la section 3.2.1.

Une fois la non-réponse créée dans tous les échantillons, nous avons remplacé les valeurs manquantes de $y$ avec la méthode d'imputation multi-robuste présentée dans la section 2.2. Pour ce faire, nous nous sommes placés dans le cadre d'une imputation doublement robuste et nous avons fixé un modèle de non-réponse et un modèle d'imputation de la forme

$$p(\mathbf{v}_i, \boldsymbol{\alpha}) = \frac{\exp(\mathbf{v}_i^\top \boldsymbol{\alpha})}{1 + \exp(\mathbf{v}_i^\top \boldsymbol{\alpha})} \quad \text{et} \quad m(\mathbf{v}_i, \boldsymbol{\beta}) = \mathbf{v}_i^\top \boldsymbol{\beta}.$$

Pour que l'estimateur du total $\widehat{t}_{MR}$ soit convergent, il suffit qu'un seul modèle soit correctement spécifié. Alors, 3 scénarios sont possibles :

(i) les deux modèles sont correctement spécifiés, notés $m$ ■, $p$ ■ ;

(ii) le modèle d'imputation est correctement spécifié mais le modèle de non-réponse n'est pas correctement spécifié, notés $m$ ■, $p$ □ ;

(iii) le modèle d'imputation n'est pas correctement spécifié et le modèle de non-réponse est correctement spécifié, notés $m$ □, $p$ ■.

Lorsqu'un modèle est correctement spécifié, ce sont les bonnes variables auxiliaires $v_1$ et $v_1^2$ qui sont utilisées, dans le modèle de régression linéaire pour $m$ et dans le modèle logistique pour $p$. Lorsqu'un modèle n'est pas correctement spécifié, la variable auxiliaire $v_1^2$ est omise et remplacée par $v_2 \sim \mathcal{U}(0, 4)$, qui n'a aucun lien avec $y$ ni avec l'indicatrice de réponse, $r$. Les différents cas sont résumés dans le tableau 3.4.

| | $v_1$ | $v_1^2$ | $v_2$ |
|---|---|---|---|
| $m$ ■ $p$ ■ | ✓ | ✓ | X |
| $m$ □ $p$ □ | ✓ | X | ✓ |

**Tableau 3.4.** Résumé des variables auxiliaires utilisées selon les modèles spécifiés

Une fois les valeurs manquantes imputées, nous calculons $\widehat{t}_{MR}$ et le biais conditionnel estimé de toutes les unités échantillonnées, $\widehat{B}_{1i}^{MR}$ (2.4.8), afin d'estimer le total avec l'estimateur efficace $\widehat{t}_{MR}^{R}$.

3.3.1.2. *Résultats*

Les estimateurs $\widehat{t}_{MR}$ et $\widehat{t}^{R}_{MR}$ sont comparés au moyen du biais relatif Monte-Carlo (1.4.3) et de l'efficacité relative en pourcentage de l'estimateur proposé par rapport à l'estimateur multi-robuste, $ER$, définie par le ratio des erreurs quadratiques moyennes Monte-Carlo, $EQM_{MC}$ en (1.4.5), tel que :

$$ER = \frac{EQM_{MC}(\widehat{t}^{R}_{MR})}{EQM_{MC}(\widehat{t}_{MR})} \times 100. \tag{3.3.1}$$

| Distribution | $\boldsymbol{\beta}$ | Scénario | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}^{R}_{MR})$ | $ER$ |
|---|---|---|---|---|---|---|
| Normale | $\boldsymbol{\beta}_{\mathcal{N}}$ | $m$ ■ , $p$ ■ | 50 | 0,08 | −0,30 | 103 |
| | | | 100 | −0,07 | −0,26 | 101 |
| | | $m$ ■ , $p$ □ | 50 | −0,05 | −0,39 | 103 |
| | | | 100 | −0,01 | −0,17 | 102 |
| | | $m$ □ , $p$ ■ | 50 | 2,25 | 1,85 | 101 |
| | | | 100 | 2,23 | 2,03 | 100 |

**Tableau 3.5.** Résultats $1m - 1p$ pour la distribution normale

| Distribution | $\boldsymbol{\beta}$ | Scénario | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}^{R}_{MR})$ | $ER$ |
|---|---|---|---|---|---|---|
| Gamma | $\boldsymbol{\beta}^1_{\mathcal{G}}$ | $m$ ■ , $p$ ■ | 50 | 0,65 | −21,66 | 68 |
| | | | 100 | −0,17 | −17,36 | 76 |
| | | $m$ ■ , $p$ □ | 50 | 0,24 | −21,78 | 69 |
| | | | 100 | −0,32 | −17,04 | 78 |
| | | $m$ □ , $p$ ■ | 50 | 0,13 | −21,43 | 72 |
| | | | 100 | 1,22 | −15,55 | 78 |
| Gamma | $\boldsymbol{\beta}^2_{\mathcal{G}}$ | $m$ ■ , $p$ ■ | 50 | −0,05 | −10,29 | 83 |
| | | | 100 | 0,06 | −7,25 | 86 |
| | | $m$ ■ , $p$ □ | 50 | 0,37 | −9,63 | 83 |
| | | | 100 | 0,60 | −6,55 | 86 |
| | | $m$ □ , $p$ ■ | 50 | 2,07 | −7,80 | 82 |
| | | | 100 | 2,22 | −5,01 | 83 |
| Gamma | $\boldsymbol{\beta}^3_{\mathcal{G}}$ | $m$ ■ , $p$ ■ | 50 | −0,20 | −3,75 | 94 |
| | | | 100 | −0,10 | −2,60 | 95 |
| | | $m$ ■ , $p$ □ | 50 | 0,16 | −3,41 | 95 |
| | | | 100 | 0,02 | −2,54 | 95 |
| | | $m$ □ , $p$ ■ | 50 | 1,14 | −2,27 | 92 |
| | | | 100 | 1,70 | −0,74 | 90 |

**Tableau 3.6.** Résultats $1m - 1p$ pour la distribution Gamma

| Distribution | $\boldsymbol{\beta}$ | Scénario | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}^{R}_{MR})$ | $ER$ |
|---|---|---|---|---|---|---|
| Lognormale | $\boldsymbol{\beta}^1_{\mathcal{L}}$ | $m$ ■ , $p$ ■ | 50 | 0,63 | −8,74 | 64 |
| | | | 100 | 0,09 | −6,97 | 68 |
| | | $m$ ■ , $p$ □ | 50 | 0,23 | −9,13 | 68 |
| | | | 100 | −0,02 | −6,91 | 71 |
| | | $m$ □ , $p$ ■ | 50 | 0,47 | −8,34 | 70 |
| | | | 100 | 1,06 | −5,84 | 67 |
| Lognormale | $\boldsymbol{\beta}^2_{\mathcal{L}}$ | $m$ ■ , $p$ ■ | 50 | 0,16 | −6,10 | 71 |
| | | | 100 | −0,06 | −4,65 | 79 |
| | | $m$ ■ , $p$ □ | 50 | −0,41 | −6,57 | 79 |
| | | | 100 | 0,05 | −4,45 | 78 |
| | | $m$ □ , $p$ ■ | 50 | 1,40 | −4,65 | 68 |
| | | | 100 | 1,75 | −2,78 | 76 |
| Lognormale | $\boldsymbol{\beta}^1_{\mathcal{L}}$ | $m$ ■ , $p$ ■ | 50 | 0,15 | −1,75 | 92 |
| | | | 100 | 0,11 | −1,22 | 94 |
| | | $m$ ■ , $p$ □ | 50 | 0,09 | −1,80 | 93 |
| | | | 100 | −0,09 | −1,42 | 95 |
| | | $m$ □ , $p$ ■ | 50 | 0,37 | −1,44 | 92 |
| | | | 100 | 0,58 | −0,72 | 93 |

**Tableau 3.7.** Résultats $1m - 1p$ pour la distribution lognormale

| Distribution | $\boldsymbol{\beta}$ | Scénario | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}^{R}_{MR})$ | $ER$ |
|---|---|---|---|---|---|---|
| Pareto | $\boldsymbol{\beta}^1_{\mathcal{P}}$ | $m$ ■ , $p$ ■ | 50 | −0,17 | −4,68 | 56 |
| | | | 100 | −0,14 | −3,62 | 63 |
| | | $m$ ■ , $p$ □ | 50 | 0,34 | −4,26 | 56 |
| | | | 100 | 0,10 | −3,44 | 59 |
| | | $m$ □ , $p$ ■ | 50 | 0,99 | −3,38 | 53 |
| | | | 100 | 1,55 | −2,07 | 53 |
| Pareto | $\boldsymbol{\beta}^2_{\mathcal{P}}$ | $m$ ■ , $p$ ■ | 50 | 0,01 | −3,86 | 68 |
| | | | 100 | 0,05 | −2,91 | 67 |
| | | $m$ ■ , $p$ □ | 50 | 0,26 | −3,65 | 66 |
| | | | 100 | −0,25 | −3,07 | 77 |
| | | $m$ □ , $p$ ■ | 50 | 1,79 | −1,94 | 67 |
| | | | 100 | 1,68 | −1,18 | 67 |
| Pareto | $\boldsymbol{\beta}^3_{\mathcal{P}}$ | $m$ ■ , $p$ ■ | 50 | 0,04 | −1,68 | 91 |
| | | | 100 | −0,01 | −1,24 | 91 |
| | | $m$ ■ , $p$ □ | 50 | 0,07 | −1,59 | 92 |
| | | | 100 | −0,02 | −1,21 | 93 |
| | | $m$ □ , $p$ ■ | 50 | 1,34 | −0,31 | 88 |
| | | | 100 | 1,46 | 0,28 | 88 |

