Journées de méthodologie statistique de l'Insee 2022

Une approche multi-robuste dans le cadre de l'intégration de données

Sixia Chen (*), David Haziza (**)

(*) University of Oklahoma
(**) University of Ottawa

Sixia-Chen@ouhsc.edu
dhaziza@uottawa.ca

Mots-clés. Équations estimantes, Estimation par score de propension, Estimation de la variance, Imputation fractionelle, Modèle de participation.

Domaines. Intégration de données.

Résumé

Ces dernières années, l'intégration d'échantillons probabilistes et non probabilistes a suscité un intérêt croissant. Les échantillons non probabilistes sont moins chers et plus rapides à réaliser, mais les estimateurs qui en résultent sont vulnérables à un biais de sélection, car les probabilités de participation sont inconnues. Pour corriger ce biais potentiel, des procédures d'estimation basées sur des modèles paramétriques ou non paramétriques ont été discutées dans la littérature. Cependant, la validité des estimateurs résultants dépend fortement de la validité des modèles sous-jacents. En outre, les approches non paramétriques peuvent souffrir du fléau de la dimension et d'une faible efficacité. Nous proposons une approche d'intégration des données en combinant plusieurs modèles de régression et plusieurs modèles de score de propension. L'approche proposée peut être utilisée pour estimer des paramètres généraux, notamment les totaux, les moyennes, les fonctions de répartition et les percentiles. Les estimateurs qui en résultent sont multi-robustes au sens où ils restent convergents si tous les modèles, sauf un, sont mal spécifiés. Les propriétés asymptotiques des estimateurs ponctuels et de la variance sont établies. Les résultats d'une étude de simulation suggèrent que la méthode proposée se comporte bien en termes de biais et d'efficacité.

Introduction

Traditionnellement, les instituts nationaux de statistique (INS) ont collecté les données au moyen de procédures d'échantillonnage probabiliste et les inférences ont été effectuées par rapport au plan de sondage. Dans le cas d'une inférence basée sur le plan de sondage, les propriétés des estimateurs ponctuels et de variance sont évaluées par rapport au plan de sondage et des inférences valides peuvent être tirées sans s'appuyer sur la validité d'un modèle, à condition que les erreurs non dues à l'échantillonnage soient négligeables. Cela ne veut pas dire que les modèles ne peuvent pas jouer un rôle dans le cadre inférentiel basé sur le plan de sondage. En effet, l'efficacité des estimateurs ponctuels peut être améliorée en utilisant une information auxiliaire au stade de l'estimation, en capitalisant sur la relation entre une variable d'intérêt et un ensemble de prédicteurs. Les procédures d'estimation qui en résultent, appelées procédures assistées par un modèle, utilisent un modèle de travail comme véhicule afin de construire des estimateurs ponctuels, mais les estimateurs qui en résultent restent convergents par rapport plan de sondage même si le modèle est mal spécifié ; voir, par exemple, Särndal et col. (1992) et Bredit et Opsomer (2017).

Ces dernières années, les méthodes d'intégration des données provenant d'échantillons probabilistes et non probabilistes ont fait l'objet d'une grande attention, car la diminution des taux de réponse et l'augmentation des coûts de collecte des données sont devenues une préoccupation majeure. De nos jours, divers types de sources de données non probabilistes sont à la disposition des praticiens d'enquêtes, notamment les panels opt-in, les médias sociaux et les informations satellitaires. Bien que ces sources de données fournissent des données actuelles pour un grand nombre de variables et d'unités de la population, elles ne parviennent souvent pas à représenter la population cible d'intérêt en raison de biais de sélection inhérents. La présence (ou la participation) d'une unité sur une source non probabiliste est inconnue, contrairement aux procédures d'échantillonnage probabiliste, pour lesquelles les probabilités d'inclusion sont connues, en général. La manière d'intégrer des données provenant d'échantillons non probabilistes a suscité beaucoup d'attention ces dernières années. Le lecteur est invité à consulter les articles suivants : Rivers (2007), Bethlehem (2016), Elliot et Vaillant (2017), Lohr et Raghunathan (2017), Chen et aol. (2019), Beaumont (2020), et Rao (2020) pour des discussions sur les méthodes d'intégration de données.

Les procédures d'estimation de l'intégration des données peuvent être classées en trois grandes catégories : (i) pondération par calage d'un échantillon non probabiliste sur des totaux estimés à partir d'une enquête probabiliste, par exemple, Elliot et Vaillant (2017) ; (ii) appariement statistique ou imputation de masse, par exemple, Rivers (2007) ; et (iii) pondération par score de propension d'un échantillon non probabiliste, par exemple, Chen et col. (2019). Cependant, quelle que soit l'approche utilisée, la validité des estimateurs ponctuels dépend fortement de la validité du modèle supposé. Par conséquent, les estimateurs ponctuels sont vulnérables à une mauvaise spécification du modèle. Afin de fournir une certaine robustesse contre la mauvaise spécification du modèle, Chen et col. (2019) ont proposé un estimateur doublement robuste de la moyenne dUne population, incorporant les prédictions obtenues par l'ajustement d'un modèle dd'imputation et les estimations des probabilités de participation obtenues par l'ajustement d'un modèle de participation. Les procédures doublement robustes sont intéressantes car elles offrent une certaine protection contre la mauvaise spécification de l'un ou l'autre modèle. Cependant, les procédures doublement robustes ont tendance à exhiber des performances numériques médiocres lorsque les deux modèles sont mal spécifiés.

Les principales contributions de ce projet sont les suivantes : (i) nous proposons un cadre inférentiel général multi-robuste dans le contexte de l'intégration de données. Dans ce contexte, cet article constitue, à notre connaissance, la première tentative de développer des procédures d'imputation de masse et d'estimation par score de propension qui peuvent être basées sur de multiples modèles d'imputation et/ou de multiples modèles de participation. Chaque modèle peut être basé sur des fonctionnels différentes et/ou des ensembles de variables explicatives différents. Les estimateurs qui en résultent sont dits multi-robustes au sens où ils restent convergents si tous les modèles sauf un sont mal spécifiés. Dans le contexte des données manquantes, les procédures multi-robustes ont été étudiées par Han et Wang (2013), Chan et Yam (2014), Han (2014a), Han (2014b), Chen et Haziza (2017), Duan et Yin (2017) et Chen et Haziza (2019). Un certain nombre d'études empiriques ont suggéré que, contrairement aux procédures doublement robustes, les procédures multi-robustes ont tendance à exhiber de bonnes performances numériques même lorsque tous les modèles sont mal spécifiés, ce qui est une caractéristique intéressante ; voir Han (2014a) et Chen et Haziza (2107). L'incorporation de multiples modèles de participation est particulièrement intéressante dans le contexte de l'intégration de données car, contrairement au cas des données manquantes où les variables explicatives sont observées à la fois pour les répondants et les non-répondants, ces dernières ne sont généralement pas observées pour les unités qui n'ont pas participé à l'échantillon non probabiliste. (ii) Les méthodes proposées peuvent s'appliquer à tout paramètre qui peut être exprimé comme solution d'une équation estimante. Cela comprend les moyennes/totaux de population, les fonctions de répartition et les quantiles. (iii) Des procédures multi-robustes pour les quantiles ont été proposées par Han et col. (2019) dans un contexte de données manquantes. Cependant, leurs procédures étaient basées sur des modèles de régression des résultats complètement paramétriques. Nos méthodes sont basées sur des modèles d'imputation semi-paramétriques qui ne nécessitent pas d'hypothèses de distribution sur les erreurs du modèle, contrairement à Han et al. (2019).

Cet article est organisé comme suit. Dans la section 2, nous introduisons le cadre de travail. Les méthodes proposées sont présentées dans la section 3 et leurs propriétés asymptotiques sont établies dans la section 4. Les procédures d'estimation de la variance sont discutées dans la section 5. Dans la section 6, les résultats d'une étude de simulation, évaluant la performance des méthodes proposées en termes de biais, d'efficacité et de probabilité de couverture des intervalles de confiance à base normale, sont présentés. Dans la section 7, nous formulons quelques remarques finales. Les preuves techniques sont reléguées à l'annexe.

1 Cadre de travail

Considérons une population finie $\mathscr{F}_N = (\mathbf{x}_i, y_i), i \in U$, de taille N, générée à partir d'un modèle de superpopulation $\mathscr{F}$, où y désigne la variable d'intérêt et $\mathbf{x}$, un vecteur de variables entièrement observées. Nous sommes intéressés par l'estimation d'un paramètre $\boldsymbol{\theta}_0$, qui est défini comme une solution des équations d'estimation

$$\mathbb{E}\left\{\mathbf{U}(\mathbf{x}, y; \boldsymbol{\theta}_0)\right\} = \mathbf{0}, \tag{1}$$

où $\mathbf{U}(\mathbf{x}, y; \boldsymbol{\theta}_0)$ est un vecteur de fonctions d'estimation impliquant la covariable $\mathbf{x}$, la variable d'intérêt y et le paramètre d'intérêt $\boldsymbol{\theta}_0$. Nous supposons que la solution de (1) est unique et que les dimensions de $\boldsymbol{\theta}_0$ et de $\mathbf{U}(\mathbf{x}, y; \boldsymbol{\theta}_0)$ sont égales. Les cas particuliers de (1) incluent la moyenne de population de y avec $U(\mathbf{x}, y; \theta_0) = y - \theta_0$, le coefficient de régression avec $\mathbf{U}(\mathbf{x}, y; \boldsymbol{\theta}_0) = \mathbf{x}(y - \mathbf{x}^\top \boldsymbol{\theta}_0)$, le τ-ième percentile de la population avec $U(\mathbf{x}, y; \theta_0) = I(y < \theta_0) - \tau$ et le τ-ième coefficient de régression quantile de la population avec $\mathbf{U}(\mathbf{x}, y; \boldsymbol{\theta}_0) = \left\{\tau - I(y - \mathbf{x}^\top \boldsymbol{\theta}_0 < 0)\right\} \mathbf{x}$.

Considérons un échantillon S_A, de taille n_A, appelé échantillon de référence, sélectionné parmi U selon un plan de sondage probabiliste, avec des probabilités d'inclusion du premier ordre et du second ordre connues $\pi_i = \mathbb{E}(I_i)$ et $\pi_{ij} = \mathbb{E}(I_i I_j)$ pour $i \neq j$, respectivement, où I_i désigne la variable indicatrice dans l'échantillon pour l'unité $i \in U$ tel que $I_i = 1$ si $i \in S_A$ et $I_i = 0$, sinon. Nous supposons que $\pi_i > 0$ pour tout $i \in U$. Les données disponibles dans S_A sont $\{(\mathbf{x}_i, d_i), i \in S_A\}$, où $d_i = 1/\pi_i$ désigne le poids de sondage pour l'unité i.

Nous considérons également un échantillon non probabiliste S_B, de taille n_B, provenant de U. Soit δ_i un indicateur de participation associé à l'unité i tel que $\delta_i = 1$ si $i \in S_B$ et $\delta_i = 0$, sinon. Nous supposons que $\mathbf{x}$ et y sont observés pour $i \in S_B$. Ainsi, les données disponibles à partir de S_B sont $\{(\mathbf{x}_i, y_i), i \in S_B\}$. La probabilité de participation associée à l'unité i est

$$\Pr(\delta_i = 1 | \mathbf{x}_i, y_i) = \Pr(\delta_i = 1 | \mathbf{x}_i) \triangleq p(\mathbf{x}_i; \boldsymbol{\alpha}_0), \tag{2}$$

où $p(\mathbf{x}_i; \boldsymbol{\alpha}_0)$ est une fonction inconnue avec un paramètre inconnu $\boldsymbol{\alpha}_0$. C'est-à-dire que nous supposons que, conditionnellement à $\mathbf{x}$, il n'y a pas de relation résiduelle entre δ et y. Ceci est essentiellement équivalent à l'hypothèse MAR (Missing At Random), telle que définie par Rubin (1976). En outre, nous supposons que $p(\mathbf{x}_i; \boldsymbol{\alpha}_0) > 0$ pour tous les $i \in U$. Cette hypothèse, souvent appelée hypothèse de positivité, peut être violée en pratique. En effet, il est possible que $p(\mathbf{x}_i; \boldsymbol{\alpha}_0) = 0$ pour une petite fraction de la popuation ; voir Chen et al. (2020) pour une discussion sur l'hypothèse de positivité.

Nous supposons que la relation entre y et $\mathbf{x}$ peut être décrite par le modèle de régression des résultats suivant :

$$y_i = m(\mathbf{x}_i; \boldsymbol{\beta}_0) + \epsilon_i, \tag{3}$$

où $m(\mathbf{x}_i; \boldsymbol{\beta}_0)$ est une fonction inconnue avec un paramètre inconnu $\boldsymbol{\beta}_0$. Les erreurs ϵ_i sont supposées indépendantes avec $\mathbb{E}(\epsilon_i|\mathbf{x}_i) = 0$ et $\mathbb{V}(\epsilon_i|\mathbf{x}_i) = \sigma^2$. Bien que nous supposions une structure de variance homoscédastique, les méthodes proposées peuvent être naturellement étendues au cas de variances inégales.

