Nº 12-001-X au catalogue
ISSN 1712-5685

Techniques d'enquête

Estimation et inférence des moyennes de domaine soumises à des contraintes qualitatives

par Cristian Oliva-Aviles, Mary C. Meyer et Jean D. Opsomer

Date de diffusion : le 15 décembre 2020

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-514-283-9350

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Techniques d'enquête, décembre 2020
Vol. 46, nº 2, pp. 155-191
Statistique Canada, nº 12-001-X au catalogue

Estimation et inférence des moyennes de domaine soumises à des contraintes qualitatives

Cristian Oliva-Aviles, Mary C. Meyer et Jean D. Opsomer[1]

Résumé

Dans de nombreuses enquêtes à grande échelle, des estimations sont produites pour un grand nombre de petits domaines définis par des classifications croisées de variables démographiques, géographiques et autres. Bien que la taille globale de l'échantillon de ces enquêtes puisse être très grande, la taille des échantillons des domaines est parfois trop petite pour permettre une estimation fiable. Nous proposons une méthode d'estimation améliorée qui s'applique quand il est possible de formuler des relations « naturelles » ou qualitatives (comme des ordonnancements ou des contraintes d'inégalité) pour les moyennes des domaines au niveau de la population. Nous restons dans un cadre inférentiel fondé sur le plan, mais nous imposons des contraintes représentant ces relations sur les estimations échantillonnales. Nous démontrons que l'estimateur de domaine contraint qui en résulte est convergent par rapport au plan et a une distribution asymptotique normale tant que les contraintes sont asymptotiquement satisfaites au niveau de la population. L'estimateur et l'estimateur de la variance connexe sont facilement mis en œuvre en pratique. L'applicabilité de la méthode est illustrée par les données de la *National Survey of College Graduates* des États-Unis (NSCG, Enquête nationale sur les diplômés des collèges) de 2015.

Mots-clés : Estimation fondée sur le plan; estimation monotone; *National Survey of College Graduates*.

1 Introduction

De nombreuses enquêtes à grande échelle visent notamment à produire des estimations pour un grand nombre de domaines, dont beaucoup peuvent avoir une petite taille d'échantillon. En général, ces domaines sont créés par la classification croisée de variables catégoriques comme les caractéristiques démographiques, géographiques ou autres caractéristiques similaires d'intérêt. Ainsi, la *Current Population Survey* des États-Unis (Enquête sur la population actuelle) publie des estimations pour des domaines définis selon le sexe, l'âge, la race ou le niveau de scolarité. De même, l'*American Community Survey* des États-Unis (Enquête sur les collectivités américaines) produit des estimations détaillées par sexe, âge, race ou origine ethnique pour différents niveaux géographiques (selon la publication). Dans un autre exemple que nous examinerons plus en détail ci-dessous, la *National Survey of College Graduates* des États-Unis s'intéresse aux estimations obtenues par croisement du niveau et du domaine d'études, de la profession et du genre. Dans certains programmes d'enquête, ces estimations « granulaires » sont souvent aussi importantes que les estimations de niveau supérieur ou de la population.

Or, bien que la taille globale de l'échantillon de ces enquêtes puisse être très grande, la taille des échantillons de nombreux domaines est souvent trop petite pour permettre des estimations fiables. Afin d'éviter ce problème, on pourrait agréger de petits domaines à de plus grandes échelles afin de produire des estimateurs directs plus fiables pour ces échelles, ce qui conduirait à la production d'informations plus agrégées que l'échelle réellement souhaitée. Plutôt que de produire des estimations pour de petits

1. Cristian Oliva-Aviles, Genentech, Inc.; Mary C. Meyer, Colorado State University; Jean D. Opsomer, Westat, Inc. Courriel : jopsomer@mac.com.

domaines, une autre solution consisterait à passer d'une méthodologie d'estimation fondée sur le plan de sondage à une méthodologie d'estimation fondée sur un modèle, comme des modèles pour petits domaines. Bien que cette méthode soit tout à fait valide sur le plan statistique pour créer des estimations précises à petite échelle, elle est laborieuse et sensible à une éventuelle spécification erronée du modèle. De plus, elle remplace l'erreur d'échantillonnage par une erreur de modèle, de sorte que le mode d'inférence change. C'est pourquoi les organismes statistiques préfèrent s'en tenir à l'approche fondée sur le plan de sondage, qui offre de la robustesse et permet de conserver le mode d'inférence standard des enquêtes.

Dans notre article, nous présentons une approche d'estimation qui s'applique quand on s'attend à ce que des relations « naturelles » ou qualitatives se vérifient pour les moyennes des domaines au niveau de la population. Ces relations peuvent servir à stabiliser les estimations pour un domaine de l'échantillon, tout en conservant le mode d'estimation et d'inférence fondé sur le plan de sondage. Le type de relations que nous étudions ici entraîne des inégalités entre les moyennes des domaines de population. On peut par exemple s'attendre à ce que certains types d'emplois soient mieux rémunérés que d'autres, ou à ce que les titulaires d'un diplôme d'études supérieures dans une discipline donnée aient un salaire plus élevé que les personnes n'en possédant pas. Toutefois, les petits domaines ayant tendance à produire des estimations présentant une grande variabilité, souvent les relations attendues de ce type au niveau de la population ne sont pas respectées au niveau de l'échantillon. Les utilisateurs de données doivent s'attendre au non-respect des relations en raison de la variabilité statistique, mais cela pourrait les amener à remettre en question la fiabilité globale de l'enquête par la production d'estimations « absurdes ».

Il existe une littérature abondante sur les statistiques d'enquête portant sur le calage des estimations d'enquête; on trouve notamment dans Särndal, Swensson et Wretman (1992) une vue d'ensemble de la question. Bien que ces estimateurs reposent également sur des contraintes, il existe d'importantes différences, y compris le fait que les contraintes sont des contraintes d'égalité et qu'elles sont appliquées aux poids d'enquête et non aux estimations elles-mêmes. Nous n'étudierons pas cette question ici, mais il serait possible de combiner calage et estimation contrainte, puisque cette dernière pourrait utiliser les estimations de domaine calées comme point de départ de la construction d'estimations de domaine contraintes. Dans un contexte fondé sur un modèle, Rueda et Lombardía (2012) ont adapté des méthodes d'estimation sur petits domaines pour les cas de moyennes de domaines monotoniquement ordonnées.

Wu, Meyer et Opsomer (2016) ont proposé une méthodologie d'estimation de la moyenne de domaine qui repose sur l'hypothèse de moyennes de domaines de population monotones avec une seule variable catégorique définissant le domaine (par exemple, classes d'âge). En combinant l'information de monotonicité des moyennes de domaine et des estimateurs fondés sur le plan à l'étape de l'estimation, ils ont proposé un estimateur *contraint* respectant l'hypothèse de monotonie. Il a été montré que cet estimateur améliore la précision et la variabilité des estimations de la moyenne de domaine par rapport aux estimateurs directs, si l'hypothèse de la monotonicité est raisonnable.

Dans cet article, nous généralisons cette constatation en permettant une classe de contraintes beaucoup plus grande entre les moyennes de domaine, qui s'applique dans les configurations multidimensionnelles. De nombreux autres types de contraintes autres que la monotonicité devraient se vérifier entre les moyennes de domaine de population dans les enquêtes réelles, surtout en présence de domaines définis par les classifications croisées de nombreuses variables catégoriques. En général, tout ensemble de contraintes linéaires peut être représenté par une *matrice de contraintes*, dans laquelle chaque ligne définit une contrainte et chaque colonne une moyenne de domaine. Comme illustration d'une matrice de contraintes, supposons que la variable d'intérêt est le salaire annuel moyen des professeurs des universités d'État instaurées par donation foncière d'une certaine taille. Considérons de plus les domaines générés par la classification croisée des variables du poste (x_1; 1 = non permanent et 2 = permanent) et de trois départements donnés (1 = anthropologie, 2 = anglais et 3 = génie). Si l'on suppose qu'en moyenne, dans une discipline, le corps professoral permanent a des salaires plus élevés que le corps professoral non permanent et que, au sein du corps professoral permanent et non permanent, les membres du corps professoral en génie devraient avoir des salaires plus élevés que ceux des départements d'anthropologie ou d'anglais, nous pouvons alors exprimer les restrictions correspondantes comme suit :

$$\mathbf{A}\boldsymbol{\mu} \geq \mathbf{0}, \quad \text{où} \quad \mathbf{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}, \tag{1.1}$$

$\boldsymbol{\mu} = \left(\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22}, \mu_{13}, \mu_{23}\right)^{\mathrm{T}}$, avec μ_{ij} qui représente la moyenne du domaine correspondant à $x_1 = i$ et $x_2 = j$; $\mathbf{0}$ étant le vecteur zéro, et l'inégalité étant par élément. Ce document décrit un nouvel estimateur contraint pour les moyennes de domaine de population respectant les contraintes qu'on peut exprimer avec les inégalités matricielles ayant la forme donnée en (1.1). En combinant les estimateurs de moyennes de domaines fondés sur le plan avec ces contraintes de forme, nous proposons un estimateur d'application large qui améliore la précision et la variabilité des estimateurs directs les plus courants.

Le plan de l'article est le suivant. Dans la section 2, nous présenterons formellement l'estimateur contraint et proposerons une méthode fondée sur la linéarisation pour l'estimation de la variance. La section contient également certains scénarios d'intérêt où des contraintes de forme peuvent surgir naturellement dans les données d'enquête. La section 3 énonce les principales propriétés théoriques de l'estimateur contraint. Les hypothèses nécessaires utilisées dans ces calculs théoriques sont également énoncées dans cette section. Les démonstrations des principaux théorèmes et des lemmes auxiliaires sont données en annexe. La section 4 montre par des simulations que l'estimateur contraint améliore l'estimation de la moyenne de domaine et la variabilité par rapport à l'estimateur non contraint, y compris quand la forme supposée ne se vérifie qu'approximativement au niveau de la population. La section 5

montre les avantages de la méthodologie proposée pour les données d'enquête réelles en l'appliquant à la *National Survey of College Graduates* de 2015. En guise de conclusion, on trouve quelques remarques additionnelles à la section 6.

2 Estimation contrainte et inférence sur des moyennes de domaine

2.1 Notation et préliminaires

Soit U_N l'ensemble des éléments dans une population de taille N. Considérons un échantillon s_N de taille n_N tiré de U_N au moyen d'un plan de sondage probabiliste $p_N(\cdot)$. Soit $\pi_{k,N} = \Pr(k \in s_N)$ et $\pi_{kl,N} = \Pr(k \in s_N, l \in s_N)$ respectivement les probabilités d'inclusion du premier et du second ordre. Supposons que $\pi_{k,N} > 0$, $\pi_{kl,N} > 0$ pour $k, l \in U_N$. Pour simplifier la notation, nous adopterons la convention habituelle de suppression de l'indice N à moins qu'il ne soit nécessaire à des fins de clarification. Désignons par $\{U_d\}_{d=1}^{D}$ une partition de domaine de U, où D est le nombre de domaines et chaque U_d est de taille N_d. De plus, soit s_d le sous-ensemble de taille n_d de s qui appartient à U_d.

Pour toute variable étudiée y, $\bar{\mathbf{y}}_U = \left(\bar{y}_{U_1}, \ldots, \bar{y}_{U_D}\right)^{\mathrm{T}}$ désigne le vecteur du domaine de population, où

$$\bar{y}_{U_D} = \frac{\sum_{k \in U_d} y_k}{N_d}. \tag{2.1}$$

Nous nous concentrerons sur l'estimateur de Hájek de $\bar{y}_{U_D}$, donné par

$$\tilde{y}_{s_d} = \frac{\sum_{k \in s_d} y_k / \pi_k}{\hat{N}_d} \tag{2.2}$$

avec $\hat{N}_d = \sum_{k \in s_d} 1/\pi_k$, et soit $\tilde{\mathbf{y}}_s$ le vecteur des estimateurs. Les résultats se vérifieront aussi pour l'estimateur de Horvitz-Thompson avec des modifications mineures, mais cette question ne sera pas traitée explicitement dans ce qui suit.

2.2 Estimateur proposé

Supposons qu'on dispose d'informations concernant les relations entre les moyennes de domaine de population qui peuvent être exprimées avec m contraintes au moyen d'une matrice de contraintes $m \times D$ *irréductible* $\mathbf{A}$. Une matrice $\mathbf{A}$ est irréductible si aucune de ses lignes n'est une combinaison linéaire positive d'autres lignes, et si l'origine n'est pas non plus une combinaison linéaire positive de ses lignes (Meyer, 1999). En pratique, cela signifie qu'il n'y a pas de contraintes redondantes dans $\mathbf{A}$. Pour tirer parti de $\tilde{\mathbf{y}}_s$ afin d'obtenir un estimateur qui respecte ces contraintes de forme, nous proposons que l'estimateur contraint $\tilde{\boldsymbol{\theta}}_s = \left(\tilde{\theta}_{s_1}, \ldots, \tilde{\theta}_{s_D}\right)^{\mathrm{T}}$ soit le vecteur unique résolvant le problème contraint des moindres carrés pondérés suivant,

$$\min_{\boldsymbol{\theta}} \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}\right)^{\top} \mathbf{W}_s \left(\tilde{\mathbf{y}}_s - \boldsymbol{\theta}\right) \text{ sujet à } \mathbf{A}\boldsymbol{\theta} \geq \mathbf{0}; \tag{2.3}$$

où $\mathbf{W}_s$ est la matrice diagonale avec les éléments $\hat{N}_1/\hat{N}$, $\hat{N}_2/\hat{N}$, …, $\hat{N}_D/\hat{N}$, et $\hat{N} = \sum_{d=1}^{D} \hat{N}_d$. On peut écrire autrement le problème contraint de l'équation (2.3) pour trouver le vecteur unique $\tilde{\boldsymbol{\phi}}_s$ qui résout

$$\min_{\boldsymbol{\phi}} \left\| \tilde{\mathbf{z}}_s - \boldsymbol{\phi} \right\|^2 \text{ sujet à } \mathbf{A}_s\boldsymbol{\phi} \geq \mathbf{0}, \tag{2.4}$$

où $\tilde{\mathbf{z}}_s = \mathbf{W}_s^{1/2} \tilde{\mathbf{y}}_s$, $\boldsymbol{\phi} = \mathbf{W}_s^{1/2} \boldsymbol{\theta}$, et $\mathbf{A}_s = \mathbf{A}\mathbf{W}_s^{-1/2}$. La matrice contrainte transformée $\mathbf{A}_s$ est également irréductible si $\mathbf{A}$ l'est et elle dépend de l'échantillon bien que $\mathbf{A}$ n'en dépende pas. La solution $\tilde{\boldsymbol{\phi}}_s$ est la *projection* de $\tilde{\mathbf{z}}_s$ sur l'ensemble des vecteurs $\boldsymbol{\phi}$ qui satisfont la condition $\mathbf{A}_s\boldsymbol{\phi} \geq \mathbf{0}$. Cet ensemble est un cône convexe polyédrique, appelé le *cône de contrainte* Ω_s défini par $\mathbf{A}_s$, plus particulièrement

$$\Omega_s = \left\{\boldsymbol{\phi} \in \mathbb{R}^D : \mathbf{A}_s\boldsymbol{\phi} \geq \mathbf{0}\right\}. \tag{2.5}$$

Nous utilisons la notation $\tilde{\boldsymbol{\phi}}_s = \Pi\left(\tilde{\mathbf{z}}_s \,|\, \Omega_s\right)$, où $\Pi\left(\mathbf{u} \,|\, \mathcal{S}\right)$ représente la projection de $\mathbf{u}$ sur l'ensemble $\mathcal{S}$, c'est-à-dire le vecteur le plus proche dans $\mathcal{S}$ de $\mathbf{u}$.

On connaît bien les projections sur ces cônes (voir Rockafellar (1970) ou Meyer (1999) pour en savoir plus). Pour ce qui est des travaux présentés dans l'article, les principaux résultats de la théorie de la projection des cônes sont résumés ici. Le cône peut être caractérisé par un ensemble *d'arêtes* générant le cône, c'est-à-dire qu'un vecteur se trouve dans le cône si et seulement s'il s'agit d'une combinaison linéaire des arêtes avec des coefficients non négatifs. (Imaginons une pyramide avec un sommet à l'origine, s'étendant indéfiniment.) Les sous-ensembles des arêtes définissent les *faces* du cône, et la projection de $\tilde{\mathbf{z}}_s$ sur le cône, sur l'une des faces. Après détermination des arêtes définissant cette face, la projection peut être caractérisée comme une projection par la méthode des moindres carrés ordinaires sur l'espace linéaire couvert par ce sous-ensemble d'arêtes. Cette propriété est cruciale pour l'algorithme de projection et pour l'inférence, car la projection sur le cône peut être caractérisée comme une projection linéaire.