**Tableau 3.8.** Résultats $1m - 1p$ pour la distribution Pareto

3.3.1.3. *Discussion*

Remarquons d'abord que pour chaque distribution, nous avons des résultats similaires entre les scénarios où le modèle d'imputation est correctement spécifié : l'estimateur multi-robuste, $\widehat{t}_{MR}$, est approximativement sans biais. Cependant, il semble très légèrement biaisé pour le scénario $m$ □ $p$ ■, le biais relatif variant entre 0,1% et 2,3%. Comme prévu, à l'exception de la configuration avec la distribution normale où les biais relatifs des deux estimateurs évalués sont sensiblement les mêmes, un biais est introduit avec l'estimateur efficace, $\widehat{t}^{R}_{MR}$. C'est avec la distribution Gamma que les plus hauts biais relatifs en valeur

absolue sont observés : jusqu'à 21,78%, alors que pour la lognormale, le biais relatif maximum est de 9,13% en valeur absolue et pour la distribution Pareto, le maximum atteint en valeur absolue est de seulement 4,68%. À nouveau, c'est avec la distribution Pareto que le biais introduit est le plus faible, avec des valeurs de $BR_{MC}$ entre 0,28% et $-4{,}68\%$.

En termes d'$EQM$, nous observons encore une très légère perte d'efficacité relative lorsque la population ne présente pas de données influentes, on a : $100\% \leq ER \leq 103\%$ pour la distribution normale. Pour les distributions asymétriques, le gain d'efficacité est net, allant jusqu'à une efficacité relative de 53% dans le cas de la Pareto. En effet, pour les trois distributions propices aux données influentes considérées, nous observons des valeurs de $ER$ plus petites que 100% : dans le cas de la Gamma, selon le paramètre $\boldsymbol{\beta}_{\mathcal{G}}$ fixé, l'efficacité relative de l'estimateur proposé par rapport à l'estimateur multi-robuste varie entre 68% et 95% ; avec une distribution lognormale, elle varie entre 64% et 95% ; enfin, pour la Pareto, l'erreur quadratique moyenne de Monte-Carlo de l'estimation est à nouveau presque divisée par deux avec l'estimateur efficace puisque les valeurs de $ER$ se situent entre 53% et 93%.

Comme pour les résultats comparant $\widehat{t}_I$ et $\widehat{t}_I^R$, lorsque la taille échantillonnale augmente, $ER$ a tendance à augmenter, signifiant que les deux estimateurs tendent à être aussi performants l'un que l'autre lorsque tous les individus de la population sont échantillonnés et qu'alors, aucune unité n'est influente.

### 3.3.2. Imputation multi-robuste avec 2 modèles d'imputation

#### 3.3.2.1. *Populations, échantillons et modèles d'imputation simulés*

Afin de comparer $\widehat{t}_{MR}^R$ et $\widehat{t}_{MR}$ lorsque 2 modèles d'imputation sont considérés, nous utilisons une fois encore les 10 000 répétitions des populations présentées à la sous-section 3.2.1. Les mêmes méthodes sont appliquées pour sélectionner les échantillons et pour y générer la non-réponse.

Ensuite, nous imputons les valeurs manquantes avec la méthode d'imputation multi-robuste présentée dans la section 2.2 en considérant 2 modèles d'imputation linéaires et aucun modèle de non-réponse. On a alors

$$m^1(\mathbf{v}_i^1, \boldsymbol{\beta}^1) = \mathbf{v}_i^{1\top}\boldsymbol{\beta}^1 \quad \text{et} \quad m^2(\mathbf{v}_i^2, \boldsymbol{\beta}^2) = \mathbf{v}_i^{2\top}\boldsymbol{\beta}^2,$$

où $\mathbf{v}_i^\ell$ est le vecteur de variables auxiliaires utilisé dans le modèle $\ell$, $\ell = 1, 2$. Pour que $\widehat{t}_{MR}$ soit convergent il suffit qu'un seul modèle soit correctement spécifié alors nous avons un seul scénario à considérer : un des modèles est correctement spécifié, noté $m^1$ ■, et l'autre est mal spécifié, noté $m^2$ □. Nous utilisons, comme plus haut (section 3.3.1.1), $v_2 \sim \mathcal{U}(0, 4)$ pour le modèle mal spécifié. Les deux cas sont résumés dans le tableau 3.9.

| | $v_1$ | $v_2$ | $v_3$ |
|---|---|---|---|
| $m^1$■ | ✓ | ✓ | X |
| $m^2$□ | ✓ | X | ✓ |

**Tableau 3.9.** Résumé des variables auxiliaires utilisées selon les deux modèles d'imputation spécifiés

Avec le fichier de données complété nous estimons le total par $\widehat{t}_{MR}$ et calculons l'influence estimée de chaque unité dans l'échantillon par $\widehat{B}_{1i}^{MR}$ (2.4.10). Enfin, nous estimons le total par la version efficace proposée de l'estimateur multi-robuste, $\widehat{t}_{MR}^{R}$.

3.3.2.2. *Résultats*

Le biais relatif Monte-Carlo, $BR_{MC}$ (1.4.3), est encore une fois utilisé pour comparer $\widehat{t}_{MR}$ et $\widehat{t}_{MR}^{R}$, ainsi que l'efficacité relative, $ER$ (3.3.1).

| Distribution | $\boldsymbol{\beta}$ | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}_{MR}^{R})$ | $ER$ |
|---|---|---|---|---|---|
| Normale | $\boldsymbol{\beta}_{\mathcal{N}}$ | 50 | 0,03 | −0,50 | 103 |
| | | 100 | −0,04 | −0,31 | 102 |
| Gamma | $\boldsymbol{\beta}_{\mathcal{G}}^1$ | 50 | 0,70 | −15,87 | 76 |
| | | 100 | 0,87 | −11,38 | 81 |
| Gamma | $\boldsymbol{\beta}_{\mathcal{G}}^2$ | 50 | 0,15 | −7,38 | 88 |
| | | 100 | 0,52 | −4,53 | 90 |
| Gamma | $\boldsymbol{\beta}_{\mathcal{G}}^3$ | 50 | 0,55 | −2,21 | 95 |
| | | 100 | 0,30 | −1,44 | 97 |

**Tableau 3.10.** Résultats $2m - 0p$ pour les distributions normale et Gamma

| Distribution | $\boldsymbol{\beta}$ | $n$ | $BR_{MC}(\widehat{t}_{MR})$ | $BR_{MC}(\widehat{t}^{R}_{MR})$ | $ER$ |
|---|---|---|---|---|---|
| Lognormale | $\boldsymbol{\beta}^1_{\mathcal{L}}$ | 50 | 0,36 | −6,89 | 75 |
| | | 100 | 0,17 | −5,04 | 78 |
| Lognormale | $\boldsymbol{\beta}^2_{\mathcal{L}}$ | 50 | 0,54 | −4,39 | 81 |
| | | 100 | 0,31 | −3,06 | 86 |
| Lognormale | $\boldsymbol{\beta}^3_{\mathcal{L}}$ | 50 | 0,13 | −1,33 | 97 |
| | | 100 | −0,13 | −1,07 | 98 |
| Pareto | $\boldsymbol{\beta}^1_{\mathcal{P}}$ | 50 | 0,42 | −3,27 | 63 |
| | | 100 | 0,37 | −2,38 | 69 |
| Pareto | $\boldsymbol{\beta}^2_{\mathcal{P}}$ | 50 | 0,39 | −2,79 | 73 |
| | | 100 | 0,06 | −2,23 | 76 |
| Pareto | $\boldsymbol{\beta}^3_{\mathcal{P}}$ | 50 | 0,21 | −1,14 | 93 |
| | | 100 | 0,01 | −0,92 | 92 |

**Tableau 3.11.** Résultats $2m - 0p$ pour les distributions lognormale et Pareto

3.3.2.3. *Discussion*

Les résultats de cette étude par simulation abondent dans le même sens que ceux obtenus précédemment. Pour toutes les tailles d'échantillon et les distributions considérées, l'estimateur multi-robuste, $\widehat{t}_{MR}$, est approximativement sans biais. Nous introduisons généralement un biais avec l'estimateur $\widehat{t}^{R}_{MR}$, à l'exception du cas de la distribution normale, sans valeurs influentes, où $\widehat{t}^{R}_{MR}$ reste sans biais. Avec les distributions asymétriques, nous notons un biais relatif certain, entre −1,44% et −15,87% pour la distribution Gamma, bien que plus faible pour la distribution Pareto, entre −0,92% et −3,27%.

