

Université de Montréal

COMPARAISON EMPIRIQUE DES MÉTHODES
BOOTSTRAP DANS UN CONTEXTE
D'ÉCHANTILLONNAGE EN POPULATION
FINIE

par

Oussama DABDOUBI

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

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

août 2019

© Oussama DABDOUBI, 2019

Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé

**COMPARAISON EMPIRIQUE DES MÉTHODES
BOOTSTRAP DANS UN CONTEXTE
D'ÉCHANTILLONNAGE EN POPULATION
FINIE**

présenté par

Oussama DABDOUBI

a été évalué par un jury composé des personnes suivantes :

David Haziza

(président-rapporteur)

Christian Léger

(directeur de recherche)

Florian Maire

(membre du jury)

Mémoire accepté le :

23 juillet 2019

Sommaire

Dans ce travail, nous comparons par simulation diverses méthodes bootstrap d'évaluation de la précision d'une estimation d'enquête pour trois plans d'échantillonnage: le plan aléatoire simple sans remise, le plan de Poisson et le plan de Poisson conditionnel. La plupart des méthodes bootstrap ont été dérivées de manière à reproduire l'estimation habituelle de la variance dans le cas d'un paramètre linéaire comme la moyenne ou le total d'une population. Nous évaluons la méthode d'une part pour estimer la variance et d'autre part pour construire des intervalles de confiance bootstrap en utilisant quatre techniques différentes, à savoir la technique asymptotique, percentile, bootstrap de base et t -bootstrap pour un total, mais aussi un ratio, un coefficient de corrélation, une médiane et un indice de Gini. La manière de mettre en œuvre le t -bootstrap pour plusieurs méthodes bootstrap est une contribution originale de ce mémoire.

Mots clés: Estimation de la variance, intervalles de confiance, t -bootstrap, poids bootstrap, pseudo-population, plan d'échantillonnage, échantillonnage à probabilité inégale.

Summary

In this thesis, we compare by simulation various bootstrap methods for evaluating the accuracy of a survey estimate for three sampling designs: simple random sampling, Poisson sampling and conditional Poisson sampling. Most bootstrap methods have been derived to reproduce the usual variance estimate for a linear parameter such as the average or the total of a population. We evaluate the method on their capacity to estimate the variance of survey estimates and to build bootstrap confidence intervals using four different techniques, namely asymptotic intervals, bootstrap percentile, basic bootstrap and t -bootstrap intervals for totals, but also ratios, correlation coefficients, medians, and Gini indices. The implementation of t -bootstrap intervals for several bootstrap methods is an original contribution of this dissertation.

Keywords: Variance estimation, confidence intervals, t -bootstrap, bootstrap weights, pseudo-population, sampling design, unequal probability sampling.

Table des matières

Sommaire	v
Summary	vii
Liste des tableaux	xv
Liste des figures	xxv
Remerciements	xxvii
Introduction	1
Chapitre 1. Quelques rappels en théorie d'échantillonnage et le calcul de précision	5
1.1. Concepts de base en échantillonnage	5
1.1.1. Population, variable et paramètre d'intérêt	5
1.1.2. Plan d'échantillonnage	7
1.1.3. Estimation de paramètres	8
1.2. Méthodes de calcul de précision	12
1.2.1. Approche analytique	12
1.2.1.1. Calcul et estimation de la variance de l'estimateur d'Horvitz Thompson	12

1.2.1.2.	La linéarisation	14
1.2.2.	Approche par la technique du bootstrap.....	15
1.2.2.1.	Les intervalles de confiance bootstrap	17
Chapitre 2.	Bootstrap pour un plan d'échantillonnage aléatoire	
	simple sans remise (EASSR)	21
2.1.	Plan d'échantillonnage aléatoire simple sans remise.....	21
2.2.	Les méthodes de bootstrap existantes	22
2.2.1.	Le bootstrap par pseudo-population	22
2.2.1.1.	Méthode de Booth et coll. (1994)	23
2.2.1.2.	Méthode de Chao et Lo (1994)	24
2.2.1.3.	Méthode de Bickel et Freedman (1984) et Chao et Lo (1985) ..	25
2.2.1.4.	Méthode de Sitter (1992a)	26
2.2.1.5.	Algorithme général pour les méthodes bootstrap par pseudo- population	28
2.2.2.	Le bootstrap direct.....	29
2.2.2.1.	Le bootstrap avec remise de McCarthy et Snowden (1985).....	30
2.2.2.2.	Le bootstrap rééchelonné de Rao et Wu (1988)	30
2.2.2.3.	Le bootstrap à concordance-miroir de Sitter (1992b).....	31
2.2.2.4.	Algorithme général pour les méthodes bootstrap direct	33
2.2.3.	Le bootstrap pondéré	34
2.2.3.1.	Le bootstrap pondéré de Rao et coll. (1992).....	36
2.2.3.2.	Chipperfield et Preston (2007)	37

2.2.3.3.	Le bootstrap généralisé de Bertail et Combris (1997) et Beaumont et Patak (2012)	37
2.2.3.4.	Les méthodes d'Antal et Tillé (2011a) et Antal et Tillé (2014) .	39
2.3.	Les intervalles de confiance bootstrap.....	41
Chapitre 3. Bootstrap pour un plan d'échantillonnage à probabilités inégaux sans remise		49
3.1.	Rappels sur l'échantillonnage à probabilités inégales	50
3.1.1.	Plan de Poisson	51
3.1.2.	Plan de Poisson conditionnel	52
3.1.3.	Formules simplifiées d'estimation de variance pour les plans à forte entropie	53
3.2.	Méthodes bootstrap par pseudo-population	54
3.2.1.	Holmberg (1998)	54
3.2.2.	Chauvet (2007)	55
3.2.3.	Barbiero et Mecatti (2010)	57
3.3.	Méthodes bootstrap pondéré	58
3.3.1.	Bertail et Combris (1997) et Beaumont et Patak (2012)	58
3.3.2.	Antal et Tillé (2011a) et Antal et Tillé (2014)	58
Chapitre 4. Étude par simulation.....		61
4.1.	Introduction	61
4.2.	Populations et échantillons simulés	62