2 Méthode proposée

Considérons une classe de modèles de régression pour la variable y, $\mathscr{M}_1 = \left\{m^{(j)}(\mathbf{x}; \boldsymbol{\beta}^{(j)}), j = 1, 2, \ldots, J\right\}$ et une classe de modèles de participation $\mathscr{M}_2 = \left\{p^{(k)}(\mathbf{x}; \boldsymbol{\alpha}^{(k)}), k = 1, 2, \ldots, K\right\}$. Les modèles dans $\mathscr{M}_1$ (respectivement $\mathscr{M}_2$) peuvent être basés sur différentes fonctionnelles et/ou différents vecteurs de variables explicatives.

À l'aide des données $\{(\mathbf{x}_i, y_i), i \in S_B\}$, les estimateurs $\widehat{\boldsymbol{\beta}}^{(j)}$ correspondant aux paramètres inconnus $\boldsymbol{\beta}^{(j)}, j = 1, 2, \ldots, J$, sont obtenus en résolvant les équations d'estimation suivantes :

$$\widehat{\mathbf{U}}_1^{(j)}(\boldsymbol{\beta}^{(j)}) = \frac{1}{N} \sum_{i \in S_B} \left\{y_i - m^{(j)}(\mathbf{x}_i; \boldsymbol{\beta}^{(j)})\right\} \left\{\frac{\partial m^{(j)}(\mathbf{x}_i; \boldsymbol{\beta}^{(j)})}{\partial \boldsymbol{\beta}^{(j)}}\right\}^\top = \mathbf{0}. \tag{4}$$

Pour chaque $i \in S_B$, nous définissons le vecteur J des valeurs prédites

$$\mathbf{v}_1(\mathbf{x}; \widehat{\boldsymbol{\eta}}) = (m^{(1)}(\mathbf{x}; \widehat{\boldsymbol{\beta}}^{(1)}), m^{(2)}(\mathbf{x}; \widehat{\boldsymbol{\beta}}^{(2)}), \ldots, m^{(J)}(\mathbf{x}; \widehat{\boldsymbol{\beta}}^{(J)}))^\top, \tag{5}$$

$\widehat{\boldsymbol{\eta}} = (\widehat{\boldsymbol{\beta}}^{(1)\top}, \widehat{\boldsymbol{\beta}}^{(2)\top}, \ldots, \widehat{\boldsymbol{\beta}}^{(J)\top})^\top$. L'estimation des paramètres du modèle $\boldsymbol{\alpha}^{(k)}$ dans $\mathscr{M}_2$ est plus délicate. Si le vecteur $\mathbf{x}_i$ était observé pour tous les $i \in U$, nous obtiendrions les estimateurs $\widehat{\boldsymbol{\alpha}}^{(j)}$ en résolvant les équations d'estimation au niveau de la population

$$\begin{aligned} \widehat{\mathbf{S}}^{(k)}(\boldsymbol{\alpha}^{(k)}) &= \frac{1}{N} \sum_{i \in U} \frac{\delta_i - p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)})}{p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)}) \left\{1 - p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)})\right\}} \left\{\frac{\partial p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)})}{\partial \boldsymbol{\alpha}^{(k)}}\right\}^\top \\ &= \mathbf{0}. \end{aligned} \tag{6}$$

Cependant, puisque $\mathbf{x}_i$ n'est pas disponible pour $i \in U - S_B$, nous exprimons (6) comme suit

$$\begin{aligned} \widehat{\mathbf{S}}^{(k)}(\boldsymbol{\alpha}^{(k)}) &= \frac{1}{N} \sum_{i \in S_B} \frac{1}{p_i^{(k)}(1 - p_i^{(k)})} \left\{\frac{\partial p_i^{(k)}}{\partial \boldsymbol{\alpha}^{(k)}}\right\}^\top - \frac{1}{N} \sum_{i \in U} \frac{1}{1 - p_i^{(k)}} \left\{\frac{\partial p_i^{(k)}}{\partial \boldsymbol{\alpha}^{(k)}}\right\}^\top \\ &= \mathbf{0}, \end{aligned} \tag{7}$$

où $p_i^{(k)} \equiv p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)})$; voir Chen et al. (2020). Le dernier terme du côté droit de (7) peut être estimé en utilisant l'échantillon de probabilité S_A. Par conséquent, les estimateurs $\widehat{\boldsymbol{\alpha}}^{(k)}$ de $\boldsymbol{\alpha}^{(k)}, k = 1, 2, \ldots, K$ peuvent être obtenus en résolvant

$$\widehat{\mathbf{U}}_2^{(k)}(\boldsymbol{\alpha}^{(k)}) = \frac{1}{N} \sum_{i \in S_B} \frac{1}{p_i^{(k)}(1 - p_i^{(k)})} \left\{\frac{\partial p_i^{(k)}}{\partial \boldsymbol{\alpha}^{(k)}}\right\}^\top - \frac{1}{N} \sum_{i \in S_A} \frac{d_i}{1 - p_i^{(k)}} \left\{\frac{\partial p_i^{(k)}}{\partial \boldsymbol{\alpha}^{(k)}}\right\}^\top = \mathbf{0}. \tag{8}$$

Pour chaque $i \in S_B$, nous définissons le vecteur des probabilités de participation estimées de taille K :

$$\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}}) = (p^{(1)}(\mathbf{x};\widehat{\boldsymbol{\alpha}}^{(1)}), p^{(2)}(\mathbf{x};\widehat{\boldsymbol{\alpha}}^{(2)}), \ldots, p^{(K)}(\mathbf{x};\widehat{\boldsymbol{\alpha}}^{(K)}))^{top}, \tag{9}$$

où $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\alpha}}^{(1)\top}, \widehat{\boldsymbol{\alpha}}^{(2)\top}, \ldots, \widehat{\boldsymbol{\alpha}}^{(K)\top})^\top$.

Nous comprimons maintenant l'information contenue dans les J modèles de régression dans $\mathscr{M}_1$, à savoir les informations contenues dans le vecteur $\mathbf{v}_1(\mathbf{x};\widehat{\boldsymbol{\eta}})$, en ajustant un modèle de régression linéaire basé sur les unités de S_B avec y comme variable indépendante et $\mathbf{v}_1(\mathbf{x};\widehat{\boldsymbol{\eta}})$ comme vecteur de variables explicatives. Le score comprimé $\widehat{m}(\mathbf{x})$ de $m(\mathbf{x};\boldsymbol{\beta})$ est

$$\widehat{\boldsymbol{\tau}}_1 = \left\{\sum_{i\in S_B} \mathbf{v}_1(\mathbf{x}_i;\widehat{\boldsymbol{\eta}})\mathbf{v}_1^\top(\mathbf{x}_i;\widehat{\boldsymbol{\eta}})\right\}^{-1} \sum_{i\in S_B} \mathbf{v}_1(\mathbf{x}_i;\widehat{\boldsymbol{\eta}}) y_i.$$

Ensuite, nous résumons l'information contenue dans les K modèles de participation en $\mathscr{M}_2$, c'est-à-dire l'information contenue dans le vecteur $\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$, en ajustant un modèle de régression linéaire avec δ comme variable indépendante et $\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$ comme vecteur de variables explicatives. Si $\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$ était disponible pour tous les $i \in U$, le score comprimé serait de la forme $\widetilde{p}(\mathbf{x}) = \widetilde{\boldsymbol{\tau}}_2^\top \mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$ avec

$$\widetilde{\boldsymbol{\tau}}_2 = \left\{\sum_{i\in U} \mathbf{v}_2(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}})\mathbf{v}_2^\top(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}})\right\}^{-1} \sum_{i\in U} \mathbf{v}_2(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}})\delta_i. \tag{10}$$

Cependant, étant donné que $\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$ n'est pas disponible pour tous les $i \in U - S_B$, l'estimateur des moindres carrés (10) ne peut pas être calculé. Nous suggérons de l'estimer par

$$\widehat{\boldsymbol{\tau}}_2 = \left\{\sum_{i\in S_A} d_i \mathbf{v}_2(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}})\mathbf{v}_2^\top(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}})\right\}^{-1} \sum_{i\in S_B} \mathbf{v}_2(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}}).$$

Le score comprimé $\widehat{p}(\mathbf{x})$ de $p(\mathbf{x};\boldsymbol{\alpha})$ is $\widehat{p}(\mathbf{x}) = \widehat{\boldsymbol{\tau}}_2^\top \mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\gamma}})$.

Remark 2.1. *Comme alternative aux scores comprimés $\widehat{m}(\mathbf{x})$ et $\widehat{p}(\mathbf{x})$, on peut utiliser leur version normalisée $\widehat{m}(\mathbf{x}) = \frac{\widehat{\boldsymbol{\tau}}_1^2}{\widehat{\boldsymbol{\tau}}_1^\top \widehat{\boldsymbol{\tau}}_1}\mathbf{v}_1(\mathbf{x};\widehat{\boldsymbol{\eta}})$ et $\widehat{p}(\mathbf{x}) = \frac{\widehat{\boldsymbol{\tau}}_2^2}{\widehat{\boldsymbol{\tau}}_2^\top \widehat{\boldsymbol{\tau}}_2}\mathbf{v}_2(\mathbf{x};\widehat{\boldsymbol{\eta}})$, respectivement. Ici, si $\mathbf{a} = (a_1, \ldots, a_h)^\top$ est un vecteur de taille h, $\mathbf{a}^2$ désigne le vecteur des coefficients élevés au carré $(a_1^2, \ldots, a_h^2)^\top$. L'utilisation des versions normalisées garantit que les prédictions se situent toujours dans les intervalles appropriées. En particulier, cette normalisation garantit que les scores de propension estimés se situent dans l'intervalle $(0,1)$. Par souci de simplicité, dans la suite, nous utilisons les versions non normalisées.*

Nous considérons maintenant trois procédures d'estimation du paramètre $\boldsymbol{\theta}_0$ défini dans (1).

(i) Un estimateur par pondération de probabilité inverse $\widehat{\boldsymbol{\theta}}_{IPW}$ de $\boldsymbol{\theta}_0$ peut être obtenu en résolvant les équations d'estimation au niveau de l'échantillon :

$$\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}_2, \widehat{\boldsymbol{\gamma}}) = \frac{1}{N}\sum_{i\in S_B} \frac{1}{\widehat{p}(\mathbf{x}_i)}\mathbf{U}(\mathbf{x}_i, y_i;\boldsymbol{\theta}) = \mathbf{0}. \tag{11}$$

Nous montrons dans la section 4 que $\widehat{\boldsymbol{\theta}}_{IPW}$ est multi-robuste au sens où il reste convergent si l'un des modèles de participation dans $\mathscr{M}_2$ est correctement spécifié.

(ii) Un estimateur par imputation massive fractionnelle $\widehat{\boldsymbol{\theta}}_{FMI}$ peut être obtenu comme suit :

(Étape 1). Obtenir les poids fractionnels $\widehat{w}_{ij} = \widehat{w}_j$ en maximisant la fonction de vraisemblance empirique suivante

$$l = \sum_{i \in S_B} \log(w_i), \tag{12}$$

sous les contraintes suivantes :

$$\sum_{i \in S_B} w_i = 1, \quad \sum_{i \in S_B} w_i \widehat{\epsilon}_i = 0, \tag{13}$$

où $\widehat{\epsilon}_i = y_i - \widehat{m}(\mathbf{x}_i)$.

(Étape 2). Obtenir l'estimateur imputé $\widehat{\boldsymbol{\theta}}_{FMI}$ en résolvant les équations d'estimation au niveau de l'échantillon

$$\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\eta}}, \widehat{\lambda}) = \frac{1}{N} \sum_{i \in S_A} d_i \sum_{j \in S_B} \widehat{w}_{ij} \mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}; \boldsymbol{\theta}) = \mathbf{0}, \tag{14}$$

avec $\widehat{w}_{ij} = \widehat{w}_j = n_B^{-1}(1 + \widehat{\lambda}\widehat{\epsilon}_j)^{-1}$ et $\widehat{y}_i^{(j)} = \widehat{m}(\mathbf{x}_i) + \widehat{\epsilon}_j$, où $\widehat{\lambda}$ est solution de la deuxième contrainte dans (13).

Nous montrons dans la section 4 que $\widehat{\boldsymbol{\theta}}_{FMI}$ est multi-robuste au sens où il reste convergent si l'un des modèles de régression dans $\mathscr{M}_1$ est correctement spécifié. Les étapes 1 et 2 de la procédure ci-dessus sont liées à la procédure de Chen et Kim (2017) proposée dans le contexte de l'imputation pour des données manquantes.

(iii) Enfin, une protection supplémentaire contre la mauvaise spécification du modèle peut être obtenue en utilisant un estimateur augmenté $\widehat{\boldsymbol{\theta}}_{AMR}$ qui peut être obtenu en résolvant

$$\begin{aligned} \widehat{\mathbf{U}}_{AMR}(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\tau}}_2, \widehat{\gamma}, \widehat{\boldsymbol{\eta}}, \widehat{\lambda}) &= \frac{1}{N} \sum_{i \in S_B} \frac{1}{\widehat{p}(\mathbf{x}_i)} \mathbf{U}(\mathbf{x}_i, y_i; \boldsymbol{\theta}) \\ &+ \frac{1}{N} \sum_{i \in S_A} d_i \sum_{j \in S_B} \widehat{w}_{ij} \mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}; \boldsymbol{\theta}) \\ &- \frac{1}{N} \sum_{i \in S_B} \frac{1}{\widehat{p}(\mathbf{x}_i)} \sum_{j \in S_B} \widehat{w}_{ij} \mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}; \boldsymbol{\theta}) = \mathbf{0}, \end{aligned} \tag{15}$$

avec $\widehat{w}_{ij}$ et $\widehat{y}_i^{(j)}$ définis précédemment. Nous montrons dans la section 4 que $\widehat{\boldsymbol{\theta}}_{AMR}$ est multi-robuste au sens où il reste convergent si l'un des modèles de $\mathscr{M}_1$ ou de $\mathscr{M}_2$ est correctement spécifié.