Concernant l'efficacité relative, $ER$, lorsque la population n'est pas propice aux unités influentes, comme la normale, nous observons une très légère perte d'efficacité avec des valeurs de $ER$ d'environ 103%. Lorsque la population présente des valeurs potentiellement influentes, nous notons une fois de plus une amélioration en termes d'erreur quadratique moyenne : les résultats de la loi Gamma et de la loi lognormale sont semblables, avec une

efficacité relative entre 75% et 98%. C'est avec la Pareto que les gains sont encore les plus flagrants où les valeurs de l'efficacité relative vont aussi bas que 63%.

# Conclusion

Dans ce mémoire, nous avons proposé une méthode d'estimation après imputation efficace en présence de données influentes. Nous n'avons travaillé qu'avec un échantillonnage aléatoire simple et sans remise mais cette méthode peut facilement être appliquée à tout plan de sondage. De plus, cette méthode est concluante pour tout type d'imputation : une imputation basée sur un seul modèle d'imputation, mais aussi une imputation multi-robuste, où plusieurs modèles d'imputation et/ou de non-réponse sont spécifiés.

Nous avons d'abord expliqué la méthode simple, présentée par Beaumont et al. (2013), pour rendre efficace un estimateur usuel en présence de données influentes. Ensuite, nous avons détaillé le fonctionnement de l'imputation multi-robuste telle que proposée par Chen et Haziza (2019). Puis, nous avons proposé la méthode d'estimation efficace après imputation, d'abord pour une imputation basée sur un seul modèle d'imputation, puis pour l'imputation multi-robuste présentée.

Dans la première étude par simulation, nous avons constaté que la méthode d'imputation multi-robuste de Chen et Haziza (2019) est numériquement efficace malgré tous les modèles spécifiés incorrectement. Dans la deuxième étude par simulation, nous avons comparé l'efficacité entre un estimateur après imputation simple et sa version efficace proposée. Les résultats obtenus montraient que la méthode proposée rendait effectivement l'estimateur plus efficace, mais aussi plus biaisé. Il s'agit donc d'un compromis biais-variance. Dans les troisième et quatrième études par simulation, nous avons comparé l'efficacité entre un estimateur après imputation multi-robuste – nous avons considéré deux cas : la double robustesse et la multi-robustesse avec deux modèles d'imputation mais aucun modèle de non-réponse – et sa version efficace proposée. Des résultats sensiblement équivalents à ceux

de l'imputation basée sur un seul modèle d'imputation ont été notés : la méthode avec le biais conditionnel rend effectivement l'estimateur plus efficace tout en introduisant un certain biais.

Dans chaque méthode d'imputation considérée, nous avons utilisé des classes de modèles d'imputation et/ou de non-réponse paramétriques. Il serait intéressant de vérifier si la méthode proposée fonctionnerait toujours pour des classes non-paramétriques. Par exemple, nous pourrions envisager d'estimer la probabilité de non-réponse et de prédire les valeurs manquantes de la variable d'intérêt avec des arbres de régression ou autres méthodes non-paramétriques. Il serait également utile d'en étudier les propriétés théoriques.

Une autre piste pour un travail futur a trait au problème difficile d'estimation théorique de l'erreur quadratique moyenne de l'estimateur efficace. On pourrait d'ailleurs commencer par omettre la non-réponse dans ce problème afin d'en simplifier la solution.

# Bibliographie


[1] Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika*, **100** :555–569.

[2] Chan, K. C. G. and Yam, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statistical Science*, **29** :380–396.

[3] Chen, S. and Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika*, **104** :439–453.

[4] Chen, S. and Haziza, D. (2019). Recent developments in dealing with item non-response in surveys : A critical review. *International Statistical Review*, **87** :S192–S218.

[5] Dongmo Jiongo, V. (2015). *Inférence robuste à la présence des valeurs aberrantes dans les enquêtes*. PhD thesis, Université de Montréal.

[6] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, **14** :153–158.

[7] Han, P. (2014a). A further study of the multiply robust estimator in missing data analysis. *Journal of Statistical Planning and Inference*, **148** :101–110.

[8] Han, P. (2014b). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, **109** :1159–1173.

[9] Han, P. and Wang, L. (2013). Estimation with missing data : beyond double robustness. *Biometrika*, **100** :417–430.

[10] Haziza, D. (2009). Imputation and inference in the presence of missing data. *Handbook of statistics*, **29** :215–246.

[11] Haziza, D. and Rao, J. N. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodology*, **32** :53–64.

[12] Kim, J. K. and Haziza, D. (2014). Doubly robust inference with missing data in survey sampling. *Statistica Sinica*, **24** :375–394.

[13] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63** :581–592.

[14] Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer Series in Statistics.

# Annexe A

Nous définissons le cadre asymptotique au moyen duquel nous établirons les propriétés asymptotiques des estimateurs. En théorie de l'échantillonnage, nous ne pouvons pas simplement considérer que $n \to \infty$ car nous avons $n < N$ et $N$ est fixe et finie. Nous considérons alors une suite de populations $\{U_t\}$. La population $U_1$ contient les $N_1$ premières unités, $U_2 \supset U_1$ contient les $N_2$ premières unités telles que $N_1 < N_2$, et ainsi de suite. Nous avons donc $U_1 \subset U_2 \subset U_3 \subset \ldots$ et $N_1 < N_2 < N_3 < \ldots$. Dans la population $U_t$ nous sélectionnons un échantillon $S_t$ de taille $n_t$ telle que $n_1 < n_2 < n_3 < \ldots$. Notons que les échantillons ne sont pas nécessairement imbriqués les uns dans les autres. Les propriétés des estimateurs sont établies pour $t \to \infty$, ce qui revient à laisser $n$, $N$ et $N - n$ tendre vers l'infini. Pour simplifier la notation, nous ne gardons pas l'indice $t$ dans ce qui suit.

## A.1. Démonstration de la proposition 1

Soient les conditions de régularité suivantes :

$(c_1)$ $\lim_{N\to\infty} \frac{n}{N} = \pi \in (0,1)$ ;

$(c_2)$ $\frac{1}{N}\sum_{k\in U} |y_k| = O(1)$ et $\frac{1}{N}\sum_{k\in U} y_k^2 = O(1)$ pour toute variable $y$ ;

$(c_3)$ $\min_{k\in U} \pi_k \geq \lambda > 0$, $\min_{k,l\in U} \pi_{kl} \geq \lambda^* > 0$ et $\overline{\lim_{N\to\infty}}\, n \max_{k\neq l} |\pi_{kl} - \pi_k\pi_l| < \infty$..

On veut montrer que sous ces conditions, on a

$$\frac{\widehat{t}_{y,HT} - t_y}{N} = O_P\left(1/\sqrt{n}\right).$$

**Résultat 2.** *Si* $\mathbb{V}_p\left(\frac{\hat{\theta}-\theta}{N}\right) = O\left(\frac{1}{n}\right)$ *et* $\mathbb{E}_p\left(\frac{\hat{\theta}-\theta}{N}\right) = O\left(\frac{1}{\sqrt{n}}\right)$, *alors*

$$\frac{\hat{\theta}-\theta}{N} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

**Résultat 3.** *Soient $X_n$ et $Y_n$ deux suites de variables aléatoires telles que*

$$X_n = O_P(a_n) \quad et \quad Y_n = O_P(b_n).$$

*Pour $c \in \mathbb{R}$ et $\alpha > 0$ des constantes, on a*

*(i)* $cX_n = O_P(a_n)$ ;

*(ii)* $|X_n|^\alpha = O_P(a_n^\alpha)$ ;

*(iii)* $X_n Y_n = O_P(a_n b_n)$ ;

*(iv)* $X_n + Y_n = O_P(\max\{a_n, b_n\})$.

Premièrement, $N^{-1}\mathbb{E}_p(\widehat{t}_{y,HT} - t_y) = 0$ alors la seconde condition du résultat 2 est nécessairement vérifiée. Ensuite, travaillons sur la variance. Puisqu'une variance est toujours positive, on a :

$$\begin{aligned}
\mathbb{V}_p\left(\frac{\widehat{t}_{y,HT} - t_y}{N}\right) &= \left|\mathbb{V}_p\left(\frac{\widehat{t}_{y,HT} - t_y}{N}\right)\right| \\
&= \frac{1}{N^2}\left|\mathbb{V}_p(\widehat{t}_{y,HT})\right| \\
&= \frac{1}{N^2}\left|\sum_{i\in U}\frac{1-\pi_i}{\pi_i}y_i^2 + \sum_{i\neq j}\Delta_{ij}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}\right| \\
&\leq \frac{1}{N^2}\left(\sum_{i\in U}\frac{1-\pi_i}{\pi_i}y_i^2 + \sum_{i\neq j}|\pi_{ij} - \pi_i\pi_j|\frac{|y_i|}{\pi_i}\frac{|y_j|}{\pi_j}\right) \quad \text{(Inégalité du triangle)} \\
&\leq \frac{1}{N^2}\sum_{i\in U}\frac{1-\lambda}{\lambda}y_i^2 + \frac{1}{N^2}\sum_{i\neq j}|\pi_{ij} - \pi_i\pi_j|\frac{|y_i|}{\pi_i}\frac{|y_j|}{\pi_j} \quad (c_3) \\
&\leq \frac{1-\lambda}{\lambda}\frac{1}{N}\frac{1}{N}\sum_{i\in U}y_i^2 + \max_{k\neq l}|\pi_{kl} - \pi_k\pi_l|\frac{1}{N^2}\sum_{i\neq j}\frac{|y_i|}{\pi_i}\frac{|y_j|}{\pi_j} \\
&= \frac{1-\lambda}{\lambda}\frac{1}{N}\frac{1}{N}\sum_{i\in U}y_i^2 + \max_{k\neq l}|\pi_{kl} - \pi_k\pi_l|\frac{1}{N^2}\left\{\left(\sum_{i\in U}\frac{|y_i|}{\pi_i}\right)^2 - \sum_{i\in U}\frac{y_i^2}{\pi_i^2}\right\} \\
&\leq \frac{1-\lambda}{\lambda}\frac{1}{N}\frac{1}{N}\sum_{i\in U}y_i^2 + \max_{k\neq l}|\pi_{kl} - \pi_k\pi_l|\frac{1}{N^2}\left\{\sum_{i\in U}\frac{|y_i|}{\pi_i}\right\}^2 \\
&\leq \frac{1-\lambda}{\lambda}\frac{1}{N}\frac{1}{N}\sum_{i\in U}y_i^2 + \max_{k\neq l}|\pi_{kl} - \pi_k\pi_l|\frac{1}{\lambda^2}\left\{\frac{1}{N}\sum_{i\in U}|y_i|\right\}^2
\end{aligned}$$