4.3.	Critères de comparaison des estimateurs de variance.....	67
4.4.	Étude 1: Méthodes bootstrap pour le plan aléatoire simple sans remise	68
4.4.1.	Résultats: cas de paramètres linéaires	70
4.4.2.	Discussion: cas de paramètres linéaires	82
4.4.3.	Résultats: cas des paramètres lisses de totaux	85
4.4.4.	Discussion: cas des paramètres lisses de totaux	97
4.4.5.	Résultats: cas des paramètres non lisses	99
4.4.6.	Discussion: cas des paramètres non lisses.....	111
4.5.	Étude 2: Plan de Poisson.....	113
4.5.1.	Résultats: cas des paramètres linéaires	114
4.5.2.	Discussion: cas des paramètres linéaires	121
4.5.3.	Résultats: cas des paramètres lisses de totaux	121
4.5.4.	Discussion: cas des paramètres lisses de totaux	128
4.5.5.	Résultats: cas des paramètres non lisses de totaux	130
4.5.6.	Discussion: cas des paramètres non lisses de totaux	137
4.6.	Étude 3: Plan de Poisson conditionnel.....	138
4.6.1.	Résultats: cas des paramètres linéaires	139
4.6.2.	Discussion: cas des paramètres linéaires	146
4.6.3.	Résultats: cas des paramètres lisses de totaux	147
4.6.4.	Discussion: cas des paramètres lisses de totaux	154
4.6.5.	Résultats: cas des paramètres non lisses de totaux	155
4.6.6.	Discussion: cas des paramètres non lisses de totaux	162
Conclusion	163

Bibliographie 169

Liste des tableaux

2.1	Méthodes bootstrap par pseudo-population pour un plan EASSR.....	28
2.2	Méthodes bootstrap direct pour un plan EASSR	35
2.3	Description du premier et second ré-échantillonnage pour chaque méthode bootstrap direct ou pondéré pour calculer un intervalle de confiance t -bootstrap	45
4.1	Populations et scénarios de simulation.....	63
4.2	Étude 1: Précision et stabilité des méthodes bootstrap (paramètres linéaires t_y et t_z)	71
4.3	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (t_y)	72
4.4	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (t_y) ...	73
4.5	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (t_y)	74
4.6	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (t_y)..	75

4.7	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (t_y)	76
4.8	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (t_z)	77
4.9	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (t_z) ...	78
4.10	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (t_z)	79
4.11	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (t_z)..	80
4.12	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (t_z)	81
4.13	Biais relatif théorique de la variance de l'estimateur d'Horvitz-Thompson des estimateurs bootstrap par pseudo-population de la variance, $BRT = 100(E_p(V^*(\hat{t}_{HT}^*) - V_p(\hat{t}_{HT}))/V_p(\hat{t}_{HT}))$	82
4.14	Étude 1: Biais relatif et stabilité des méthodes bootstrap (paramètres lisses de totaux $r_{y,z}$ et $\rho_{y,z}$).....	86
4.15	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$).....	87

4.16	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$) ..	88
4.17	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)	89
4.18	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)	90
4.19	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (le ratio $r_{y,z}$)	91
4.20	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)	92
4.21	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$) ..	93
4.22	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)	94
4.23	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)	95
4.24	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (Le coefficient de corrélation $\rho_{y,z}$)	96

4.25	Étude 1: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses de totaux, m_y et G_y)	100
4.26	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (m_y)	101
4.27	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (m_y) ..	102
4.28	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (m_y)	103
4.29	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (m_y) .	104
4.30	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (La médiane m_y)	105
4.31	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (G_y)	106
4.32	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (G_y) ..	107
4.33	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (G_y)	108
4.34	Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (G_y) .	109

4.35	Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (Gini G_y)	110
4.36	Étude 2: Biais relatif et stabilité des méthodes bootstrap (paramètres linéaires, t_y et t_z).	115
4.37	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres linéaires, t_y et t_z).	116
4.38	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires, t_y et t_z).	117
4.39	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires, t_y et t_z).	118
4.40	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres linéaires, t_y et t_z).	119
4.41	Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (Totaux t_y et t_z)	120
4.42	Étude 2: Biais relatif et stabilité des méthodes bootstrap, (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	122

4.43	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	123
4.44	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	124
4.45	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	125
4.46	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	126
4.47	Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) ($r_{y,z}$ et $\rho_{y,z}$)	127
4.48	Étude 2: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses m_y et G_y)	131
4.49	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)	132
4.50	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)	133

4.51	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)	134
4.52	Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)	135
4.53	Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (m_y et G_y)	136
4.54	Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres linéaires t_y et t_z)	140
4.55	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres linéaires t_y et t_z)	141
4.56	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires t_y et t_z)	142
4.57	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires t_y et t_z)	143
4.58	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres linéaires t_y et t_z)	144

4.59	Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (Totaux t_y et t_z)	145
4.60	Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	148
4.61	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	149
4.62	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	150
4.63	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres lisses des totaux, $r_{y,z}$ et $\rho_{y,z}$)	151
4.64	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)	152
4.65	Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) ($r_{y,z}$ et $\rho_{y,z}$)	153
4.66	Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses, m_y et G_y)	156

4.67	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres non lisses, m_y et G_y)	157
4.68	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres non lisses, m_y et G_y)	158
4.69	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres non lisses, m_y et G_y)	159
4.70	Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres non lisses, m_y et G_y)	160
4.71	Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (m_y et G_y)	161

Liste des figures

2.1	Méthode bootstrap par pseudo-population unifiée pour un plan EASSR .	29
2.2	Méthode bootstrap direct unifiée pour un plan EASSR	34
4.1	Répartition des quatre populations selon la variable y	64
4.2	Répartition des quatre populations selon la variable z	65

Remerciements

Ce mémoire n'aurait pas été possible sans l'intervention, consciente, d'un grand nombre de personnes. Je souhaite ici les en remercier.

Je tiens d'abord à remercier très chaleureusement mon directeur de recherche, Christian LÉGER qui m'a permis de bénéficier de son encadrement. Les conseils qu'il m'a prodigués, la patience qu'il a montrée ont été déterminants dans la réalisation de ce travail de recherche.

Mes remerciements s'étendent également à tous mes enseignants durant les années des études.

Enfin, je remercie ma famille pour leurs encouragements et leur soutien tout au long de mes études. Un merci particulier à ma bien aimée, ma femme Loubna et mes trois filles, Tasnim, Nouha et Lina pour leur patience et leur amour.

Introduction

Les organismes de statistique officielle comme le Bureau du recensement des États-Unis et Statistique Canada, réalisent des centaines d'enquêtes par année pour estimer des paramètres descriptifs de la population comme la moyenne, la médiane et le ratio, de même que des paramètres de modèle comme le coefficient de corrélation ou de régression et les indices d'inégalités de revenu, sur la base d'un échantillon sélectionné d'une manière aléatoire à partir d'une population d'intérêt.

Pour estimer un paramètre les statisticiens d'enquête utilisent les données collectées et les poids d'échantillonnage associés aux individus sélectionnés dans l'échantillon après les avoir ajustés pour améliorer les estimations et/ou pour remédier au problème de la non réponse. Un poids d'échantillonnage peut être interprété comme le nombre d'individus dans la population que chaque individu échantillonné représente.