Dans les présents travaux, nous projetterons $\tilde{\mathbf{z}}_s$ sur le *cône dual négatif (ou cône polaire)* Ω_s^0 (Rockafellar, 1970, page 121), défini comme suit :

$$\Omega_s^0 = \left\{\boldsymbol{\rho} \in \mathbb{R}^D : \left\langle \boldsymbol{\rho}, \boldsymbol{\phi} \right\rangle \leq 0, \ \forall \boldsymbol{\phi} \in \Omega_s\right\}, \tag{2.6}$$

où $\left\langle \mathbf{u}, \mathbf{v} \right\rangle = \mathbf{u}^{\top}\mathbf{v}$. Autrement dit, le cône polaire est l'ensemble des vecteurs qui forment des angles obtus avec tous les vecteurs dans Ω_s. Le cône polaire est analogue à l'espace orthogonal dans les projections linéaires par la méthode des moindres carrés, en ce sens que la projection d'un vecteur sur le cône polaire est le résidu de sa projection sur le cône des contraintes, et vice versa. Meyer (1999) a montré que les lignes négatives d'une matrice irréductible sont les *arêtes* (générateurs) du cône polaire, ce qui conduit à la caractérisation suivante du cône polaire dans (2.6) :

$$\Omega_s^0 = \left\{ \boldsymbol{\rho} \in \mathbb{R}^D : \boldsymbol{\rho} = \sum_{j=1}^{m} a_j \boldsymbol{\gamma}_{s_j},\ a_j \geq 0,\ j = 1, 2, \ldots, m \right\}, \tag{2.7}$$

où $\boldsymbol{\gamma}_{s_1}, \boldsymbol{\gamma}_{s_2}, \ldots, \boldsymbol{\gamma}_{s_m}$ sont les lignes de $-\mathbf{A}_s$. Robertson, Wright et Dykstra (1988, page 17) ont établi les conditions nécessaires et suffisantes pour qu'un vecteur $\tilde{\boldsymbol{\phi}}_s$ soit la projection de $\tilde{\mathbf{z}}_s$ sur Ω_s. Ainsi, $\tilde{\boldsymbol{\phi}}_s \in \Omega_s$ résout le problème contraint de (2.4) si et seulement si

$$\left\langle \tilde{\mathbf{z}}_s - \tilde{\boldsymbol{\phi}}_s, \tilde{\boldsymbol{\phi}}_s \right\rangle = 0, \quad \text{et} \quad \left\langle \tilde{\mathbf{z}}_s - \tilde{\boldsymbol{\phi}}_s, \boldsymbol{\phi} \right\rangle \leq 0,\ \forall \boldsymbol{\phi} \in \Omega_s.$$

De plus, les conditions ci-dessus peuvent être adaptées au cône polaire comme suit : le vecteur $\tilde{\boldsymbol{\rho}}_s \in \Omega_s^0$ minimise $\left\| \tilde{\mathbf{z}}_s - \boldsymbol{\rho} \right\|^2$ sur Ω_s^0 si et seulement si

$$\left\langle \tilde{\mathbf{z}}_s - \tilde{\boldsymbol{\rho}}_s, \tilde{\boldsymbol{\rho}}_s \right\rangle = 0, \quad \text{et} \quad \left\langle \tilde{\mathbf{z}}_s - \tilde{\boldsymbol{\rho}}_s, \boldsymbol{\gamma}_{s_j} \right\rangle \leq 0 \quad \text{pour} \quad j = 1, 2, \ldots, m. \tag{2.8}$$

On peut utiliser les conditions de (2.8) pour montrer que la projection de $\tilde{\mathbf{z}}_s$ sur le cône polaire Ω_s^0 coïncide avec la projection sur l'espace linéaire généré par les arêtes $\boldsymbol{\gamma}_{s_j}$, de sorte que $\left\langle \tilde{\mathbf{z}}_s - \tilde{\boldsymbol{\rho}}_s, \boldsymbol{\gamma}_{s_j} \right\rangle = 0$. Cet ensemble d'arêtes peut être vide, ce qui signifie que la projection sur Ω_s^0 est égale à la projection sur le vecteur nul. Dans ce cas, le minimum non contraint respecte toutes les contraintes. L'ensemble d'arêtes peut aussi ne pas être unique. Pour mettre en forme ces idées, nous notons $V_{s,J} = \left\{ \boldsymbol{\gamma}_{s_j} : j \in J \right\}$ pour tout $J \subseteq \{1, 2, \ldots, m\}$ Nous définissons l'ensemble $\bar{\mathcal{F}}_{s,J}$ comme étant

$$\bar{\mathcal{F}}_{s,J} = \left\{ \boldsymbol{\rho} \in \mathbb{R}^D : \boldsymbol{\rho} = \sum_{j \in J} a_j \boldsymbol{\gamma}_{s_j},\ a_j \geq 0,\ j \in J \right\}, \tag{2.9}$$

où $\bar{\mathcal{F}}_{s,\varnothing} = \mathbf{0}$ par convention. (Techniquement, l'ensemble est la fermeture d'une face du cône.) Autrement dit, $\bar{\mathcal{F}}_{s,J}$ est un sous-cône polyédrique fermé de Ω_s^0 qui commence à l'origine et est défini par les arêtes dans $V_{s,J}$. De plus, soit $\mathcal{L}\left(V_{s,J}\right)$ l'espace linéaire généré par les vecteurs dans $V_{s,J}$. Dans Meyer (1999), il est démontré que la projection sur Ω_s^0 équivaut à la projection sur $\mathcal{L}\left(V_{s,J}\right)$ pour un ensemble approprié J. Si les lignes de la matrice de contraintes $\mathbf{A}$ sont linéairement indépendantes, alors l'ensemble minimal J est unique. Sinon, plus d'un J peut définir l'espace linéaire. Dans ce dernier cas, la projection est tout de même unique (voir le théorème 1 de la section suivante).

Wu et coll. (2016) ont examiné la solution à (2.3) dans le cas particulier d'une relation monotone entre des domaines définis avec une seule variable catégorique. Dans ce cas, la solution équivaut à l'algorithme PAVA (Pool Adjacent Violator Algorithm), qui a une expression explicite en termes de regroupement des domaines voisins. Les résultats théoriques présentés dans Wu et coll. (2016) ont été obtenus au moyen de cette expression explicite et ne s'appliquent donc pas au cas plus général considéré ici. Néanmoins, comme dans le cas de l'exemple simple à six domaines de la section 1 et dans de nombreuses situations d'intérêt pratique, la matrice spécifique $\mathbf{A}$ correspondra souvent à un *ordonnancement partiel* multivarié des moyennes de domaine. Selon un ordonnancement partiel, la solution à la minimisation contrainte de

(2.3) équivaut encore une fois à un regroupement de domaines voisins qui respecte les contraintes d'ordonnancement partiel. Voir par exemple dans Robertson et coll. (1988, page 23) une expression explicite de cette expression de domaine regroupée selon un ordonnancement partiel, comprenant la définition du regroupement. Cependant, contrairement à l'algorithme PAVA dans le cas univarié, cela ne donne pas d'algorithme de calcul général pratique. Dans l'article, nous allons permettre l'utilisation d'une matrice de contraintes arbitraire et irréductible $\mathbf{A}$, qui inclura l'ordonnancement partiel et la monotonicité univariée comme cas particuliers.

Une des méthodes possibles de calcul de $\tilde{\boldsymbol{\phi}}_s$ se fonde sur les arêtes du cône de contrainte Ω_s. Cependant, le nombre d'arêtes peut être considérablement plus grand que le nombre de contraintes pour les grandes valeurs de D, surtout quand il y a plus de contraintes que de domaines (voir Meyer, 1999). En outre, étant donné l'absence de solution sous forme fermée générale pour les arêtes de Ω_s (quand $m > D$), les arêtes doivent être calculées numériquement dans ce cas. En raison des ressources de calcul importantes nécessaires à cette tâche, la méthode est inefficace pour calculer $\tilde{\boldsymbol{\phi}}_s$. Un algorithme plus efficace basé sur le calcul de la projection sur le cône polaire a été mis au point : l'algorithme de projection du cône (APC) (Meyer, 2013). Cette autre méthode tire parti des arêtes facilement trouvables du cône polaire $\boldsymbol{\gamma}_{s_j}$, des conditions de (2.8) et du fait que $\Pi\left(\tilde{\mathbf{z}}_s \,|\Omega_s\right) = \tilde{\mathbf{z}}_s - \Pi\left(\tilde{\mathbf{z}}_s \,|\Omega_s^0\right)$. Ce dernier fait est un élément essentiel des démonstrations des principaux résultats théoriques présentés dans notre article. L'APC a été mis en œuvre sur le logiciel R dans le module `coneproj`. Pour des précisions, voir Liao et Meyer (2014).

Dans les situations où les contraintes correspondent à un ordonnancement complet ou partiel, la solution de l'APC correspond encore une fois au regroupement de domaines. On peut ensuite calculer explicitement les estimations de moyennes de domaine comme moyennes de domaine fondées sur l'échantillon pour les domaines regroupés déterminés par l'APC. Cela facilite grandement l'intégration de cette méthodologie dans les procédures d'estimation des enquêtes, parce que les définitions des domaines regroupés peuvent être facilement communiquées dans les instructions accompagnant la publication d'un ensemble de données d'enquête, et qu'on peut calculer les estimations sans nécessiter d'accès à un logiciel spécialisé.

2.3 Estimation de la variance de $\tilde{\theta}_{s_d}$

Il est compliqué d'estimer correctement la variance de $\tilde{\theta}_{s_d}$, car la projection de $\tilde{\mathbf{z}}_s$ sur Ω_s^0 (ou sur $\Omega_s\partial$) peut ne pas toujours se faire sur le même espace linéaire $\mathcal{L}\left(V_{s,J}\right)$ pour différents échantillons s. Pour mieux le comprendre, nous définissons $\mathcal{G}_s$ comme étant l'ensemble de tous les sous-ensembles $J \subseteq \{1, 2, \ldots, m\}$ de sorte que $\Pi\left(\tilde{\mathbf{z}}_s \,|\, \Omega_s^0\right) = \Pi\left(\tilde{\mathbf{z}}_s \,|\mathcal{L}\left(V_{s,J}\right)\right) \in \bar{\mathcal{F}}_{s,J}$, selon la définition qui se trouve dans (2.9). Comme nous l'avons indiqué plus haut, il pourrait y avoir différents ensembles J_1 et J_2 tels que la projection sur le cône polaire Ω_s^0 soit égale à la projection sur $\mathcal{L}\left(V_{s,J_1}\right)$ ou $\mathcal{L}\left(V_{s,J_2}\right)$. Toutefois, quel que soit l'ensemble choisi, la projection $\tilde{\boldsymbol{\rho}}_s$ est unique.

Pour illustrer ce qui précède, considérons les restrictions suivantes avec trois domaines seulement : la première moyenne de domaine doit être inférieure ou égale à la deuxième moyenne de domaine et la troisième moyenne de domaine doit être supérieure ou égale à la moyenne des deux premières moyennes de domaine. Par conséquent, la matrice de contraintes $\mathbf{A}$ peut être exprimée comme suit :

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & -1 & 2 \end{pmatrix}.$$

Supposons qu'on observe que $\tilde{y}_{s_1} = \tilde{y}_{s_2} < \tilde{y}_{s_3}$. Le vecteur transformé $\tilde{\mathbf{z}}_s$ possède des éléments de forme

$$\tilde{z}_{s_1} = \sqrt{\frac{\hat{N}_1}{\hat{N}}}\,\tilde{y}_{s_1},\ \tilde{z}_{s_2} = \sqrt{\frac{\hat{N}_2}{\hat{N}}}\,\tilde{y}_{s_2},\ \tilde{z}_{s_3} = \sqrt{\frac{\hat{N}_3}{\hat{N}}}\,\tilde{y}_{s_3}.$$

Dans ce cas, on voit aisément que $\Pi\left(\tilde{\mathbf{z}}_s \,|\Omega_s^0\right) = \mathbf{0}$. Dans le processus de calcul à l'aide de l'algorithme général, nous projetons $\tilde{\mathbf{z}}_s$ sur chacun des $2^2 = 4$ espaces linéaires générés par les arêtes du cône polaire

$$\boldsymbol{\gamma}_{s_1} = \left(\sqrt{\frac{\hat{N}}{\hat{N}_1}},\ -\sqrt{\frac{\hat{N}}{\hat{N}_2}},\ 0\right)^{\mathrm{T}},\ \boldsymbol{\gamma}_{s_2} = \left(\sqrt{\frac{\hat{N}}{\hat{N}_1}},\ \sqrt{\frac{\hat{N}}{\hat{N}_2}},\ -2\sqrt{\frac{\hat{N}}{\hat{N}_3}}\right)^{\mathrm{T}}.$$

On constate ainsi que les conditions $\Pi\left(\tilde{\mathbf{z}}_s \,|\Omega_s^0\right) = \mathbf{0} = \Pi\left(\tilde{\mathbf{z}}_s \,|\mathcal{L}\left(V_{s,J}\right)\right) \in \bar{\mathcal{F}}_{s,J}$ sont satisfaites uniquement pour $J = \varnothing$ et $J = \{1\}$, ce qui implique que $\mathcal{G}_s = \{\varnothing, \{1\}\}$. De plus, notons que $V_{s,\varnothing}$ et $V_{s,\{1\}}$ ne couvrent pas les mêmes espaces linéaires, ce qui complique l'estimation de la variance de $\tilde{\theta}_{s_d}$. Dans le cas fondé sur un modèle avec des variables continues, l'ensemble des vecteurs de l'échantillon où ces scénarios se produisent est nul. Ils ne peuvent cependant pas être exclus de la configuration fondée sur le plan.

Nous proposons un estimateur de la variance pour $\tilde{\theta}_{s_d}$ qui repose sur les ensembles dans $\mathcal{G}_s$ et est fondé sur des méthodes de linéarisation. Considérons tout ensemble fixe $J \in \mathcal{G}_s$ et supposons que $\mathbf{P}_{s,J}$ soit la matrice de projection correspondant à l'espace linéaire $\mathcal{L}\left(V_{s,J}\right)$, où $\mathbf{P}_{s,\varnothing}$ est la matrice de zéros par convention. En sélectionnant J, on peut alors exprimer $\tilde{\boldsymbol{\rho}}_s$ sous la forme $\mathbf{P}_{s,J}\tilde{\mathbf{z}}_s$, ce qui implique que $\tilde{\boldsymbol{\theta}}_s$ peut être écrit comme étant $\tilde{\boldsymbol{\theta}}_{s,J} = \tilde{\mathbf{y}}_s - \mathbf{W}_s^{-1/2}\mathbf{P}_{s,J}\mathbf{W}_s^{1/2}\tilde{\mathbf{y}}_s$, où l'on ajoute l'indice J dans $\tilde{\boldsymbol{\theta}}_s$ pour tenir compte du fait que l'expression dépend du J choisi.

Nous constatons alors que $\tilde{\boldsymbol{\theta}}_{s,J}$ est une fonction non linéaire lisse des $\hat{t}_d$ et des $\hat{N}_d$, où $\hat{t}_d$ est l'estimateur de Horvitz-Thompson de $t_d = \sum_{k \in U_d} y_k$. Par conséquent, en traitant J comme une valeur fixe, nous obtenons la variance asymptotique de $\tilde{\theta}_{s_d,J}$ par linéarisation en séries de Taylor (Särndal et coll., 1992, page 175) comme suit :

$$\mathrm{VA}\left(\tilde{\theta}_{s_d,J}\right) = \sum_{k \in U}\sum_{l \in U}\Delta_{kl}\frac{u_k}{\pi_k}\frac{u_l}{\pi_l}, \tag{2.10}$$

où $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$ et

$$u_k = \sum_{i=1}^{D} \alpha_i y_k 1_{k \in U_i} + \sum_{i=1}^{D} \beta_i 1_{k \in U_i} \quad \text{pour} \quad k = 1, 2, \ldots, N,$$

1_A étant la variable d'indicateur de l'événement A, et

$$\alpha_i = \frac{\partial \tilde{\theta}_{s_d, J}}{\partial \hat{t}_i}\bigg|_{(\hat{t}_1, \ldots, \hat{t}_D, \hat{N}_1, \ldots, \hat{N}_D) = (t_1, \ldots, t_D, N_1, \ldots, N_D)}; \quad \beta_i = \frac{\partial \tilde{\theta}_{s_d, J}}{\partial \hat{N}_i}\bigg|_{(\hat{t}_1, \ldots, \hat{t}_D, \hat{N}_1, \ldots, \hat{N}_D) = (t_1, \ldots, t_D, N_1, \ldots, N_D)}.$$

De plus, un estimateur convergent de la variance asymptotique dans (2.10) est donné par

$$\hat{V}\left(\tilde{\theta}_{s_d, J}\right) = \sum_{k \in s} \sum_{l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{\hat{u}_k}{\pi_k} \frac{\hat{u}_l}{\pi_l}, \tag{2.11}$$

où

$$\hat{u}_k = \sum_{i=1}^{D} \hat{\alpha}_i y_k 1_{k \in s_i} + \sum_{i=1}^{D} \hat{\beta}_i 1_{k \in s_i} \quad \text{pour} \quad k = 1, 2, \ldots, N,$$

et où l'on obtient $\hat{\alpha}_i$, $\hat{\beta}_i$ à partir de α_i, β_i en substituant les estimateurs de Horvitz-Thompson appropriés pour chaque total de population. Nous proposons l'estimateur dans (2.11), calculé à la valeur de J obtenue dans l'échantillon, comme estimateur de la variance de $\tilde{\theta}_{s_d}$.

Afin de donner un exemple clair de l'estimateur de la variance proposé pour $\tilde{\theta}_{s_d}$, considérons le scénario présenté au début de la sous-section. Étant donné que $\mathcal{G}_s = \{\varnothing, \{1\}\}$, il peut être intéressant de calculer la variance estimée de $\tilde{\theta}_{s_d, J}$ pour $J = \{1\}$ et une certaine valeur d. La matrice $\mathbf{P}_{s,\{1\}}$ est la matrice de projection correspondant à l'espace linéaire généré par $\boldsymbol{\gamma}_{s_1}$, donné par

$$\mathbf{P}_{s,\{1\}} = \left(\hat{N}_1 + \hat{N}_2\right)^{-1} \begin{pmatrix} \hat{N}_2 & -\sqrt{\hat{N}_1 \hat{N}_2} & 0 \\ -\sqrt{\hat{N}_1 \hat{N}_2} & \hat{N}_1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notons que $\mathbf{P}_{s,\{1\}}$ est une fonction de $\left(\hat{N}_1, \hat{N}_2, \hat{N}_3\right)$ parce que $\boldsymbol{\gamma}_{s_1}$ l'est. Au moyen de l'équation ci-dessus, on peut simplifier $\tilde{\boldsymbol{\theta}}_{s,\{1\}}$ sous la forme suivante,

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_{s,\{1\}} = \left(\tilde{\theta}_{s_1,\{1\}}, \tilde{\theta}_{s_2,\{1\}}, \tilde{\theta}_{s_3,\{1\}}\right)^{\mathsf{T}} &= \left(\frac{\hat{N}_1 \tilde{y}_{s_1} + \hat{N}_2 \tilde{y}_{s_2}}{\hat{N}_1 + \hat{N}_2}, \frac{\hat{N}_1 \tilde{y}_{s_1} + \hat{N}_2 \tilde{y}_{s_2}}{\hat{N}_1 + \hat{N}_2}, \tilde{y}_{s_3}\right)^{\mathsf{T}} \\ &= \left(\frac{\hat{t}_1 + \hat{t}_2}{\hat{N}_1 + \hat{N}_2}, \frac{\hat{t}_1 + \hat{t}_2}{\hat{N}_1 + \hat{N}_2}, \frac{\hat{t}_3}{\hat{N}_3}\right)^{\mathsf{T}}. \end{aligned}$$

Ainsi, si l'on a un domaine d, on peut dériver les α et les β en prenant les dérivées partielles de $\tilde{\theta}_{s_d,\{1\}}$ par rapport aux $\hat{t}$ et aux $\hat{N}$, et en évaluant ces dérivés aux t et aux N. Si $d = 2$, on obtient

$$\begin{aligned} \alpha_1 &= \alpha_2 = \frac{1}{N_1 + N_2}, \qquad \alpha_3 = 0, \\ \beta_1 &= \beta_2 = -\frac{t_1 + t_2}{\left(N_1 + N_2\right)^2}, \quad \beta_3 = 0. \end{aligned}$$

Les $\hat{\alpha}$ et les $\hat{\beta}$ sont calculés par substitution des estimateurs de Horvitz-Thompson dans les équations ci-dessus, qui servent ensuite à évaluer $\hat{u}_k$ pour chaque k dans l'échantillon s. On peut alors enfin calculer l'estimateur de la variance proposé dans (2.11).

3 Propriétés de l'estimateur contraint

3.1 Hypothèses

Pour obtenir nos résultats théoriques, nous établissons des hypothèses sur le comportement asymptotique de la population U_N et sur le plan de sondage p_N :

A1. Le nombre de domaines D est fixe.

A2. $\limsup_{N\to\infty} N^{-1}\sum_{k\in U}|y_k|^r < \infty$, pour $r = 1, 2$.

A3. Pour $d = 1, \ldots, D$, il existe des constantes μ_d et $r_d > 0$ de sorte que $\bar{y}_{U_d,N} - \mu_d = O(N^{-1/2})$ et $N_{d,N}/N - r_d = O(N^{-1/2})$, pour toutes les valeurs de d.

A4. La taille de l'échantillon n_N est non aléatoire et elle satisfait $0 < \lim_{N\to\infty} n_N/N < 1$. De plus, il existe des valeurs ε, $0 < \varepsilon < 1$, de sorte que $n_{d,N} \geq \varepsilon n_N/D$ pour tous les d et tous les N.