Finalement, puisqu'une constante est toujours $O(1)$, que $\frac{1}{N}\sum_{i\in U} y_i^2$ l'est aussi par $c_2$, on a que le premier terme de la somme est $O(1/n)$ en appliquant également $c_1$. Ensuite, le facteur du deuxième terme est $O(1/n)$ par $c_3$, tout comme le terme au carré, par $c_2$. Ainsi, en appliquant le résultat 3, on a que $\mathbb{V}_p\left(\frac{\hat{t}_y - t_y}{N}\right)$ est $O(1/n)$. Alors, par le résultat 2, on a bien que l'erreur due à l'échantillonnage est $O_P(1/\sqrt{n})$.

## A.2. Démonstration de la proposition 4

On veut montrer que $\mathbb{E}_{mpq}(\widehat{t}_I - t_y) = 0$. Cela revient à montrer que

$$\mathbb{E}_{qpm}(\widehat{t}_I - t_y) = \mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\widehat{t}_I - t_y)\}] = 0,$$

où $\widehat{t}_I = \sum_{i\in S} w_i\tilde{y}_i = \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i(1-r_i)y_i^*$ avec la valeur imputée par la régression linéaire $y_i^* = \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}$. On peut alors écrire $\widehat{t}_I$ de manière plus détaillée :

$$\begin{aligned}
\widehat{t}_I &= \sum_{i\in S_r} w_i y_i + \sum_{i\in S_{nr}} w_i y_i^* \\
&= \sum_{i\in S} w_i r_i y_i + \sum_{i\in S} w_i(1-r_i)y_i^* \\
&= \sum_{i\in U} w_i r_i I_i y_i + \sum_{i\in U} w_i(1-r_i)I_i y_i^* \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i} y_i + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \widehat{\boldsymbol{\beta}} \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i} y_i + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \left(\sum_{i\in S_r} \frac{w_i}{c_i}\mathbf{v}_i\mathbf{v}_i^\top\right)^{-1} \sum_{i\in S_r} \frac{w_i}{c_i}\mathbf{v}_i y_i \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i} y_i + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \left(\sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i\mathbf{v}_i^\top\right)^{-1} \sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i y_i
\end{aligned}$$

On calcule d'abord l'espérance selon le modèle $m$ :

$$\begin{aligned}
\mathbb{E}_m(\widehat{t}_I) &= \mathbb{E}_m\left(\sum_{i\in U} \frac{r_i I_i}{\pi_i} y_i + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \left\{\sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i\mathbf{v}_i^\top\right\}^{-1} \sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i y_i\right) \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i}\mathbb{E}_m(y_i) + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \left(\sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i\mathbf{v}_i^\top\right)^{-1} \sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i \mathbb{E}_m(y_i) \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top \left(\sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i\mathbf{v}_i^\top\right)^{-1} \sum_{i\in U} \frac{r_i I_i}{\pi_i c_i}\mathbf{v}_i\mathbf{v}_i^\top\boldsymbol{\beta} \\
&= \sum_{i\in U} \frac{r_i I_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U} \frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta}
\end{aligned}$$

Ensuite, l'espérance sous le plan d'échantillonnage :

$$\begin{aligned}
\mathbb{E}_p\{\mathbb{E}_m(\widehat{t_I})\} &= \mathbb{E}_p\left\{\sum_{i\in U}\frac{r_i I_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U}\frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta}\right\} \\
&= \sum_{i\in U}\frac{r_i\mathbb{E}_p\{I_i\}}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U}\frac{(1-r_i)\mathbb{E}_p\{I_i\}}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} \\
&= \sum_{i\in U}\frac{r_i\pi_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U}\frac{(1-r_i)\pi_i}{\pi_i}\mathbf{v}_i^\top\boldsymbol{\beta} \\
&= \sum_{i\in U} r_i\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U}(1-r_i)\mathbf{v}_i^\top\boldsymbol{\beta} \\
&= \sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta}
\end{aligned}$$

Finalement, l'espérance sous le modèle de non-réponse :

$$\begin{aligned}
\mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\widehat{t_I})\}] &= \mathbb{E}_q\left[\sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta}\right] \\
&= \sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta}
\end{aligned}$$

D'un autre côté, on a aussi

$$\begin{aligned}
\mathbb{E}_m(\mathbb{E}_p\{\mathbb{E}_q[t_y]\}) &= \mathbb{E}_m\left(\mathbb{E}_p\left\{\mathbb{E}_q\left[\sum_{i\in U} y_i\right]\right\}\right) \\
&= \mathbb{E}_m\left(\sum_{i\in U} y_i\right) = \sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta}.
\end{aligned}$$

D'où $\mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\widehat{t_I} - t_y)\}] = 0$.

Avec une imputation aléatoire par la régression linéaire, l'hypothèse $\mathbb{E}_I(e_i^*) = 0$ est formulée. En plus des espérances au regard du modèle d'imputation, de la non-réponse et du plan d'échantillonnage, on calcule l'espérance selon le modèle d'imputation utilisé pour générer

les $e_i^*$ :

$$\begin{aligned}\widehat{t}_I &= \sum_{i\in U}\frac{r_i I_i}{\pi_i}y_i + \sum_{i\in U}\frac{(1-r_i)I_i}{\pi_i}(\mathbf{v}_i^\top\widehat{\boldsymbol{\beta}} + \widehat{\sigma}\sqrt{c_i}e_i^*)\\ &= \sum_{i\in U}\frac{r_i I_i}{\pi_i}y_i + \sum_{i\in U}\frac{(1-r_i)I_i}{\pi_i}\mathbf{v}_i^\top\widehat{\boldsymbol{\beta}} + \sum_{i\in U}\frac{(1-r_i)I_i}{\pi_i}\widehat{\sigma}\sqrt{c_i}e_i^*\end{aligned}$$

$$\begin{aligned}\mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\mathbb{E}_I(\widehat{t}_I))\}] &= \sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta} + \sum_{i\in U}\frac{(1-r_i)I_i}{\pi_i}\mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\widehat{\sigma}\sqrt{c_i}\mathbb{E}_I(e_i^*))\}]\\ &= \sum_{i\in U}\mathbf{v}_i^\top\boldsymbol{\beta}\end{aligned}$$

On a à nouveau $\mathbb{E}_q[\mathbb{E}_p\{\mathbb{E}_m(\mathbb{E}_I(\widehat{t}_I - t_y))\}] = 0$.

## A.3. Démonstration de la proposition 5

Soient les conditions de régularité suivantes :

($c_1$) $\lim\limits_{N\to\infty} \dfrac{n}{N} = \pi \in (0,1)$ ;

($c_2$) $\dfrac{1}{N}\sum\limits_{k\in U} y_k = O(1)$, $\dfrac{1}{N}\sum\limits_{k\in U} |y_k| = O(1)$ et $\dfrac{1}{N}\sum\limits_{k\in U} y_k^2 = O(1)$ pour tout $y$ ;

($c_3$) $\min\limits_{k\in U} \pi_k \geq \lambda > 0$, $\min\limits_{k,l\in U} \pi_{kl} \geq \lambda^* > 0$ et $\overline{\lim\limits_{N\to\infty}}\, n \max\limits_{k\neq l} |\pi_{kl} - \pi_k\pi_l| < \infty.$ ;

($c_4$) $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_P(1/\sqrt{n})$.

On veut montrer que sous ces conditions, on a

$$\frac{\widehat{t}_I - t_y}{N} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

On décompose l'erreur $\widehat{t}_I - t_y$ entre l'erreur dûe à l'échantillonnage et l'erreur dûe à l'imputation telle que

$$\widehat{t}_I - t_y = (\widehat{t}_{y,HT} - t_y) + (\widehat{t}_I - \widehat{t}_{y,HT}).$$

On veut montrer que

$$\frac{\widehat{t}_{y,HT} - t_y}{N} = O_P\left(\frac{1}{\sqrt{n}}\right) \quad \text{et} \quad \frac{\widehat{t}_I - \widehat{t}_{y,HT}}{N} = O_P\left(\frac{1}{\sqrt{n}}\right).$$

La première partie est déjà prouvée (A.1).