Dans ce mémoire on s'intéresse aux méthodes bootstrap de calcul de la précision des estimations d'une enquête probabiliste. La précision est mesurée par la variance estimée ou par intervalle de confiance.

Pour mettre en œuvre le bootstrap dans un contexte d'échantillonnage (population finie) plusieurs adaptations ont été proposées. Mashreghi et coll. (2016) ont donné un excellent aperçu du sujet. Ils y ont présenté les méthodes bootstrap dans un cadre unifié en les regroupant en trois classes: le bootstrap par pseudo-population,

le bootstrap direct et le bootstrap pondéré. Ils ont étudié l'estimation de la variance et la construction d'intervalles de confiance pour l'échantillonnage aléatoire simple stratifié ainsi que certains plans d'échantillonnage à probabilités inégales.

Le présent mémoire s'inscrit dans la continuité d'une partie de ce travail en voulant d'une part comparer via des simulations les différentes méthodes bootstrap proposées pour calculer la précision dans le contexte d'échantillonnage et d'autre part de mettre la lumière sur la construction des intervalles de confiance bootstrap pour certaines méthodes, notamment une technique comme le t -bootstrap. En effet, dans la littérature sur le bootstrap pour population finie la construction d'intervalles de confiance a reçu beaucoup moins d'attention que le calcul d'estimateur de la variance. De plus, à notre connaissance, la construction d'intervalles de confiance t -bootstrap n'a pas vraiment été discutée dans cette littérature. Nous allons donc introduire ces intervalles pour les méthodes bootstrap pour lesquelles ce sera possible.

Pour ce faire, nous organisons notre travail en quatre chapitres. Dans le premier chapitre nous introduisons la notation ainsi que des généralités sur la théorie d'échantillonnage et nous introduisons deux méthodes pour l'estimation de la précision (la technique de linéarisation et la technique du bootstrap d'Efron). Dans le deuxième chapitre, nous reprenons la même classification proposée par Mashreghi et coll. (2016) pour présenter et décrire les méthodes bootstrap existantes dans la littérature pour un plan d'échantillonnage aléatoire simple. Nous étudions l'estimation de la variance et la construction d'intervalles de confiance bootstrap.

Dans le chapitre 3 nous introduisons quelques rappels sur les plans d'échantillonnage à probabilités inégales et nous présentons les méthodes bootstrap pour le plan de Poisson et certains plans à grande entropie.

Dans le chapitre 4 nous décrivons et présentons les résultats de nos études par simulations dont l'objet est de comparer les différentes méthodes bootstrap pour les plans d'échantillonnage étudiés.

Et en conclusion nous donnons des recommandations pratiques basées sur les résultats issus de ces comparaisons empiriques.

Chapitre 1

Quelques rappels en théorie d'échantillonnage et le calcul de précision

Dans ce chapitre, nous rappelons les principales notions élémentaires utilisées en théorie de l'échantillonnage, présentons deux approches pour estimer la précision et introduisons des notations qui nous seront utiles dans la suite.

La rédaction d'une grande partie de ce chapitre est basée sur les deux ouvrages suivants, Särndal et coll. (1992) et Tillé (2001). Ainsi, pour une présentation détaillée de la théorie d'échantillonnage et de ses principaux résultats, nous conseillons les lecteurs intéressés de consulter ces deux références.

1.1. Concepts de base en échantillonnage

1.1.1. Population, variable et paramètre d'intérêt

En théorie de l'échantillonnage (ou des sondages), on s'intéresse à **une population** finie U composée de N unités ou individus dénotés $u_1, \dots, u_i, \dots, u_N$. On suppose que les unités de la population sont identifiables par un numéro d'ordre et on représentera par simplicité les unités par leur numéro d'ordre. On notera alors

$$U = \{1, \dots, i, \dots, N\}.$$

On appelle donc i l'étiquette de l'unité et on suppose que cette étiquette permet d'identifier l'unité sans ambiguïté. On note par y **une variable d'intérêt** (éventuellement vectorielle), qu'on suppose mesurable sur chaque unité de U . La valeur prise par y sur la i -ème unité d'observation de U est notée y_i , et notons par $\theta = \theta(y_i, i \in U)$ **un paramètre d'intérêt** qui s'exprime en fonction des valeurs prises par y sur l'ensemble de la population.

Parmi les paramètres d'intérêt, on trouve le total t_y de la variable y sur la population U ;

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

et la moyenne

$$\mu_y = \frac{1}{N} \sum_{i \in U} y_i.$$

On peut s'intéresser à des paramètres plus complexes, comme le ratio de deux totaux t_y et t_z de deux variables d'intérêt y et z

$$r_{y,z} = \frac{t_y}{t_z},$$

le coefficient de corrélation de y et z

$$\rho_{y,z} = \frac{\sum_{i \in U} (y_i - \mu_y)(z_i - \mu_z)}{\sqrt{\sum_{i \in U} (y_i - \mu_y)^2 \sum_{i \in U} (z_i - \mu_z)^2}},$$

le quantile q_α d'ordre $\alpha \in [0, 1]$ de la distribution de la variable y

$$q_\alpha = F_N^{-1}(\alpha),$$

où $F_N^{-1}(\cdot)$ est la fonction inverse de $F_N(\cdot)$ définie par $F_N(x) = \frac{1}{N} \sum_{i \in U} I(y_i \leq x)$, $\forall x \in \mathbb{R}$ et $I(A)$ est la fonction indicatrice de l'événement A prenant la valeur 1 quand A

se réalise et 0 sinon et l'indice de Gini de la variable y

$$G = \frac{\sum_{i \in U} y_{(i)} (2F_N(y_{(i)}) - 1)}{t_y} - \frac{1}{N},$$

où les $y_{(i)}$ sont les statistiques d'ordre $y_{(1)} < \dots < y_{(i)} < \dots < y_{(N)}$ de y . L'indice de Gini est utilisé comme une mesure d'inégalité des revenus.

1.1.2. Plan d'échantillonnage

Pour estimer un paramètre d'intérêt θ , nous allons tirer un **échantillon** $s \subset U$, de taille n , selon un certain plan d'échantillonnage, ce qui nous permettra d'observer dans le cas idéal (absence de données manquantes), $\mathbf{y}_s = (y_i, i \in s)^T$ le vecteur (éventuellement la matrice) constitué par les valeurs prises par la variable d'intérêt y (éventuellement vectorielle) sur les unités échantillonnées. À partir de \mathbf{y}_s et d'autres informations qui se rapportent au plan d'échantillonnage on peut fournir une estimation du paramètre d'intérêt θ . Le plan d'échantillonnage que nous désignons par p est une loi de probabilité $p(\cdot)$ associant à chaque échantillon $s \in \Omega$ une probabilité d'être sélectionné $p(s)$, où Ω est l'ensemble de tous les échantillons possibles. La loi de probabilité p vérifie donc $p(s) \geq 0$ pour tous les $s \in \Omega$ et $\sum_{s \in \Omega} p(s) = 1$.