A5. Pour toutes les valeurs de N, $\min_{k\in U_N}\pi_k \geq \lambda > 0$, $\min_{k,l\in U_N}\pi_{kl} \geq \lambda^* > 0$, et

$$\limsup_{N\to\infty} n_N \max_{k,l\in U_N:\, k\neq l}|\Delta_{kl}| < \infty.$$

A6. L'estimateur de Horvitz-Thompson $\hat{\mathbf{x}}_{s_N}$ du vecteur bidimensionnel $2D$ des moyennes de la population $\bar{\mathbf{x}}_{U_N} = N^{-1}(t_1, \ldots, t_D, N_1, \ldots, N_D)^\top$ satisfait

$$\mathrm{var}_{p_N}\left(\hat{\mathbf{x}}_{s_N}\right)^{-1/2}\left(\hat{\mathbf{x}}_{s_N} - \bar{\mathbf{x}}_{U_N}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_{2D}),$$

et

$$\hat{\mathrm{var}}\left(\hat{\mathbf{x}}_{s_N}\right) - \mathrm{var}_{p_N}\left(\hat{\mathbf{x}}_{s_N}\right) = o_p\left(n_N^{-1}\right);$$

où $\mathbf{I}_q$ désigne la matrice d'identité de dimension q, la matrice de variance-covariance par rapport au plan $\mathrm{var}_{p_N}\left(\hat{\mathbf{x}}_{s_N}\right)$ est définie positive et $\hat{\mathrm{var}}\left(\hat{\mathbf{x}}_{s_N}\right)$ est l'estimateur de Horvitz-Thompson de var_{p_N}.

L'hypothèse A1 établit que le nombre de domaines demeure constant quand la taille de la population change. La condition de l'hypothèse A2 vise à assurer la convergence par rapport au plan de sondage des estimateurs de Horvitz-Thompson aux niveaux de la population et du domaine. Notons en particulier que cette condition est satisfaite quand la variable y est limitée, ce qui peut être supposé naturellement pour de nombreux types de variables d'enquête. L'hypothèse A3 garantit que les moyennes et les tailles de

domaine de population convergent respectivement vers les valeurs limites μ_d et r_d. Par ailleurs, les valeurs μ peuvent être considérées comme des espérances de superpopulation pour une distribution générant les éléments de population y_k sous forme de tirages indépendants. Nos résultats théoriques dépendent en fait de la validité des contraintes supposées pour ces espérances de superpopulation et non pas pour les moyennes de domaine de population. Bien que cela puisse sembler inadéquat compte tenu de notre intérêt pour l'utilisation des contraintes au niveau de la population, l'hypothèse A3 garantit que la forme des moyennes de domaine soit raisonnablement près de la forme des moyennes de superpopulation. Selon l'hypothèse A4, la taille de l'échantillon dans chaque domaine ne peut pas être inférieure à une fraction du ratio n/D, ce qu'on obtiendrait en divisant également la taille de l'échantillon sur tous les domaines. Cette hypothèse vise à ce que les moments des fonctions lisses de $N^{-1}\hat{t}_d$ et que les $N^{-1}\hat{N}_d$ soient bornés. De plus, elle établit que la taille de l'échantillon est non aléatoire. Cela peut être adapté à une taille d'échantillon aléatoire en imposant certaines conditions à la taille d'échantillon espérée $\mathbb{E}_p(n)$. L'hypothèse A5 établit que les bornes inférieures sont non nulles pour les probabilités d'inclusion du premier et du second ordre, et que les covariances du plan Δ_{kl} doivent converger vers zéro au moins aussi rapidement que n^{-1}. L'hypothèse A6 garantit une normalité asymptotique pour $\hat{\mathbf{x}}_{s_N}$, nécessaire au maintien des propriétés de normalité sur les estimateurs non linéaires qui sont exprimés comme des fonctions lisses de $\hat{\mathbf{x}}_{s_N}$. Elle sert également à établir les conditions de convergence de l'estimateur de variance-covariance. On trouve dans la littérature les résultats de normalité asymptotique pour des plans en particulier, notamment le résultat classique de Hájek (1960) pour l'échantillonnage de Poisson et l'échantillonnage aléatoire simple sans remise. Comme autres démonstrations du théorème central limite pour un échantillonnage stratifié, citons Krewski et Rao (1981), qui ont étudié des échantillons stratifiés à probabilités inégales avec remise, Bickel et Freedman (1984), qui ont considéré un échantillonnage aléatoire simple sans remise stratifié, et Breidt, Opsomer et Sanchez-Borrego (2016), qui ont examiné des plans à généraux à probabilités inégales, avec ou sans remise.

3.2 Principaux résultats

Nous dérivons les propriétés théoriques de l'estimateur contraint en nous concentrant sur la projection sur Ω_s^0 au lieu de Ω_s. Rappelons que les arêtes du cône polaire Ω_s^0 sont tout simplement les m lignes de $-\mathbf{A}_s$, notées par $\boldsymbol{\gamma}_{s_j}$, et que $\tilde{\boldsymbol{\rho}}_s$, la projection sur Ω_s^0, peut être décrite par les ensembles $J \in \mathcal{G}_s$. Le fait de pouvoir caractériser la propriété selon laquelle $J \in \mathcal{G}_s$ sous la forme de vecteurs dans $V_{s,J}$ nous permet d'obtenir des taux de convergence théoriques, qui servent à développer les propriétés d'inférence de l'estimateur contraint. Quand l'ensemble $J \in \mathcal{G}_s$ produit un ensemble de vecteurs linéaires indépendants $V_{s,J}$, il est alors simple d'écrire $\tilde{\boldsymbol{\rho}}_s$ comme étant $\mathbf{P}_{s,J}\,\tilde{\mathbf{z}}_s = \mathbf{A}_{s,J}^{\top}\left(\mathbf{A}_{s,J}\mathbf{A}_{s,J}^{\top}\right)^{-1}\mathbf{A}_{s,J}\,\tilde{\mathbf{z}}_s$, où $\mathbf{A}_{s,J}$ désigne la matrice formée par les lignes de $\mathbf{A}_s$ dans les positions J. Par conséquent, selon les conditions décrites dans (2.8), $J \in \mathcal{G}_s$ si et seulement si

$$\left\langle \tilde{\mathbf{z}}_s - \mathbf{P}_{s,J}\,\tilde{\mathbf{z}}_s,\, \boldsymbol{\gamma}_{s_j} \right\rangle \le 0 \quad \text{pour} \quad j \notin J, \quad \text{et} \quad \left(\mathbf{A}_{s,J}\,\mathbf{A}_{s,J}^{\top}\right)^{-1}\mathbf{A}_{s,J}\,\tilde{\mathbf{z}}_s \ge \mathbf{0} \tag{3.1}$$

dans ce cas, où cette dernière condition permet que $\Pi\left(\tilde{\mathbf{z}}_s \left| \mathcal{L}\left(V_{s,J}\right)\right.\right) \in \bar{\mathcal{F}}_{s,J}$. Cependant, il est possible que l'ensemble $J \in \mathcal{G}_s$ produise un ensemble de vecteurs linéairement dépendants $V_{s,J}$. Dans ce cas, le théorème 1 ci-dessous garantit qu'il est toujours possible de trouver un sous-ensemble $J^* \subset J$ de sorte que V_{s,J^*} soit un ensemble linéairement indépendant qui couvre le même espace linéaire que $V_{s,J}$ et qui satisfasse $J^* \in \mathcal{G}_s$. On peut ainsi établir des conditions analogues comme dans (3.1) au moyen de J^* plutôt que J.

Théorème 1. *Soit* $\mathbf{A}$ *une matrice irréductible* $m \times D$ *avec des lignes* $-\boldsymbol{\gamma}_j$. *Soit* Ω^0 *le cône polaire correspondant. Pour tout ensemble* $J \subseteq \{1, 2, \ldots, m\}$, *définissons* $V_J = \{\boldsymbol{\gamma}_j : j \in J\}$. *Ensuite, soit* $\bar{\mathcal{F}}_J$ *le sous-cône de* Ω^0 *généré par les arêtes données par l'ensemble* J. *Pour un vecteur* $\mathbf{z}$, *définissons son ensemble* $\mathcal{G}$ *formé par tous les ensembles* $J \subseteq \{1, 2, \ldots, m\}$ *de sorte que* $\Pi\left(\mathbf{z} \left| \Omega^0\right.\right) = \Pi\left(\mathbf{z} \left| \mathcal{L}\left(V_J\right)\right.\right) \in \bar{\mathcal{F}}_J$. *Supposons que* J *est un ensemble non vide de sorte que* V_J *soit qui un ensemble linéairement dépendant et* $J \in \mathcal{G}$. *Ensuite, on a* $J^* \subset J$ *de sorte que* V_{J^*} *soit un ensemble linéairement indépendant,* $\mathcal{L}\left(V_{J^*}\right) = \mathcal{L}\left(V_J\right)$, *et* $J^* \in \mathcal{G}$.

Tous les concepts ci-dessus qui ont été définis au niveau de l'échantillon peuvent être définis de manière analogue au niveau de la superpopulation. En particulier, soit $\mathcal{G}_\mu$ l'ensemble de tous les sous-ensembles $J \subseteq \{1, \ldots, m\}$ de sorte que $\Pi\left(\mathbf{z}_\mu \left| \Omega_\mu^0\right.\right) = \Pi\left(\mathbf{z}_\mu \left| \mathcal{L}\left(V_{\mu,J}\right)\right.\right) \in \bar{\mathcal{F}}_{\mu,J}$, où $\mathbf{z}_\mu$, Ω_μ^0, $V_{\mu,J}$ et $\bar{\mathcal{F}}_{\mu,J}$ soient les versions analogues de $\tilde{\mathbf{z}}_s$, Ω_s^0, $V_{s,J}$ et $\bar{\mathcal{F}}_{s,J}$ qu'on obtient en remplaçant $\tilde{\mathbf{y}}_s$ et $\mathbf{W}_s$ par $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_D)$ et $\mathbf{W}_\mu = \text{diag}(r_1, r_2, \ldots, r_D)$. On peut établir des conditions nécessaires et suffisantes comme celles de (2.8) de manière analogue pour caractériser le vecteur $\boldsymbol{\rho}_\mu$ comme étant la projection sur Ω_μ^0.

Rappelons que l'ensemble $\mathcal{G}_s$ peut varier selon les échantillons. Ajoutons que les petits échantillons très variables sont susceptibles de choisir des ensembles $J \in \mathcal{G}_s$ qui ne sont pas choisis dans $\mathcal{G}_\mu$ qui est « asymptotiquement correct ». Cependant, à mesure que la taille de l'échantillon augmente, ces choix incorrects sont moins susceptibles de se produire, car les moyennes de domaine de l'échantillon se rapprochent des moyennes limites de domaine de la population. Le théorème 2 précise cette idée en établissant que les ensembles qui ne sont pas dans $\mathcal{G}_\mu$ ont une probabilité asymptotiquement négligeable d'être choisis dans l'échantillon.

Théorème 2. *Considérons tout ensemble* $J \subseteq \{1, 2, \ldots, m\}$ *de sorte que* $J \notin \mathcal{G}_\mu$. *Alors,* $P\left(J \in \mathcal{G}_s\right) = O\left(n^{-1}\right)$.

Le théorème 3 ci-dessous montre la normalité asymptotique de l'estimateur contraint et justifie l'utilisation de l'estimateur de la variance par linéarisation pour la projection observée (ou le regroupement en cas d'ordonnancement partiel) pour l'inférence asymptotique pour la moyenne d'une population finie. Il généralise le théorème 2 de Wu et coll. (2016), qui considérait uniquement les restrictions monotones. Nous constatons la présence du terme de biais B dans la moyenne de la

distribution asymptotique. Cette situation non souhaitable se produit quand on a plus d'un ensemble $J \in \mathcal{G}_\mu$ de sorte que les arêtes correspondantes dans $V_{\mu,J}$ couvrent des espaces linéaires différents, ou de façon équivalente, que la projection sur le cône polaire Ω_μ^0 appartienne à l'intersection de ces espaces linéaires différents. Cependant, quand les contraintes sont strictes, c'est-à-dire $\mathbf{A}\boldsymbol{\mu} > \mathbf{0}$, le vecteur $\mathbf{z}_\mu$ est strictement à l'intérieur du cône de contrainte Ω_μ, et dans ce cas, on n'a pas d'ensemble $J \neq \varnothing$ de sorte que $\Pi\left(\mathbf{z}_\mu \,\middle|\, \mathcal{L}\left(V_{\mu,J}\right)\right) = \mathbf{0}$. Dans ce cas, le terme de biais disparaît.

Théorème 3. *Supposons que* $\boldsymbol{\mu}$ *satisfait* $\mathbf{A}\boldsymbol{\mu} \geq \mathbf{0}$. *Considérons tout ensemble* J *de sorte que* $J \in \mathcal{G}_s$. *Alors,*

$$\hat{V}\left(\tilde{\theta}_{s_d,J}\right)^{-1/2}\left(\tilde{\theta}_{s_d} - \bar{y}_{U_d}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(B, 1),$$

pour tout $d = 1, 2, \ldots, D$, *où* $B = O\left(\sqrt{\frac{n}{N}}\right)$ *est un terme de biais qui disparaît quand* $\mathbf{A}\boldsymbol{\mu} > \mathbf{0}$.

Le théorème 3 s'appuie sur le fait que les contraintes de forme supposées se vérifient pour le vecteur des moyennes de domaine limites $\boldsymbol{\mu}$ plutôt que pour le vecteur des moyennes de domaine de population $\bar{\mathbf{y}}_U$. Dans la section suivante, nous montrons par des simulations que l'estimateur contraint améliore à la fois l'estimation et la variabilité quand les domaines de population sont approximativement près de la forme supposée, comparativement aux estimateurs non contraints.

4 Performances de l'estimateur contraint

4.1 Simulations

Nous exécutons des expériences de simulation pour mesurer les performances de la méthodologie proposée dans le calcul de l'estimation et de l'inférence des moyennes de domaine de la population. Au moyen d'une paire de nombres naturels D_1 et D_2, nous générons les moyennes de domaine limites μ_d à partir de la fonction bivariée monotone $\mu(x_1, x_2)$ donnée par

$$\mu(x_1, x_2) = \sqrt{1 + 4x_1/D_1} + \frac{4\exp(0{,}5 + 2x_2/D_2)}{1+\exp(0{,}5 + 2x_2/D_2)}.$$

On crée les valeurs μ_d en évaluant $\mu(x_1, x_2)$ à chaque combinaison de $x_1 = 1, 2, \ldots, D_1$ et $x_2 = 1, 2, \ldots, D_2$, ce qui produit un nombre total de domaines égal à $D = D_1 D_2$. Nous établissons $D_1 = 6$ et $D_2 = 4$. Notons que bien que la fonction $\mu(x_1, x_2)$ produise une matrice plutôt qu'un vecteur de moyennes de domaine, elle peut être vectorisée afin de représenter les moyennes de domaine limites sous la forme du vecteur $\boldsymbol{\mu}$. Pour chaque domaine d, nous générons ses $N_d = N/D = 400$ éléments en ajoutant le bruit indépendant et normalement distribué de moyenne 0 et de variance σ^2 à μ_d. Une fois que les éléments de la population ont été simulés, les moyennes de domaine de population $\bar{\mathbf{y}}_U$ sont

calculées. Les moyennes de domaine de population utilisées dans les simulations quand $\sigma = 1$ sont illustrées dans la figure 4.1. Nous observons que ces moyennes de domaine sont raisonnablement (et non strictement) monotones par rapport à x_1 et x_2.

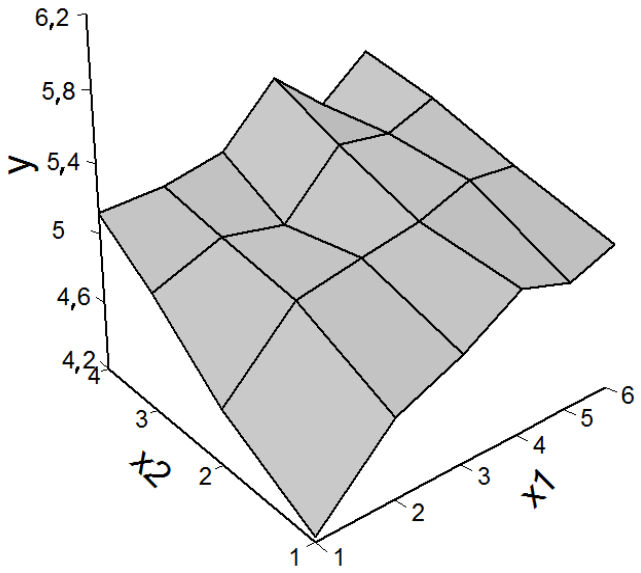


Figure 4.1 Moyennes de domaine de population pour les simulations quand $\sigma = 1$.

Les échantillons sont tirés d'un plan de sondage stratifié sans remise, comprenant 4 strates qui recoupent les domaines D. Les strates sont construites au moyen d'une variable auxiliaire v qui est corrélée à la variable d'intérêt y. Le vecteur v est créé par l'ajout d'un bruit standard indépendant normalement distribué à $\sigma d/D$ pour chaque élément du domaine d. On attribue ensuite l'appartenance à une strate en triant le vecteur v et en créant 4 blocs de $N/4 = 2\,400$ éléments, chacun étant fondé sur le vecteur v trié. Pour rendre le plan informatif, nous échantillonnons $n = 480$ éléments répartis entre les strates (60, 120, 120, 180). Ce plan de sondage probabiliste ressemble à celui décrit dans Wu et coll. (2016).

Nous examinons quatre scénarios différents obtenus à partir de la combinaison de deux types possibles de contraintes de forme et de $\sigma = 1$ ou 2. Le premier type de contraintes suppose que les moyennes de domaine de population sont monotones et augmentent pour ce qui est de x_1 et x_2 (doublement monotone), tandis que le deuxième type de contraintes suppose la monotonicité uniquement pour ce qui est de x_1 (seulement monotone pour ce qui est de x_1). Si σ est fixe, on considère exactement la même population pour les deux types de contraintes possibles. Pour chaque scénario, les estimations non contraintes $\tilde{\mathbf{y}}_s$ et contraintes $\tilde{\boldsymbol{\theta}}_s$ sont calculées avec leurs estimations de la variance par linéarisation (voir (2.11)). Les estimations contraintes sont calculées au moyen de l'APC, et leurs estimations de la variance sont calculées à partir de l'ensemble sélectionné de l'échantillon $J \in \mathcal{G}_s$. En outre, on construit des intervalles de confiance de 95 % de Wald fondés sur la distribution normale pour les deux estimateurs.

Pour mesurer la précision de $\tilde{\mathbf{y}}_s$ et $\tilde{\boldsymbol{\theta}}_s$ en tant qu'estimateurs des moyennes de domaine de population $\bar{\mathbf{y}}_U$, nous considérons l'erreur quadratique moyenne pondérée (EQMP) donnée par

$$\text{EQMP}\left(\tilde{\boldsymbol{\varphi}}_s\right) = \mathrm{E}\left[\left(\tilde{\boldsymbol{\varphi}}_s - \bar{y}_U\right)^{\mathrm{T}} \mathbf{W}_U \left(\tilde{\boldsymbol{\varphi}}_s - \bar{y}_U\right)\right],$$

où $\tilde{\boldsymbol{\varphi}}_s$ pourrait être l'estimateur non contraint ou contraint et $\mathbf{W}_U$ est la matrice diagonale avec les éléments N_d/N, $d = 1, \ldots, D$. Les valeurs de l'EQMP sont obtenues approximativement par des simulations comme suit :

$$\frac{1}{B}\sum_{b=1}^{B}\left(\tilde{\boldsymbol{\varphi}}_s^{(b)} - \bar{\mathbf{y}}_U\right)^{\mathrm{T}} \mathbf{W}_U \left(\tilde{\boldsymbol{\varphi}}_s^{(b)} - \bar{\mathbf{y}}_U\right),$$

où B est le nombre de simulations et $\tilde{\boldsymbol{\varphi}}_s^{(b)}$ est l'estimateur pour l'échantillon b^{e}.