Travaillons à présent avec l'erreur due à la non-réponse, $\widehat{t}_I - \widehat{t}_{y,HT}$. On a

$$\begin{aligned}
\frac{\widehat{t}_I - \widehat{t}_{y,HT}}{N} &= -\frac{1}{N}\sum_{i\in S} w_i(1-r_i)(y_i - \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}) \\
&= -\frac{1}{N}\sum_{i\in S} w_i(1-r_i)(y_i - \mathbf{v}_i\top\boldsymbol{\beta} + \mathbf{v}_i\top\boldsymbol{\beta} - \mathbf{v}_i\top\widehat{\boldsymbol{\beta}}) \\
&= -\frac{1}{N}\sum_{i\in S} w_i(1-r_i)(y_i - \mathbf{v}_i\top\boldsymbol{\beta}) + \frac{1}{N}\sum_{i\in S} w_i(1-r_i)\mathbf{v}_i\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= T_1 + T_2(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),
\end{aligned}$$

où $T_1 = -N^{-1}\sum_{i\in S} w_i(1-r_i)(y_i - \mathbf{v}_i\top\boldsymbol{\beta})$ et $T_2 = N^{-1}\sum_{i\in S} w_i(1-r_i)\mathbf{v}_i\top$.

Pour montrer que $T_1$ est $O_P(1/\sqrt{n})$, on utilise à nouveau le résultat 2. D'abord, la deuxième condition est respectée puisque

$$\begin{aligned}\mathbb{E}_{mpq}(T_1) &= \mathbb{E}_{qpm}\left(-\frac{1}{N}\sum_{i\in S} w_i(1-r_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta})\right) \\ &= \mathbb{E}_{qp}\left(-\frac{1}{N}\sum_{i\in S} w_i(1-r_i)\mathbb{E}_m(y_i - \mathbf{v}_i^\top \boldsymbol{\beta})\right) = 0.\end{aligned}$$

Ensuite, la variance $mpq$ se développe telle que

$$\mathbb{V}_{mpq}(T_1) = \mathbb{V}_m[\mathbb{E}_p\{\mathbb{E}_q(T_1)\}] + \mathbb{E}_m[\mathbb{V}_p\{\mathbb{E}_q(T_1)\}] + \mathbb{E}_m[\mathbb{E}_p\{\mathbb{V}_q(T_1)\}] = T_{11} + T_{12} + T_{13},$$

avec $T_{11} = \mathbb{V}_m[\mathbb{E}_p\{\mathbb{E}_q(T_1)\}]$, $T_{12} = \mathbb{E}_m[\mathbb{V}_p\{\mathbb{E}_q(T_1)\}]$ et $T_{13} = \mathbb{E}_m[\mathbb{E}_p\{\mathbb{V}_q(T_1)\}]$.

D'abord, pour $T_{11}$, on a l'espérance selon le modèle de non-réponse :

$$\begin{aligned}\mathbb{E}_q(T_1) &= -\frac{1}{N}\sum_{i\in U} w_i I_i(1-\mathbb{E}_q(r_i))(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \\ &= -\frac{1}{N}\sum_{i\in U} w_i I_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}).\end{aligned}$$

Ensuite, l'espérance sous le plan d'échantillonnage :

$$\begin{aligned}\mathbb{E}_p\{\mathbb{E}_q(T_1)\} &= -\frac{1}{N}\sum_{i\in U} w_i \mathbb{E}_p\{I_i\}(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \\ &= -\frac{1}{N}\sum_{i\in U} w_i \pi_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \\ &= -\frac{1}{N}\sum_{i\in U} (1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}).\end{aligned}$$

Finalement, on prend la variance par rapport au modèle $m$ pour obtenir $T_{11}$ :

$$\begin{aligned}T_{11} = \mathbb{V}_m[\mathbb{E}_p\{\mathbb{E}_q(T_1)\}] &= \frac{1}{N^2}\sum_{i\in U}(1-p_i)^2 \mathbb{V}_m[y_i - \mathbf{v}_i^\top \boldsymbol{\beta}] \quad \text{(par indépendance)} \\ &= \frac{1}{N^2}\sum_{i\in U}(1-p_i)^2 \sigma^2 c_i.\end{aligned}$$

Ensuite, pour $T_{12}$, on commence par l'espérance selon le modèle de non-réponse :

$$\begin{aligned}\mathbb{E}_q(T_1) &= -\frac{1}{N}\sum_{i\in U} w_i I_i(1-\mathbb{E}_q(r_i))(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \\ &= -\frac{1}{N}\sum_{i\in U} w_i I_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}).\end{aligned}$$

Puis, on prend la variance sous le plan d'échantillonnage :

$$\begin{aligned}
\mathbb{V}_p\{\mathbb{E}_q(T_1)\} &= \frac{1}{N^2}\left(\sum_{i\in U} w_i^2 \mathbb{V}_p\{I_i\}(1-p_i)^2(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2\right. \\
&\qquad \left. + \sum_{i\neq j} w_i w_j \mathrm{Cov}_p\{I_i,I_j\}(1-p_i)(1-p_j)(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})(y_j-\mathbf{v}_j^\top\boldsymbol{\beta})\right) \\
&= \frac{1}{N^2}\left(\sum_{i\in U} \frac{\pi_i(1-\pi_i)}{\pi_i^2}(1-p_i)^2(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2\right. \\
&\qquad \left. + \sum_{i\neq j} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}(1-p_i)(1-p_j)(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})(y_j-\mathbf{v}_j^\top\boldsymbol{\beta})\right) \\
&= \frac{1}{N^2}\left(\sum_{i\in U} \frac{1-\pi_i}{\pi_i}(1-p_i)^2(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2\right. \\
&\qquad \left. + \sum_{i\neq j} \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}(1-p_i)(1-p_j)(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})(y_j-\mathbf{v}_j^\top\boldsymbol{\beta})\right)..
\end{aligned}$$

Finalement, on obtient $T_{12}$ en calculant l'espérance selon le modèle $m$ :

$$\begin{aligned}
T_{12} = \mathbb{E}_m[\mathbb{V}_p\{\mathbb{E}_q(T_1)\}] &= \frac{1}{N^2}\sum_{i\in U}\frac{1-\pi_i}{\pi_i}(1-p_i)^2\mathbb{E}_m[(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2] \\
&\qquad + \frac{1}{N^2}\sum_{i\neq j}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}(1-p_i)(1-p_j)\mathbb{E}_m[y_i-\mathbf{v}_i^\top\boldsymbol{\beta}]\mathbb{E}_m[y_j-\mathbf{v}_j^\top\boldsymbol{\beta}] \\
&= \frac{1}{N^2}\sum_{i\in U}\frac{1-\pi_i}{\pi_i}(1-p_i)^2\mathbb{V}_m[y_i] \\
&= \frac{1}{N^2}\sum_{i\in U}\frac{1-\pi_i}{\pi_i}(1-p_i)^2\sigma^2 c_i.
\end{aligned}$$

Enfin, pour $T_{13}$, on commence par calculer la variance selon le modèle de non-réponse. On a, par indépendance :

$$\begin{aligned}
\mathbb{V}_q(T_1) &= \frac{1}{N^2}\sum_{i\in U} w_i^2 I_i^2 \mathbb{V}_q(1-r_i)(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2 \\
&= \frac{1}{N^2}\sum_{i\in U} w_i^2 I_i^2 p_i(1-p_i)(y_i-\mathbf{v}_i^\top\boldsymbol{\beta})^2.
\end{aligned}$$

On prend ensuite l'espérance sous le plan d'échantillonnage :

$$\begin{aligned}
E_p\{\mathbb{V}_q(T_1)\} &= \frac{1}{N^2}\sum_{i\in U} w_i^2 \mathbb{E}_p\{I_i^2\} p_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta})^2 \\
&= \frac{1}{N^2}\sum_{i\in U} w_i^2 (\pi_i(1-\pi_i) + \pi_i^2) p_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta}) \\
&= \frac{1}{N^2}\sum_{i\in U} w_i p_i(1-p_i)(y_i - \mathbf{v}_i^\top \boldsymbol{\beta})^2.
\end{aligned}$$

Et on obtient $T_{13}$ avec l'espérance par rapport au modèle $m$ :

$$\begin{aligned}
T_{13} = \mathbb{E}_m[\mathbb{E}_p\{\mathbb{V}_q(T_1)\}] &= \frac{1}{N^2}\sum_{i\in U} w_i p_i(1-p_i)\mathbb{E}_m[(y_i - \mathbf{v}_i^\top \boldsymbol{\beta})^2] \\
&= \frac{1}{N^2}\sum_{i\in U} w_i p_i(1-p_i)\mathbb{V}_m[y_i] \\
&= \frac{1}{N^2}\sum_{i\in U} w_i p_i(1-p_i)\sigma^2 c_i.
\end{aligned}$$