3 Résultats asymptotiques

Dans cette section, nous établissons les propriétés asymptotiques des estimateurs ponctuels présentés dans la section 3. En particulier, nous donnons les expressions de leur variance asymptotique qui peuvent être utile pour construire des intervalles de confiance.

Le théorème 1 ci-dessous établit les propriétés asymptotiques de l'estimateur de pondération de probabilité inverse (IPW) multi-robuste défini comme solution de (11). La preuve du théorème 1 est présentée à l'annexe B.

Theorem 1. *Sous les conditions de régularité (C1)-(C9) de l'annexe A, l'estimateur IPW $\widehat{\boldsymbol{\theta}}_{IPW}$ défini dans (11) est convergent pour $\boldsymbol{\theta}_0$ lorsque l'un des modèles de participation dans $\mathscr{M}_2$ est* *correctement spécifié. De plus, l'estimateur $\widehat{\boldsymbol{\theta}}_{IPW}$ a l'expansion asymptotique suivante :*

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{IPW} - \boldsymbol{\theta}^* &= -\mathbf{A}_0^{-1}\Big[\frac{1}{N}\sum_{i\in S_B}\left\{\frac{\mathbf{U}(\mathbf{x}_i, y_i; \boldsymbol{\theta}^*)}{p(\mathbf{x}_i; \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)} - \mathbf{A}_1\mathbf{v}_2(\mathbf{x}_i; \boldsymbol{\gamma}^*) + \mathbf{A}_2\mathbf{z}_{1i}\right\} \\ &+ \frac{1}{N}\sum_{i\in S_A} d_i\left\{\mathbf{A}_1\mathbf{v}_2(\mathbf{x}_i; \boldsymbol{\gamma}^*)\mathbf{v}_2^\top(\mathbf{x}_i; \boldsymbol{\gamma}^*)\boldsymbol{\tau}_2^* - \mathbf{A}_2\mathbf{z}_{2i}\right\}\Big] \\ &+ \mathbf{o}_p(n^{*-1/2}), \end{aligned} \tag{16}$$

où $n^* = \min(n_A, n_B)$, $\boldsymbol{\theta}^*$, $\boldsymbol{\tau}_2^*$ et $\boldsymbol{\gamma}^*$ correspondent aux limites de $\widehat{\boldsymbol{\theta}}_{IPW}$, $\widehat{\boldsymbol{\tau}}_2$ et $\widehat{\boldsymbol{\gamma}}$,

$$\mathbf{A}_0 = \frac{\partial E(\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*))}{\partial \boldsymbol{\theta}}, \quad \mathbf{A}_1 = -E\left(\frac{\partial \widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\tau}_2}\right)\left\{E\left(\frac{\partial \widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\tau}_2}\right)\right\}^{-1},$$

$$\begin{aligned}\mathbf{A}_2 = \Bigg[E\left(\frac{\partial \widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\tau}_2}\right)\left\{E\left(\frac{\partial \widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\tau}_2}\right)\right\}^{-1} E\left(\frac{\partial \widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}}\right) \\ -E\left(\frac{\partial \widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}}\right)\Bigg] \times \left\{E\left(\frac{\partial \widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}}\right)\right\}^{-1},\end{aligned} \tag{17}$$

$$\mathbf{z}_{1i} = \left(\frac{1}{p_i^{(1)*}(1-p_i^{(1)*})}\frac{\partial p_i^{(1)*}}{\partial \boldsymbol{\alpha}^{(1)*}}, \ldots, \frac{1}{p_i^{(K)*}(1-p_i^{(K)*})}\frac{\partial p_i^{(K)*}}{\partial \boldsymbol{\alpha}^{(K)*}}\right)^\top, \tag{18}$$

et

$$\mathbf{z}_{2i} = \left(\frac{1}{1-p_i^{(1)*}}\frac{\partial p_i^{(1)*}}{\partial \boldsymbol{\alpha}^{(1)*}}, \ldots, \frac{1}{1-p_i^{(K)*}}\frac{\partial p_i^{(K)*}}{\partial \boldsymbol{\alpha}^{(K)*}}\right)^\top, \tag{19}$$

avec $p_i^{(k)*} = p^{(k)}(\mathbf{x}_i; \boldsymbol{\alpha}^{(k)*})$, $\widehat{\mathbf{U}}_2(\boldsymbol{\gamma})$ et $\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2, \boldsymbol{\gamma})$ défini à l'Annexe A.

Le théorème 2 ci-dessous établit les propriétés asymptotiques de l'estimateur multi-robuste de l'estimateur imputé par imputation massive fractionnelle (IMF) défini comme solution de (14). La preuve esquissée du théorème 2 est présentée à l'annexe C. Le théorème 2 qui suit présente les résultats asymptotiques de l'estimateur imputé (FMI) proposé dans la section précédente. La preuve à du théorème 2 est présentée à l'annexe C.

Theorem 2. *Sous les conditions de régularité (C1)-(C9) de l'annexe A, l'estimateur FMI $\widehat{\boldsymbol{\theta}}_{FMI}$ est convergent pour $\boldsymbol{\theta}_0$ si l'un des modèles de régression dans $\mathscr{M}_1$ est correctement spécifié. En outre, on a expansion asymptotique suivante :*

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{FMI} - \boldsymbol{\theta}^* &= -\mathbf{B}_0^{-1}\Big[\frac{1}{N}\sum_{i\in S_B}\Big[\frac{1}{E(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\boldsymbol{g}_{2,i}^* + \boldsymbol{B}_1\mathbf{z}_{3i} \\ &+ \boldsymbol{B}_2\mathbf{v}_1(\mathbf{x}_i; \boldsymbol{\eta}^*)\left\{y_i - \mathbf{v}_1^\top(\mathbf{x}_i; \boldsymbol{\eta}^*)\boldsymbol{\tau}_1^*\right\} + \boldsymbol{B}_3\frac{\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\Big] \\ &+ \frac{1}{N}\sum_{i\in S_A}\frac{d_i}{E(\delta)}\boldsymbol{g}_{1,i}^*\Big] \\ &+ \mathbf{o}_p(n^{*-1/2}), \end{aligned} \tag{20}$$

où $\boldsymbol{\theta}^*$, $\boldsymbol{\eta}^*$, $\boldsymbol{\tau}_1^*$, et λ^* sont les limites de $\widehat{\boldsymbol{\theta}}_{FMI}$, $\widehat{\boldsymbol{\eta}}$, $\widehat{\boldsymbol{\tau}}_1$, et $\widehat{\lambda}$, $\boldsymbol{g}_{1,i}^*$ et $\boldsymbol{g}_{2,i}^*$ sont définis en (C.10) et (C.11) à l'Annexe C,

$$\boldsymbol{B}_0 = \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{B}_3 = \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \lambda} A_{0,l},$$

$$
\begin{aligned}
\boldsymbol{B}_1 &= \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\tau}_1}\left[E\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}\right]^{-1} E\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\} \\
&\times \left[E\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}\right]^{-1} - \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\eta}}\left[E\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}\right]^{-1} \\
&+ \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \lambda} A_{0,l}\boldsymbol{A}_{2,l},
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{B}_2 &= \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \lambda} A_{0,l}\boldsymbol{A}_{1,l} \\
&- \frac{\partial E\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\tau}_1}\left[E\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}\right]^{-1},
\end{aligned}
$$

avec $A_{0,l}$, $\boldsymbol{A}_{1,l}$, et $\boldsymbol{A}_{2,l}$ defined in (C.6), et

$$
\mathbf{z}_{3i} = \left((y_i - m_i^{(1)*})\frac{\partial m_i^{(1)*}}{\partial \boldsymbol{\beta}^{(1)}}, \ldots, (y_i - m_i^{(J)*})\frac{\partial m_i^{(J)*}}{\partial \boldsymbol{\beta}^{(J)}}\right)^\top, \tag{21}
$$

où $m_i^{(j)*} = m^{(j)}(\mathbf{x}_i; \boldsymbol{\beta}^{(j)})$ for $j = 1, 2, \ldots, J$.

Le théorème 3 ci-dessous établit les propriétés asymptotiques de l'estimateur augmenté (AMR) défini comme une solution de (15). La preuve du théorème 3 est présentée à l'annexe D. Le théorème 3 suivant présente les résultats asymptotiques de l'estimateur augmenté proposé dans la section précédente. La preuve du théorème 3 est présentée à l'annexe D.

Theorem 3. *Sous les conditions de régularité (C1)-(C9) de l'annexe A, l'estimateur augmenté $\widehat{\boldsymbol{\theta}}_{AMR}$ défini dans (15) est convergent pour $\boldsymbol{\theta}_0$ lorsqu'un des modèles de régression dans $\mathscr{M}_1$ ou un des modèles de participation dans $\mathscr{M}_2$ est correctement spécifié. De plus, il a l'expansion asymptotique suivante :*

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_{AMR} - \boldsymbol{\theta}^* &= -\boldsymbol{C}_0^{-1}\left(\widehat{\mathbf{U}}^*_{AMR} + \boldsymbol{C}_1\widehat{\mathbf{U}}^*_1 + \boldsymbol{C}_2\widehat{\mathbf{U}}^*_2 + \boldsymbol{C}_3\widehat{\mathbf{U}}^*_3 + \boldsymbol{C}_4\widehat{\mathbf{U}}^*_4 + \boldsymbol{C}_5\widehat{U}^*_5\right) \\
&+ \mathbf{o}_p(n^{*-1/2}),
\end{aligned} \tag{22}
$$

avec

$$
\begin{aligned}
\boldsymbol{C}_1 &= \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\tau}_1}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_3}{\partial \boldsymbol{\tau}_1}\right)\right\}^{-1} E\left(\frac{\partial \widehat{\mathbf{U}}^*_3}{\partial \boldsymbol{\eta}}\right)\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_1}{\partial \boldsymbol{\eta}}\right)\right\}^{-1} \\
&- \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\eta}}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_1}{\partial \boldsymbol{\eta}}\right)\right\}^{-1} + \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \lambda} A_{0,l}\boldsymbol{A}_{2,l},
\end{aligned}
$$

$$
\boldsymbol{C}_0 = \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{C}_3 = -\frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\tau}_1}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_3}{\partial \boldsymbol{\tau}_1}\right)\right\}^{-1} + \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \lambda} A_{0,l}\boldsymbol{A}_{1,l},
$$

$$\begin{aligned}
\boldsymbol{C}_2 &= \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\tau}_2}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_4}{\partial \boldsymbol{\tau}_2}\right)\right\}^{-1} E\left(\frac{\partial \widehat{\mathbf{U}}^*_4}{\partial \gamma}\right)\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_2}{\partial \gamma}\right)\right\}^{-1} \\
&- \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \gamma}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_2}{\partial \gamma}\right)\right\}^{-1},
\end{aligned}$$

$$\boldsymbol{C}_4 = -\frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \boldsymbol{\tau}_2}\left\{E\left(\frac{\partial \widehat{\mathbf{U}}^*_4}{\partial \boldsymbol{\tau}_2}\right)\right\}^{-1}, \quad \boldsymbol{C}_5 = \frac{\partial E\left(\widehat{\mathbf{U}}^*_{AMR}\right)}{\partial \lambda} A_{0,l},$$

où $\widehat{\mathbf{U}}^*_{AMR}$, $\widehat{\mathbf{U}}^*_1$, $\widehat{\mathbf{U}}^*_2$, $\widehat{\mathbf{U}}^*_3$, $\widehat{\mathbf{U}}^*_4$, et $\widehat{U}^*_5$ définis dans l'annexe A et $A_{0,l}$, $\boldsymbol{A}_{1,l}$, et $\boldsymbol{A}_{2,l}$ sont définis dans l'annexe C.

Les développements asymptotiques donnés dans les théorèmes 1 à 3 peuvent être utilisés pour obtenir des estimateurs de variance de $\widehat{\boldsymbol{\theta}}_{IPW}$, $\widehat{\boldsymbol{\theta}}_{FMI}$ et $\widehat{\boldsymbol{\theta}}_{AMR}$.

4 Estimation de la variance par bootstrap

La variance des estimateurs proposés dans la section 3 peut être estimée par des procédures basées sur une linéarisaton de Taylor du premier ordre. Cependant, cela implique des calculs relativement fastidieux. Nous proposons une procédure bootstrap, qui est décrite ci-après.

Nous commençons par générer L échantillons bootstrap, $S_B^{*(l)}$, $l = 1, 2, \ldots, L$, à partir de S_B en utilisant un plan aléatoire simple sans remise. Soit $d_i^{(l)}$ le poids bootstrap associé à l'unité i $dansS_A$ pour $l = 1, 2, \ldots, L$.