Les résultats de la simulation sont résumés dans les figures 4.2 à 4.5 et sont fondés sur $R = 10\,000$ rééchantillonnages. Les figures montrent les 24 domaines divisés en groupes de 6, et chaque groupe est supposé monotone. Il est possible de représenter le scénario doublement monotone dans des graphiques similaires comprenant des groupes de quatre domaines monotones. Comme l'illustrent les ajustements d'un seul échantillon dans ces figures, on constate que les estimations contraintes peuvent être exactement égales aux estimations non contraintes pour certains domaines. Dans ces cas, leurs estimations de la variance sont égales aussi. Dans l'ensemble, les intervalles de confiance de l'estimateur contraint ont tendance à être plus étroits que ceux de l'estimateur non contraint. En moyenne, l'estimateur contraint se comporte légèrement différemment des moyennes de domaine de population, en raison de la monotonicité non stricte de ces dernières. L'estimateur contraint présente en effet l'avantage d'avoir des centiles plus étroits, ce qui montre que la distribution de l'estimateur proposé est plus étroite que la distribution de l'estimateur non contraint. Pour les petites valeurs de σ, les estimations non contraintes sont plus susceptibles de satisfaire les restrictions supposées, ce qui apporte de petites améliorations à l'estimateur contraint par rapport à l'estimateur non contraint. En revanche, les hypothèses de forme tendent à être plus gravement non respectées dans les estimations non contraintes pour les valeurs plus grandes de σ, ce qui permet à l'estimateur proposé de gagner beaucoup plus d'efficacité dans ces cas. On peut constater cette dernière propriété en observant que la bande des centiles de l'estimateur contraint s'éloigne de plus en plus de la bande de l'estimateur non contraint à mesure que σ augmente.

Pour ce qui est de la variabilité, l'estimateur contraint a une plus petite variance que l'autre estimateur. De manière intéressante, elle est surestimée par l'estimation de la variance correspondante obtenue par linéarisation. En revanche, l'estimation de la variance de l'estimateur non contraint sous-estime la variance réelle, ce qui est un inconvénient connu et souvent observé des variances obtenues par linéarisation. Malgré cette différence, les intervalles de confiance des deux estimateurs présentent un bon taux de couverture similaire quand $\sigma = 1$, alors que cette couverture est légèrement améliorée par l'estimateur contraint quand $\sigma = 2$.

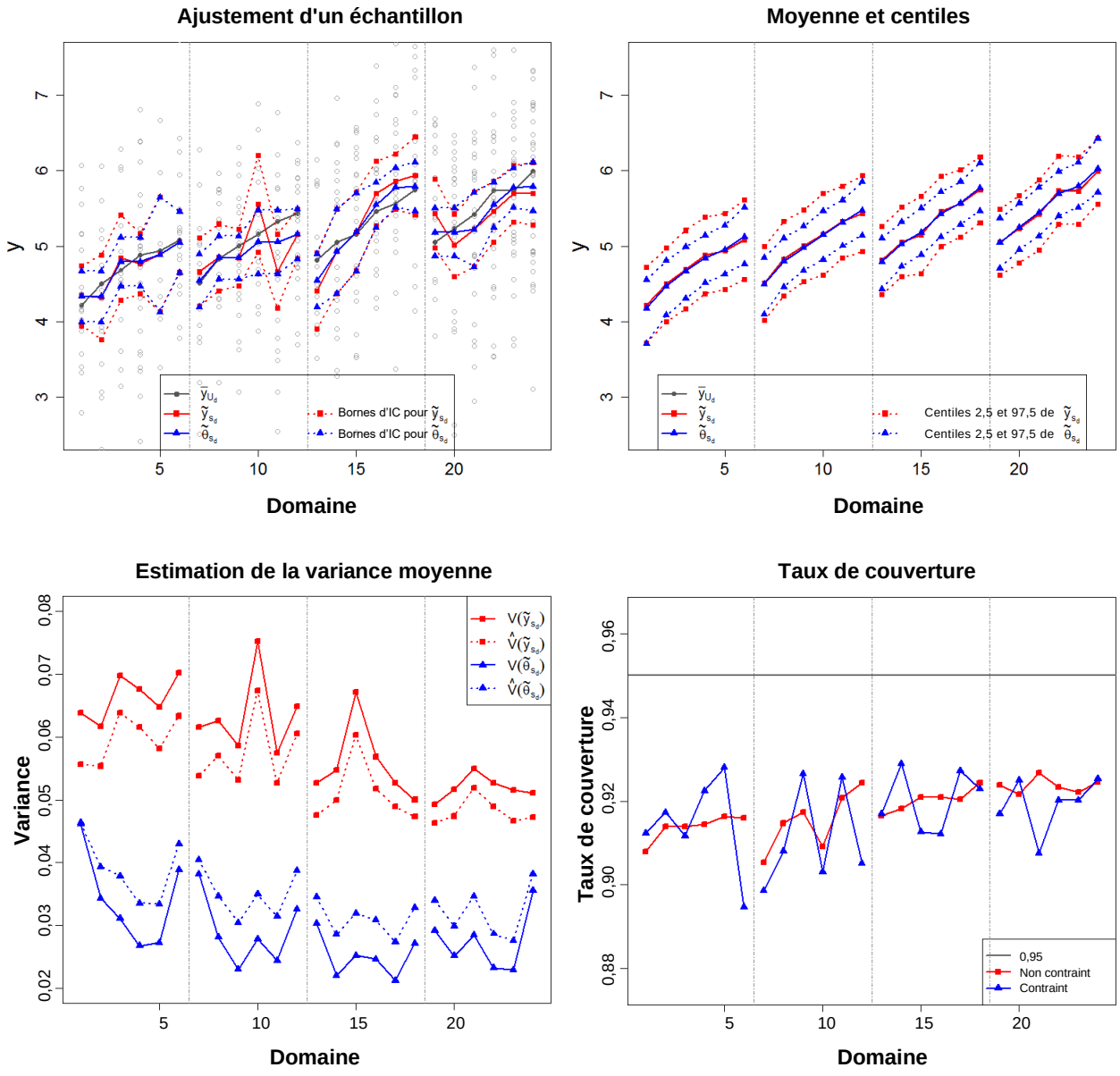


Figure 4.2 Graphiques des résultats de simulation pour les estimateurs non contraint et contraint dans le scénario doublement monotone avec $\sigma = 1$. Dans le graphique « Moyenne et centiles », $\bar{y}_{U_d}$ est masqué par $\tilde{y}_{s_d}$.

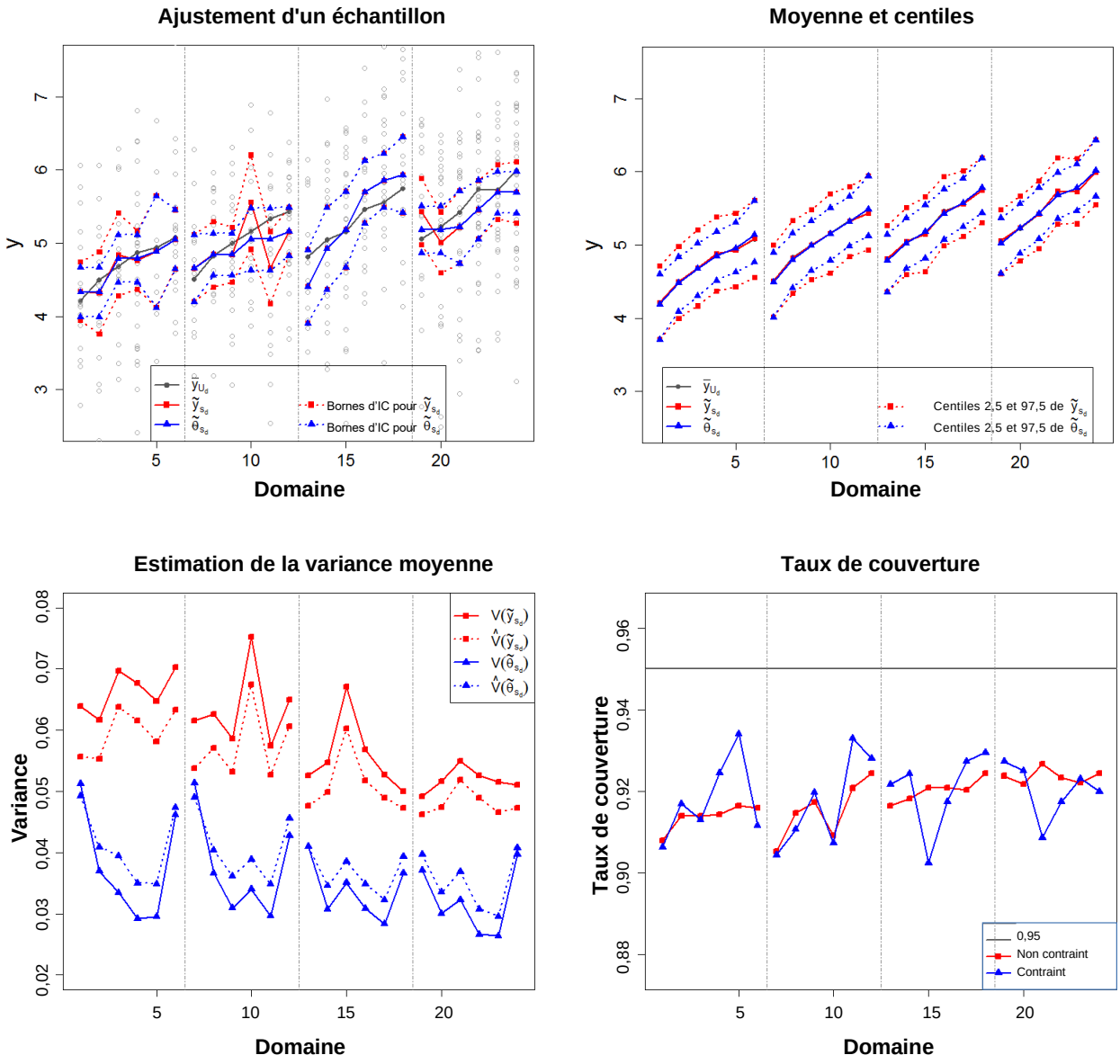


Figure 4.3 Graphiques des résultats de simulation pour les estimateurs non contraint et contraint dans le scénario monotone seulement pour ce qui est de x_1 avec $\sigma = 1$. Dans le graphique « Moyenne et centiles », $\bar{y}_{U_d}$ est masqué par $\tilde{y}_{s_d}$.

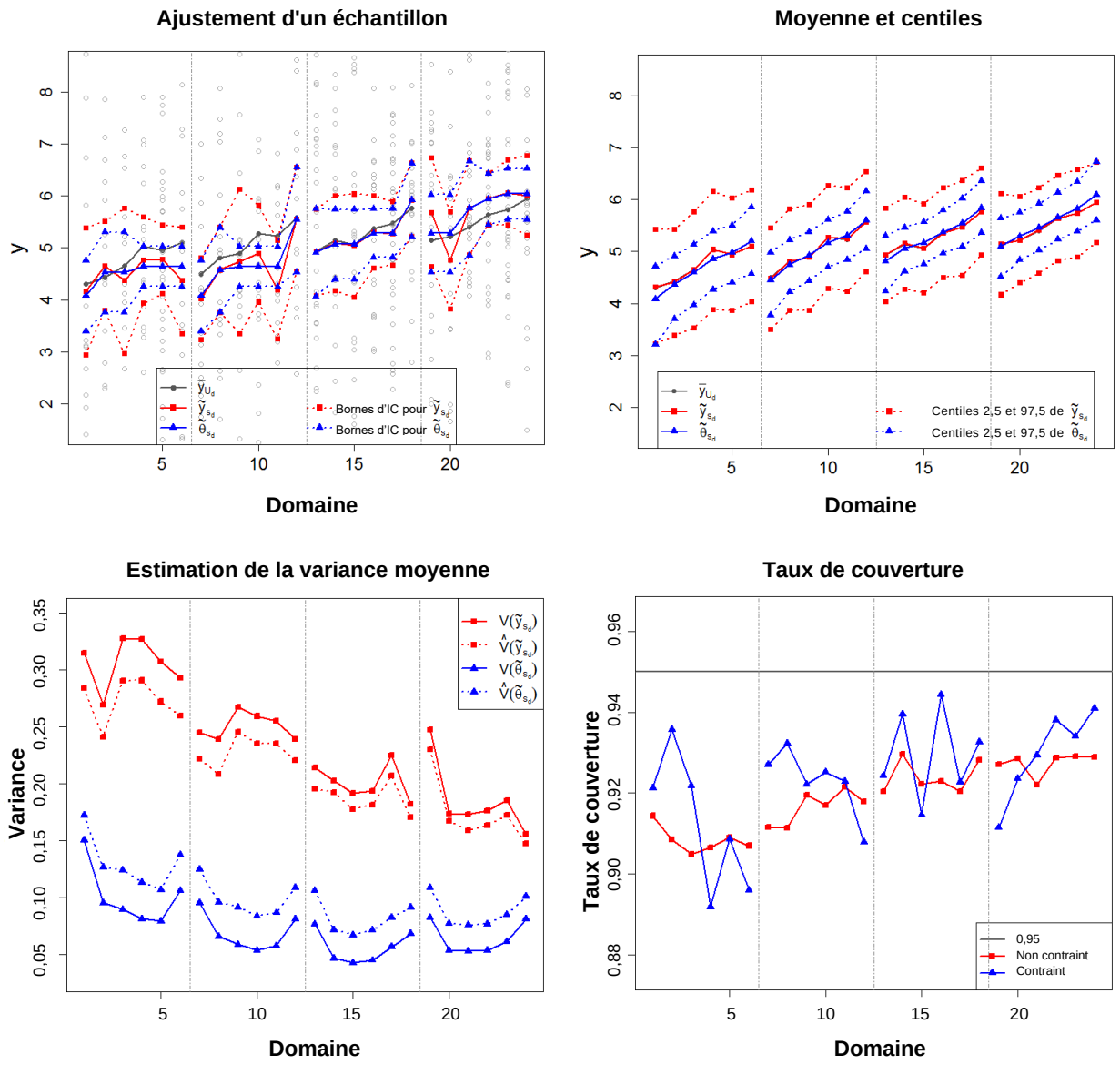


Figure 4.4 Graphiques des résultats de simulation pour les estimateurs non contraint et contraint dans le scénario doublement monotone avec $\sigma = 2$. Dans le graphique « Moyenne et centiles », $\bar{y}_{U_d}$ est masqué par $\tilde{y}_{s_d}$.

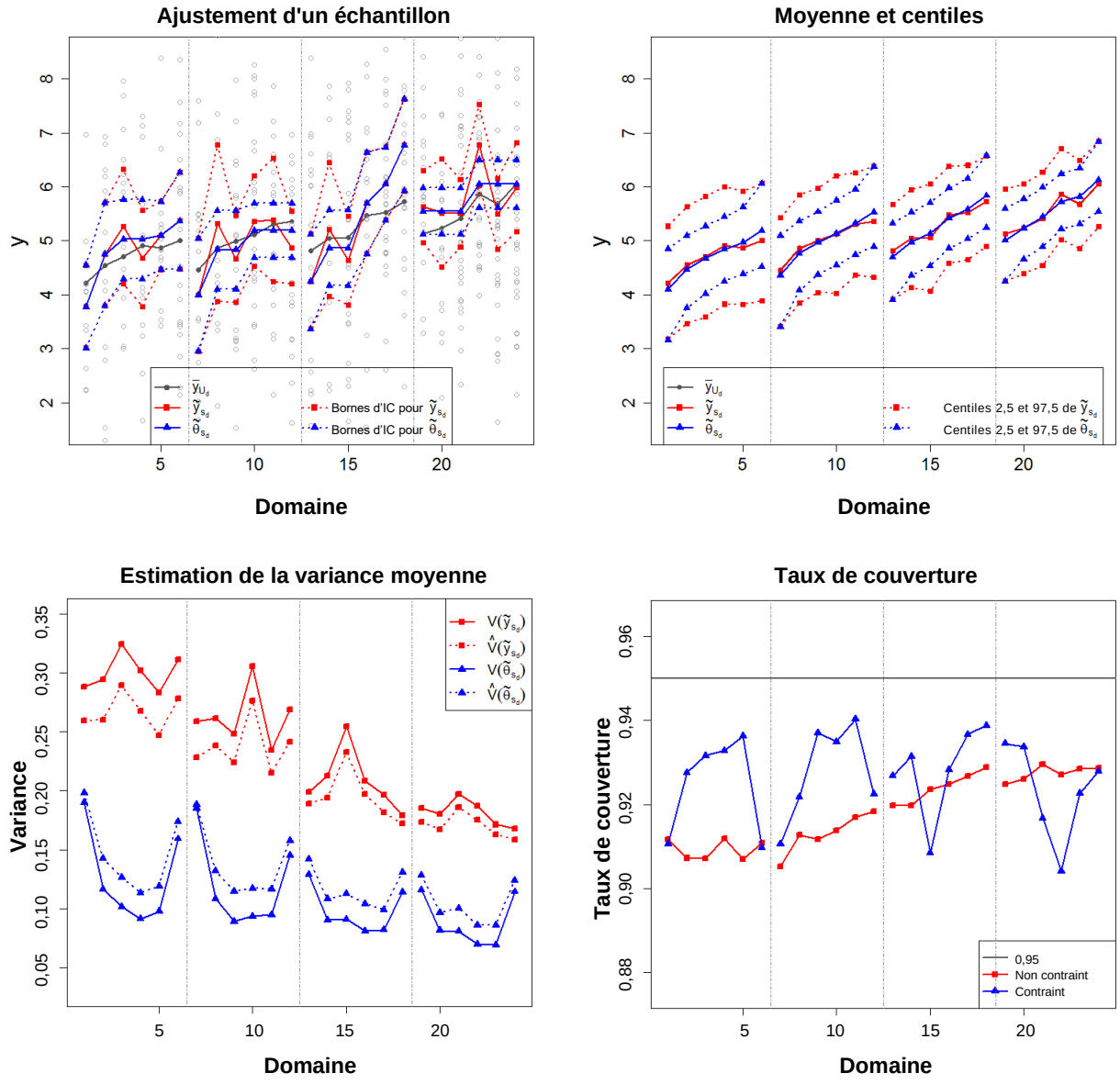


Figure 4.5 Graphiques des résultats de simulation pour les estimateurs non contraint et contraint dans le scénario monotone seulement pour ce qui est de x_1 avec $\sigma = 2$. Dans le graphique « Moyenne et centiles », $\bar{y}_{U_d}$ est masqué par $\tilde{y}_{s_d}$.

Le tableau 4.1 montre que l'estimateur contraint est plus précis en moyenne que l'estimateur non contraint. La précision de l'estimateur contraint s'améliore quand on suppose la monotonicité pour ce qui est des deux variables plutôt que seulement pour x_1. On s'y attend ici, car la surface sous-jacente est effectivement doublement monotone, de sorte que l'estimateur bénéficie du fait que la contrainte la plus forte est imposée.

Tableau 4.1
Valeurs empiriques de l'EQMP

	Non contraint	Monotone seulement pour x_1	Doublement monotone
$\sigma = 1$	0,0593	0,0362	0,0298
$\sigma = 2$	0,2384	0,1175	0,0832

4.2 Méthodes de rééchantillonnage aux fins d'estimation de la variance

En pratique, il est courant dans les enquêtes à grande échelle d'utiliser des méthodes de rééchantillonnage aux fins d'estimation de la variance. Les dernières éditions de la NHANES et la National Survey of College Graduates (NSCG) en sont des exemples. Pour étudier les performances des estimateurs de la variance par rééchantillonnage selon la méthodologie contrainte proposée, nous réalisons des études de simulation fondées sur l'estimateur de la variance Jackknife avec suppression de groupe (DAGJK) proposé par Kott (2001).