Finalement, en additionnant $T_{11}$, $T_{12}$ et $T_{13}$, on obtient la variance souhaitée

$$\begin{aligned}
\mathbb{V}_{mpq}(T_1) &= \frac{1}{N^2}\sum_{i\in U}(1-p_i)^2\sigma^2 c_i + \frac{1-\pi_i}{\pi_i}(1-p_i)^2\sigma^2 c_i + \frac{1}{\pi_i}p_i(1-p_i)\sigma^2 c_i \\
&= \frac{1}{N^2}\sum_{i\in U}\frac{1-p_i}{\pi_i}(\pi_i(1-p_i) + (1-\pi_i)(1-p_i) + p_i)\sigma^2 c_i \\
&= \frac{1}{N^2}\sum_{i\in U}\frac{1-p_i}{\pi_i}\sigma^2 c_i \\
&\leq \frac{1}{N^2}\frac{\sigma^2}{\lambda}\sum_{i\in U}(1-p_i)c_i \quad (c_3) \\
&= \frac{1}{N}\frac{\sigma^2}{\lambda}\frac{1}{N}\sum_{i\in U}(1-p_i)c_i \\
&\leq \frac{1}{N}\frac{\sigma^2}{\lambda}\frac{1}{N}\sum_{i\in U} c_i.
\end{aligned}$$

Ainsi, puisque $1/N = O(1/n)$, que la constante $\sigma^2/\lambda$ et la moyenne $N^{-1}\sum_{i\in U} c_i$ sont $O(1)$, alors en appliquant les résultats 2 et 3, $T_1$ est $O_P(1/\sqrt{n})$.

Montrons que $T_2$ est $O_P(1/\sqrt{n})$ plus directement :

$$
\begin{aligned}
T_2 &= \frac{1}{N}\sum_{i\in S} w_i(1-r_i)\mathbf{v}_i^\top \\
&= \frac{1}{N}\left(\sum_{i\in S} w_i\mathbf{v}_i^\top - \sum_{i\in S} w_i r_i \mathbf{v}_i^\top\right) \\
&= \frac{1}{N}\left(\widehat{\mathbf{t}}_{\mathbf{v}}^\top - \mathbf{t}_{\mathbf{v}}^\top + \mathbf{t}_{\mathbf{v}}^\top - \sum_{i\in S} w_i r_i \mathbf{v}_i^\top\right) \\
&= \frac{\widehat{\mathbf{t}}_{\mathbf{v}}^\top - \mathbf{t}_{\mathbf{v}}^\top}{N} - \frac{\widehat{\mathbf{t}}_{\mathbf{z}}^\top - \mathbf{t}_{\mathbf{z}}^\top}{N} + \frac{1}{N}\left(\sum_{i\in U}\mathbf{v}_i^\top - \sum_{i\in U} r_i\mathbf{v}_i^\top\right) \quad \text{(en posant } \mathbf{z}_i = r_i\mathbf{v}_i\text{)} \\
&= \frac{\widehat{\mathbf{t}}_{\mathbf{v}}^\top - \mathbf{t}_{\mathbf{v}}^\top}{N} - \frac{\widehat{\mathbf{t}}_{\mathbf{z}}^\top - \mathbf{t}_{\mathbf{z}}^\top}{N} + \frac{1}{N}\sum_{i\in U}(1-r_i)\mathbf{v}_i^\top \\
&\leq \frac{\widehat{\mathbf{t}}_{\mathbf{v}}^\top - \mathbf{t}_{\mathbf{v}}^\top}{N} - \frac{\widehat{\mathbf{t}}_{\mathbf{z}}^\top - \mathbf{t}_{\mathbf{z}}^\top}{N} + \frac{1}{N}\sum_{i\in U}\mathbf{v}_i^\top
\end{aligned}
$$

On a déjà montré qu'une erreur due à l'échantillonnage est $O_P(1/\sqrt{n})$, alors les deux premiers termes à droite le sont et le troisième terme étant une moyenne, il est $O_P(1)$ par $c_2$. Alors, en appliquant le résultat 3, on a que $T_2 = O_P(1/\sqrt{n})$ puisque les deux premiers termes multipliés par $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ deviennent $O_P(1/n)$ et le troisième $O_P(1/\sqrt{n})$. Finalement, l'erreur due à la non-réponse est, comme l'erreur due à l'échantillonnage, $O_P(1/\sqrt{n})$, donc l'erreur globale $N^{-1}(\widehat{t_I} - t_y)$ l'est également.

# Annexe B

## B.1. Dérivées nécessaires à l'estimation du biais conditionnel (2.4.4)

Soient $\widehat{\mathbf{A}}_{\boldsymbol{\alpha}} = (\widehat{\mathbf{A}}_{\boldsymbol{\alpha}}^1, \ldots, \widehat{\mathbf{A}}_{\boldsymbol{\alpha}}^J)^\top$ et $\widehat{\mathbf{A}}_{\boldsymbol{\beta}} = (\widehat{\mathbf{A}}_{\boldsymbol{\beta}}^1, \ldots, \widehat{\mathbf{A}}_{\boldsymbol{\beta}}^L)^\top$. On a la variable $\widehat{\psi}_i$ définie par

$$\begin{aligned}
\widehat{\psi}_j &= \frac{r_j}{\widehat{p}_j} y_j + \left(1 - \frac{r_j}{\widehat{p}_j}\right) \mathbf{h}_j^\top \widehat{\boldsymbol{\tau}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\alpha}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\alpha}} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}} \\
&\quad + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_p}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_p} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\tau}} \\
&= \frac{r_i}{\widehat{p}_i} y_i + \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}} + \sum_{j=1}^{J} \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\alpha}^j}\right) \widehat{\mathbf{A}}_{\boldsymbol{\alpha}}^j + \sum_{\ell=1}^{L} \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}^\ell}\right) \widehat{\mathbf{A}}_{\boldsymbol{\beta}}^\ell \\
&\quad + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_p}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_p} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m}\right) \widehat{\mathbf{A}}_{\boldsymbol{\eta}_m} + \left(\frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}}\right) \widehat{\mathbf{A}}_{\boldsymbol{\tau}}.
\end{aligned}$$

On note $[\cdot]_{k=1,\ldots,K}$ un vecteur ligne de taille $K$. Les dérivées de l'estimateur $\widehat{t}_{MR}$ selon les $J + L + 3$ paramètres sont les suivantes :

$$(i) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\alpha}^j} = -\sum_{i \in S} w_i r_i \left(y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}\right) \frac{\widehat{\eta}_{pj}^2}{\widehat{\boldsymbol{\eta}}_p{}^\top \widehat{\boldsymbol{\eta}}_p} \frac{1}{\widehat{p}_i^2} \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j};$$

$$(ii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}^\ell} = \sum_{i \in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \widehat{\boldsymbol{\tau}}^\top \begin{bmatrix} \mathbf{0} \\ \dfrac{\widehat{\eta}_{m\ell}^2}{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m} \dfrac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \end{bmatrix};$$

$$(iii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_p} = -\sum_{i \in S} w_i r_i \left(y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}\right) \frac{1}{\widehat{p}_i^2} \left[2\widehat{\eta}_{pj} p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j) \frac{\widehat{\boldsymbol{\eta}}_p{}^\top \widehat{\boldsymbol{\eta}}_p - \widehat{\eta}_{pj}^2}{\widehat{\boldsymbol{\eta}}_p{}^\top \widehat{\boldsymbol{\eta}}_p}\right]_{j=1,\ldots,J};$$

$$(iv) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m} = \sum_{i \in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \widehat{\boldsymbol{\tau}}^\top \left[ \begin{array}{c} \mathbf{0} \\ 2\widehat{\eta}_{m\ell} m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell) \dfrac{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m - \widehat{\eta}^2_{m\ell}}{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m} \end{array} \right]_{\ell=1,\ldots,L};$$

$$(v) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}} = \sum_{i \in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top.$$

La dérivée nécessaire au développement de Taylor $\widehat{\mathbf{A}}^j_{\boldsymbol{\alpha}},\ j = 1, \ldots, J$ est :

$$\begin{aligned} \frac{\partial \widehat{S}^j_{\widehat{\boldsymbol{\alpha}}}}{\partial \boldsymbol{\alpha}^j} = \sum_{i \in S} w_i \Bigg( & \frac{-\dfrac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\{1 - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\}}{[p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\{1 - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\}]^2} \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} \\ & - \frac{\{r_i - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\} \left( \dfrac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} - \dfrac{\partial p^{j^2}(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} \right)}{[p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\{1 - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\}]^2} \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} \\ & + \frac{r_i - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\{1 - p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)\}} \frac{\partial^2 p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j \partial \boldsymbol{\alpha}^{j\top}} \Bigg). \end{aligned}$$

La dérivée nécessaire au développement de Taylor $\widehat{\mathbf{A}}^\ell_{\boldsymbol{\beta}},\ \ell = 1, \ldots, L$ est :

$$\frac{\partial \widehat{S}^\ell_{\widehat{\boldsymbol{\beta}}}}{\partial \boldsymbol{\beta}^\ell} = \sum_{i \in S_r} w_i \left( \left\{ y_i - m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell) \right\} \frac{\partial^2 m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell \partial \boldsymbol{\beta}^{\ell\top}} - \left\{ \frac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \right\}^2 \right).$$