Le plan d'échantillonnage joue un rôle primordial en théorie des sondages car c'est lui qui mène à l'inférence. Le vecteur $y = (y_1, \dots, y_N)^T$ est considéré comme fixe. L'aléa correspond à l'échantillon aléatoire s dont sa réalisation est modélisée par le plan d'échantillonnage p . Par souci de simplicité, nous dénotons tout simplement s pour désigner l'échantillon aléatoire et sa réalisation.

Théoriquement, un plan d'échantillonnage est déterminé entièrement par la loi de probabilité qui associe à chaque partie de U , s , une probabilité $p(s)$ d'être sélectionnée. En pratique, il peut être difficile de calculer la distribution induite par des plans d'échantillonnage complexes, toutefois il serait toujours possible de vérifier que ces plans respectent bien les probabilités d'inclusion du premier ordre des unités $i \in U$

dans l'échantillon qui sont définies par $\pi_i = Pr(i \in s) = \sum_{s \ni i} p(s)$ et qui sont fixées à priori. De manière similaire on définit les probabilités d'inclusion du deuxième ordre des unités $i, j \in U$ c'est-à-dire la probabilité que deux individus apparaissent dans le même échantillon par $\pi_{ij} = Pr(i \in s \text{ et } j \in s) = \sum_{s \ni i, j} p(s)$.

Ces probabilités d'inclusion sont fondamentales car elles assurent le lien entre la population et l'échantillon. On peut distinguer les plans d'échantillonnage selon les probabilité d'inclusion du premier ordre en deux catégories: ceux dits à probabilités égales et ceux dits à probabilités inégales. Dans la première catégorie on trouvera par exemple le plan aléatoire simple sans remise (voir chapitre 2). Dans la seconde catégorie on trouvera par exemple le plan proportionnel à la taille (voir chapitre 3).

Une autre distinction des plans d'échantillonnage repose sur le critère de la taille d'échantillon. Si la taille est aléatoire, le plan est dit à taille aléatoire, comme le plan de Poisson (voir chapitre 3). Par contre si la taille est fixe le plan est dit à taille fixe, comme le plan aléatoire simple sans remise.

1.1.3. Estimation de paramètres

L'estimation de paramètres de la population finie est fondée sur les données de n unités échantillonnées et sur leurs probabilités d'inclusion dans le plan d'échantillonnage. L'inverse de la probabilité d'inclusion d'une unité i représente le poids d'échantillonnage. Ce dernier est souvent redressé pour améliorer la précision, auquel cas il sera appelé poids d'estimation. Pour bien déceler les obstacles rencontrés lors de l'évaluation de la précision de l'estimateur $\hat{\theta}$ du paramètre θ , on classe les paramètres d'intérêt en deux grandes catégories: ceux qui sont linéaires, et ceux qui ne le sont pas. On commence tout d'abord par un estimateur linéaire général du total t de y

$$\hat{t} = \sum_{i \in s} w_i(s) y_i, \quad (1.1.1)$$

où $w_i(s)$ est le poids d'échantillonnage ou d'estimation associé à l'individu i . Si on utilise le poids d'échantillonnage $w_i(s) = \pi_i^{-1}$, l'estimateur qui en résulte s'appelle l'estimateur d'Horvitz et Thompson (1952):

$$\hat{t}_{HT} = \sum_{i \in s} \pi_i^{-1} y_i. \quad (1.1.2)$$

D'autres estimateurs comme l'estimateur linéaire GREG (*generalized regression*), voir chapitre 6 de Särndal et coll. (1992), utilisent des poids redressés en se basant sur l'information auxiliaire disponible pour obtenir des propriétés souhaitables comme la cohérence des estimations entre les différentes sources et l'amélioration de la précision des estimations. Les poids d'estimation du GREG se basent sur l'information disponible des individus échantillonnés concernant l variables auxiliaires $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{il})^T$ et du vecteur de totaux $\mathbf{t}_x = (t_{x1}, t_{x2}, \dots, t_{xl})^T$ de ces variables sur l'ensemble de la population. Les poids d'estimation GREG sont

$$w_i(s) = \pi_i^{-1} \left\{ 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{xHT})^T \hat{\mathbf{T}}^{-1} c_i^{-1} \mathbf{x}_i \right\}$$

où $\hat{\mathbf{t}}_{xHT} = \sum_{i \in s} \pi_i^{-1} \mathbf{x}_i$, $\hat{\mathbf{T}} = \sum_{i \in s} \pi_i^{-1} \mathbf{x}_i c_i^{-1} \mathbf{x}_i^T$ et c_i sont des facteurs d'échelle choisis par le statisticien.

Considérons maintenant le cas où le paramètre θ n'est pas linéaire mais a l'avantage d'être exprimé comme une fonction régulière de K totaux de variables individuelles v_1, \dots, v_K (les variables individuelles peuvent être différentes des variables d'intérêt; c'est pour cette raison que nous avons changé la notation)

$$\theta = f(t_1, \dots, t_K),$$

où $t_h = \sum_{i \in U} v_{hi}$.

Un estimateur par substitution de θ est donné par

$$\hat{\theta} = f(\hat{t}_{1HT}, \dots, \hat{t}_{KHT}) \quad (1.1.3)$$

où chaque total est remplacé par son estimateur Horvitz-Thompson.

Le ratio des totaux t_y et t_z de deux variables d'intérêt y et z est un exemple avec $K = 2$. Dans cet exemple, les variables individuelles sont les variables d'intérêt $y = v_1$ et $z = v_2$, ainsi son estimateur par substitution est donné par;

$$\hat{r}_{y,z} = \frac{\sum_{i \in s} \pi_i^{-1} y_i}{\sum_{i \in s} \pi_i^{-1} z_i} = \frac{\hat{t}_{1HT}}{\hat{t}_{2HT}}. \quad (1.1.4)$$

La fonction de répartition de la variable y au point x dans le cas où la taille N de la population est inconnue peut être vue comme le ratio de totaux de deux variables individuelles v_1 et v_2 où, $v_{1i} = I(y_i \leq x)$ et $v_{2i} = 1$:

$$\hat{F}_N(x) = \frac{\sum_{i \in s} \pi_i I(y_i \leq x)}{\sum_{i \in s} \pi_i}.$$

Le coefficient de corrélation de y et z est un paramètre qui s'exprime comme une fonction régulière de $K = 5$ totaux de variables individuelles v_1, \dots, v_5 , où $v_{1i} = y_i$, $v_{2i} = z_i$, $v_{3i} = y_i^2$, $v_{4i} = z_i^2$ et $v_{5i} = y_i z_i$. Ainsi son estimateur par substitution est donné par

$$\hat{\rho}_{y,z} = \frac{\hat{t}_{5HT} - N^{-1} \hat{t}_{1HT} \hat{t}_{2HT}}{\sqrt{(\hat{t}_{3HT} - N^{-1} \hat{t}_{1HT}^2) (\hat{t}_{4HT} - N^{-1} \hat{t}_{2HT}^2)}}. \quad (1.1.5)$$

Étudions maintenant le cas des paramètres qui peuvent être définis implicitement comme des solutions d'équations estimantes.