Nous considérons d'abord l'estimateur de pondération de probabilité inverse $\widehat{\boldsymbol{\theta}}_{IPW}$ défini comme une solution de (11). Soit $\widehat{\boldsymbol{\theta}}^{*(l)}_{IPW}$ l'estimateur $\widehat{\boldsymbol{\theta}}_{IPW}$ basé sur le $l$ième échantillon bootstrap $S_B^{*(l)}$ et les poids bootstrap $d_i^{(l)}$ pour l'échantillon i $dansS_A$; c'est-à-dire que $\widehat{\boldsymbol{\theta}}^{*(l)}_{IPW}$ est obtenu en résolvant

$$\widehat{\mathbf{U}}^{*(l)}_{IPW}(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}_2^{*(l)}, \widehat{\boldsymbol{\gamma}}^{*(l)}) = \frac{1}{N}\sum_{i \in S_B^{*(l)}} \frac{1}{\widehat{p}^{*(l)}(\mathbf{x}_i)}\mathbf{U}(\mathbf{x}_i, y_i; \boldsymbol{\theta}) = \mathbf{0}, \tag{23}$$

où $\widehat{p}^{*(l)}(\mathbf{x}) = \widehat{\boldsymbol{\tau}}_2^{*(l)\top}\mathbf{v}_2(\mathbf{x}; \widehat{\boldsymbol{\gamma}}^{*(l)})$ avec

$$\widehat{\boldsymbol{\tau}}_2^{*(l)} = \left\{\sum_{i \in S_A} d_i^{(l)}\mathbf{v}_2(\mathbf{x}_i; \widehat{\boldsymbol{\gamma}}^{*(l)})\mathbf{v}_2^\top(\mathbf{x}_i; \widehat{\boldsymbol{\gamma}}^{*(l)})\right\}^{-1}\sum_{i \in S_B^{*(l)}}\mathbf{v}_2(\mathbf{x}_i; \widehat{\boldsymbol{\gamma}}^{*(l)}),$$

et $\widehat{\boldsymbol{\gamma}}^{*(l)} = (\widehat{\boldsymbol{\alpha}}^{(1)(*(l))\top}, \widehat{\boldsymbol{\alpha}}^{(2)(*(l))\top}, \ldots, \widehat{\boldsymbol{\alpha}}^{(K)(*(l))\top})^\top$ qui est obtenu comme solution de (8) en remplaçant S_B et d_i par $S_B^{*(l)}$ et $d_i^{(l)}$. De même, l'estimateur par imputation massive fractionnelle $\widehat{\boldsymbol{\theta}}^{*(l)}_{FMI}$ et l'estimateur augmenté $\widehat{\boldsymbol{\theta}}^{*(l)}_{AMR}$, basé sur le $l$ième échantillon bootstrap $S_B^{*(l)}$ et les $l$ième poids bootstrap $d_i^{(l)}$ peuvent être obtenus en utilisant les procédures décrites dans (12)- (15) en remplaçant S_B et d_i par $S_B^{*(l)}$ et $d_i^{(l)}$, respectivement.

Soit $\widehat{\boldsymbol{\theta}}$ une notation générique pour $\widehat{\boldsymbol{\theta}}^{(l)}_{IPW}$, $\widehat{\boldsymbol{\theta}}^{(l)}_{FMI}$, ou $\widehat{\boldsymbol{\theta}}^{(l)}_{AMR}$. Ensuite, un estimateur de variance bootstrap de $\widehat{\boldsymbol{\theta}}$ est donné par

$$\hat{V}_{boot} = \frac{1}{L}\sum_{l=1}^{L}\left(\widehat{\boldsymbol{\theta}}^{*(l)} - \widehat{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}}^{*(l)} - \widehat{\boldsymbol{\theta}}\right)^{top}.$$

5 Étude par simulation

Afin d'évaluer les performances des estimateurs proposés en termes de biais et d'efficacité, nous avons réalisé une étude par simulation. Nous avons généré $B = 1,000$ populations finies de taille $N = 20,000$, chacune consistant en une variable d'intérêt y et deux variables auxiliaires x_1 et x_2. Dans chaque population, les variables x_1 et x_2 ont d'abord été générées indépendamment à partir d'une distribution normale avec une moyenne égale à 1 et une variance égale à 2. Étant donné x_1 et x_2, la variable d'intérêt y a été générée selon le modèle de régression linéaire suivant

$$y = 0.3 + 2x_1 + 2x_2 + \epsilon, \tag{24}$$

où les erreurs ϵ ont été générées à partir d'une distribution normale standard.

À partir de chaque population finie, un échantillon S_A, de taille n_A, a été sélectionné par échantillonnage aléatoire simple sans remise. Nous avons utilisé $n_A = 500$ et $n_A = 1000$. En outre, un échantillon non probabiliste S_B a été généré en utilisant un plan de sondage Poissonnien. Autrement dit, les variables indicatrices de participation δ_i ont été générées indépendamment à partir d'une distribution de Bernoulli avec la probabilité suivante

$$p(\mathbf{x}_i; \boldsymbol{\alpha}) = \frac{\exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i})}{1 + \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i})}, \tag{25}$$

où $\mathbf{x}_i = (x_{1i}, x_{2i})^{top}$ et $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^{top}$. Les valeurs de α_0, α_1, et α_2 ont été choisies de manière à conduire à une taille d'échantillon n_B approximativement égale à 500 et 1000, respectivement.

Pour évaluer les performances des méthodes proposées en présence d'une mauvaise spécification du modèle, nous avons défini les variables explicatives transformées z_1 et z_2 comme étant $z_1 = \exp(x_1/2)$ et $z_2 = x_2 \{1 + \exp(x_1)\}^{-1}$. Une configuration similaire a été utilisée dans l'étude de Kang et Schaffer (2007).

Les modèles de régression et de participation correctement spécifiés ont été ajustés au moyen d'un modèle de régression linéaire et d'un modèle logistique, respectivement, sur la base de l'ensemble des variables explicatives $\mathbf{x} = (x_1, x_2)^\top$. Les modèles de régression et de participation mal spécifiés ont été ajustés respectivement au moyen d'un modèle de régression linéaire et d'un modèle logistique, sur la base de l'ensemble des variables explicatives transformées $\mathbf{z} = (z_1, z_2)^{top}$.

Nous avons supposé que seules les variables $\mathbf{x}$ et $\mathbf{z}$ étaient disponibles dans l'échantillon probabiliste S_A, alors que les variables $\mathbf{x}$, $\mathbf{z}$ et y étaient disponibles dans l'échantillon non probabiliste S_B.

Nous nous sommes intéressés à l'estimation de la moyenne de population $\theta_1 = E(y)$ et du 25ième percentile de population de y, θ_2. Nous avons calculé les estimateurs ponctuels suivants de θ_1 et θ_2 :

(1) Les estimateurs pondérés de réferece (obtenus en supposant que la variable y est observée pour tout $i \in S_A$, Benchmark), obtenus comme solution des équations d'estimation suivantes

$$\widehat{U}_{Ben}(\theta) = \frac{1}{N} \sum_{i \in S_A} d_i U(\mathbf{x}_i, y_i; \theta) = 0.$$

(2) Les estimateurs naïfs (Naive) basés sur S_B obtenus comme solution des équations d'estimation suivantes

$$\widehat{U}_{naive}(\theta) = \frac{1}{N} \sum_{i \in S_B} U(\mathbf{x}_i, y_i; \theta) = 0.$$

(3) Les estimateurs imputés par imputation massive paramétriques de Kim et al. (2018) basés sur le modèle de régression correctement spécifié (PMI(1000)) et le modèle de régression mal spécifié (PMI(0100)). Nous utilisons quatre chiffres entre parenthèses pour distinguer les estimateurs construits à partir des différents modèles. Les deux premiers chiffres correspondent aux modèles de régression corrects et incorrects, respectivement. Les deux derniers chiffres correspondent aux modèles de participation corrects et incorrects, respectivement. Par exemple, le quadruplet (1000) signifie que la procédure d'estimation est basée sur le modèle de régression correct et qu'aucun modèle de participation n'a été utilisé, tandis que le quadruplet (0100) signifie que la procédure est basée sur le modèle de régression incorrect et qu'aucun modèle de participation n'a été utilisé. Le premier chiffre définit l'utilisation (1) ou la non-utilisation (0) du modèle d'imputation correct. Le deuxième chiffre définit l'utilisation (1) ou la non-utilisation (0) du modèle d'imputation incorrect. Le troisième chiffre définit l'utilisation (1) ou la non-utilisation (0) du modèle de score de propension correct. Le quatrième chiffre définit l'utilisation (1) ou la non-utilisation (0) du modèle de score de propension incorrect.
(4) Les estimateurs doublement robustes proposés par Chen et al. (2020) : DR(1010), DR(1001), DR(0110) et DR(0101).
(5) L'estimateur de pondération de probabilité inverse donné par (11) : MRIPW(0011).
(6) L'estimateur imputé par imputation massive fractionnelle donné par (14) : MRFMI(1100).
(7) Les estimateurs robustes multipliés augmentés donnés par (15) : AMR(1110), AMR(1101), AMR(1011), AMR(0111) et AMR(1111).

Pour mesurer le biais d'un estimateur ponctuel, nous avons calculé le biais relatif Monte Carlo (RB) en pourcentage. Nous avons également calculé l'erreur standard relative Monte Carlo (RSE) et l'erreur quadratique moyenne relative Monte Carlo (RRMSE). Les résultats sont présentés dans le tableau 1. Les estimateurs doublement robustes de Chen et al. (2020) ayant été développés pour la moyenne de la population uniquement, ils ne sont pas inclus dans la partie du tableau 1 présentant les résultats pour le 25ième percentile de la population.

Comme prévu, les estimateurs de référence ont montré une RB négligeable dans tous les scénarios. D'autre part, les estimateurs naïfs ont montré un biais significatif dans tous les scénarios en raison du biais de sélection. Par exemple, pour $n_A = n_B = 500$, l'estimateur naïf a montré un biais relatif d'environ 9,4% pour la moyenne et d'environ 12,5% pour le 25ième percentile. Lorsque le modèle de régression était correctement spécifié, les estimateurs PMI présentaient un faible biais (moins de 2%). Cependant, ils étaient considérablement biaisés lorsque le modèle de régression était incorrectement spécifié. Cela suggère que les estimateurs PMI sont vulnérables à la mauvaise spécification du modèle de régression. Comme prévu, les estimateurs doublement robustes ont montré un biais négligeable lorsque le modèle de régression ou le modèle de participation était correctement spécifié. Cependant, ils ont montré un biais significatif lorsque les deux modèles étaient mal spécifiés. Les estimateurs proposés MRIPW, MRFMI et AMR ont montré un biais négligeable dans tous les scénarios, ce qui est conforme à la théorie. En termes de RSE, les estimateurs proposés ont donné de bons résultats. Les résultats suggèrent que l'incorporation d'un modèle de régression supplémentaire et/ou d'un modèle de participation supplémentaire dans la procédure n'a pas eu d'effet significatif sur l'efficacité de l'estimateur résultant. Par exemple, pour $n_A = n_B = 500$, l'estimateur AMR(1111) a montré une valeur de RRMSE de 1,59 pour la moyenne et de 3,08 pour le 25ième percentile de la population. D'autre part, l'estimateur PMI(1000) présente également une valeur de RRMSE égale à 1,59 pour la moyenne de la population et une valeur légèrement supérieure, 3,45, pour le 25ième percentile.

En outre, nous avons évalué les performances de la procédure bootstrap décrite dans la section 5 en termes de biais, de couverture et de longueur d'intervalle. Nous avons considéré le cas de $n_A = n_B = 500$ uniquement avec une taille d'échantillon bootstrap égale à 500. Les résultats concernant le biais relatif Monte Carlo en pourcentage, le taux de couverture et la longueur moyenne des intervalles de confiance sont présentés dans le tableau 2. D'après le tableau 2, nous constatons que le biais relatif de l'estimateur de variance bootstrap était relativement faible dans tous les scénarios avec un RB absolu inférieur à 10%. Les intervalles de confiance étaient bons dans tous les scénarios avec un taux de couverture proche du taux nominal.

TABLE 1 – Biais relatif Monte Carlo ($\times 10^{-2}$), Erreur standard relative (RSE) ($\times 10^{-2}$) et erreur quadratique moyenne relative (RRMSE) ($\times 10^{-2}$) pour 5 procédures d'estimation.