Les dérivées nécessaires au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\eta}_p}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{p}}}{\partial \boldsymbol{\eta}_p} = -\sum_{i \in S} w_i \widehat{\mathbf{U}}_{pi} \widehat{\mathbf{U}}_{pi}^\top;$$

$$\begin{aligned} (ii) \qquad \frac{\partial \widehat{U}_{\widehat{p}}}{\partial \boldsymbol{\alpha}^j} = \sum_{i \in S} w_i \Bigg( & -\widehat{\eta}_{pj} \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} \widehat{\mathbf{U}}_{pi} \\ & + (y_i - \widehat{\mathbf{U}}_{pi}^\top \widehat{\boldsymbol{\eta}}_p) \left[ \mathbf{0}_{1\times(j-1)} \quad \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j} \quad \mathbf{0}_{1\times(J-j)} \right]^\top \Bigg). \end{aligned}$$

Les dérivées nécessaires au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\eta}_m}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\eta}_m} = -\sum_{i \in S_r} w_i \widehat{\mathbf{U}}_{mi} \widehat{\mathbf{U}}_{mi}^\top;$$

$$(ii) \qquad \frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\beta}^\ell} = \sum_{i \in S_r} w_i \left( -\widehat{\eta}_{m\ell} \frac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \widehat{\mathbf{U}}_{mi} \right.$$
$$\left. + (y_i - \widehat{\mathbf{U}}_{mi}^\top \widehat{\boldsymbol{\eta}}_m) \left[ \mathbf{0}_{1\times(\ell-1)} \quad \frac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \quad \mathbf{0}_{1\times(L-\ell)} \right]^\top \right).$$

Les dérivées nécessaires au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\tau}}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}} = -\sum_{i \in S} w_i r_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \mathbf{h}_i \mathbf{h}_i^\top;$$

$$(ii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\alpha}^j} = -\sum_{i \in S_r} w_i (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) \mathbf{h}_i \frac{\widehat{\eta}_{pj}^2}{\widehat{\boldsymbol{\eta}}_p^\top \widehat{\boldsymbol{\eta}}_p} \frac{1}{\widehat{p}_i} \frac{\partial p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j)}{\partial \boldsymbol{\alpha}^j};$$

$$(iii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}^\ell} = \sum_{j \in S_r} w_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \begin{bmatrix} -\widehat{\tau}_1 \dfrac{\widehat{\eta}_{m\ell}^2}{\widehat{\boldsymbol{\eta}}_m^\top \widehat{\boldsymbol{\eta}}_m} \dfrac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \\ (y_i - \widehat{\tau}_0 - 2\widehat{\tau}_1 \widehat{m}_i) \dfrac{\widehat{\eta}_{m\ell}^2}{\widehat{\boldsymbol{\eta}}_m^\top \widehat{\boldsymbol{\eta}}_m} \dfrac{\partial m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell)}{\partial \boldsymbol{\beta}^\ell} \end{bmatrix};$$

$$(iv) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_p} = -\sum_{i \in S_r} w_i (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) \mathbf{h}_i \frac{1}{\widehat{p}_i^2} \left[ 2\widehat{\eta}_{pj} p^j(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}^j) \frac{\widehat{\boldsymbol{\eta}}_p{}^\top \widehat{\boldsymbol{\eta}}_p - \widehat{\eta}_{pj}^2}{\widehat{\boldsymbol{\eta}}_p{}^\top \widehat{\boldsymbol{\eta}}_p} \right]_{j=1,\ldots,J};$$

$$(v) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_m} = \sum_{j \in S_r} w_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \begin{bmatrix} \widehat{\tau}_1 2\widehat{\eta}_{m\ell} m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell) \dfrac{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m - \widehat{\eta}_{m\ell}^2}{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m} \\ (y_i - \widehat{\tau}_0 - 2\widehat{\tau}_1 \widehat{m}_i) 2\widehat{\eta}_{m\ell} m^\ell(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}^\ell) \dfrac{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m - \widehat{\eta}_{m\ell}^2}{\widehat{\boldsymbol{\eta}}_m{}^\top \widehat{\boldsymbol{\eta}}_m} \end{bmatrix}_{\ell=1,\ldots,L}.$$

## B.2. Dérivées nécessaires à l'estimation du biais conditionnel (2.4.8) pour la double robustesse $1m - 1p$

On fixe les modèles

$$\widehat{p}_i = p(\mathbf{v}_i, \widehat{\boldsymbol{\alpha}}) = \frac{\exp(\mathbf{v}_i^\top \widehat{\boldsymbol{\alpha}})}{1 + \exp(\mathbf{v}_i^\top \widehat{\boldsymbol{\alpha}})} \quad \text{et} \quad \widehat{m}_i = m(\mathbf{v}_i, \widehat{\boldsymbol{\beta}}) = \mathbf{v}_i^\top \widehat{\boldsymbol{\beta}}.$$

Les dérivées de l'estimateur $\widehat{t}_{MR}$ selon les 3 paramètres sont les suivantes :

$$(i) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\alpha}} = -\sum_{i \in S} w_i r_i (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) \frac{1 - \widehat{p}_i}{\widehat{p}_i} \mathbf{v}_i^\top;$$

$$(ii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}} = \sum_{i \in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \widehat{\boldsymbol{\tau}} \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_i^\top \end{bmatrix};$$

$$(iii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}} = \sum_{j \in S} w_i \left(1 - \frac{r_i}{\widehat{p}_i}\right) \mathbf{h}_i^\top.$$

La dérivée nécessaire au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\alpha}}$ est :

$$\frac{\partial \widehat{S}_{\widehat{\boldsymbol{\alpha}}}}{\partial \boldsymbol{\alpha}} = -\sum_{i \in S} w_i \widehat{p}_i (1 - \widehat{p}_i) \mathbf{v}_i \mathbf{v}_i^\top.$$

La dérivée nécessaire au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\beta}}$ est :

$$\frac{\partial \widehat{S}_{\widehat{\boldsymbol{\beta}}}}{\partial \boldsymbol{\beta}} = -\sum_{i \in S_r} w_i \mathbf{v}_i \mathbf{v}_i^\top.$$

Les dérivées nécessaire au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\tau}}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}} = -\sum_{j \in S} w_i r_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \mathbf{h}_i \mathbf{h}_i^\top;$$

$$(ii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\alpha}} = -\sum_{j \in S} w_i r_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} (y_i - \mathbf{h}_i^\top \widehat{\boldsymbol{\tau}}) \mathbf{h}_i \mathbf{v}_i^\top;$$

$$(iii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}} = \sum_{j \in S} w_i r_i \frac{1 - \widehat{p}_i}{\widehat{p}_i} \begin{bmatrix} -\widehat{\tau}_1 \\ y_i - \widehat{\tau}_0 - 2\widehat{\tau}_1 \widehat{m}_i \end{bmatrix} \mathbf{v}_i^\top.$$

## B.3. Dérivées nécessaires à l'estimation du biais conditionnel (2.4.10) pour la multi-robustesse $2m - 0p$

On fixe les modèles

$$m^1(\mathbf{v}_i^1, \widehat{\boldsymbol{\beta}}^1) = \mathbf{v}_i^{1\top}\widehat{\boldsymbol{\beta}}^1 \quad \text{et} \quad m^2(\mathbf{v}_i^2, \widehat{\boldsymbol{\beta}}^2) = \mathbf{v}_i^{2\top}\widehat{\boldsymbol{\beta}}^2.$$

Les dérivées de l'estimateur $\widehat{t}_{MR}$ selon les 4 paramètres sont les suivantes :

$$(i) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\beta}^\ell} = \sum_{i\in S} w_i\,(1-r_i)\,\widehat{\boldsymbol{\tau}} \begin{bmatrix} \mathbf{0} \\ \dfrac{\widehat{\eta}^2_{m\ell}}{\widehat{\eta}^2_{m1}+\widehat{\eta}^2_{m2}}\mathbf{v}_i^{\ell\top} \end{bmatrix};$$

$$(ii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\eta}_m} = \sum_{i\in S} w_i\,(1-r_i)\,\widehat{\boldsymbol{\tau}}^\top \begin{bmatrix} \mathbf{0} \\ 2\widehat{\eta}_{m\ell}\dfrac{\widehat{\eta}^2_{m1}+\widehat{\eta}^2_{m2}-\widehat{\eta}^2_{m\ell}}{\widehat{\eta}^2_{m1}+\widehat{\eta}^2_{m2}}\mathbf{v}_i^{\ell\top}\widehat{\boldsymbol{\beta}}^\ell \end{bmatrix}_{\ell=1,2};$$

$$(iii) \qquad \frac{\partial \widehat{t}_{MR}}{\partial \boldsymbol{\tau}} = \sum_{i\in S} w_i(1-r_i)\mathbf{h}_i^\top.$$

La dérivée nécessaire au développement de Taylor $\widehat{\mathbf{A}}^\ell_{\boldsymbol{\beta}},\ \ell = 1,\ldots,L$ est :

$$\frac{\partial \widehat{S}^\ell_{\widehat{\boldsymbol{\beta}}}}{\partial \boldsymbol{\beta}^\ell} = -\sum_{i\in S_r} w_i \mathbf{v}_i^\ell \mathbf{v}_i^{\ell\top}$$