Nous effectuons des expériences de simulation par rééchantillonnage sur la configuration décrite dans la section 4.1. Pour calculer l'estimateur de la variance Jackknife avec suppression de groupe, nous créons d'abord aléatoirement des groupes de taille égale G dans chacune des quatre strates. Puis, pour chaque rééchantillonnage $g = 1, \ldots, G$, nous supprimons le groupe g^{e} dans chaque strate, nous ajustons les poids restants par $w_k^{(g)} = \left(\frac{G}{G-1}\right) w_k$, où $w_k = \pi_k^{-1}$; et nous calculons l'estimation contrainte de rééchantillonnage $\tilde{\boldsymbol{\theta}}_s^{(g)}$ au moyen des poids ajustés. On obtient l'estimation de la variance Jackknife avec suppression de groupe de $\tilde{\theta}_{s_d}$, $\hat{V}_{\text{JK}}\left(\tilde{\theta}_{s_d}\right)$, en calculant

$$\hat{V}_{\text{JK}}\left(\tilde{\theta}_{s_d}\right) = \frac{G-1}{G} \sum_{g=1}^{G} \left(\tilde{\theta}_{s_d}^{(g)} - \tilde{\theta}_{s_d}\right)^2.$$

On obtient un estimateur de la variance par rééchantillonnage de $\tilde{y}_{s_d}$ en remplaçant $\tilde{\boldsymbol{\theta}}_s$ par $\tilde{\mathbf{y}}_s$.

Nos simulations tiennent compte seulement du scénario doublement monotone, avec $\sigma = 1$ ou 2, et $G = 10$, 20 ou 30. La taille de l'échantillon est fixée à $n = 480$ ou $n = 960$, et le dernier cas est obtenu par doublement de la taille de l'échantillon original dans chaque strate. Les figures 4.6 à 4.9 contiennent les résultats de simulation fondés sur 10 000 rééchantillonnages. Contrairement au comportement des estimations de la variance fondées sur la linéarisation, on constate que les estimations Jackknife avec suppression de groupe ont tendance à surestimer la variance de l'estimateur non contraint, comme on l'observe souvent en pratique. Qu'elles soient fondées sur un rééchantillonnage ou sur la linéarisation, les estimations de la variance de l'estimateur contraint surestiment la variance réelle, de sorte que les résultats sont plus cohérents dans les différentes méthodes d'estimation de la variance. Quand le nombre de groupes G augmente, les estimations Jackknife avec suppression de groupe tendent à augmenter, surtout

pour les petites valeurs de σ. Ces incréments des estimations Jackknife avec suppression de groupe ont comme conséquence directe d'accroître le taux de couverture quand G augmente. De plus, le taux de couverture des deux estimateurs s'améliore (plus près de 0,95) quand la taille de l'échantillon augmente. Dans l'ensemble, il semble que l'estimation de la variance par répliques soit une solution de rechange pratique à la linéarisation.

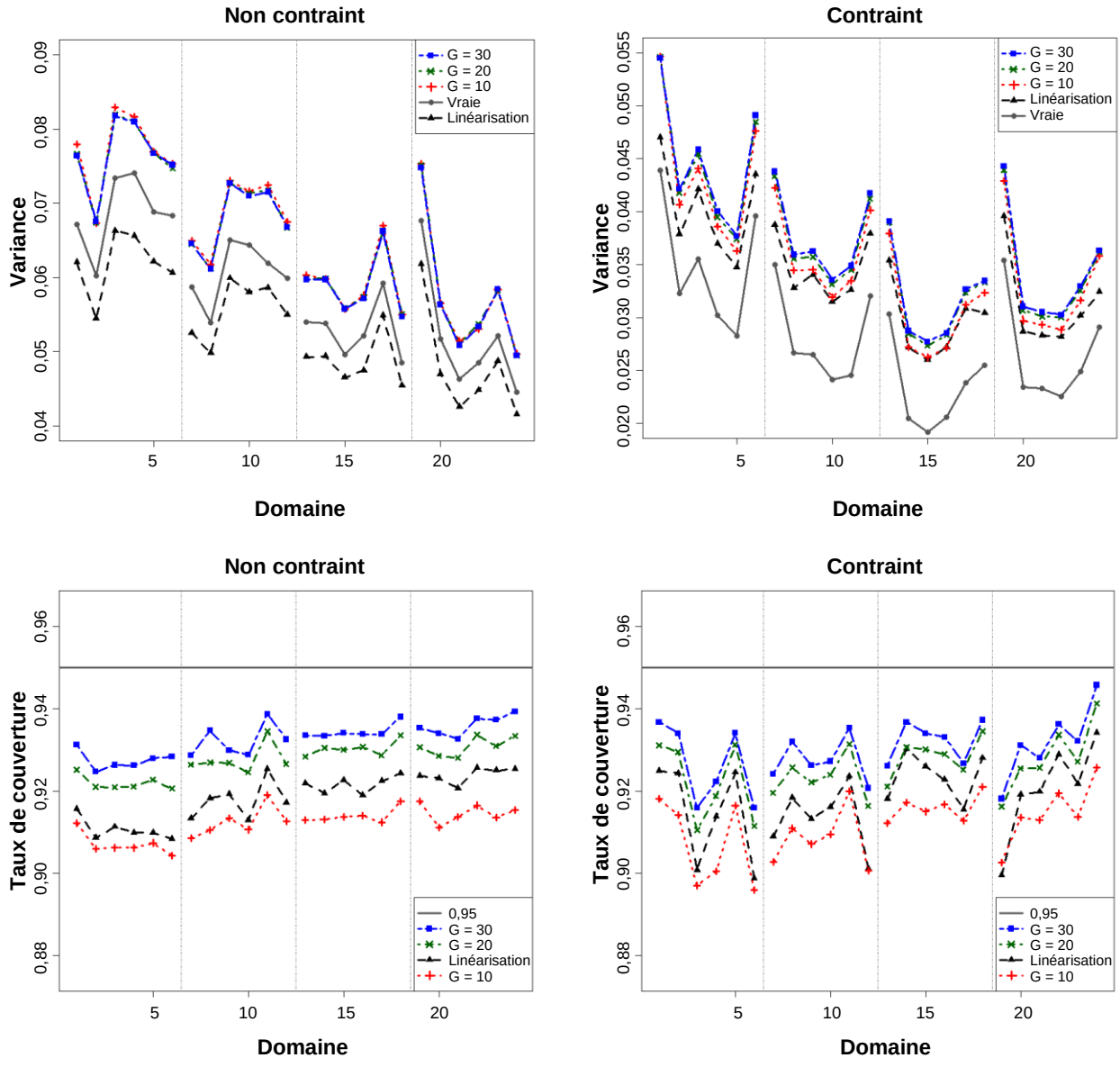


Figure 4.6 Estimation de la variance (en haut) et résultats de simulation du taux de couverture (en bas) fondés sur les méthodes de linéarisation et Jackknife avec suppression de groupe pour les estimateurs non contraint (à gauche) et contraint (à droite), dans le scénario doublement monotone avec $n_N = 480$ et $\sigma = 1$.

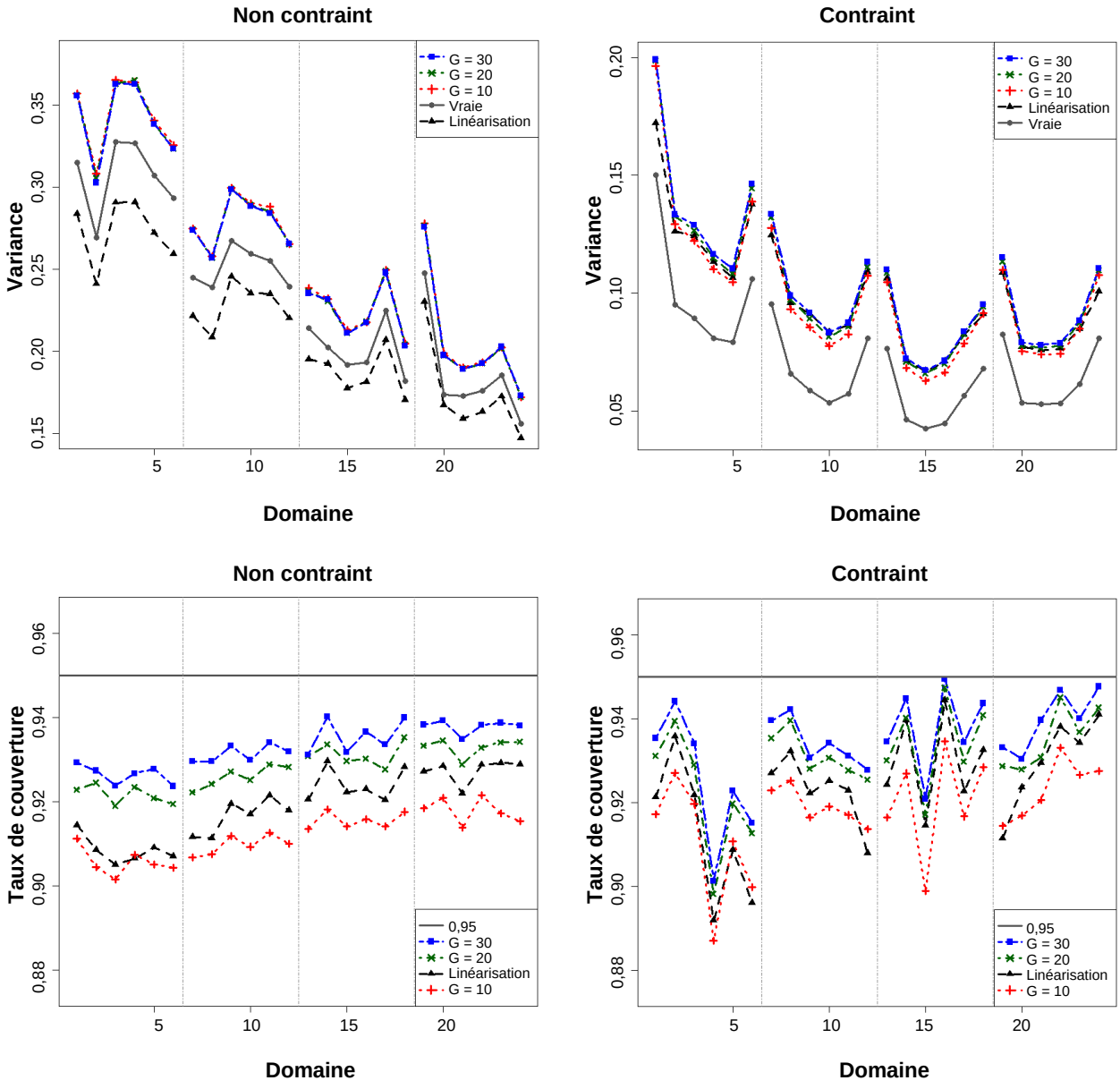


Figure 4.7 Estimation de la variance (en haut) et résultats de simulation du taux de couverture (en bas) fondés sur les méthodes de linéarisation et Jackknife avec suppression de groupe pour les estimateurs non contraint (à gauche) et contraint (à droite), dans le scénario doublement monotone avec $n_N = 480$ et $\sigma = 2$.

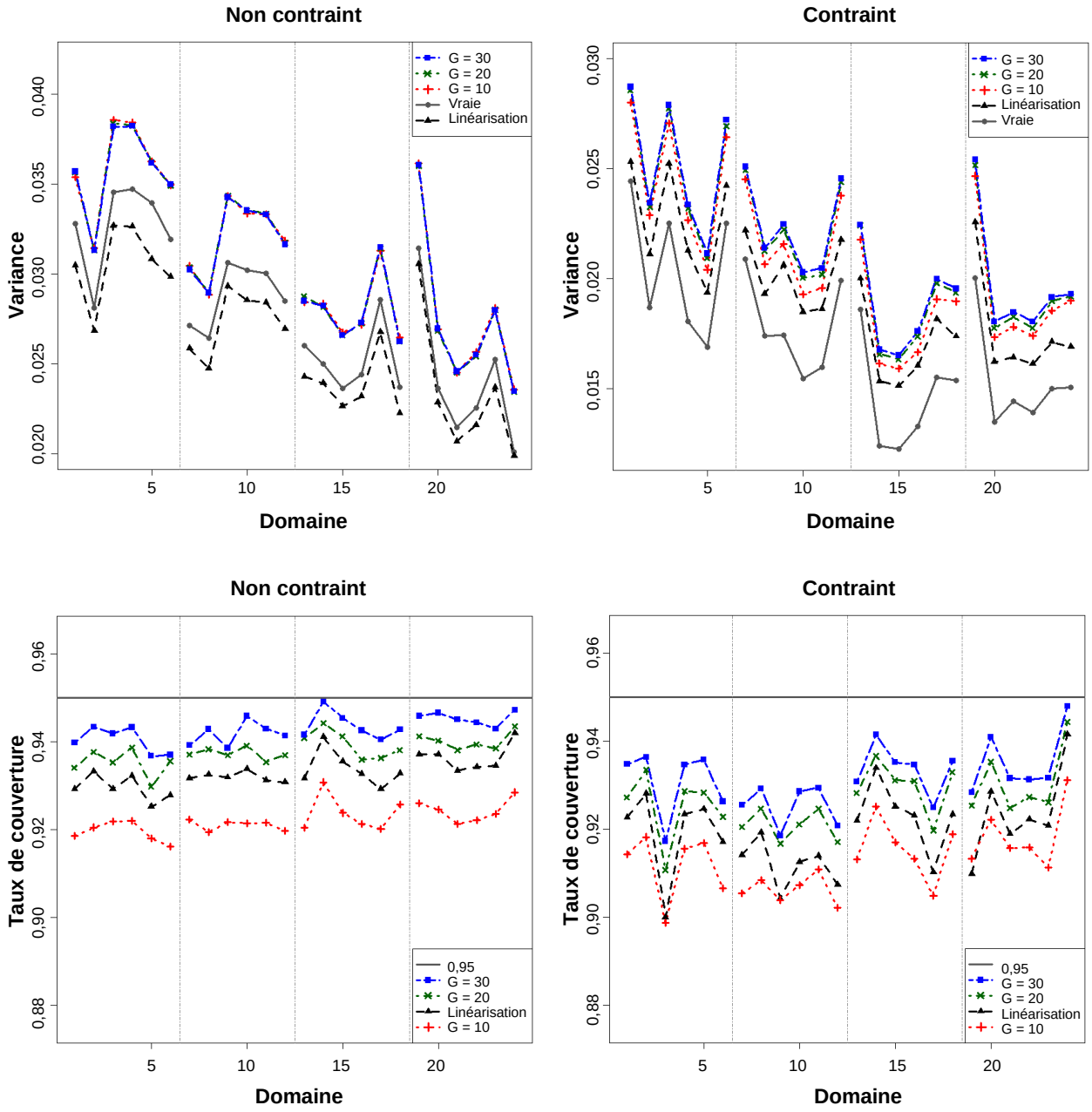


Figure 4.8 Estimation de la variance (en haut) et résultats de simulation du taux de couverture (en bas) fondés sur les méthodes de linéarisation et Jackknife avec suppression de groupe pour les estimateurs non contraint (à gauche) et contraint (à droite), dans le scénario doublement monotone avec $n_N = 960$ et $\sigma = 1$.

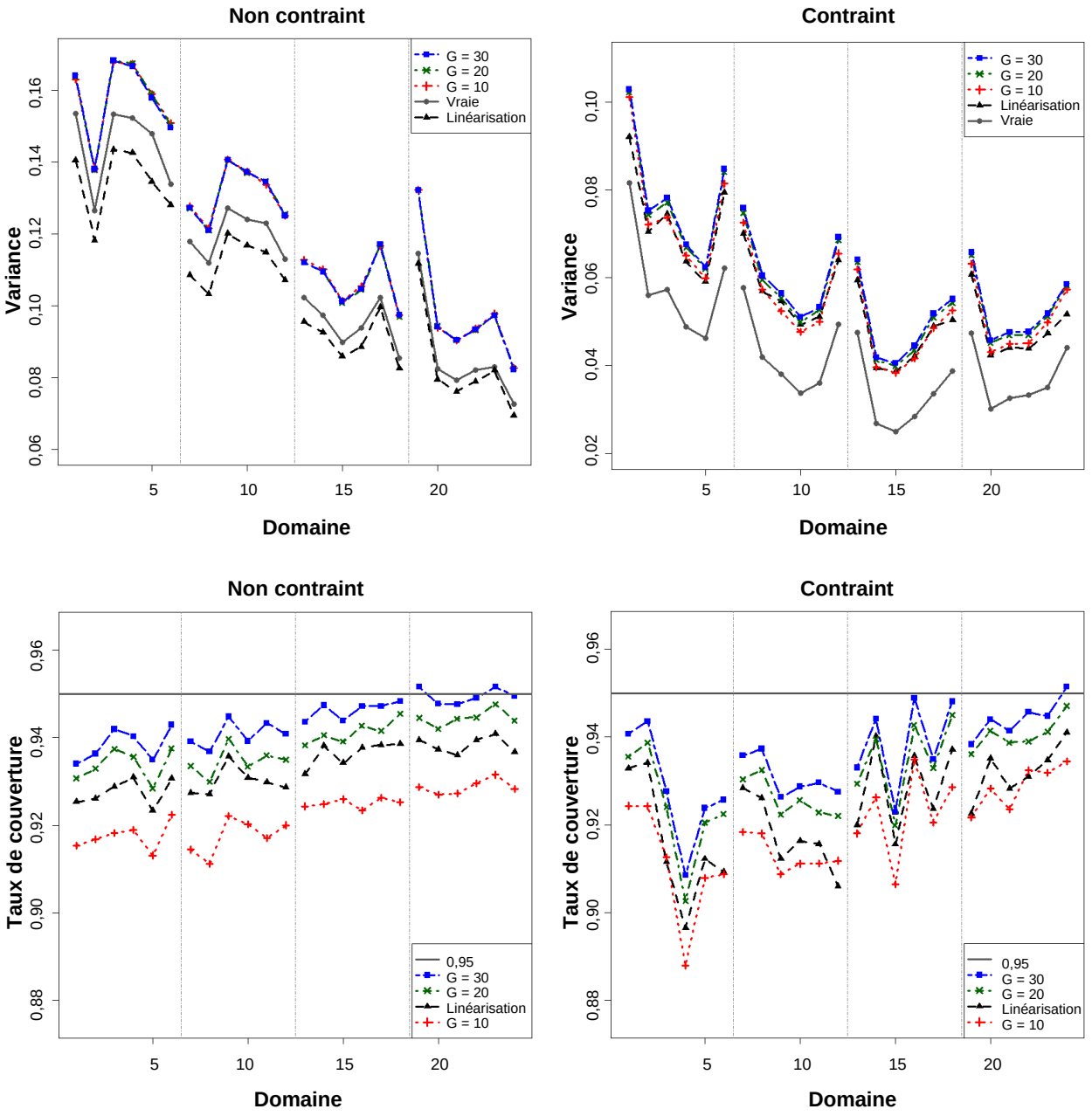


Figure 4.9 Estimation de la variance (en haut) et résultats de simulation du taux de couverture (en bas) fondés sur les méthodes de linéarisation et Jackknife avec suppression de groupe pour les estimateurs non contraint (à gauche) et contraint (à droite), dans le scénario doublement monotone avec $n_N = 960$ et $\sigma = 2$.