		$(n_A, n_B = 500)$			$(n_A, n_B = 1000)$		
Paramètre	Méthode	RB	RSE	RRMSE	RB	RSE	RRMSE
Moyenne	Benchmark	-0.04	1.59	1.59	-0.04	1.14	1.14
	Naive	9.37	1.56	9.50	9.14	1.16	9.21
	PMI(1000)	0.02	1.59	1.59	-0.03	1.13	1.13
	PMI(0100)	5.06	1.57	5.29	4.84	1.11	4.96
	DR(1010)	0.01	1.59	1.59	-0.03	1.13	1.13
	DR(1001)	0.02	1.60	1.60	-0.03	1.13	1.13
	DR(0110)	0.10	1.89	1.89	-0.04	1.34	1.34
	DR(0101)	4.60	1.69	4.90	4.42	1.19	4.58
	MRIPW(0011)	0.48	1.98	2.04	0.12	1.37	1.37
	MRFMI(1100)	0.01	1.59	1.59	-0.03	1.12	1.12
	AMR(1110)	0.01	1.59	1.59	-0.03	1.13	1.13
	AMR(1101)	0.02	1.60	1.60	-0.03	1.13	1.13
	AMR(1011)	0.01	1.59	1.59	-0.03	1.13	1.13
	AMR(0111)	0.66	1.93	2.04	0.21	1.35	1.37
	AMR(1111)	0.01	1.59	1.59	-0.03	1.13	1.13
Quantile	Benchmark	0.02	2.85	2.85	-0.05	2.04	2.04
	Naive	12.54	2.89	12.87	12.15	2.1	12.33
	PMI(1000)	1.96	2.84	3.45	1.84	2.02	2.73
	PMI(0100)	21.98	2.48	22.12	21.74	1.81	21.82
	MRIPW(0011)	0.44	3.69	3.72	-0.06	2.55	2.55
	MRFMI(1100)	0.03	2.52	2.52	-0.04	1.78	1.78
	AMR(1110)	0.83	3.03	3.14	0.19	2.14	2.15
	AMR(1101)	0.39	2.88	2.91	0.05	2.08	2.08
	AMR(1011)	0.78	2.99	3.09	0.19	2.14	2.15
	AMR(0111)	1.46	3.24	3.55	0.5	2.3	2.35
	AMR(1111)	0.77	2.99	3.08	0.2	2.15	2.16

6 Remarques finales

Dans cet article, nous avons proposé un ensemble de procédures d'estimation multi-robustes dans le contexte de l'intégration des données dans les enquêtes. Nos méthodes, qui peuvent être appliquées à tout paramètre défini comme solution d'une équation d'estimation, offrent une certaine robustesse face à la mauvaise spécification du modèle.

Nous avons établi les propriétés théoriques des estimateurs ponctuels basés sur des modèles de régression paramétriques et/ou des modèles de participation paramétriques. Bien que la mise en oeuvre des méthodes proposées dans le cas de modèles de régression non paramétriques (y

Table 2 – Biais relatif Monte Carlo (RB) des estimateurs de variance bootstrap, taux de couverture (CR) %, et longueur moyenne (AL) des intervalles de confiance pour $n_A = n_B = 500$.

Paramètre	Méthode	RB	CR	AL
Moyenne	MRIPW(0011)	6.56	95.7	0.66
	MRFMI(1100)	2.28	95.3	0.52
	AMR(1110)	2.29	95.3	0.52
	AMR(1101)	2.25	95.4	0.52
	AMR(1011)	2.30	95.3	0.52
	AMR(0111)	9.00	96.4	0.66
	AMR(1111)	2.30	95.3	0.52
Quantile	MRIPW(0011)	9.37	95.5	0.94
	MRFMI(1100)	-0.42	94.7	0.63
	AMR(1110)	7.79	95.0	0.78
	AMR(1101)	4.41	95.3	0.74
	AMR(1011)	7.58	95.1	0.77
	AMR(0111)	9.28	96.3	0.89
	AMR(1111)	7.87	95.1	0.77

compris les procédures d'apprentissage automatique) soit simple, établir les propriétés des estimateurs ponctuels en résultant est un sujet de recherche future. En outre, l'utilisation de modèles de participation non paramétriques présente certains défis et nécessite des recherches supplémentaires, car l'extension de (8) au cas de la méthode non paramétrique n'est pas triviale.

Dans cet article, nous avons considéré le problème de l'estimation d'un paramètre de population fini défini comme une fonction d'une seule variable d'intérêt y. Dans la pratique, on peut être intéressé à préserver la relation entre plusieurs variables d'intérêt. L'utilisation de procédures d'estimation multi-robustes dans le cas de paramètres mesurant la relation entre les variables (par exemple, un coefficient de corrélation) est actuellement à l'étude. Dans cet article, nous avons utilisé une procédure de régression linéaire pour combiner les scores $\widehat{m}(\mathbf{x})$ et $\widehat{p}(\mathbf{x})$. La recherche de la meilleure façon de combiner ces scores fera l'objet d'un article séparé.

Bibliographie

[1] Beaumont, J. (2020). Are probability surveys bound to disappear for the production of official statistics ?, *Survey Methodology* 46(1), 1-28.

[2] Bethlehem, J. (2016). Solving the nonresponse problem with sample matching ?, *Social Science Computer Review* 34(1), 59-77.

[3] Breidt, F. J. et Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques, *Statistical Science* 32(2), 190-205.

[4] Chan, K. C. G. et Yam, S. C. P. (2014). Oracle, multiple robust et multipurpose calibration in a missing response problem, *Statistical Science* 29(3), 380-396.

[5] Chen, S. et Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys, *Biometrika* 104(2), 439-453.

[6] Chen, S. et Haziza, D. (2019). Multiply robust nonparametric multiple imputation for the treatment of missing data, *Statistica Sinica* 29(4), 2035-2053.

[7] Chen, S. et Kim, J. K. (2017). Semiparametric fractional imputation using empirical likelihood in survey sampling, *Statistical theory et related fields* 1(1), 69-81.

[8] Chen, Y., Li, P. et Wu, C. (2019). Doubly robust inference with nonprobability survey samples, Journal of the American Statistical Association pp. DOI : 10.1080/01621459.2019.1677241, https ://www.tandfonline.com/doi/abs/10.1080/01621459.2019.1677241.

[9] Duan, X. et Yin, G. (2017). Ensemble approaches to estimating the population mean with missing response, *Scandinavian Journal of Statistics* 44(4), 899-917.

[10] Elliott, M. R. et Valliant, R. (2017). Inference for nonprobability samples, *Statistical Science* 32(2), 249-264.

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

[12] Han, P. (2014b). Multiply robust estimation in regression analysis with missing data, *Journal of American Statistical Association* 109(507), 1159-1173.

[13] Han, P., Kong, L., Zhao, J. et Zhou, X. (2019). A general framework for quantile estimation with incomplete data, *Journal of the Royal Statistical Society : Series B* (Statistical Methodology) 81(2), 305-333.

[14] Han, P. et Wang, L. (2013). Estimation with missing data : beyond double robustness, *Biometrika* 100(2), 417-430.

[15] Lohr, S. L. et Raghunathan, T. E. (2017). Combining survey data with other data sources, *Statistical Science* 32(2), 293-312.

[16] Rao, J. N. K. (2020). On making valid inferences by integrating data from surveys et other sources, *Sankhya B*, In Process.

[17] Rivers, D. (2007). Sampling for web surveys, *Proceedings of the Survey Research Methods Section of the American Statistical Association.*

[18] Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model-Assisted Survey Sampling.* Springer.

7 Annexes

A Conditions de régularité

Soit

$$\boldsymbol{\eta} = (\boldsymbol{\beta}^{(1)\top}, \ldots, \boldsymbol{\beta}^{(J)\top})^\top, \quad \boldsymbol{\gamma} = (\boldsymbol{\alpha}^{(1)\top}, \ldots, \boldsymbol{\alpha}^{(K)\top})^\top.$$

On définit

$$\widehat{\mathbf{U}}_1(\boldsymbol{\eta}) = (\widehat{\mathbf{U}}_1^{(1)\top}(\boldsymbol{\beta}^{(1)}), \ldots, \widehat{\mathbf{U}}_1^{(J)\top}(\boldsymbol{\beta}^{(J)}))^\top, \quad \widehat{\mathbf{U}}_2(\boldsymbol{\gamma}) = (\widehat{\mathbf{U}}_2^{(1)\top}(\boldsymbol{\alpha}^{(1)}), \ldots, \widehat{\mathbf{U}}_2^{(K)\top}(\boldsymbol{\alpha}^{(K)}))^\top,$$

$$\widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1,\boldsymbol{\eta}) = \frac{1}{N}\sum_{i\in S_B}\mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta})\left\{y_i - \mathbf{v}_1^\top(\mathbf{x}_i;\boldsymbol{\eta})\boldsymbol{\tau}_1\right\},$$

$$\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2,\gamma) = \frac{1}{N}\sum_{i\in S_A} d_i\mathbf{v}_2(\mathbf{x}_i;\gamma)\mathbf{v}_2^\top(\mathbf{x}_i;\gamma)\boldsymbol{\tau}_2 - \frac{1}{N}\sum_{i\in S_B}\mathbf{v}_2(\mathbf{x}_i;\gamma),$$

et

$$\widehat{U}_5(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda) = \frac{1}{N}\sum_{i\in S_B}\frac{\widehat{\epsilon}_i(\boldsymbol{\tau}_1,\boldsymbol{\eta})}{1+\lambda\widehat{\epsilon}_i(\boldsymbol{\tau}_1,\boldsymbol{\eta})},$$

avec $\widehat{\epsilon}_i(\boldsymbol{\tau}_1,\boldsymbol{\eta}) = y_i - \boldsymbol{\tau}_1^\top\mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta})$. De plus, on définit $\mathbf{U}_1^*(\boldsymbol{\eta}) = \mathbb{E}\left\{\widehat{\mathbf{U}}_1(\boldsymbol{\eta})\right\}$, $\mathbf{U}_2^*(\gamma) = \mathbb{E}\left\{\widehat{\mathbf{U}}_2(\gamma)\right\}$, $\mathbf{U}_3^*(\boldsymbol{\tau}_1,\boldsymbol{\eta}) = \mathbb{E}\left\{\widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1,\boldsymbol{\eta})\right\}$, $\mathbf{U}_4^*(\boldsymbol{\tau}_2,\gamma) = \mathbb{E}\left\{\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2,\gamma)\right\}$, $U_5^*(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda) = \mathbb{E}\left\{\widehat{U}_5(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda)\right\}$,

$$\mathbf{U}_{IPW}^*(\boldsymbol{\theta},\boldsymbol{\tau}_2,\gamma) = \mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^\top\mathbf{v}_2(\mathbf{x}_i;\gamma)}\mathbf{U}(\mathbf{x}_i,y_i;\boldsymbol{\theta})\right\}$$

,

$$\mathbf{U}_{FMI}^*(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda) = \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\left\{1+\lambda\widehat{\epsilon}_j(\boldsymbol{\tau}_1,\boldsymbol{\eta})\right\}^{-1}\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)}(\boldsymbol{\tau}_1,\boldsymbol{\eta});\boldsymbol{\theta})\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}},$$

où $\widehat{y}_i^{(j)}(\boldsymbol{\tau}_1,\boldsymbol{\eta}) = \boldsymbol{\tau}_1^\top\mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta}) + \widehat{\epsilon}_j(\boldsymbol{\tau}_1,\boldsymbol{\eta})$,

$$\begin{aligned}\mathbf{U}_{AMR}^*(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\gamma,\boldsymbol{\eta},\lambda) &= \mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^\top\mathbf{v}_2(\mathbf{x}_i;\gamma)}\mathbf{U}(\mathbf{x}_i,y_i;\boldsymbol{\theta})\right\}\\ &+ \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\left\{1+\lambda\widehat{\epsilon}_j(\boldsymbol{\tau}_1,\boldsymbol{\eta})\right\}^{-1}\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)}(\boldsymbol{\tau}_1,\boldsymbol{\eta});\boldsymbol{\theta})\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}}\\ &- \frac{1}{p(\boldsymbol{\alpha}_0)}\mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)p(\mathbf{x}_j;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^\top\mathbf{v}_2(\mathbf{x}_i;\gamma)\left\{1+\lambda\widehat{\epsilon}_j(\boldsymbol{\tau}_1,\boldsymbol{\eta})\right\}}\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)}(\boldsymbol{\tau}_1,\boldsymbol{\eta});\boldsymbol{\theta})\right\},\end{aligned}$$

où $p(\boldsymbol{\alpha}_0) = \mathbb{E}\left\{p(\mathbf{x};\boldsymbol{\alpha}_0)\right\}$. Soit

$$\widehat{\mathbf{U}}(\boldsymbol{\phi}) = \left(\widehat{\mathbf{U}}_2^\top(\gamma),\widehat{\mathbf{U}}_4^\top(\boldsymbol{\tau}_2,\gamma),\widehat{\mathbf{U}}_{IPW}^\top(\boldsymbol{\theta},\boldsymbol{\tau}_2,\gamma)\right)^\top,$$

et

$$\mathbf{U}^*(\boldsymbol{\phi}) = \left(\mathbf{U}_2^{*\top}(\gamma),\mathbf{U}_4^{*\top}(\boldsymbol{\tau}_2,\gamma),\mathbf{U}_{IPW}^{*\top}(\boldsymbol{\theta},\boldsymbol{\tau}_2,\gamma)\right)^\top,$$

avec $\boldsymbol{\phi} = \left(\boldsymbol{\theta}^\top,\boldsymbol{\tau}_2^\top,\gamma^\top\right)^\top$, or

$$\widehat{\mathbf{U}}(\boldsymbol{\phi}) = \left(\widehat{\mathbf{U}}_1^\top(\boldsymbol{\eta}),\widehat{\mathbf{U}}_3^\top(\boldsymbol{\tau}_1,\boldsymbol{\eta}),\widehat{U}_5^\top(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda),\widehat{\mathbf{U}}_{FMI}^\top(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda)\right)^\top,$$

et

$$\mathbf{U}^*(\boldsymbol{\phi}) = \left(\mathbf{U}_1^{*\top}(\boldsymbol{\eta}),\mathbf{U}_3^{*\top}(\boldsymbol{\tau}_1,\boldsymbol{\eta}),U_5^{*\top}(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda),\mathbf{U}_{FMI}^{*\top}(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda)\right)^\top,$$

avec $\boldsymbol{\phi} = \left(\boldsymbol{\theta}^\top,\boldsymbol{\tau}_1^\top,\boldsymbol{\eta}^\top,\lambda\right)^\top$, or

$$\widehat{\mathbf{U}}(\boldsymbol{\phi}) = \left(\widehat{\mathbf{U}}_1^\top(\boldsymbol{\eta}),\widehat{\mathbf{U}}_2^\top(\gamma),\widehat{\mathbf{U}}_3^\top(\boldsymbol{\tau}_1,\boldsymbol{\eta}),\widehat{\mathbf{U}}_4^\top(\boldsymbol{\tau}_2,\gamma),\widehat{U}_5^\top(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda),\widehat{\mathbf{U}}_{AMR}^\top(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\gamma,\boldsymbol{\eta},\lambda)\right)^\top,$$

et

$$\mathbf{U}^*(\boldsymbol{\phi}) = \left(\mathbf{U}_1^{*\top}(\boldsymbol{\eta}),\mathbf{U}_2^{*\top}(\gamma),\mathbf{U}_3^{*\top}(\boldsymbol{\tau}_1,\boldsymbol{\eta}),\mathbf{U}_4^{*\top}(\boldsymbol{\tau}_2,\gamma),U_5^{*\top}(\boldsymbol{\tau}_1,\boldsymbol{\eta},\lambda),\mathbf{U}_{AMR}^{*\top}(\boldsymbol{\theta},\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\gamma,\boldsymbol{\eta},\lambda)\right)^\top,$$