Prenons l'exemple du quantile q_α de la distribution de la variable y qui peut être défini comme solution de l'équation estimante

$$F_N(q_\alpha) = \alpha \Leftrightarrow \sum_{i \in U} I(y_i \leq q_\alpha) = \sum_{i \in U} \alpha$$

$$\sum_{i \in U} (I(y_i \leq q_\alpha) - \alpha) = 0.$$

Son estimateur par équation estimante \hat{q}_α est solution de l'équation

$$\sum_{i \in s} \pi_i^{-1} (I(y_i \leq \hat{q}_\alpha) - \alpha) = 0 \Leftrightarrow \hat{F}_n(\hat{q}_\alpha) = \alpha \Leftrightarrow \hat{q}_\alpha = \hat{F}_n^{-1}(\alpha),$$

où, $\hat{F}_n^{-1}(\cdot)$, l'inverse de $\hat{F}_n(\cdot)$ est définie ainsi

$$\hat{F}_n^{-1}(b) = \inf\{y_i | \hat{F}_n(y_i) \geq b; i \in s\}, \text{ avec } 0 \leq b \leq 1.$$

Dans le cas particulier $\alpha = 0,5$ on trouve l'estimateur de la médiane de la variable y

$$\hat{m}_y = \hat{F}_n^{-1}(0,5). \quad (1.1.6)$$

L'indice de Gini est un paramètre plus complexe, il requiert la résolution d'un système de plusieurs équations estimantes. En effet, dans le cas de ce paramètre il est nécessaire d'estimer de façon simultanée les quantiles de la variable d'intérêt. Dans notre étude par simulation nous utilisons l'estimateur \hat{G} utilisé par Antal et coll. (2011):

$$\hat{G} = \frac{\sum_{i \in s} \pi_{(i)}^{-1} y_{(i)} (2\hat{F}_N(y_{(i)}) - 1)}{\sum_{i \in s} \pi_{(i)}^{-1} y_{(i)}} - \frac{\sum_{i \in s} \pi_{(i)}^{-2} y_{(i)}}{\sum_{i \in s} \pi_{(i)}^{-1} \sum_{i \in s} \pi_{(i)}^{-1} y_{(i)}} \quad (1.1.7)$$

où les $y_{(i)}$ sont les statistiques d'ordre $y_{(1)} < \dots < y_{(i)} < \dots < y_{(n)}$ et $\pi_{(i)}^{-1}$ désigne le poids associé au i -ème statistique d'ordre.

Pour un estimateur $\hat{\theta}$ son espérance par rapport au plan d'échantillonnage (respectivement sa variance) sont donnés par:

$$E_p(\hat{\theta}) = \sum_{s \in \Omega} p(s) \hat{\theta}(s),$$

$$V_p(\hat{\theta}) = \sum_{s \in \Omega} p(s) \left(\hat{\theta}(s) - E_p(\hat{\theta}) \right)^2.$$

1.2. Méthodes de calcul de précision

Pour évaluer la précision d'un estimateur dans un contexte d'échantillonnage, il y a en général deux approches : une analytique et l'autre par des méthodes de rééchantillonnage. La première approche consiste à établir des formules explicites de variance dans lesquelles chacun des termes peut être évalué soit exactement, soit approximativement. A partir de la variance estimée on pourra également construire un intervalle de confiance en s'appuyant sur l'approximation normale de la loi de $\hat{\theta}$. Dans la seconde, on ne cherche pas à établir d'expression analytique de la variance, mais à manipuler les données quantitatives issues de l'échantillon s en suivant une démarche carrément calculatoire, mais basée sur le plan d'échantillonnage, pour obtenir une estimation de la variance et un intervalle de confiance.

1.2.1. Approche analytique

1.2.1.1. Calcul et estimation de la variance de l'estimateur d'Horvitz Thompson

Soit s un échantillon sélectionné selon un plan d'échantillonnage sans remise p qui respecte les probabilités d'inclusion d'ordre 1 fixées à priori et supposées strictement positives $\pi_i > 0$, $i \in U$, alors l'estimateur de Horvitz Thompson \hat{t}_{HT} (1.1.2) du total t de la variable y est sans biais et sa variance est donnée par

$$V_p(\hat{t}_{HT}) = \sum_{i \in U} \sum_{j \in U} \frac{y_i y_j}{\pi_i \pi_j} \Delta_{ij}, \quad (1.2.1)$$

où

$$\Delta_{ij} = \pi_{ij} - \pi_i \pi_j \text{ et } \Delta_{ii} = \pi_i (1 - \pi_i).$$

Un estimateur sans biais de cette variance, pourvu que $\pi_{ij} > 0$ pour tout $i \neq j$, est donné par:

$$\hat{V}_{HT} = \sum_{i \in s} \sum_{j \in s} \frac{y_i y_j}{\pi_i \pi_j} \frac{\Delta_{ij}}{\pi_{ij}}. \quad (1.2.2)$$

Cet estimateur peut être utilisé lorsque le plan d'échantillonnage est à taille fixe ou aléatoire. Pour un plan à taille fixe, l'expression de la variance (1.2.1) peut également s'écrire comme

$$V_p(\hat{t}_{HT}) = -\frac{1}{2} \sum_{i \in U} \sum_{j \in U} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \Delta_{ij}. \quad (1.2.3)$$

L'expression (1.2.3) suggère un autre estimateur sans biais de la variance: l'estimateur de Sen-Yates-Grundy donné par

$$\hat{V}_{SYG} = -\frac{1}{2} \sum_{i \in s} \sum_{j \in s} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \frac{\Delta_{ij}}{\pi_{ij}}. \quad (1.2.4)$$

Une condition suffisante pour que \hat{V}_{SYG} soit positif est que $\pi_{ij} - \pi_i \pi_j < 0$ pour tout $i \neq j$. Cette condition s'appelle la condition de Sen-Yates-Grundy. En revanche, même si cette condition est satisfaite, \hat{V}_{HT} peut être négatif. Les deux formules d'estimateur de la variance (1.2.2) et (1.2.4) utilisent les probabilités d'inclusion double. Or, en dehors de quelques plans d'échantillonnage (échantillonnage aléatoire simple, voir chapitre 2 ; plans de Poisson et de Poisson conditionnel, voir chapitre 3), ces probabilités sont généralement très difficiles à calculer. Pour contourner ces difficultés, il est fréquent de passer par des approximations des probabilités d'inclusion double ou bien par la construction de formules de variances simplifiées dans lesquelles les probabilités d'inclusion double n'apparaissent pas. Ces approximations sont valides sous des conditions dites asymptotiques, c'est-à-dire que n et N sont supposés grands et n négligeable par rapport à N et/ou pour des plans d'échantillonnage à grande entropie par exemple, Deville (1993), Hájek (1964) et Brewer et Donadio (2003).