5 Application de l'estimateur contraint à l'enquête NSCG

Afin de montrer l'utilité de la méthodologie contrainte proposée sur des données d'enquête réelles, nous examinerons la *National Survey of College Graduates* (NSCG) de 2015, qui est parrainée par le *National Center for Science and Engineering Statistics* (NCSES) au sein de la *National Science Foundation* (NSF) et est menée par le *U.S. Census Bureau*. Les données et la documentation de la NSCG de 2015 sont disponibles sur le site Web de la NSF (www.nsf.gov/statistics/srvygrads). La NSCG cherche à fournir des données sur les caractéristiques des diplômés des collèges américains, en mettant particulièrement l'accent sur les diplômés en sciences et en génie.

Nous considérons le revenu gagné total avant les retenues de l'année précédente (2014) comme la variable d'intérêt (notée EARN). Pour éviter l'asymétrie élevée de cette variable, on effectue une transformation logarithmique. De plus, nous tenons compte uniquement des personnes qui ont déclaré un revenu positif. Au total, notre analyse examine 76 389 observations. En outre, elle étudie 252 domaines. Ils sont déterminés par la classification croisée de quatre prédicteurs. Ces variables et leurs contraintes supposées sont :

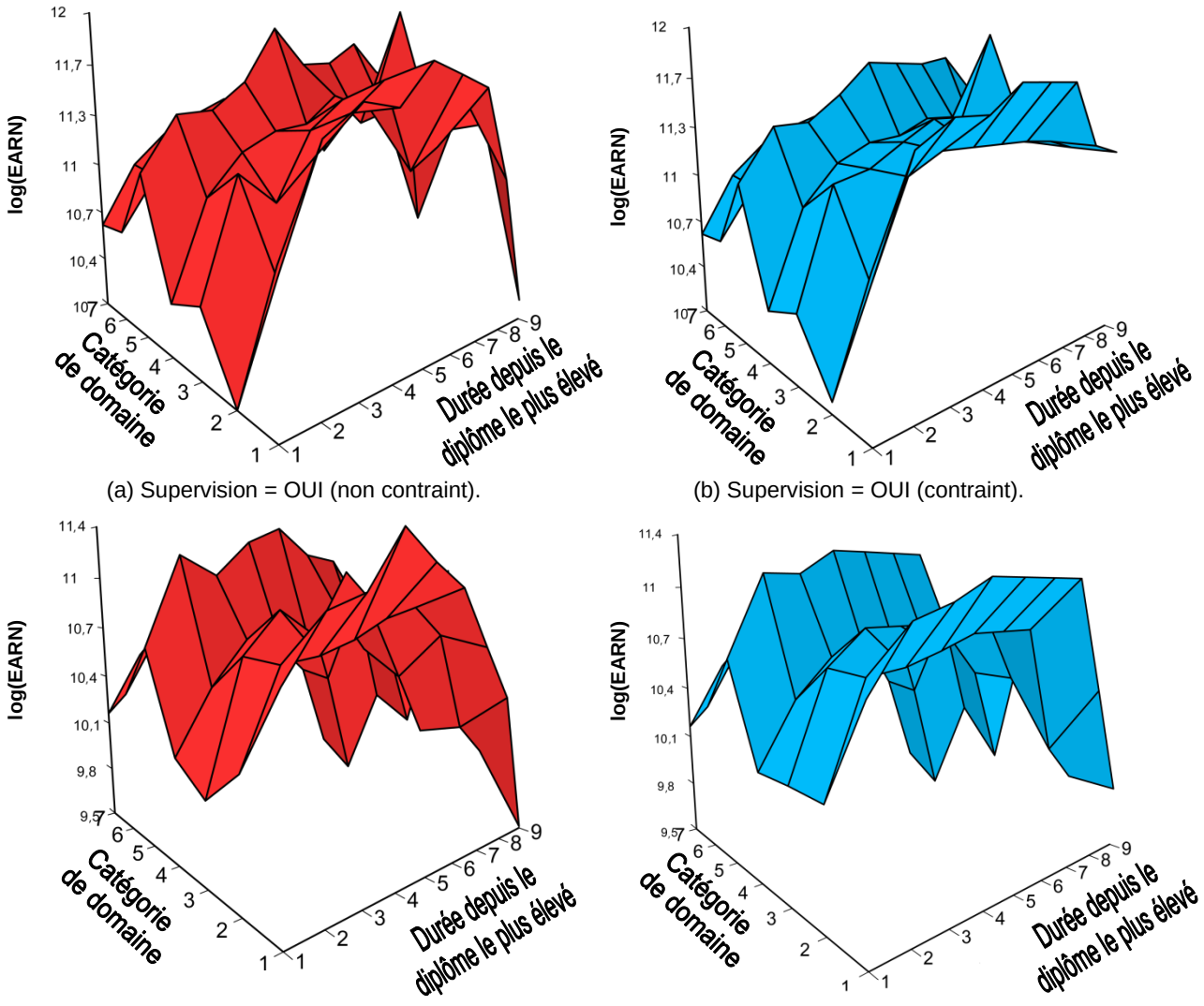
- *Durée depuis le diplôme le plus élevé*. Cette variable définit l'année d'attribution du diplôme le plus élevé. La période de 2015 à 1959 est divisée en 9 catégories : les 8 premières catégories (notées de 1 à 8) sont d'une durée de 6 ans chacune, et la dernière catégorie (notée 9) est de 9 ans. *Contrainte :* étant donné les autres prédicteurs, le revenu total gagné moyen augmente selon la durée écoulée depuis le plus haut diplôme dans les catégories d'années de 1 à 7. On ne formule aucune hypothèse concernant les catégories 8 et 9, car ces personnes sont susceptibles d'être à la retraite (au moins 42 années se sont écoulées depuis leur diplôme le plus élevé).
- *Domaine d'études*. Cette variable nominale définit le domaine d'études pour le diplôme le plus élevé, en fonction d'une catégorisation de grand groupe fournie dans la NSCG de 2015. Les 7 catégories pour cette variable sont :

  1 : Informatique et sciences de l'information;
  2 : Sciences biologiques, sciences agronomiques et sciences de l'environnement et de la vie;
  3 : Sciences physiques et connexes;
  4 : Sciences sociales et connexes;
  5 : Génie;
  6 : Domaines liés aux sciences et au génie;
  7 : Domaines non liés aux sciences et au génie.

  Contrainte : étant donné les autres prédicteurs, le revenu gagné total moyen pour chacun des champs 2 et 4 est inférieur à celui des champs 1, 3 et 5. Aucune hypothèse n'est formulée concernant les catégories 6 et 7, car elles couvrent de nombreux domaines pour lesquels il pourrait être compliqué d'imposer une restriction d'ordre raisonnable.
- *Cycle supérieur*. Cette variable binaire détermine si le diplôme le plus élevé est au niveau des études supérieures (OUI) ou au niveau du baccalauréat (NON). *Contrainte :* étant donné les autres prédicteurs, le revenu gagné total moyen est plus élevé chez les titulaires d'un diplôme de cycle supérieur.

- *Supervision.* Cette variable binaire détermine si la supervision fait partie des responsabilités de l'emploi principal (OUI) ou pas (NON). *Contrainte :* étant donné les autres prédicteurs, le revenu total moyen est plus élevé chez les personnes ayant des fonctions de supervision dans leur emploi principal.

Les figures 5.1 et 5.2 montrent les estimations non contraintes et limitées pour chacun des quatre groupes obtenus par la classification croisée des variables binaires Cycle supérieur et Supervision. Notons que puisque les contraintes supposées constituent un ordonnancement partiel, les estimations contraintes sont obtenues par le regroupement de domaines. Ces figures montrent que l'estimateur contraint a un comportement plus lisse que l'estimateur non contraint. De plus, il tend à corriger certains des « pics » produits par l'estimateur non contraint, qui sont habituellement la conséquence d'une très petite taille d'échantillon.



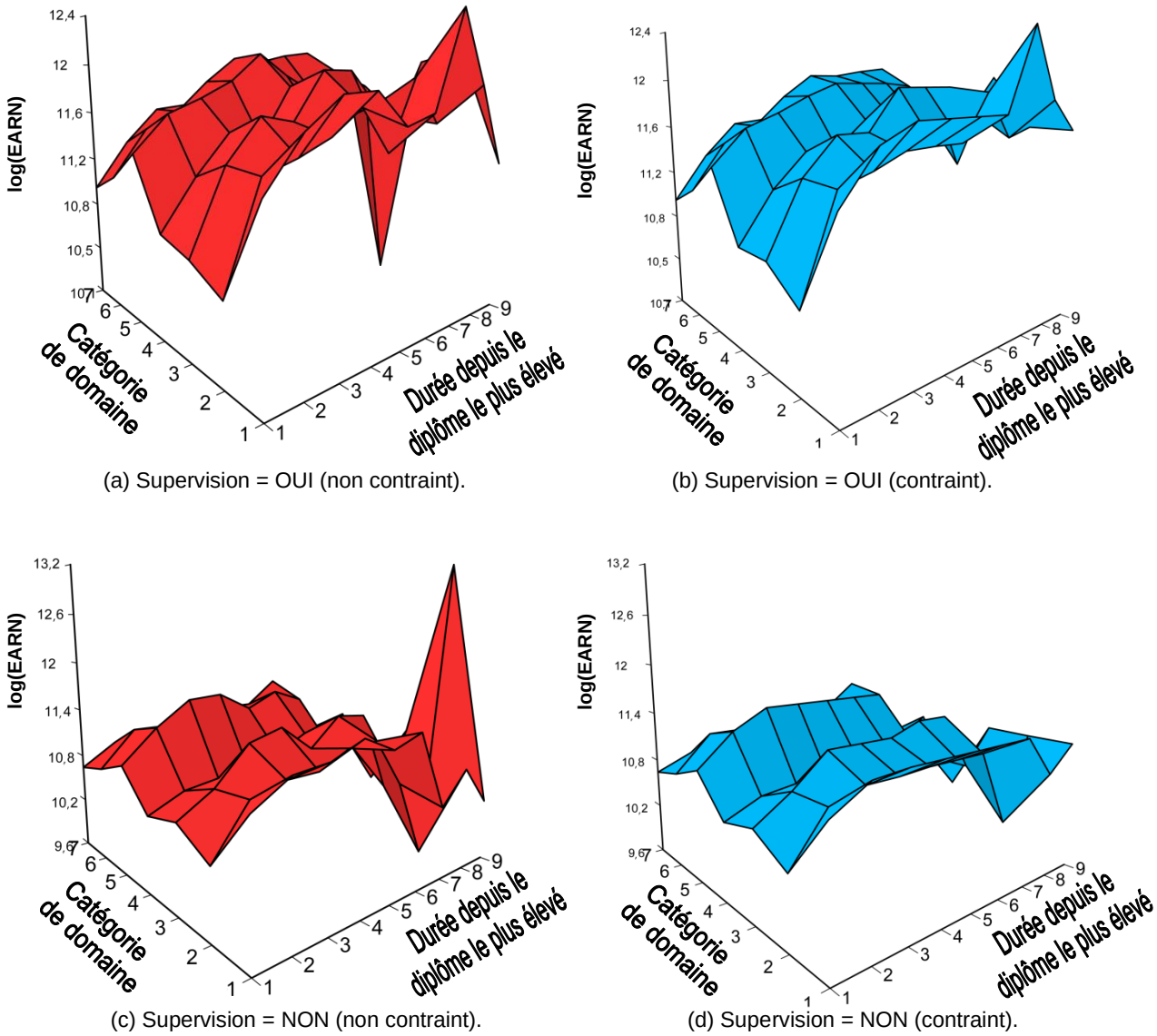
(a) Supervision = OUI (non contraint).

(b) Supervision = OUI (contraint).

(c) Supervision = NON (non contraint).

(d) Supervision = NON (contraint).

Figure 5.1 Estimations de la moyenne de domaine non contraintes (à gauche) et contraintes (à droite) pour les données de la NSCG de 2015, étant donné que Cycle supérieur = NON est fixe.



(a) Supervision = OUI (non contraint).

(b) Supervision = OUI (contraint).

(c) Supervision = NON (non contraint).

(d) Supervision = NON (contraint).

Figure 5.2 Estimations de la moyenne de domaine non contraintes (à gauche) et contraintes (à droite) pour les données de la NSCG de 2015, étant donné que Cycle supérieur = OUI est fixe.

Les erreurs-types pour les estimations non contraintes et contraintes sont calculées au moyen des poids de rééchantillonnage de la NSCG de 2015, qui sont fondés sur la méthode des répliques des différences successives (Opsomer, Breidt, White et Li, 2016). Les poids de rééchantillonnage et les facteurs d'ajustement ont été fournis par le directeur de programme du Programme de la statistique des ressources humaines du NCSES et sont disponibles sur demande.

La figure 5.3 présente le ratio de ces estimations pour chacun des 252 domaines. Dans la grande majorité des cas, les estimations de l'erreur-type de l'estimateur proposé sont inférieures à celles de l'estimateur non contraint, avec des améliorations la réduisant jusqu'à sept fois. Dans certains cas, le

comportement contraire se produit. Ils sont étudiés dans la figure 5.4, qui montre les graphiques de deux « tranches » différentes de domaines : l'une concernant la variable de la période depuis le diplôme le plus élevé et l'autre concernant la catégorie de domaine. Ces graphiques comprennent les estimations non contraintes et contraintes, les intervalles de confiance de Wald et les tailles d'échantillon. Chacune de ces deux tranches contient l'un des deux domaines qui peuvent être facilement reconnus dans la figure 5.3, car ils ont les ratios les plus petits.

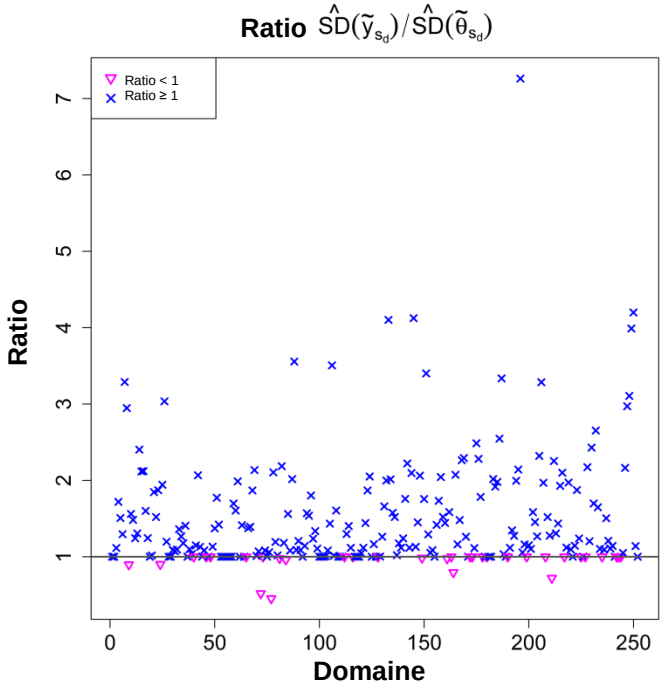
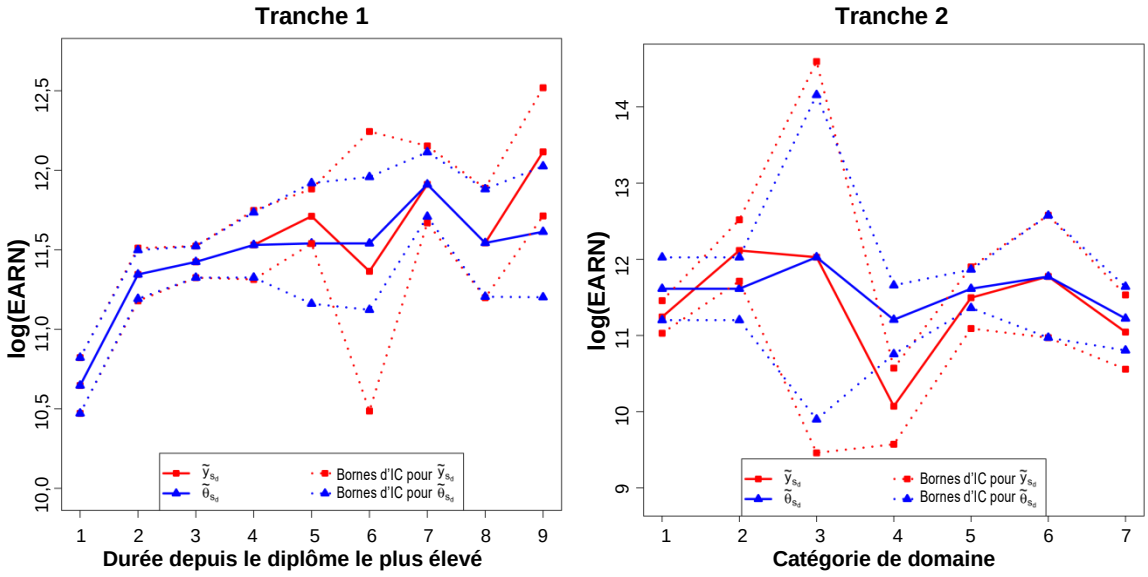


Figure 5.3 Ratio des erreurs-types estimées des estimations non contraintes par rapport à celles produites par des estimations contraintes pour les données de la NSCG de 2015.

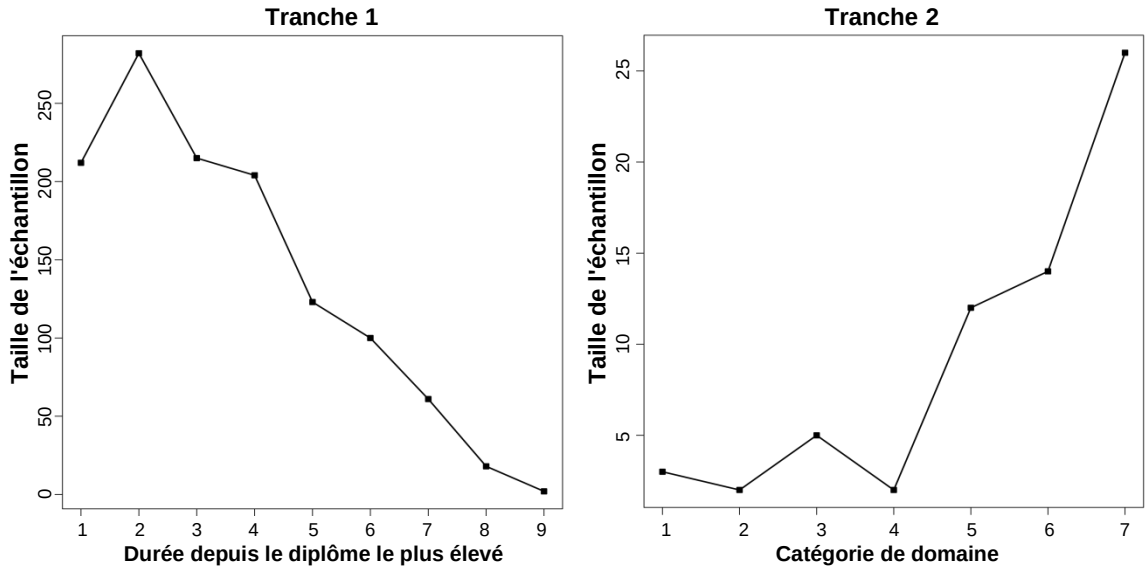
Le premier des domaines est illustré dans les figures 5.4(a) et 5.4(c), indiqué par 5. Les estimations non contraintes pour les domaines indiqués par 5 et 6 ne respectent pas l'hypothèse de monotonicité et sont donc regroupées pour l'obtention des estimations contraintes (un autre regroupement avec des domaines dans d'autres « tranches » est également effectué, mais il n'est pas visible dans ce graphique). Comme l'illustre la figure 5.4(a), l'intervalle de confiance est plus étroit pour les estimations non contraintes. Cependant, l'erreur type estimée de l'estimateur non contraint du domaine 6 est très grande, et le regroupement avec le domaine 5 stabilise grandement à la fois l'estimateur et les erreurs-types estimées pour ce domaine. La figure 5.4(c) montre que la taille des échantillons dans ces domaines est raisonnablement grande, comptant approximativement 100 observations chacun, ce qui implique que le non-respect de la monotonicité observé pourrait en fait être vrai dans la population. La décision finale sur l'équilibre entre l'amélioration de la stabilité de certains domaines et la possibilité d'un biais dû à des contraintes incorrectes doit être soigneusement évaluée.