$\boldsymbol{\phi} = \left(\boldsymbol{\theta}^\top,\boldsymbol{\tau}_1^\top,\boldsymbol{\tau}_2^\top,\gamma^\top,\boldsymbol{\eta}^\top,\lambda\right)^\top$. Denote by $\mathbb{E}\left\{\widehat{\mathbf{U}}(\boldsymbol{\phi})|S_B,\mathbf{x},y\right\} = N^{-1}\sum_{i=1}^N\tilde{\mathbf{U}}(\boldsymbol{\phi};\mathbf{x}_i,y_i,\delta_i)+o_p(1)$.

Nous supposons les conditions de régularité suivantes pour établir les théorèmes 1-3 dans la section 4.

(C1). $m^{(j)}(\mathbf{x};\boldsymbol{\beta}^{(j)})$ a une dérivée première continue $\partial m^{(j)}(\mathbf{x};\boldsymbol{\beta}^{(j)})/\partial\boldsymbol{\beta}^{(j)}$ et une dérivée seconde $\partial^2 m^{(j)}(\mathbf{x};\boldsymbol{\beta}^{(j)})/\partial\boldsymbol{\beta}^{(j)}\partial\boldsymbol{\beta}^{(j)\top}$ dans le voisinage de $\boldsymbol{\beta}^*$, $\mathbb{E}\left\{m^{(j)2}(\mathbf{x};\boldsymbol{\beta}^{(j)})\right\}$, $\mathbb{E}\left\{\partial m^{(j)}(\mathbf{x};\boldsymbol{\beta})/\partial\boldsymbol{\beta}\right\}$, $\mathbb{E}||\left\{\partial m^{(j)}(\mathbf{x};\boldsymbol{\beta})/\partial\boldsymbol{\beta}\right\}||^2$, et $\mathbb{E}||\partial^2 m^{(j)}(\mathbf{x};\boldsymbol{\beta}^{(j)})/\partial\boldsymbol{\beta}^{(j)}\partial\boldsymbol{\beta}^{(j)\top}||$ sont bornés dans ce voisinage, où $j = 1, 2, \ldots, J$.

(C2). Les erreurs du modèle en (3) satisfont $\mathbb{E}(\epsilon^2) < \infty$ et $\max\{||\epsilon_i|| : i \in S_B\} = o_p(n_B^{1/2})$.

(C3). On suppose que $\widehat{\mathbf{U}}(\boldsymbol{\phi})$ converge en probabilité vers $\mathbf{U}^*(\boldsymbol{\phi})$ uniformément dans le voisinage de $\boldsymbol{\phi}^*$, qui est la solution unique de $\mathbf{U}^*(\boldsymbol{\phi}) = 0$. Pour tout $a > 0$, $\inf_{\boldsymbol{\phi}:||\boldsymbol{\phi}-\boldsymbol{\phi}^*||\geq a} ||\mathbf{U}^*(\boldsymbol{\phi})|| > 0$.

(C4). Il existe une fonction mesurable $L(\mathbf{x}, y, \delta)$ avec $\mathbb{E}\left\{L^2(\mathbf{x}, y, \delta)\right\} < \infty$ et pour tout $\boldsymbol{\phi}_1$ et $\boldsymbol{\phi}_2$ dans le voisinage de $\boldsymbol{\phi}^*$, $||\tilde{\mathbf{U}}(\boldsymbol{\phi}_1;\mathbf{x}, y, \delta) - \tilde{\mathbf{U}}(\boldsymbol{\phi}_2;\mathbf{x}, y, \delta)|| \leq L(\mathbf{x}, y, \delta)||\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2||$.

(C5). On suppose que $\mathbb{E}\left\{||\tilde{\mathbf{U}}(\boldsymbol{\phi};\mathbf{x}, y, \delta)||^2\right\} < \infty$ et $\mathbb{E}\left\{\tilde{\mathbf{U}}(\boldsymbol{\phi};\mathbf{x}, y, \delta)\right\}$ on des dérivées continues et inversibles par rapport à $\boldsymbol{\phi}$ et les dérivées sont bornées par une fonction intégrable dans le voisinage de $\boldsymbol{\phi}^*$.

(C6). $n^{*1/2}\widehat{\mathbf{U}}(\boldsymbol{\phi}^*) \to^d N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}^*})$ as $n^*, N \to \infty$, où $\Sigma_{\boldsymbol{\phi}^*} = \mathbb{V}\left\{n^{*1/2}\widehat{\mathbf{U}}(\boldsymbol{\phi}^*)\right\}$ désigne la variance par rapport à la distribution conjointe du modèle et du plan de sondage et $n^* = \min(n_A, n_B)$.

(C7). Les probabilités d'inclusion dupremier ordre π_i satisfont $K_L \leq N n_A^{-1}\pi_i \leq K_U$ pour tout i, où K_L et K_U sont des constantes positives.

(C8). Les probabilités d'inclusion de second ordre satisfont $\max_{i,j}|\pi_{ij}\pi_i^{-1}\pi_j^{-1} - 1| = o(1)$ pour $i \neq j$.

(C9). Les modèles de participation satisfont $1 > p^{(k)}(\mathbf{x};\boldsymbol{\alpha}^{(k)*}) > C > 0$ presque surement. $p^{(k)*}(\mathbf{x};\boldsymbol{\alpha}^{(k)*})$ possède une dérivée partielle $\partial p^{(k)}(\mathbf{x};\boldsymbol{\alpha}^{(k)*})/\partial\boldsymbol{\alpha}^{(k)}$ continue avec $\mathbb{E}\left\{\partial p^{(k)}(\mathbf{x};\boldsymbol{\alpha}^{(k)*})/\partial\boldsymbol{\alpha}^{(k)}\right\} < \infty$, où $k = 1, 2, \ldots, K$.

B Preuve du Théorème 1

Supposons qu'un des modèles de score de propension dans $\mathscr{M}_2$ soit correctement spécifié. Sans perte de généralité, nous supposons que $p^{(1)}(\mathbf{x};\boldsymbol{\alpha}^{(1)})$ est correctement spécifié. On peut montrer que $\widehat{\boldsymbol{\tau}}_2 \to^p (1, 0, \ldots, 0)^\top$ et $\widehat{p}(\mathbf{x}) \to^p p(\mathbf{x};\boldsymbol{\alpha}_0)$ lorsque n_A, n_B et N tendent vers l'infini. Sous la condition (C3) de l'annexe A, on peut montrer que $\widehat{U}_{IPW}(\boldsymbol{\theta}, \widehat{\boldsymbol{\tau}}_2, \widehat{\gamma}) \to^p \mathbb{E}\{U(\mathbf{x}, y;\boldsymbol{\theta})\}$ uniformément lorsque n_A, n_B et N tendent vers l'infini. De plus, $\mathbb{E}\{U(\mathbf{x}, y;\boldsymbol{\theta}_0)\} = \mathbf{0}$. On peut donc montrer que $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$, ce qui implique que l'estimateur IPW $\widehat{\boldsymbol{\theta}}_{IPW}$ est convergent. Selon (C3), on peut montrer que $\boldsymbol{\phi}^* = \left(\boldsymbol{\theta}^{*\top}, \boldsymbol{\tau}_2^{*\top}, \boldsymbol{\gamma}^{*\top}\right)^\top$ is the probability limit of $\widehat{\boldsymbol{\phi}} = \left(\widehat{\boldsymbol{\theta}}_{IPW}^\top, \widehat{\boldsymbol{\tau}}_2^\top, \widehat{\gamma}^\top\right)^\top$. A first-order Taylor linearization leads to

$$
\begin{aligned}
\mathbf{0} &= \widehat{\mathbf{U}}_2(\widehat{\gamma}) \\
&= \widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*) + \mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}(\widehat{\gamma} - \boldsymbol{\gamma}^*) + \mathbf{o}_p(n^{*-1/2}), \qquad \text{(B.1)}
\end{aligned}
$$

où

$$
\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\} = \begin{bmatrix}
\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2^{(1)}(\boldsymbol{\alpha}^{(1)*})}{\partial\boldsymbol{\alpha}^{(1)}}\right\} & \mathbf{0} & \ldots & \mathbf{0} \\
\mathbf{0} & \mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2^{(2)}(\boldsymbol{\alpha}^{(2)*})}{\partial\boldsymbol{\alpha}^{(2)}}\right\} & \ldots & \mathbf{0} \\
\ldots & \ldots & \ldots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \ldots & \mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2^{(K)}(\boldsymbol{\alpha}^{(K)*})}{\partial\boldsymbol{\alpha}^{(K)}}\right\}
\end{bmatrix}.
$$

Il en découle que

$$
\widehat{\gamma} - \boldsymbol{\gamma}^* = -\left[\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}\right]^{-1}\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*) + \mathbf{o}_p(n^{*-1/2}). \qquad \text{(B.2)}
$$

De manière similaire, on a

$$\begin{aligned}\mathbf{0} &= \widehat{\mathbf{U}}_4(\widehat{\boldsymbol{\tau}}_2,\widehat{\boldsymbol{\gamma}})\\ &= \widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)+\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\tau}_2}\right\}(\widehat{\boldsymbol{\tau}}_2-\boldsymbol{\tau}_2^*)+\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*)\\ &+ \mathbf{o}_p(n^{*-1/2}),\end{aligned}\tag{B.3}$$

où

$$\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\tau}_2} = \frac{1}{N}\sum_{i\in S_A} d_i\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)\mathbf{v}_2^\top(\mathbf{x}_i;\boldsymbol{\gamma}^*),$$

et

$$\begin{aligned}\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}} &= \frac{1}{N}\sum_{i\in S_A} d_i\frac{\partial\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\mathbf{v}_2^\top(\mathbf{x}_i;\boldsymbol{\gamma}^*)\boldsymbol{\tau}_2^*+\frac{1}{N}\sum_{i\in S_A} d_i\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)\boldsymbol{\tau}_2^{*\top}\frac{\partial\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\\ &- \frac{1}{N}\sum_{i\in S_B}\frac{\partial\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}.\end{aligned}$$

Des calculs algébriques conduisent à

$$\begin{aligned}\widehat{\boldsymbol{\tau}}_2-\boldsymbol{\tau}_2^* &= -\left[\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\tau}_2}\right\}\right]^{-1}\\ &\times \left[\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)-\mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_4(\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}\mathbb{E}^{-1}\left\{\frac{\partial\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}\widehat{\mathbf{U}}_2(\boldsymbol{\gamma}^*)\right]\\ &+ \mathbf{o}_p(n^{*-1/2}).\end{aligned}\tag{B.4}$$

Par souci de simplicité, désignons $\widehat{\boldsymbol{\theta}}_{IPW}$ par $\widehat{\boldsymbol{\theta}}$. Au moyen d'une linéarisation de Taylor du premier ordre, on a

$$\begin{aligned}0 &= \widehat{\mathbf{U}}_{IPW}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\tau}}_2,\widehat{\boldsymbol{\gamma}})\\ &= \widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)+\frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)\right\}}{\partial\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\\ &+ \mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\tau}_2}\right\}(\widehat{\boldsymbol{\tau}}_2-\boldsymbol{\tau}_2^*)\\ &+ \mathbb{E}\left\{\frac{\partial\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}}\right\}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*)+o_p(n^{*-1/2}),\end{aligned}\tag{B.5}$$

où

$$\frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)\right\}}{\partial\boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^N\frac{\partial\mathbb{E}\left\{p(\mathbf{x}_i)p^{*-1}(\mathbf{x}_i;\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)U(\mathbf{x}_i,y_i;\boldsymbol{\theta}^*)\right\}}{\partial\boldsymbol{\theta}},$$

avec $p(\mathbf{x}_i)=p(\mathbf{x}_i;\boldsymbol{\alpha}_0)$ et $p^*(\mathbf{x}_i;\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)=\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)$.

$$\frac{\partial\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\tau}_2} = -\frac{1}{N}\sum_{i\in S_B}\frac{\mathbf{U}(\mathbf{x}_i,y_i;\boldsymbol{\theta}^*)\mathbf{v}_2^\top(\mathbf{x}_i;\boldsymbol{\gamma}^*)}{\left\{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)\right\}^2},\tag{B.6}$$

et

$$\frac{\partial\widehat{\mathbf{U}}_{IPW}(\boldsymbol{\theta}^*,\boldsymbol{\tau}_2^*,\boldsymbol{\gamma}^*)}{\partial\boldsymbol{\gamma}} = -\frac{1}{N}\sum_{i\in S_B}\frac{\mathbf{U}(\mathbf{x}_i,y_i;\boldsymbol{\theta}^*)\boldsymbol{\tau}_2^{*\top}\partial\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)/\partial\boldsymbol{\gamma}}{\left\{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)\right\}^2}.\tag{B.7}$$

Selon (B.5) to (B.7), on obtient le résultat donné au Théorème 1.