1.2.1.2. La linéarisation

Considérons maintenant le cas où le paramètre d'intérêt est non linéaire mais s'exprime comme fonction régulière de K totaux, $\theta = f(t_1, \dots, t_K)$, où t_k est le total d'une variable v_k et supposons qu'on l'estime à l'aide de son estimateur par substitution (1.1.3).

Dans ce cas, il y a une technique permettant des calculs analytiques approchés de la variance de cet estimateur: il s'agit de la technique de linéarisation dont le principe est d'approcher l'estimateur non linéaire $\hat{\theta}$ par un estimateur linéaire $\hat{\theta}_l$ obtenu à l'aide de la linéarisation de Taylor de la fonction f au point (t_1, \dots, t_K) . Plus précisément :

$$\hat{\theta} \simeq \hat{\theta}_l = \theta - \sum_{i \in U} z_i + \sum_{i \in s} \pi_i^{-1} z_i,$$

où

$$z_i = \sum_{k=1}^K v_{ki} \frac{\partial f(\hat{t}_{1HT}, \dots, \hat{t}_{KHT})}{\partial \hat{t}_{kHT}} \Big|_{(\hat{t}_{1HT}, \dots, \hat{t}_{KHT}) = (t_1, \dots, t_K)},$$

avec v_{ki} la valeur pour l'unité i de la variable v_k dont le total est t_k . La variance approchée de $\hat{\theta}$ est donnée par

$$AV_p(\hat{\theta}) = V_p(\hat{\theta}_l) = \sum_{i \in U} \sum_{j \in U} \frac{z_i z_j}{\pi_i \pi_j} \Delta_{ij}. \quad (1.2.5)$$

La variable z est appelée la linéarisée de θ et un estimateur de (1.2.5) est donné par

$$\hat{V}_{LIN} = \sum_{i \in s} \sum_{j \in s} \frac{\hat{z}_i \hat{z}_j}{\pi_i \pi_j} \frac{\Delta_{ij}}{\pi_j} \quad (1.2.6)$$

avec \hat{z}_i s'obtenant de z_i en remplaçant chaque total par l'estimateur de Horvitz Thompson correspondant.

Notons que le cadre de la linéarisation est un cadre dit **asymptotique** c'est-à-dire qu'il concerne des tailles d'échantillon et de population qui tendent vers l'infini.

En général on construit un cadre de modélisation dans lequel on peut faire tendre la taille de la population N et la taille d'échantillon n vers l'infini. On considère une suite croissante de populations finies indexées par δ et de taille $N_\delta : U_\delta = \{1, \dots, N_\delta\}$. La séquence des populations $(y_{i,\delta}, i \in U_\delta), \delta = 1, 2, \dots$ peut être modélisée de deux façons différentes. Soit comme un vecteur de valeurs fixées avec certaines conditions qui garantissent, par exemple, que la variance de la population converge vers une constante. Soient que les valeurs $y_{i,\delta}$ sont des réalisations de variables aléatoires $Y_{i,\delta}$, mais dans ce cas-ci l'inférence se fait dans un autre cadre où l'on tient également compte du caractère aléatoire des valeurs de la population. Par conséquent, différents auteurs utilisent des cadres asymptotiques différents en fonction des plans d'échantillonnage, des estimateurs et de la modélisation de la séquence des populations; voir par exemple la section 1.3 de Fuller (2009).

1.2.2. Approche par la technique du bootstrap

Le bootstrap d'Efron (1979) est un outil très largement utilisé dans le cas d'une analyse statistique en population infinie, où les observations sont supposées indépendantes et identiquement distribuées (i.i.d.) selon une distribution F généralement inconnue. Une telle supposition n'est généralement pas valable dans le contexte d'échantillonnage, puisque dans la plupart des plans d'échantillonnage on interdit d'échantillonner avec remise les unités statistiques pour améliorer l'estimation des paramètres de la population. Efron et Tibshirani (1993) présentent une introduction au bootstrap.

Soient X_1, \dots, X_n un échantillon d'observations i.i.d. selon une loi F inconnue, et $\theta(F)$ un paramètre estimé par $\hat{\theta}$ une fonction de l'échantillon. Pour bien illustrer le principe du bootstrap nous adoptons la notation $\hat{\theta} = \theta(X_1, \dots, X_n; F)$ de telle sorte que $\hat{\theta}$ est une fonction d'un échantillon X_1, \dots, X_n i.i.d. selon une loi F .

Donc, le bootstrap estime la distribution de $\hat{\theta}$ par celle de $\hat{\theta}^* = \theta \left(X_1^*, \dots, X_n^*, \hat{F}_n \right)$, où X_1^*, \dots, X_n^* est un échantillon i.i.d. distribué selon \hat{F}_n la fonction de répartition expérimentale calculée à partir de X_1, \dots, X_n .

La fonction de répartition expérimentale pour un réel z est donnée par:

$$\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq z).$$

L'estimation bootstrap de la variance de $\hat{\theta}$ est $V^* = V \left(\hat{\theta}^* | X_1, \dots, X_n \right)$, la variance conditionnelle de $\hat{\theta}^*$ étant donné X_1, \dots, X_n . En général V^* ne peut pas être calculé directement, on procède donc par simulation et une approximation de V^* est obtenue par la méthode de Monte Carlo via l'algorithme bootstrap suivant:

- (1) Générer un échantillon (X_1^*, \dots, X_n^*) i.i.d. selon la loi \hat{F}_n , ce qui est équivalent à tirer un échantillon avec remise de (X_1, \dots, X_n) . Calculer $\hat{\theta}^*$ à partir de l'échantillon résultant.
- (2) Répéter l'étape 1 un grand nombre de fois (disons B). On obtient ainsi B statistiques $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$.
- (3) Calculer l'estimateur de V^* :

$$\hat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^* \right)^2$$

$$\text{où } \hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*.$$

Si on applique cet algorithme pour estimer la variance de \hat{t}_{HT} , l'estimation bootstrap obtenue est une approximation Monte Carlo de la variance de l'estimateur bootstrap $\hat{t}_{HT}^* = \sum_{i \in s^*} \pi_i^{*-1} y_i^* = \sum_{i \in s} \pi_i^{-1} m_i y_i$, où s^* est un échantillon sélectionné dans s avec remise et m_i est un nombre aléatoire qui désigne le nombre de fois où l'individu $i \in s$

apparaît dans l'échantillon bootstrap s^* . La variance est

$$\begin{aligned}
V^*(\hat{t}_{HT}^*) &= V\left(\sum_{i \in s} \frac{y_i}{\pi_i} m_i\right) \\
&= \sum_{i \in s} \left(\frac{y_i}{\pi_i}\right)^2 V(m_i) + \sum_{i \neq j \in s} y_i y_j \pi_i^{-1} \pi_j^{-1} Cov(m_i, m_j) \\
&= \sum_{i \in s} \left(\frac{y_i}{\pi_i}\right)^2 \left(1 - \frac{1}{n}\right) + \sum_{i \neq j \in s} y_i y_j \pi_i^{-1} \pi_j^{-1} \left(-\frac{1}{n}\right) \\
&= \sum_{i \in s} \left(\frac{y_i}{\pi_i} - \frac{\hat{t}_{HT}}{n}\right)^2 = v_{naive}.
\end{aligned}$$

Donc le v_{naive} est identique à l'estimateur habituel de la variance de \hat{t}_{HT} à un facteur près $(n-1)/n$, dans le cas d'un plan d'échantillonnage aléatoire simple *avec remise*.

1.2.2.1. Les intervalles de confiance bootstrap

Pour une présentation des intervalles de confiance bootstrap, voir Hall (1992). La construction d'intervalles de confiance d'un paramètre θ suppose de connaître, même de façon approchée, la fonction quantile de la distribution de $(\hat{\theta} - \theta) / \sqrt{V(\hat{\theta})}$, $J_n^{-1}(\cdot)$ ou celle de $(\hat{\theta} - \theta) / \sqrt{\hat{V}}$, $K_n^{-1}(\cdot)$, où \hat{V} est un estimateur de $V(\hat{\theta})$ la variance de $\hat{\theta}$. Une première solution consiste à définir l'intervalle de confiance asymptotique en utilisant les quantiles de la distribution normale et une estimation bootstrap \hat{V}^* de la variance de $\hat{\theta}$

$$\left[\hat{\theta} - z_{1-\alpha/2} \sqrt{\hat{V}^*}, \hat{\theta} - z_{\alpha/2} \sqrt{\hat{V}^*} \right],$$

où z_β est le quantile d'ordre β de la distribution normale centrée réduite et $1 - \alpha$ est le niveau de confiance retenu.

En utilisant la fonction quantile bootstrap de la distribution de $(\hat{\theta} - \theta) / \sqrt{\hat{V}}$, il sera possible d'obtenir d'autres limites de confiance pour θ . Plus précisément, soit

$$\hat{K}_{n,*}(x) = Prob^* \left(\frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\hat{V}^*}} \leq x \right)$$

où $(\hat{\theta}^* - \hat{\theta}) / \sqrt{\hat{V}^*}$ est la version bootstrap de $(\hat{\theta} - \theta) / \sqrt{\hat{V}}$. Notons que \hat{V}^* est un estimateur de V^* , la variance de $\hat{\theta}^*$, qui dépend de (X_1^*, \dots, X_n^*) . \hat{V}^* peut être calculé par une formule théorique, lorsqu'une telle formule existe, ou à partir du rééchantillonnage du rééchantillon (X_1^*, \dots, X_n^*) utilisé pour calculer $\hat{\theta}^*$. Il s'agit donc dans ce cas d'un double bootstrap, puisque chaque rééchantillon fait lui-même l'objet d'un rééchantillonnage bootstrap tel que décrit précédemment, mais à partir des valeurs X_1^*, \dots, X_n^* plutôt qu'à partir de X_1, \dots, X_n .

Disposant de la fonction quantile bootstrap $\hat{K}_{n,*}^{-1}(\cdot)$, on détermine les quantiles $\alpha/2$ et $1 - \alpha/2$, et on obtient l'intervalle de confiance t -bootstrap de θ

$$\left[\hat{\theta} - \sqrt{\hat{V}} \hat{K}_{n,*}^{-1}(1 - \alpha/2), \hat{\theta} - \sqrt{\hat{V}} \hat{K}_{n,*}^{-1}(\alpha/2) \right]. \quad (1.2.7)$$

Pour estimer la fonction quantile, $J_n^{-1}(\cdot)$, de la distribution $(\hat{\theta} - \theta) / \sqrt{V}$, nous utilisons la distribution bootstrap de $(\hat{\theta}^* - \hat{\theta}) / \sqrt{V^*}$, où, $V^* = V(\hat{\theta}^*)$ supposée connue et qui est généralement égale à \hat{V} .

Disposant de la fonction quantile bootstrap $\hat{J}_{n,*}^{-1}(\cdot)$, on détermine les quantiles $\alpha/2$ et $1 - \alpha/2$, et on obtient l'intervalle de confiance bootstrap de base de θ

$$\left[\hat{\theta} - \sqrt{\hat{V}} \hat{J}_{n,*}^{-1}(1 - \alpha/2), \hat{\theta} - \sqrt{\hat{V}} \hat{J}_{n,*}^{-1}(\alpha/2) \right].$$

En développant, on peut facilement voir que cet intervalle est équivalent à

$$\left[\hat{\theta} - \hat{L}_{n,*}^{-1}(1 - \alpha/2), \hat{\theta} - \hat{L}_{n,*}^{-1}(\alpha/2) \right]. \quad (1.2.8)$$

où $\hat{L}_{n,*}(x) = Prob^*(\hat{\theta}^* - \hat{\theta} \leq x)$ est une estimation bootstrap de $L_n(x) = Prob(\hat{\theta} - \theta \leq x)$. Une façon plus directe d'utiliser le bootstrap pour construire les intervalles de confiance (la méthode des percentiles), consiste à déterminer les percentiles $\alpha/2$ et $1 - \alpha/2$ de la distribution d'échantillonnage expérimentale de $\hat{\theta}^*$. Soient $\hat{\theta}_{(1)}^* \leq \hat{\theta}_{(2)}^* \leq \dots \leq \hat{\theta}_{(B)}^*$ les statistiques bootstrap ordonnées par ordre croissant. L'intervalle de confiance bootstrap de type percentile est donné par:

$$\left[\hat{\theta}_{(B\alpha/2)}^*, \hat{\theta}_{(B(1-\alpha/2))}^* \right]. \quad (1.2.9)$$

Cette approche est équivalente à l'intervalle de confiance suivant

$$\left[\hat{\theta} + \hat{L}_{n,*}^{-1}(\alpha/2), \hat{\theta} + \hat{L}_{n,*}^{-1}(1 - \alpha/2) \right]. \quad (1.2.10)$$

Notons que cet intervalle consiste à utiliser les mêmes quantiles que pour l'intervalle de confiance de base, mais à les inverser et à changer leur signe.