Le deuxième domaine où les estimations non contraintes produisent des estimations d'écart-type plus petites est présenté dans les figures 5.4(b) et 5.4(d), indiqué par 1. Ici, on regroupe ce domaine avec le domaine voisin pour obtenir l'estimation contrainte. Toutefois, étant donné que les tailles d'échantillon de ces deux domaines sont très petites, les estimations non contraintes pourraient être considérées comme peu fiables, si bien que leurs erreurs-types estimées ne seraient pas une bonne indication de leur précision. L'estimateur contraint semble préférable ici en raison de l'augmentation de la taille effective des cellules.



(a) Catégorie de domaine = 2.

(b) Durée depuis le diplôme le plus élevé = 9.



(c) Catégorie de domaine = 2

(d) Durée depuis le diplôme le plus élevé = 9.

Figure 5.4 Estimations non contraintes et contraintes avec intervalles de confiance de Wald (en haut) et tailles d'échantillon (en bas) pour les données de la NSCG de 2015, étant donné que Cycle supérieur = OUI et Supervision = OUI.

6 Conclusions

Nous avons proposé une méthodologie générale d'estimation des moyennes de domaines qui permet d'intégrer les restrictions naturelles entre domaines dans l'estimation fondée sur le plan de sondage. Il a été démontré qu'elle améliore l'estimation et l'inférence, surtout pour les petits domaines. Comme cette nouvelle méthodologie couvre un large éventail d'hypothèses de forme dépassant la monotonicité univariée, elle vise à tirer parti conjointement de plusieurs types d'information qualitative qui apparaissent naturellement pour les données d'enquête. Les formes supplémentaires susceptibles d'être imposées comprennent la convexité ou la log-concavité; cette dernière peut être imposée si l'on croit que les moyennes de domaine de la population augmentent puis diminuent sur un ensemble de domaines. Les travaux futurs des auteurs comprendront un estimateur « monotone relâché » qui sera utilisé quand les moyennes de domaine de population sont « à peu près » monotones dans une séquence de domaines. Pour l'estimateur monotone relâché, on utilise un type de moyenne mobile sur les domaines pour mettre en œuvre les contraintes, ce qui permet à l'estimateur d'avoir quelques écarts par rapport à la monotonicité.

Nous avons également proposé une méthode d'estimation de la variance fondée sur le plan de sondage de l'estimateur, qui nécessite seulement de connaître l'ensemble de contraintes propres à l'échantillon. Il est montré que les méthodes de rééchantillonnage se comportent de la même façon. Pour ce qui est du calcul, l'estimateur est basé sur l'algorithme de projection du cône, qui est efficacement mis en œuvre dans le module `coneproj` et disponible gratuitement. Dans le cas pratique important d'un ordonnancement partiel, l'estimateur contraint équivaut à un regroupement de domaines voisins, de sorte qu'une fois que l'ensemble de contraintes est identifié par l'algorithme de projection de cône, les calculs subséquents des estimateurs et des estimateurs de la variance peuvent être faits directement par une estimation fondée sur le plan de sondage sur les domaines pertinents.

Comme l'a illustré l'analyse de la NSCG à la section 5, il est aussi important de déterminer les fois où la contrainte imposée pourrait ne pas être valide pour une application d'enquête en particulier. Récemment, Oliva-Aviles, Meyer et Opsomer (2019) ont proposé le critère d'information du cône fondé sur un échantillon comme critère permettant de choisir entre des ajustements contraints et non contraints pour l'estimateur de Wu et coll. (2016). Une réflexion sur la généralisation de cette démarche à la configuration étudiée ici est en cours.

Annexe

La première partie de l'annexe contient les lemmes qui ont servi à obtenir les résultats théoriques présentés dans l'article. Les démonstrations des théorèmes se trouvent à la fin de l'annexe.

Lemme 1. *Si un vecteur non nul peut être écrit comme la combinaison linéaire positive de vecteurs non nuls linéairement dépendants, il peut alors être exprimé comme la combinaison linéaire positive d'un sous-ensemble linéairement indépendant de ces vecteurs.*

Démonstration. Soit $\mathbf{v}$ un vecteur non nul qui peut être écrit comme étant $\mathbf{v} = \sum_{i=1}^{k} a_i \ell_i$ où $\ell_1, \ell_2, \ldots, \ell_k$ sont des vecteurs non nuls et $a_i > 0$ pour $i = 1, 2, \ldots, k$. Si cet ensemble de vecteurs n'est pas linéairement indépendant, il existe alors des constantes $b_1, \ldots, b_k$, qui ne sont pas toutes nulles, de sorte que $\sum_{i=1}^{k} b_i \ell_i = \mathbf{0}$, et que pour tout $c \in \mathbf{R}$, $\mathbf{v} = \sum_{i=1}^{k} (a_i + cb_i)\ell_i$. Soit $c = -\min_{i: b_i \neq 0} a_i / b_i$; alors $a_i + cb_i \geq 0$ pour $i = 1, \ldots, k$ mais pour au moins une valeur i, $a_i + cb_i = 0$. Nous avons ensuite écrit $\mathbf{v}$ sous la forme d'une combinaison linéaire positive d'un sous-ensemble approprié des vecteurs. Si ce sous-ensemble reste linéairement dépendant, le processus peut être répété.

Lemme 2. *Si* $\mathbf{A}$ *est une matrice irréductible* $m \times D$ *et* $\mathbf{B}$ *est une matrice non singulière* $D \times D$, *alors* $\tilde{\mathbf{A}} = \mathbf{AB}$ *est aussi irréductible.*

Démonstration. Supposons $\tilde{\mathbf{A}}^\top \mathbf{c} = \mathbf{0}$ pour certains $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{c} \geq \mathbf{0}$. Alors $\mathbf{B}^\top \mathbf{A}^\top \mathbf{c} = \mathbf{0}$ implique que $\mathbf{A}^\top \mathbf{c} = \mathbf{0}$ par la non-singularité de $\mathbf{B}$. Parce que $\mathbf{A}$ est irréductible, nous devons avoir $\mathbf{c} = \mathbf{0}$, afin que l'origine ne soit pas une combinaison linéaire positive de lignes de $\tilde{\mathbf{A}}$. Supposons ensuite que l'une des lignes de $\tilde{\mathbf{A}}$ est une combinaison linéaire positive d'autres lignes de $\tilde{\mathbf{A}}$. Cela signifie que nous pouvons écrire $\tilde{\mathbf{A}}^\top \mathbf{b} = \mathbf{0}$, où $b_j = -1$ pour certains $j \in \{1, \ldots, m\}$ et $b_i \geq 0$, $i \neq j$. Mais $\tilde{\mathbf{A}}^\top \mathbf{b} = \mathbf{0}$ implique que $\mathbf{B}^\top \mathbf{A}^\top \mathbf{b} = \mathbf{0}$ implique que $\mathbf{A}^\top \mathbf{b} = \mathbf{0}$ par la non-singularité de $\mathbf{B}$. Parce que nous ne pouvons pas avoir $\mathbf{A}^\top \mathbf{b} = \mathbf{0}$ pour cette valeur de $\mathbf{b}$, nous ne pouvons pas avoir de ligne de $\tilde{\mathbf{A}}$ qui soit une combinaison linéaire positive d'autres lignes de $\tilde{\mathbf{A}}$. Par conséquent, $\tilde{\mathbf{A}}$ est irréductible.

Lemme 3. *Soit* $\mathbf{A}$ *une matrice* $m \times D$. *De plus, soit* $\mathbf{S}_1$ *et* $\mathbf{S}_2$ *des matrices diagonales* $D \times D$ *comportant des éléments non nuls sur la diagonale. Pour tout ensemble* $J \subseteq \{1, 2, \ldots, m\}$, *soit* $V_{i,J}$ *l'ensemble de vecteurs dans les lignes* J *de* $\mathbf{A}_i = \mathbf{AS}_i$, $i = 1, 2$. *Ensuite, pour tout* $J^* \subseteq J$,

$$\mathcal{L}(V_{1,J^*}) = \mathcal{L}(V_{1,J}) \Leftrightarrow \mathcal{L}(V_{2,J^*}) = \mathcal{L}(V_{2,J}).$$

Démonstration. Soit $\mathbf{A}_{i,J} = \mathbf{A}_J \mathbf{S}_i$, $i = 1, 2$; où $\mathbf{A}_J$ désigne la sous-matrice de $\mathbf{A}$ qui contient les lignes dans les positions J. Supposons tout d'abord que $\mathcal{L}(V_{1,J^*}) = \mathcal{L}(V_{1,J})$. Puisque $J^* \subseteq J$, on constate aisément que $\mathcal{L}(V_{2,J^*}) \subseteq \mathcal{L}(V_{2,J})$. Ensuite, considérons tout $\mathbf{v} \in \mathcal{L}(V_{2,J})$ de sorte que $\mathbf{v} = \mathbf{A}_{2,J}^\top \mathbf{a} = \mathbf{S}_2 \mathbf{A}_J^\top \mathbf{a}$ pour un certain vecteur $\mathbf{a}$. Nous obtenons alors $\mathbf{S}_1 \mathbf{S}_2^{-1} \mathbf{v} = \mathbf{S}_1 \mathbf{A}_J^\top \mathbf{a} \in \mathcal{L}(V_{1,J})$. Par hypothèse, il existe un vecteur $\mathbf{b}$ tel que $\mathbf{S}_1 \mathbf{S}_2^{-1} \mathbf{v} = \mathbf{S}_1 \mathbf{A}_{J^*}^\top \mathbf{b}$. Par conséquent, $\mathbf{v} = \mathbf{S}_2 \mathbf{A}_{J^*}^\top \mathbf{b} \in \mathcal{L}(V_{2,J^*})$. Alors, $\mathcal{L}(V_{2,J}) \subseteq \mathcal{L}(V_{2,J^*})$. De façon analogue, il s'ensuit que $\mathcal{L}(V_{2,J^*}) = \mathcal{L}(V_{2,J})$ implique $\mathcal{L}(V_{1,J^*}) = \mathcal{L}(V_{1,J})$.

Lemme 4. *Selon les hypothèses A1 à A5, les énoncés suivants s'appliquent :*

(i) *Les* $N^{-1}\hat{t}_d$ *sont bornés uniformément.*

(ii) *Les* $N^{-1}\hat{N}_d$ *possèdent une borne supérieure uniforme et une borne inférieure strictement positive uniformément.*

(iii) $var(N^{-1}\hat{t}_d) = O(n^{-1})$ *et* $var(N^{-1}\hat{N}_d) = O(n^{-1})$.

(iv) $\mathrm{E}\left[\left(N^{-1}\hat{t}_d - r_d\mu_d\right)^2\right] = O\left(n^{-1}\right)$ *et* $\mathrm{E}\left[\left(N^{-1}\hat{N}_d - r_d\right)^2\right] = O\left(n^{-1}\right)$.

Démonstration.

(i) Notons que

$$\frac{\left|\hat{t}_d\right|}{N} = \left|\frac{\sum_{k\in s_d} y_k/\pi_k}{N}\right| \le \frac{\sum_{k\in U}\left|y_k\right|}{\lambda N}$$

qui ne dépend pas de s, et qui est borné indépendamment de N selon l'hypothèse A2.

(ii) À partir des hypothèses A4 et A5, nous constatons que

$$\frac{\varepsilon n}{DN} \le \frac{n_d}{N} \le \frac{\hat{N}_d}{N} = N^{-1}\sum_{k\in s_d} 1/\pi_k \le \lambda^{-1}N^{-1}N_d \le \lambda^{-1},$$

où la borne inférieure et la borne supérieure ne dépendent pas de s, et sont bornées pour tous les N par les hypothèses A1 et A4.

(iii) Notons que

$$n\,\mathrm{var}\left(N^{-1}\hat{t}_d\right) = n\,\mathrm{var}\left(N^{-1}\sum_{k\in s_d} y_k/\pi_k\right) \le \frac{\sum_{k\in U_d} y_k^2}{\lambda^2 N}\left(\frac{n}{N} + n\max_{k,\,l\in U_d:\,k\ne l}\left|\Delta_{kl}\right|\right)$$

qui est borné selon les hypothèses A2, A4 et A5. En établissant $y_k \equiv 1$ et en suivant un argument analogue, on peut montrer que $n\,\mathrm{var}\left(N^{-1}\hat{N}_d\right) = O(1)$.

(iv) Puisque

$$\mathrm{E}\left[\left(N^{-1}\hat{t}_d - r_d\mu_d\right)^2\right] = \mathrm{var}\left(N^{-1}\hat{t}_d\right) + \left(\frac{N_d}{N}\bar{y}_{U_d} - r_d\mu_d\right)^2,$$

l'hypothèse A3 et (iii) nous conduisent à la conclusion souhaitée. De manière analogue, nous obtenons

$$\mathrm{E}\left[\left(N^{-1}\hat{N}_d - r_d\right)^2\right] = O\left(n^{-1}\right).$$

Démonstration du théorème 1. Supposons tout d'abord que $\Pi\left(\mathbf{z}\middle|\Omega^0\right) = \Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right) = \mathbf{0}$. Dans ce cas, tout sous-ensemble $J^* \subset J$ tel que V_J est linéairement indépendant satisfait $\Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_{J^*}\right)\right) = \mathbf{0} \in \bar{\mathcal{F}}_{J^*}$. Il suffit alors de choisir $J^* \subset J$ de sorte que V_{J^*} soit linéairement indépendant et couvre $\mathcal{L}\left(V_J\right)$. Supposons maintenant que $\Pi\left(\mathbf{z}\middle|\Omega^0\right) \ne \mathbf{0}$. Étant donné que $\Pi\left(\mathbf{z}\middle|\Omega^0\right) = \Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right) \in \bar{\mathcal{F}}_J$, $\Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right)$ peut être écrit comme la combinaison linéaire positive des vecteurs $\boldsymbol{\gamma}_j$, $j \in J$. De plus, $\left\langle \mathbf{z} - \Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right), \boldsymbol{\gamma}_j\right\rangle = 0$ pour $j \in J$. À partir du lemme 1, on a $J_0 \subset J$ de sorte que V_{J_0} soit linéairement indépendant et que $\Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right)$ puisse être écrit comme une combinaison linéaire positive des vecteurs dans V_{J_0}, ce qui implique que $\Pi\left(\mathbf{z}\middle|\mathcal{L}\left(V_J\right)\right) \in \bar{\mathcal{F}}_{J_0}$. En outre, puisque

$\left\langle \mathbf{z} - \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_J\right)\right), \boldsymbol{\gamma}_j \right\rangle = 0$ pour $j \in J_0, \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_0}\right)\right) = \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_J\right)\right)$. Alors, $\Pi\left(\mathbf{z} \mid \Omega^0\right) = \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_0}\right)\right)$. Si $\mathcal{L}\left(V_{J_0}\right) = \mathcal{L}\left(V_J\right)$ alors $J^* = J_0$ satisfait toutes les conditions requises. Supposons maintenant que $\mathcal{L}\left(V_{J_0}\right) \subset \mathcal{L}\left(V_J\right)$. Le fait que $\Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_0}\right)\right) = \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_J\right)\right)$ implique que $\Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_1}\right)\right) = \Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_0}\right)\right)$ pour tout ensemble J_1 de sorte que $J_0 \subseteq J_1 \subseteq J$. En outre, puisque $\Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_0}\right)\right) \in \bar{\mathcal{F}}_{J_0}$ alors $\Pi\left(\mathbf{z} \mid \mathcal{L}\left(V_{J_1}\right)\right) \in \bar{\mathcal{F}}_{J_1}$. Il suffit alors de choisir J^* de sorte que $J_0 \subset J^* \subset J$ et V_{J^*} soit linéairement indépendant et couvre $\mathcal{L}\left(V_J\right)$.

Démonstration du théorème 2. Pour prouver le théorème, nous commençons par un ensemble $J \notin \mathcal{G}_\mu$ et trouvons les conditions nécessaires pour que cet ensemble appartienne à $\mathcal{G}_s$. Ces conditions nécessaires, exprimées sous forme d'inégalités en termes de fonctions lisses et continues de $\hat{N}_d/N$ et de $\hat{t}_d/N$, servent ensuite à borner la probabilité d'intérêt. Enfin, nous utilisons le théorème 5.4.3 qui se trouve dans Fuller (1996) pour montrer que cette probabilité converge vers zéro à un taux de $O\left(n^{-1}\right)$.

Soit $\mathbf{A}_\mu$, $\mathbf{A}_{\mu,J}$ et $\boldsymbol{\gamma}_{\mu_d}$ les versions analogues de $\mathbf{A}_s$, $\mathbf{A}_{s,J}$ et $\boldsymbol{\gamma}_{s_d}$ obtenues en remplaçant $\tilde{\mathbf{y}}_s$ et $\mathbf{W}_s$ respectivement par $\boldsymbol{\mu}$ et $\mathbf{W}_\mu$. Le lemme 2 permet que $\mathbf{A}_s$ et $\mathbf{A}_\mu$ soient tous deux irréductibles puisque $\mathbf{A}$ l'est.

Supposons tout d'abord que $\varnothing \notin \mathcal{G}_\mu$ et que $J = \varnothing$. Ensuite, à partir des conditions de (2.8), $\varnothing \in \mathcal{G}_s$ si et seulement si $\left\langle \tilde{\mathbf{z}}_s, \boldsymbol{\gamma}_{s_j} \right\rangle \leq 0$ pour $j = 1, 2, \ldots, m$. Par ailleurs, supposons que $\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_j} \right\rangle \leq 0$ pour $j = 1, 2, \ldots, m$. Alors, $\varnothing \in \mathcal{G}_\mu$, ce qui contredit notre choix de J. Par conséquent, il existe une valeur j_0 telle que $\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle > 0$. Nous obtenons alors

$$
\begin{aligned}
P\left(\varnothing \in \mathcal{G}_s\right) \leq P\left(0 \geq \left\langle \tilde{\mathbf{z}}_s, \boldsymbol{\gamma}_{s_{j_0}} \right\rangle\right) &= P\left(\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle - \left\langle \tilde{\mathbf{z}}_s, \boldsymbol{\gamma}_{s_{j_0}} \right\rangle \geq \left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle\right) \\
&= P\left(\left[\frac{\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle - \left\langle \tilde{\mathbf{z}}_s, \boldsymbol{\gamma}_{s_{j_0}} \right\rangle}{\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle}\right]^2 \geq 1\right) \\
&\leq \frac{1}{\left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle^2} \mathrm{E}\left[\left(\left\langle \tilde{\mathbf{z}}_s, \boldsymbol{\gamma}_{s_{j_0}} \right\rangle - \left\langle \mathbf{z}_\mu, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle\right)^2\right]
\end{aligned}
$$

où la dernière inégalité est obtenue par l'application de l'inégalité de Markov (voir par exemple Casella et Berger (2002), section 3.6.1). Nous montrons ensuite que la valeur espérée pour le dernier terme est $O\left(n^{-1}\right)$. Notons que l'expression contenue dans la valeur espérée dans l'inégalité ci-dessus est une fonction du vecteur $\hat{\mathbf{x}}_s = \left(N^{-1}\hat{t}_1, \ldots, N^{-1}\hat{t}_D, N^{-1}\hat{N}_1, \ldots, N^{-1}\hat{N}_D\right)^\top$. Soit $f_1(\cdot)$ une fonction (qui ne dépend pas de N) et $\mathbf{x}_\mu = \left(r_1\mu_1, \ldots, r_D\mu_D, r_1, \ldots, r_D\right)$. Pour appliquer le théorème 5.4.3 de Fuller (1996) avec $\alpha = 1$, $s = 2$ et $a_N = O\left(n^{-1/2}\right)$, nous devons d'abord montrer que les conditions suivantes sont satisfaites :

(a) $\mathrm{E}\left[\left(\hat{\mathbf{x}}_s - \mathbf{x}_\mu\right)^2\right] = O\left(n^{-1}\right)$.