C Preuve du Théorème 2

Denote $\widehat{\boldsymbol{\theta}}_{FMI}$ by $\widehat{\boldsymbol{\theta}}$. Without loss of generality, suppose that one of the outcome regression models $m^{(1)}(\mathbf{x};\boldsymbol{\beta}^{(1)})$ is correctly specified. Then, on peut montrer que $\widehat{\boldsymbol{\tau}}_1 \rightarrow^p \boldsymbol{\tau}_1^* = (1,0,\ldots,0)^\top$, $\widehat{\lambda} \rightarrow^p \lambda^* = 0$, et $\widehat{\boldsymbol{\eta}} \rightarrow^p \boldsymbol{\eta}^*$ avec $\boldsymbol{\eta}^* = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}^{(2)*\top}, \ldots, \boldsymbol{\beta}^{(J)*\top})^\top$. Il en découle que $\widehat{m}(\mathbf{x}) \rightarrow^p m(\mathbf{x};\boldsymbol{\beta}_0)$, lorsque n_A, n_B et N tendent vers l'infini.Selon (C3), on peut montrer que $\boldsymbol{\phi}^* = \left(\boldsymbol{\theta}^{*\top}, \boldsymbol{\tau}_1^{*\top}, \boldsymbol{\eta}^{*\top}, \lambda^*\right)^\top$ est la limite de $\widehat{\boldsymbol{\phi}} = \left(\widehat{\boldsymbol{\theta}}^\top, \widehat{\boldsymbol{\tau}}_1^\top, \widehat{\boldsymbol{\eta}}^\top, \widehat{\lambda}\right)^\top$. Parce qu'on a

$$\begin{aligned}
\mathbf{U}^*_{FMI}(\boldsymbol{\theta}_0, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) &= \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0)\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}} \\
&= \mathbb{E}\left[\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0)\right] \\
&= \mathbf{0},
\end{aligned}$$

$\widehat{\boldsymbol{\theta}}$ est un estimateur convergent de $\boldsymbol{\theta}_0$. Au moyen d'une linéarisation de Taylor du premier ordre, on a

$$\begin{aligned}
\mathbf{0} &= \widehat{\mathbf{U}}_1(\widehat{\boldsymbol{\eta}}) \\
&= \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*) + \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + \mathbf{o}_p(n_B^{-1/2}),
\end{aligned} \tag{C.1}$$

où

$$\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\} = \begin{bmatrix}
\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1^{(1)}(\boldsymbol{\beta}^{(1)*})}{\partial \boldsymbol{\beta}^{(1)}}\right\} & \mathbf{0} & \ldots & \mathbf{0} \\
\mathbf{0} & \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1^{(2)}(\boldsymbol{\beta}^{(2)*})}{\partial \boldsymbol{\beta}^{(2)}}\right\} & \ldots & 0 \\
\ldots & \ldots & \ldots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \ldots & \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1^{(J)}(\boldsymbol{\beta}^{(J)*})}{\partial \boldsymbol{\beta}^{(J)}}\right\}
\end{bmatrix}.$$

Il s'ensuit que

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = -\left\{\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}\right\}^{-1} \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*) + \mathbf{o}_p(n_B^{-1/2}). \tag{C.2}$$

Une linéarisation de Taylor du premier ordre conduit à

$$\begin{aligned}
\mathbf{0} &= \widehat{\mathbf{U}}_3(\widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\eta}}) \\
&= \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*) + \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}(\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1^*) + \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \\
&+ \mathbf{o}_p(n_B^{-1/2}),
\end{aligned} \tag{C.3}$$

où

$$\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1} = -\frac{1}{N}\sum_{i \in S_B} \mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta}^*)\mathbf{v}_1^\top(\mathbf{x}_i;\boldsymbol{\eta}^*),$$

and

$$\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = \frac{1}{N}\sum_{i \in S_B} \frac{\partial \mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\left\{y_i - \mathbf{v}_1^\top(\mathbf{x}_i;\boldsymbol{\eta}^*)\boldsymbol{\tau}_1^*\right\} - \frac{1}{N}\sum_{i \in S_B} \mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta}^*)\boldsymbol{\tau}_1^{*\top}\frac{\partial \mathbf{v}_1(\mathbf{x}_i;\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}.$$

Après des calculs algébriques, on obtient

$$
\begin{aligned}
\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1^* &= -\left[\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}\right]^{-1} \\
&\times \left[\widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*) - \mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\} \mathbb{E}^{-1}\left\{\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\} \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)\right] \\
&+ \mathbf{o}_p(n_B^{-1/2}). \qquad \text{(C.4)}
\end{aligned}
$$

Une linéarisation de Taylor du premier ordre conduit à

$$
\begin{aligned}
\mathbf{0} &= \widehat{U}_5(\widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\eta}}, \widehat{\lambda}) \\
&= \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) + \mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\tau}_1}\right\}(\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1^*) + \mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\eta}}\right\}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \\
&+ \mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \lambda}\right\}\left(\widehat{\lambda} - \lambda^*\right) + o_p(n_B^{-1/2}), \qquad \text{(C.5)}
\end{aligned}
$$

où

$$
\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\tau}_1} = -\frac{1}{N}\sum_{i \in S_B} \frac{\mathbf{v}_1^\top(\mathbf{x}_i; \boldsymbol{\eta}^*)}{\{1 + \lambda^* \widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\}^2},
$$

$$
\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\eta}} = -\frac{1}{N}\sum_{i \in S_B} \frac{\boldsymbol{\tau}_1^{*\top} \partial \mathbf{v}_1(\mathbf{x}_i; \boldsymbol{\eta}^*)/\partial \boldsymbol{\eta}}{\{1 + \lambda^* \widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\}^2},
$$

et

$$
\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \lambda} = -\frac{1}{N}\sum_{i \in S_B} \frac{\widehat{\epsilon}_i^2(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\{1 + \lambda^* \widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\}^2}.
$$

Selon (C.2), (C.4), et (C.5), on a

$$
\begin{aligned}
\widehat{\lambda} - \lambda^* &= A_{0,l}\left\{\widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) + \boldsymbol{A}_{1,l}\widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*) + \boldsymbol{A}_{2,l}\widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)\right\} \\
&+ o_p(n_B^{-1/2}), \qquad \text{(C.6)}
\end{aligned}
$$

où

$$
A_{0,l} = -\left[\mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \lambda}\right\}\right]^{-1},
$$

$$
\boldsymbol{A}_{1,l} = -\mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\tau}_1}\right\}\left[\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}\right]^{-1},
$$

et

$$
\begin{aligned}
\boldsymbol{A}_{2,l} = &\left[\mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\tau}_1}\right\}\left\{\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\tau}_1}\right\}\right\}^{-1}\mathbb{E}\left\{\frac{\partial \widehat{\mathbf{U}}_3(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right\}\right. \\
&\left. -\mathbb{E}\left\{\frac{\partial \widehat{U}_5(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)}{\partial \boldsymbol{\eta}}\right\}\right] \times \left\{\mathbb{E}\left(\frac{\partial \widehat{\mathbf{U}}_1(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right)\right\}^{-1}.
\end{aligned}
$$

Une linéarisation de Taylor du premier ordre conduit à

$$\begin{aligned}
\mathbf{0} &= \widehat{\mathbf{U}}_{FMI}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\eta}}, \widehat{\lambda}) \\
&= \widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) + \frac{\partial \mathbb{E}\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\tau}_1}(\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1^*) \\
&+ \frac{\partial \mathbb{E}\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\eta}}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + \frac{\partial \mathbb{E}\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \lambda}\left(\widehat{\lambda} - \lambda^*\right) \\
&+ \frac{\partial \mathbb{E}\left\{\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*)\right\}}{\partial \boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) + \mathbf{o}_p(n^{*-1/2}). \qquad (C.7)
\end{aligned}$$

De plus, on a

$$\begin{aligned}
\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) &= \frac{1}{N}\sum_{i\in S_A} d_i \sum_{j\in S_B} \frac{1}{n_B}\frac{1}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*) \\
&= \frac{1}{N^2}\sum_{i\in S_A} d_i \sum_{j\in S_B} \frac{1}{N^{-1}n_B}\frac{1}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} I_i d_i \sum_{j=1}^{N} \frac{\delta_j}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*) \\
&+ \mathbf{o}_p(n^{*-1/2}) \\
&= \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \boldsymbol{H}(\mathbf{z}_i, \mathbf{z}_j) + \mathbf{o}_p(n^{*-1/2}), \qquad (C.8)
\end{aligned}$$

où $\mathbf{z}_i = (I_i, \delta_i, \mathbf{x}_i, y_i)$ et $\boldsymbol{H}(\mathbf{z}_i, \mathbf{z}_j) = \left(\boldsymbol{\zeta}_{ij} + \boldsymbol{\zeta}_{ji}\right)/2$ avec

$$\boldsymbol{\zeta}_{ij} = I_i d_i \frac{\delta_j}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*),$$

et

$$\boldsymbol{\zeta}_{ji} = I_j d_j \frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_j, \widehat{y}_j^{(i)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*).$$

Des calculs algébriques mènent à

$$\mathbb{E}\left(\boldsymbol{\zeta}_{ij}|\mathbf{z}_i\right) = \frac{I_i d_i}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^*, \quad \mathbb{E}\left(\boldsymbol{\zeta}_{ji}|\mathbf{z}_i\right) = \frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\boldsymbol{g}_{2,i}^*, \qquad (C.9)$$

où

$$\boldsymbol{g}_{1,i}^* = \mathbb{E}\left\{\frac{p(\mathbf{x}_j; \boldsymbol{\alpha}_0)\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)} \mid \mathbf{z}_i\right\}, \qquad (C.10)$$

et

$$\boldsymbol{g}_{2,i}^* = \mathbb{E}\left\{\mathbf{U}(\mathbf{x}_j, \widehat{y}_j^{(i)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*)|\mathbf{z}_i\right\}. \qquad (C.11)$$

Selon (C.8)-(C.9) et par la théorie des statistiques U , see Serfling (1980), on a

$$\begin{aligned}
\widehat{\mathbf{U}}_{FMI}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*, \lambda^*) &= \frac{1}{N}\sum_{i=1}^{N}\frac{I_i d_i}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^* \\
&+ \frac{1}{N}\sum_{i=1}^{N}\frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\boldsymbol{g}_{2,i}^* \\
&+ \mathbf{o}_p(n^{*-1/2}). \qquad (C.12)
\end{aligned}$$

Selon (C.2), (C.4), (C.6), (C.7), (C.12), on obtient le dévelopement asymptotique au Théorème 2.