Les dérivées nécessaires au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\eta}_m}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\eta}_m} = -\sum_{i\in S_r} w_i \widehat{\mathbf{U}}_{mi}\widehat{\mathbf{U}}_{mi}^\top;$$

$$(ii) \qquad \frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\beta}^1} = \sum_{i\in S_r} w_i \begin{bmatrix} y_i - 2\widehat{\eta}_{m1}\mathbf{v}_i^{1\top}\widehat{\boldsymbol{\beta}}^1 - \widehat{\eta}_{m2}\mathbf{v}_i^{2\top}\widehat{\boldsymbol{\beta}}^2 \\ -\widehat{\eta}_{m1}\mathbf{v}_i^{2\top}\widehat{\boldsymbol{\beta}}^2 \end{bmatrix} \mathbf{v}_i^{1\top};$$

$$(iii) \qquad \frac{\partial \widehat{U}_{\widehat{m}}}{\partial \boldsymbol{\beta}^2} = \sum_{i\in S_r} w_i \begin{bmatrix} -\widehat{\eta}_{m2}\mathbf{v}_i^{1\top}\widehat{\boldsymbol{\beta}}^1 \\ y_i - 2\widehat{\eta}_{m2}\mathbf{v}_i^{2\top}\widehat{\boldsymbol{\beta}}^2 - \widehat{\eta}_{m1}\mathbf{v}_i^{1\top}\widehat{\boldsymbol{\beta}}^1 \end{bmatrix} \mathbf{v}_i^{2\top}.$$

Les dérivées nécessaires au développement de Taylor $\widehat{\mathbf{A}}_{\boldsymbol{\tau}}$ sont :

$$(i) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\tau}} = -\sum_{i \in S_r} w_i \mathbf{h}_i \mathbf{h}_i^\top ;$$

$$(ii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\beta}^\ell} = \sum_{i \in S} w_i \begin{bmatrix} -\widehat{\tau}_1 \\ y_i - \widehat{\tau}_0 - 2\widehat{\tau}_1 \widehat{m}_i \end{bmatrix} \frac{\widehat{\eta}^2_{m\ell}}{\widehat{\eta}^2_{m1} + \widehat{\eta}^2_{m2}} \mathbf{v}_i^{\ell\top} ;$$

$$(iii) \qquad \frac{\partial \widehat{U}_{\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{\eta}_m} = \sum_{i \in S_r} w_i \begin{bmatrix} \widehat{\tau}_1 2\widehat{\eta}_{m\ell} \mathbf{v}_i^{\ell\top} \widehat{\boldsymbol{\beta}}^\ell \dfrac{\widehat{\eta}^2_{m1} + \widehat{\eta}^2_{m2} - \widehat{\eta}^2_{m\ell}}{\widehat{\eta}^2_{m1} + \widehat{\eta}^2_{m2}} \\ (y_i - \widehat{\tau}_0 - 2\widehat{\tau}_1 \widehat{m}_i) 2\widehat{\eta}_{m\ell} \mathbf{v}_i^{\ell\top} \widehat{\boldsymbol{\beta}}^\ell \dfrac{\widehat{\eta}^2_{m1} + \widehat{\eta}^2_{m2} - \widehat{\eta}^2_{m\ell}}{\widehat{\eta}^2_{m1} + \widehat{\eta}^2_{m2}} \end{bmatrix}_{\ell=1,2}$$

# Annexe C

## C.1. Paramétrisation des distributions utilisées à la section 3.2.1

Pour générer la variable d'intérêt telle que

$$y_i|\mathbf{v}_i \sim \mathcal{D}(\mu_i, \nu_i),$$

où $\mathcal{D}$ désigne chacune des distributions normale, Gamma, lognormale et Pareto, nous devons trouver les paramètres $\alpha_i$ et $\beta_i$ de chaque distribution qui permettront d'obtenir

$$\mathbb{E}(y_i|\mathbf{v}_i) = \mu_i \quad \text{et} \quad \mathbb{V}(y_i|\mathbf{v}_i) = \nu_i.$$

Pour la distribution normale, nous cherchons $\alpha_i$ et $\beta_i$ tels que $y_i|\mathbf{v}_i \sim \mathcal{N}(\alpha_i, \beta_i^2)$ et tels que $y_i|\mathbf{v}_i \sim \mathcal{N}(\mu_i, \nu_i)$. Trivialement, on obtient $\alpha_i = \mu_i$ et $\beta_i = \sqrt{\nu_i}$.

Dans le cas de la loi $\mathcal{G}amma(\alpha_i, \beta_i)$, nous savons que

$$\mathbb{E}(y_i|\mathbf{v}_i) = \frac{\alpha_i}{\beta_i} \quad \text{et} \quad \mathbb{V}(y_i|\mathbf{v}_i) = \frac{\alpha_i}{\beta_i^2}.$$

Nous devons alors résoudre simplement un système à deux inconnues et deux équations :

$$\begin{cases} \dfrac{\alpha_i}{\beta_i} = \mu_i \\ \dfrac{\alpha_i}{\beta_i^2} = \nu_i \end{cases} \Leftrightarrow \begin{cases} \beta_i = \dfrac{\alpha_i}{\mu_i} \\ \dfrac{\alpha_i}{\frac{\alpha_i^2}{\mu_i^2}} = \nu_i \end{cases} \Leftrightarrow \begin{cases} \beta_i = \dfrac{\alpha_i}{\mu_i} \\ \dfrac{1}{\alpha_i} = \dfrac{\nu_i}{\mu_i^2} \end{cases} \Leftrightarrow \begin{cases} \beta_i = \dfrac{\alpha_i}{\mu_i} \\ \alpha_i = \dfrac{\mu_i^2}{\nu_i} \end{cases} \Leftrightarrow \begin{cases} \beta_i = \dfrac{\mu_i}{\nu_i} \\ \alpha_i = \dfrac{\mu_i^2}{\nu_i} \end{cases}.$$

Dans le cas de la lognormale, on a $y_i|\mathbf{v}_i \sim \mathcal{LN}(\alpha_i, \beta_i)$, alors

$$\mathbb{E}(y_i|\mathbf{v}_i) = \exp\left(\alpha_i + \frac{\beta_i^2}{2}\right) \quad \text{et} \quad \mathbb{V}(y_i|\mathbf{v}_i) = \{\exp(\beta_i^2) - 1\}\exp(2\alpha_i + \beta_i^2).$$

Cela nous donne le système suivant à résoudre :

$$\begin{cases} \exp\left(\alpha_i + \dfrac{\beta_i^2}{2}\right) = \mu_i \\ \{\exp(\beta_i^2) - 1\}\exp(2\alpha_i + \beta_i^2) = \nu_i \end{cases} \Leftrightarrow \begin{cases} \exp\left(\alpha_i + \dfrac{\beta_i^2}{2}\right) = \mu_i \\ \{\exp(\beta_i^2) - 1\}\mu_i^2 = \nu_i \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_i + \dfrac{\beta_i^2}{2} = \ln(\mu_i) \\ \beta_i^2 = \ln\left(\dfrac{\nu_i}{\mu_i^2} + 1\right) \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_i = \ln(\mu_i) - \dfrac{1}{2}\ln\left(\dfrac{\nu_i}{\mu_i^2} + 1\right) \\ \beta_i = \sqrt{\ln\left(\dfrac{\nu_i}{\mu_i^2} + 1\right)} \end{cases} .$$

Finalement, pour une distribution Pareto de paramètres $\alpha_i$ et $\beta_i$, les moments sont tels que

$$\mathbb{E}(y_i|\mathbf{v}_i) = \frac{\alpha_i\beta_i}{\beta_i - 1},\ \beta_i > 1 \quad \text{et} \quad \mathbb{V}(y_i|\mathbf{v}_i) = \frac{\alpha_i^2\beta_i}{(\beta_i - 1)^2(\beta_i - 2)},\ \beta_i > 2.$$

Ainsi, nous résolvons :

$$\begin{cases} \dfrac{\alpha_i\beta_i}{\beta_i - 1} = \mu_i,\ \beta_i > 1 \\ \dfrac{\alpha_i^2\beta_i}{(\beta_i - 1)^2(\beta_i - 2)} = \nu_i,\ \beta_i > 2 \end{cases} \Leftrightarrow \begin{cases} \alpha_i = \mu_i\dfrac{\beta_i - 1}{\beta_i} \\ \nu_i = \dfrac{\mu_i^2}{\beta_i(\beta_i - 2)},\ \beta_i > 2 \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_i = \mu_i\dfrac{\beta_i - 1}{\beta_i} \\ \beta_i^2 - 2\beta_i - \dfrac{\mu_i^2}{\nu_i} = 0,\ \beta_i > 2 \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_i = \mu_i\dfrac{\beta_i - 1}{\beta_i} \\ \beta_i = 1 + \sqrt{1 + \dfrac{\mu_i^2}{\nu_i}},\ \text{car } \beta_i > 2 \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_i = \mu_i\dfrac{\sqrt{1 + \dfrac{\mu_i^2}{\nu_i}}}{1 + \sqrt{1 + \dfrac{\mu_i^2}{\nu_i}}} \\ \beta_i = 1 + \sqrt{1 + \dfrac{\mu_i^2}{\nu_i}} \end{cases} .$$