(b) f_1 est uniformément bornée dans une sphère fermée et bornée $\mathcal{S}$.

(c) $f_1^{(i_1, i_2)}\left(\mathbf{x}\right)$ est continue dans $\mathbf{x}$ sur $\mathcal{S}$, où

$$f_1^{(i_1, \ldots, i_r)}(\mathbf{x}_0) = \frac{\partial^r}{\partial_{x_{i_1}} \ldots \partial_{x_{i_r}}} f_1(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_0}.$$

(d) $\mathbf{x}_\mu$ est un point intérieur de $\mathcal{S}$.

(e) Il existe un nombre fini K tel que

$$\left| f_1^{(i_1, i_2)}(\mathbf{x}) \right| \leq K \quad \text{pour tous les} \quad \mathbf{x} \in \mathcal{S},$$

$$\left| f_1^{(i_1)}(\mathbf{x}_\mu) \right| \leq K \quad \text{et} \quad \left| f_1(\mathbf{x}_\mu) \right| \leq K.$$

La condition (a) est directement satisfaite par le lemme 4 (iv). De plus, le lemme 4 (i)-(ii) garantit qu'il existe une constante $M > 1$ telle que $\left| N^{-1}\hat{t}_d \right| \leq M$ et $M^{-1} \leq N^{-1}\hat{N}_d \leq M$. Il existe donc une sphère fermée et bornée $\mathcal{S}$ qui est contenue dans ces bornes constantes. De plus, à partir de l'hypothèse A3, nous pouvons conclure que $\mathbf{x}_\mu \in \mathcal{S}$, par conséquent la condition (d) est satisfaite. Pour montrer que la condition (b) est satisfaite, notons que f_1 est une fonction continue dans $\mathcal{S}$ puisque $W_s^{-1/2}$ et $\tilde{y}_{s_d}$ existent tous deux pour tout $\mathbf{x} \in \mathcal{S}$. Par conséquent, le théorème de la valeur extrême (voir le théorème 4.15 dans Rudin (1976)) garantit que f_1 est uniformément bornée dans $\mathcal{S}$. Les conditions (c) et (e) sont satisfaites puisque f_1 est une fonction rationnelle continue dans $\mathcal{S}$, ce qui implique que f_1 est différenciable à l'infini et que ses dérivées sont bornées dans $\mathcal{S}$. Enfin, toutes les conditions (a) à (e) sont satisfaites. Par conséquent, à partir du théorème 5.4.3 de Fuller (1996), nous pouvons conclure que $\mathrm{E}\left[f_1(\mathbf{x}) \right] = O\left(n^{-1}\right)$, puisque f_1 et sa première dérivée par rapport à $N^{-1}\hat{t}_d$ et $N^{-1}\hat{N}_d$ sont évaluées à zéro à $\mathbf{x}_\mu$.

Prenons maintenant $J \neq \varnothing$ tel que $J \notin \mathcal{G}_\mu$, et supposons que $J \in \mathcal{G}_s$. Le théorème 1 garantit que nous pouvons toujours choisir un sous-ensemble $J^* \subseteq J$ tel que $J^* \in \mathcal{G}_s$, V_{s,J^*} est linéairement indépendant, et $\mathcal{L}\left(V_{s,J^*}\right) = \mathcal{L}\left(V_{s,J}\right)$. Notons que $\Pi\left(\tilde{\mathbf{z}}_s \,|\, \mathcal{L}\left(V_{s,J^*}\right)\right) = \mathbf{A}_{s,J^*}^\top \left(\mathbf{A}_{s,J^*}\mathbf{A}_{s,J^*}^\top\right)^{-1} \mathbf{A}_{s,J^*}\tilde{\mathbf{z}}_s$. Soit $\tilde{\mathbf{b}}_{s,J^*} = \left(\mathbf{A}_{s,J^*}\mathbf{A}_{s,J^*}^\top\right)^{-1} \mathbf{A}_{s,J^*}\tilde{\mathbf{z}}_s$. Par conséquent, à partir des conditions de (2.8), nous obtenons que $J \in \mathcal{G}_s$ implique que $\tilde{\mathbf{b}}_{s,J^*} \geq \mathbf{0}$, et $\left\langle \tilde{\mathbf{z}}_s - \mathbf{A}_{s,J^*}^\top \tilde{\mathbf{b}}_{s,J^*}, \boldsymbol{\gamma}_{s_j} \right\rangle \leq 0$ pour tout j. Définissons $\mathbf{b}_{\mu,J^*} = \left(\mathbf{A}_{\mu,J^*}\mathbf{A}_{\mu,J^*}^\top\right)^{-1} \mathbf{A}_{\mu,J^*}\mathbf{z}_\mu$ et supposons que $\mathbf{b}_{\mu,J^*} \geq \mathbf{0}$, et $\left\langle \mathbf{z}_\mu - \mathbf{A}_{\mu,J^*}^\top \mathbf{b}_{\mu,J^*}, \boldsymbol{\gamma}_{\mu_j} \right\rangle \leq 0$ pour $j = 1, 2, \ldots, m$. Ces conditions impliqueraient que $J^* \in \mathcal{G}_\mu$, ce qui contredit l'hypothèse originale selon laquelle $J \notin \mathcal{G}_\mu$, puisque $\mathcal{L}\left(V_{\mu,J^*}\right) = \mathcal{L}\left(V_{\mu,J}\right)$ selon le lemme 3. Par conséquent, soit un élément de $\mathbf{b}_{\mu,J^*}$ qui est strictement négatif, soit il existe une valeur j_0 telle que $\left\langle \mathbf{z}_\mu - \mathbf{A}_{\mu,J^*}^\top \mathbf{b}_{\mu,J^*}, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle > 0$. Par conséquent, si $P\left(J \in \mathcal{G}_s\right) = O\left(n^{-1}\right)$ est démontré dans un de ces deux scénarios, la preuve sera concluante.

Supposons que le j_0^{e} élément de $\mathbf{b}_{\mu,J^*}$ est strictement négatif. C'est-à-dire que $\mathbf{e}_{j_0}^\top \mathbf{b}_{\mu,J^*} < 0$, où $\mathbf{e}_j$ désigne le vecteur indicateur qui est 1 pour l'entrée j et 0 sinon. Nous obtenons alors

$$\begin{aligned} P\left(J \in \mathcal{G}_s\right) \leq P\left(\mathbf{e}_{j_0}^\top \tilde{\mathbf{b}}_{s,J^*} \geq 0\right) &= P\left(\mathbf{e}_{j_0}^\top \tilde{\mathbf{b}}_{s,J^*} - \mathbf{e}_{j_0}^\top \mathbf{b}_{\mu,J^*} \geq -\mathbf{e}_{j_0}^\top \mathbf{b}_{\mu,J^*}\right) \\ &\leq \frac{1}{\left(\mathbf{e}_{j_0}^\top \mathbf{b}_{\mu,J^*}\right)^2} \mathrm{E}\left[\left(\mathbf{e}_{j_0}^\top \tilde{\mathbf{b}}_{s,J^*} - \mathbf{e}_{j_0}^\top \mathbf{b}_{\mu,J^*}\right)^2\right]. \end{aligned}$$

Soit $f_2(\hat{\mathbf{x}}_s)$ l'expression à l'intérieur de la valeur espérée ci-dessus. On peut appliquer un argument analogue à celui utilisé pour la fonction f_1 à la fonction rationnelle continue f_2 sur $\mathcal{S}$, pour conclure que $\mathbb{E}\left[f_2(\hat{\mathbf{x}}_s)\right] = O(n^{-1})$. Notons que nous avons également utilisé le fait que $\mathbf{A}_{s,J^*}\mathbf{A}_{s,J^*}^\top$ est une matrice inversible pour tout $\mathbf{x} \in \mathcal{S}$.

Enfin, supposons qu'il existe j_0 de sorte que $\kappa_{\mathbf{z}_\mu, j_0} = \left\langle \mathbf{z}_\mu - \mathbf{A}_{\mu,J^*}^\top \mathbf{b}_{\mu,J^*}, \boldsymbol{\gamma}_{\mu_{j_0}} \right\rangle > 0$, et soit $\kappa_{\tilde{\mathbf{z}}_\mu, j_0} = \left\langle \tilde{\mathbf{z}}_s - \mathbf{A}_{s,J^*}^\top \tilde{\mathbf{b}}_{s,J^*}, \boldsymbol{\gamma}_{s_{j_0}} \right\rangle$. Nous obtenons alors

$$\begin{aligned} P\left(J \in \tilde{\mathcal{G}}_s\right) \leq P\left(0 \geq \kappa_{\tilde{\mathbf{z}}_s, j_0}\right) &= P\left(\kappa_{\mathbf{z}_\mu, j_0} - \kappa_{\tilde{\mathbf{z}}_s, j_0} \geq \kappa_{\mathbf{z}_\mu, j_0}\right) \\ &\leq \frac{1}{\kappa_{\mathbf{z}_\mu, j_0}^2} \mathbb{E}\left[\left(\kappa_{\tilde{\mathbf{z}}_s, j_0} - \kappa_{\mathbf{z}_\mu, j_0}\right)^2\right]. \end{aligned}$$

Soit $f_3(\hat{\mathbf{x}}_s)$ l'expression à l'intérieur de la valeur espérée ci-dessus. On applique un argument analogue à celui utilisé pour les fonctions f_1, f_2 pour conclure que $\mathbb{E}\left[f_3(\hat{\mathbf{x}}_s)\right] = O(n^{-1})$.

Démonstration du théorème 3. Prenons tout $J \in \mathcal{G}_s$ et tout domaine d. Notons que la condition $\mathbf{A}\boldsymbol{\mu} \geq \mathbf{0}$ implique que $\varnothing \in \mathcal{G}_\mu$. Nous pouvons alors écrire $\tilde{\theta}_{s_d} - \bar{y}_{U_d}$ comme suit

$$\tilde{\theta}_{s_d} - \bar{y}_{U_d} = \left(\tilde{y}_{s_d} - \bar{y}_{U_d}\right)1_{J=\varnothing} + \sum_{J_G \in \mathcal{G}_\mu \setminus \varnothing} \left(\tilde{\theta}_{s_d, J_G} - \bar{y}_{U_d}\right)1_{J_G = J} + \sum_{J_G \in \mathcal{G}_\mu^c} \left(\tilde{\theta}_{s_d, J_G} - \bar{y}_{U_d}\right)1_{J_G = J},$$

où nous avons utilisé $\tilde{\theta}_{s_d,\varnothing} = \tilde{y}_{s_d}$. Nous pouvons maintenant écrire un estimateur de variance irréalisable $\mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right)$ comme suit

$$\mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right) = \mathrm{VA}\left(\tilde{y}_{s_d}\right)1_{J=\varnothing} + \sum_{J_G \in \mathcal{G}_\mu \setminus \varnothing} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)1_{J=J_G} + \sum_{J_G \in \mathcal{G}_\mu^c} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)1_{J=J_G}.$$

Par conséquent,

$$\begin{aligned} \mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right)^{-1/2}\left(\tilde{\theta}_{s_d} - \bar{y}_{U_d}\right) &= \mathrm{VA}\left(\tilde{y}_{s_d}\right)^{-1/2}\left(\tilde{y}_{s_d} - \bar{y}_{U_d}\right)1_{J=\varnothing} \\ &\quad + \sum_{J_G \in \mathcal{G}_\mu \setminus \varnothing} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\tilde{\theta}_{s_d, J_G} - \bar{y}_{U_d}\right)1_{J=J_G} \\ &\quad + \sum_{J_G \in \mathcal{G}_\mu^c} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\tilde{\theta}_{s_d, J_G} - \bar{y}_{U_d}\right)1_{J=J_G} \\ &= \Bigg[\mathrm{VA}\left(\tilde{y}_{s_d}\right)^{-1/2}\left(\tilde{y}_s - \bar{y}_{U_d}\right)1_{J=\varnothing} \\ &\qquad + \sum_{J_G \in \mathcal{G}_\mu \setminus \varnothing} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\tilde{\theta}_{s_d, J_G} - \theta_{U_d, J_G}\right)1_{J=J_G} \\ &\qquad + \sum_{J_G \in \mathcal{G}_\mu^c} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\tilde{\theta}_{s_d, J_G} - \theta_{U_d, J_G}\right)1_{J=J_G}\Bigg] \\ &\quad + \left[\sum_{J_G \in \mathcal{G}_\mu \setminus \varnothing} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right)1_{J=J_G}\right] \\ &\quad + \left[\sum_{J_G \in \mathcal{G}_\mu^c} \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right)1_{J=J_G}\right] \\ &= c_{1N} + c_{2N} + c_{3N}, \end{aligned}$$

où θ_{U_d, J_G} est la version de la population de $\tilde{\theta}_{s_d, J_G}$. Un développement en séries de Taylor du premier ordre de $\tilde{\theta}_{s_d, J_G}$ et l'hypothèse A6 permettent de conclure que chaque terme de forme

$$\mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\tilde{\theta}_{s_d, J_G} - \theta_{U_d, J_G}\right)$$

converge dans la distribution vers une distribution normale standard. Par conséquent, c_{1N} converge aussi vers une distribution normale standard. Notons que pour chaque $J_G \in \mathcal{G}_\mu^c$,

$$\mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right) = \left[n\mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)\right]^{-1/2}\left[n^{1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right)\right] = O(n^{1/2}),$$

tandis que $1_{J = J_G} = O_p\left(n^{-1}\right)$ selon le théorème 2 (puisque $J \in \mathcal{G}_s$). Alors, $c_{3N} = O_p\left(n^{-1/2}\right)$. Notons maintenant que $\theta_{U_d, J_G} - \bar{y}_{U_d} = O\left(N^{-1/2}\right)$ quand $J_G \in \mathcal{G}_\mu \backslash \varnothing$ selon l'hypothèse A3. Par conséquent, pour tout $J_G \in \mathcal{G}_\mu \backslash \varnothing$,

$$\mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)^{-1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right) = \left[n\mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right)\right]^{-1/2}\left[n^{1/2}\left(\theta_{U_d, J_G} - \bar{y}_{U_d}\right)\right] = O\left(\sqrt{\frac{n}{N}}\right),$$

ce qui implique que $c_{2N} = O\left(\sqrt{\frac{n}{N}}\right)$ (terme de biais). Ainsi, en combinant ces propriétés de c_{1N}, c_{2N} et c_{3N}, nous pouvons conclure que

$$\mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right)^{-1/2}\left(\tilde{\theta}_{s_d} - \bar{y}_{U_d}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(B, 1),$$

où $B = O\left(\sqrt{\frac{n}{N}}\right)$.

Écrivons maintenant l'estimateur de la variance réalisable $\hat{V}\left(\tilde{\theta}_{s_d, J}\right)$ comme suit

$$\hat{V}\left(\tilde{\theta}_{s_d, J}\right) = \hat{V}\left(\tilde{y}_{s_d}\right)1_{J = \varnothing} + \sum_{J_G \in \mathcal{G}_\mu \backslash \varnothing} \hat{V}\left(\tilde{\theta}_{s_d, J_G}\right)1_{J = J_G} + \sum_{J_G \in \mathcal{G}_\mu^c} \hat{V}\left(\tilde{\theta}_{s_d, J_G}\right)1_{J = J_G}.$$

Selon l'hypothèse A6, nous obtenons que $\hat{V}\left(\tilde{\theta}_{s_d, J_G}\right) - \mathrm{VA}\left(\tilde{\theta}_{s_d, J_G}\right) = o_p\left(n^{-1}\right)$ pour tout J_G, ce qui implique que $\hat{V}\left(\tilde{\theta}_{s_d, J}\right)^{1/2} - \mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right)^{1/2} = o_p\left(n^{-1/2}\right)$. Par conséquent, une application du théorème de Slutsky permet de remplacer $\mathrm{VA}\left(\tilde{\theta}_{s_d, J}\right)^{-1/2}$ par $\hat{V}\left(\tilde{\theta}_{s_d, J}\right)^{-1/2}$.

Pour prouver la dernière partie du théorème, il suffit de constater que $\mathbf{A\mu} > \mathbf{0}$ implique $\mathcal{G}_\mu = \{\varnothing\}$. Ainsi, le terme c_{2N} n'existe pas et le terme de biais disparaît.

Bibliographie

Bickel, P.J., et Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, 12, 470-482.

Breidt, F.J., Opsomer, J.D. et Sanchez-Borrego, I. (2016). Nonparametric variance estimation under fine stratification: An alternative to collapsed strata. *Journal of the American Statistical Association*, 111(514), 822-833.

Casella, G., et Berger, R.L. (2002). *Statistical Inference*. Duxbury, 2nd Edition.

Fuller, W. (1996). *Introduction to Statistical Time Series*. 2nd Edition, New York: John Wiley & Sons, Inc.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 361-374.

Kott, P.S. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, 17, 521-526.

Krewski, D., et Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9, 1010-1019.

Liao, X., et Meyer, M.C. (2014). `coneproj`: an R package for the primal or dual cone projections with routines for constrained regression. *Journal of Statistical Software*, 61, 1-22.

Meyer, M.C. (1999). An extension of the mixed primal-dual bases algorithm to the case of more constraints than dimensions. *Journal of Statistical Planning and Inference*, 81, 13-31.

Meyer, M.C. (2013). A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics*, 42, 1126-1139.

Oliva-Aviles, C., Meyer, M.C. et Opsomer, J.D. (2019). Checking validity of monotone domain mean estimators. *Canadian Journal of Statistics*, 47(2), 315-331.

Opsomer, J.D., Breidt, F.J., White, M. et Li, Y. (2016). Succesive difference replication variance estimation in two-phase sampling. *Journal of Survey Statistical Methodology*, 4(1), 43-70.

Robertson, T., Wright, F.T. et Dykstra, R.L. (1988). *Order Restricted Statistical Inference*. New York: John Wiley & Sons, Inc.

Rockafellar, R.T. (1970). *Convex Analysis*. New Jersey: Princeton University Press.

Rudin, W. (1976). *Principles of Mathematical Analysis*. New York: McGraw Hill.

Rueda, C., et Lombardía, M. (2012). Small area semiparametric additive monotone models. *Statistical Modelling*, 12(6), 527-549.

Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Wu, J., Meyer, M.C. et Opsomer, J.D. (2016). Survey estimation of domain means that respect natural orderings. *Canadian Journal of Statistics*, 44(4), 431-444.