D Preuve du Théorème 3

Désignons $\widehat{\boldsymbol{\theta}}_{AMR}$ pr $\widehat{\boldsymbol{\theta}}$. Sans perte de généralité,supposons que l'un des modèles de régression $m^{(1)}(\mathbf{x};\boldsymbol{\beta}^{(1)})$ est correctement spécifié. Alors, on peut montrer que $\widehat{\boldsymbol{\tau}}_1 \to^p \boldsymbol{\tau}_1^* = (1,0,\ldots,0)^\top$, $\widehat{\lambda} \to^p \lambda^* = 0$, et $\widehat{\boldsymbol{\eta}} \to^p \boldsymbol{\eta}^*$ avec $\boldsymbol{\eta}^* = (\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}^{(2)*\top}, \ldots, \boldsymbol{\beta}^{(J)*\top})^\top$, et donc $\widehat{m}(\mathbf{x}) \to^p m(\mathbf{x};\boldsymbol{\beta}_0)$, lorsque n_A, n_B et N tendent vers l'infini. Selon (C3), on peut montrer que $\boldsymbol{\phi}^* = \left(\boldsymbol{\theta}^{*\top}, \boldsymbol{\tau}_1^{*\top}, \boldsymbol{\tau}_2^{*\top}, \boldsymbol{\gamma}^{*\top}, \boldsymbol{\eta}^{*\top}, \lambda^*\right)^\top$ est la limite de $\widehat{\boldsymbol{\phi}} = \left(\widehat{\boldsymbol{\theta}}^\top, \widehat{\boldsymbol{\tau}}_1^\top, \widehat{\boldsymbol{\tau}}_2^\top, \widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\eta}}^\top, \widehat{\lambda}\right)^\top$. Soit $\mathbf{U}^*_{AMR}(\boldsymbol{\theta}_0)$ lorsque $\mathbf{U}^*_{AMR}(\boldsymbol{\theta}_0, \boldsymbol{\tau}_1^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*, \boldsymbol{\eta}^*, \lambda^*)$. Parce qu'on a

$$
\begin{aligned}
\mathbf{U}^*_{AMR}(\boldsymbol{\theta}_0) &= \mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}\mathbf{U}(\mathbf{x}_i, y_i;\boldsymbol{\theta}_0)\right\} \\
&+ \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0)\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}} \\
&- \frac{1}{p(\boldsymbol{\alpha}_0)}\mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)p(\mathbf{x}_j;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0)\right\} \\
&= \mathbb{E}\left\{\mathbf{U}(\mathbf{x}_i, y_i;\boldsymbol{\theta}_0)\right\} = 0,
\end{aligned}
$$

il en découle que $\widehat{\boldsymbol{\theta}}$ est un estimateur convergent de $\boldsymbol{\theta}_0$. Suppose one of the propensity score models in $\mathscr{M}_2$ is correctly specified. Sans perte de généralité, on suppose que $p^{(1)}(\mathbf{x};\boldsymbol{\alpha}^{(1)})$ est correctement spécifié, alors on peut montrer que $\widehat{\boldsymbol{\tau}}_2 \to^p (1,0,\ldots,0)^\top$, so $\widehat{p}(\mathbf{x}) \to^p p(\mathbf{x};\boldsymbol{\alpha}_0)$ as n_A, n_B et N tendent vers l'infini. Selon (C3), on peut montrer que $\boldsymbol{\phi}^* = \left(\boldsymbol{\theta}^{*\top}, \boldsymbol{\tau}_1^{*\top}, \boldsymbol{\tau}_2^{*\top}, \boldsymbol{\gamma}^{*\top}, \boldsymbol{\eta}^{*\top}, \lambda^*\right)^\top$ est la limite de $\widehat{\boldsymbol{\phi}} = \left(\widehat{\boldsymbol{\theta}}^\top, \widehat{\boldsymbol{\tau}}_1^\top, \widehat{\boldsymbol{\tau}}_2^\top, \widehat{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{\eta}}^\top, \widehat{\lambda}\right)^\top$. Parce qu'on a

$$
\begin{aligned}
\mathbf{U}^*_{AMR}(\boldsymbol{\theta}_0) &= \mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)}\mathbf{U}(\mathbf{x}_i, y_i;\boldsymbol{\theta}_0)\right\} \\
&+ \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\left\{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\right\}^{-1}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0))\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}} \\
&- \frac{1}{p(\boldsymbol{\alpha}_0)}\mathbb{E}\left\{\frac{p(\mathbf{x}_i;\boldsymbol{\alpha}_0)p(\mathbf{x}_j;\boldsymbol{\alpha}_0)}{\boldsymbol{\tau}_2^{*\top}\mathbf{v}_2(\mathbf{x}_i;\boldsymbol{\gamma}^*)\left\{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\right\}}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0)\right\} \\
&= \mathbb{E}\left\{\mathbf{U}(\mathbf{x}_i, y_i;\boldsymbol{\theta}_0)\right\} \\
&+ \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\left\{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\right\}^{-1}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0))\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}} \\
&- \frac{\mathbb{E}\left[p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\left\{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)\right\}^{-1}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*);\boldsymbol{\theta}_0))\right]}{\mathbb{E}\left\{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)\right\}} \\
&= 0,
\end{aligned}
$$

il en découle que $\widehat{\boldsymbol{\theta}}$ est un estimateur convergent de $\boldsymbol{\theta}_0$. Une approximation par série de Taylor du premier ordre conduit à

$$
\begin{aligned}
\mathbf{0} &= \widehat{\mathbf{U}}_{AMR}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\tau}}_1, \widehat{\boldsymbol{\tau}}_2, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\eta}}, \widehat{\lambda}) \\
&= \widehat{\mathbf{U}}^*_{AMR} + \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\boldsymbol{\tau}_1}(\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_1^*) + \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\boldsymbol{\tau}_2}(\widehat{\boldsymbol{\tau}}_2 - \boldsymbol{\tau}_2^*) \\
&+ \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\boldsymbol{\gamma}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) + \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\boldsymbol{\eta}}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\lambda}\left(\widehat{\lambda} - \lambda^*\right) \\
&+ \frac{\partial\mathbb{E}\left\{\widehat{\mathbf{U}}^*_{AMR}\right\}}{\partial\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right) + \mathbf{o}_p(n^{*-1/2}),
\end{aligned}
\tag{D.1}
$$

où $\widehat{\mathbf{U}}^*_{AMR} = \widehat{\mathbf{U}}_{AMR}(\boldsymbol{\theta}^*, \boldsymbol{\tau}_1^*, \boldsymbol{\tau}_2^*, \boldsymbol{\gamma}^*, \boldsymbol{\eta}^*, \lambda^*)$. De plus, on a

$$\begin{aligned}
\widehat{\mathbf{U}}^*_{AMR} &= \frac{1}{N}\sum_{i\in S_B}\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\mathbf{U}(\mathbf{x}_i, y_i; \boldsymbol{\theta}^*) + \frac{1}{N^2}\sum_{i\in S_A} d_i \sum_{j\in S_B}\frac{1}{n_B N^{-1}}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} \\
&- \frac{1}{N^2}\sum_{i\in S_B}\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\sum_{j\in S_B}\frac{1}{n_B N^{-1}}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} \\
&= \frac{1}{N}\sum_{i\in S_B}\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\mathbf{U}(\mathbf{x}_i, y_i; \boldsymbol{\theta}^*) + \frac{1}{N^2}\sum_{i\in S_A} d_i \sum_{j\in S_B}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} \\
&- \frac{1}{N^2}\sum_{i\in S_B}\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\sum_{j\in S_B}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} \\
&+ \mathbb{E}\left[\left\{\frac{p(\mathbf{x}_i)}{\widehat{p}^*(\mathbf{x}_i)} - 1\right\}\frac{p(\mathbf{x}_j)\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*}\right]\frac{1}{\{\mathbb{E}(\delta)\}^2}\left\{n_B N^{-1} - \mathbb{E}(\delta)\right\} \\
&+ \mathbf{o}_p(n^{*-1/2}).
\end{aligned} \tag{D.2}$$

Soit

$$\boldsymbol{\Delta}_1 = \frac{1}{N^2}\sum_{i\in S_A} d_i \sum_{j\in S_B}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*}.$$

On a alors

$$\begin{aligned}
\boldsymbol{\Delta}_1 &= \boldsymbol{\Delta}_1 - \mathbb{E}(\boldsymbol{\Delta}_1) + \mathbb{E}(\boldsymbol{\Delta}_1) \\
&= \frac{1}{N(N-1)}\sum_{i=1}^N\sum_{j=1,j\neq i}^N\left\{I_i d_i \delta_j \frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)*}; \boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} - \mathbb{E}(\boldsymbol{\Delta}_1)\right\} \\
&+ \mathbb{E}(\boldsymbol{\Delta}_1) + \mathbf{o}_p(n^{*-1/2}) \\
&= \frac{1}{N(N-1)}\sum_{i=1}^N\sum_{j=1,j\neq i}^N \boldsymbol{H}_1(\mathbf{z}_i, \mathbf{z}_j) + \mathbb{E}(\boldsymbol{\Delta}_1) + \mathbf{o}_p(n^{*-1/2}),
\end{aligned} \tag{D.3}$$

où $\boldsymbol{H}_1(\mathbf{z}_i, \mathbf{z}_j) = (\boldsymbol{\zeta}_{1,ij} + \boldsymbol{\zeta}_{1,ji})/2$ avec

$$\boldsymbol{\zeta}_{1,ij} = I_i d_i \frac{\delta_j}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_i, \widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*) - \mathbb{E}(\boldsymbol{\Delta}_1),$$

et

$$\boldsymbol{\zeta}_{1,ji} = I_j d_j \frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\mathbf{U}(\mathbf{x}_j, \widehat{y}_j^{(i)}(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*); \boldsymbol{\theta}^*) - \mathbb{E}(\boldsymbol{\Delta}_1).$$

Après quelques calculs, on obtient

$$\mathbb{E}\left(\boldsymbol{\zeta}_{1,ij}|\mathbf{z}_i\right) = \frac{I_i d_i}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^* - \mathbb{E}(\boldsymbol{\Delta}_1), \tag{D.4}$$

et

$$\mathbb{E}\left(\boldsymbol{\zeta}_{1,ji}|\mathbf{z}_i\right) = \frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\boldsymbol{g}_{2,i}^* - \mathbb{E}(\boldsymbol{\Delta}_1). \tag{D.5}$$

Selon (D.2)-(D.4) et par la théorie des statistiques U, voir Sefling (1980), on a

$$\begin{aligned}
\boldsymbol{\Delta}_1 &= \frac{1}{N}\sum_{i=1}^N\left\{\frac{I_i d_i}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^* - \mathbb{E}(\boldsymbol{\Delta}_1)\right\} \\
&+ \frac{1}{N}\sum_{i=1}^N\left\{\frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*, \boldsymbol{\eta}^*)}\boldsymbol{g}_{2,i}^* - \mathbb{E}(\boldsymbol{\Delta}_1)\right\} \\
&+ \mathbb{E}(\boldsymbol{\Delta}_1) + \mathbf{o}_p(n^{*-1/2}).
\end{aligned} \tag{D.6}$$

Soit

$$\boldsymbol{\Delta}_2 = \frac{1}{N^2}\sum_{i\in S_B}\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\sum_{j\in S_B}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)*};\boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*}.$$

On a alors

$$\begin{aligned}
\boldsymbol{\Delta}_2 &= \boldsymbol{\Delta}_2 - \mathbb{E}(\boldsymbol{\Delta}_2) + \mathbb{E}(\boldsymbol{\Delta}_2) \\
&= \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left\{\delta_i\delta_j\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)*};\boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j^*} - \mathbb{E}(\boldsymbol{\Delta}_2)\right\} \\
&+ \mathbb{E}(\boldsymbol{\Delta}_2) + \mathbf{o}_p(n^{*-1/2}) \\
&= \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\boldsymbol{H}_2(\mathbf{z}_i,\mathbf{z}_j) + \mathbb{E}(\boldsymbol{\Delta}_2) + \mathbf{o}_p(n^{*-1/2}), \qquad \text{(D.7)}
\end{aligned}$$

où $\boldsymbol{H}_2(\mathbf{z}_i,\mathbf{z}_j) = (\boldsymbol{\zeta}_{2,ij}+\boldsymbol{\zeta}_{2,ji})/2$ avec

$$\boldsymbol{\zeta}_{2,ij} = \delta_i\delta_j\frac{1}{\widehat{p}^*(\mathbf{x}_i)}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_i,\widehat{y}_i^{(j)}(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*);\boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_j(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*)} - \mathbb{E}(\boldsymbol{\Delta}_2),$$

et

$$\boldsymbol{\zeta}_{2,ji} = \delta_j\delta_i\frac{1}{\widehat{p}^*(\mathbf{x}_j)}\frac{1}{\mathbb{E}(\delta)}\frac{\mathbf{U}(\mathbf{x}_j,\widehat{y}_j^{(i)}(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*);\boldsymbol{\theta}^*)}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*)} - \mathbb{E}(\boldsymbol{\Delta}_2).$$

Après des calculs algébriques, on obtient

$$\mathbb{E}\left(\boldsymbol{\zeta}_{2,ij}|\mathbf{z}_i\right) = \frac{\delta_i}{\widehat{p}^*(\mathbf{x}_i)}\frac{1}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^* - \mathbb{E}(\boldsymbol{\Delta}_2), \qquad \text{(D.8)}$$

et

$$\mathbb{E}\left(\boldsymbol{\zeta}_{2,ji}|\mathbf{z}_i\right) = \frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*)}\boldsymbol{g}_{3,i}^* - \mathbb{E}(\boldsymbol{\Delta}_2), \qquad \text{(D.9)}$$

où

$$\boldsymbol{g}_{3,i}^* = \mathbb{E}\left\{\frac{p(\mathbf{x}_j;\boldsymbol{\alpha}_0)}{\widehat{p}^*(\mathbf{x}_j)}\mathbf{U}(\mathbf{x}_j,\widehat{y}_j^{(i)}(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*);\boldsymbol{\theta}^*)|\mathbf{z}_i\right\}.$$

D'après (D.7)-(D.9) et par la théorie des statistiques U, voir Sefling (1980), nous avons

$$\begin{aligned}
\boldsymbol{\Delta}_2 &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\delta_i}{\widehat{p}^*(\mathbf{x}_i)}\frac{1}{\mathbb{E}(\delta)}\boldsymbol{g}_{1,i}^* - \mathbb{E}(\boldsymbol{\Delta}_2)\right\} \\
&+ \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\delta_i}{\mathbb{E}(\delta)}\frac{1}{1+\lambda^*\widehat{\epsilon}_i(\boldsymbol{\tau}_1^*,\boldsymbol{\eta}^*)}\boldsymbol{g}_{3,i}^* - \mathbb{E}(\boldsymbol{\Delta}_2)\right\} \\
&+ \mathbb{E}(\boldsymbol{\Delta}_2) + \mathbf{o}_p(n^{*-1/2}). \qquad \text{(D.10)}
\end{aligned}$$

Selon (D.1)-(D.1) , on obtient le dévelopement asymptotique au Théorème 3.