Les formules d'intervalle de confiance que nous avons présentées dans cette sous-section sont valables dans un contexte d'échantillonnage en population infinie, où les observations sont supposées i.i.d.. Toutefois, dans un contexte d'échantillonnage en population finie il faut apporter quelques ajustements à ces formules de calcul pour obtenir le degré de confiance $1 - \alpha$.

Chapitre 2

Bootstrap pour un plan d'échantillonnage aléatoire simple sans remise (EASSR)

2.1. Plan d'échantillonnage aléatoire simple sans remise

Le plan d'échantillonnage aléatoire simple sans remise (EASSR) est fondamental en théorie des sondages car il est à la base de plusieurs plans de sondage plus complexes et très utilisés en pratique, comme le plan stratifié simple et le plan à plusieurs degrés. Il consiste à sélectionner un nombre n fixe d'unités, de telle sorte que tous les sous-ensembles de même taille n aient la même probabilité d'être pigés. Ces deux propriétés permettent de déterminer entièrement le plan (EASSR) :

$$p(s) = \begin{cases} 1/\binom{N}{n} & \text{si } \text{card}(s) = n \\ 0 & \text{sinon.} \end{cases}$$

Les probabilités d'inclusion d'ordre un sont égales à

$$\pi_i = \frac{n}{N}.$$

Les probabilités d'inclusion d'ordre deux sont égales à

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}.$$

En tenant compte des probabilités d'inclusion d'ordre un et deux, l'estimateur d'Horvitz Thompson \hat{t}_{HT} du total de y est égal à $N/n \sum_{i \in s} y_i$ et sa variance peut être calculée exactement:

$$V_p(\hat{t}_{HT}) = N^2 (1 - f) \frac{S_y^2}{n}$$

où $f = n/N$ est la fraction de sondage, $S_y^2 = (N - 1)^{-1} \sum_{i \in U} (y_i - \mu_y)^2$ et $\mu_y = N^{-1} \sum_{i \in U} y_i$ sont la variance et la moyenne de la population. Cette variance peut être estimée sans biais par

$$\hat{V}(\hat{t}_{HT}) = N^2 (1 - f) \frac{s_y^2}{n}$$

où $s_y^2 = (n - 1)^{-1} \sum_{i \in s} (y_i - \bar{y})^2$ et $\bar{y} = n^{-1} \sum_{i \in s} y_i$. Le facteur $(1 - f)$ s'appelle la correction de population finie, c'est-à-dire le gain obtenu en échantillonnant sans remise plutôt qu'avec remise.

2.2. Les méthodes de bootstrap existantes

Pour mettre en œuvre le bootstrap dans un contexte d'échantillonnage (population finie) plusieurs adaptations ont été proposées. Mashreghi et coll. (2016) ont donné un excellent aperçu du sujet. Ils ont présenté les méthodes bootstrap existantes dans un cadre unifié pour la première fois dans la littérature.

Dans la suite de ce chapitre, nous adopterons les notations suivantes. On désigne par s un échantillon de taille n sélectionné par plan d'échantillonnage aléatoire simple sans remise, $\lceil x \rceil$ désigne le plus petit entier supérieur ou égal à x et la valeur $\lfloor x \rfloor$ correspondra à la partie entière de x .

2.2.1. Le bootstrap par pseudo-population

En statistique classique le mécanisme inconnu qui modélise un jeu de données i.i.d. est une loi de probabilité F . En conséquence pour effectuer une procédure bootstrap

pour le modèle ajusté, F est substitué par \hat{F}_n la fonction de répartition expérimentale calculée à partir du jeu de données, puis des observations i.i.d. sont générées à partir de \hat{F}_n . Dans un contexte d'enquête par échantillonnage le mécanisme inconnu est la population U dont l'échantillon a été tiré. Par conséquent l'approche bootstrap par pseudo-population consiste à substituer U par une population artificielle (pseudo-population) U^* construite en répétant l'échantillon d'origine en respectant le plan d'échantillonnage d'origine. Ensuite le plan d'échantillonnage de départ est appliqué à U^* pour sélectionner les échantillons bootstrap.

Les méthodes que nous présenterons sous peu sont essentiellement des variantes de la méthode de Gross (1980) qui ont été proposées pour tenir compte du cas (général) où N/n n'est pas entier. Gross (1980) a supposé que N/n est entier. En conséquence une pseudo-population est constituée en répétant l'échantillon s , N/n fois.

2.2.1.1. *Méthode de Booth et coll. (1994)*

Booth et coll. (1994) proposent de dupliquer $k = \lfloor N/n \rfloor$ fois chaque unité de s , et de compléter les $n \times k$ unités ainsi obtenues en sélectionnant sans remise un échantillon de taille $r = N - n \times k$ dans s , pour obtenir finalement une pseudo-population U^* . On prélève ensuite un échantillon bootstrap s^* par sondage aléatoire simple sans remise de taille n dans la population U^* ainsi constituée.

Il faut noter qu'il existe deux composantes aléatoires dont nous devons tenir compte dans une telle procédure bootstrap: le plan d'échantillonnage appliqué pour compléter une pseudo-population et celui pour choisir un échantillon bootstrap que nous indexons par u^* et p^* , respectivement. Mashreghi et coll. (2016) suggèrent de prendre comme estimation bootstrap de la variance de $\hat{\theta}$, l'estimateur du paramètre

d'intérêt,

$$V^*(\hat{\theta}^*) = E_{u^*} E_{p^*} \left[\left(\hat{\theta}^* - \theta^* \right)^2 | U^* \right], \quad (2.2.1)$$

au lieu de

$$\tilde{V}^*(\hat{\theta}^*) = E_{u^*p^*} \left[\hat{\theta}^* - E_{u^*p^*} \left(\hat{\theta}^* \right) \right]^2 \quad (2.2.2)$$

où, $\hat{\theta}^*$ est l'estimateur bootstrap calculé sur s^* et θ^* est le paramètre bootstrap, soit la valeur du paramètre calculé sur la pseudo-population U^* .

En effet (2.2.2) estime $V_p(\hat{\theta})$ en calculant la variabilité de $\hat{\theta}^*$ autour de $E_{u^*p^*}(\hat{\theta}^*)$. Par contre (2.2.1) estime la même variance en calculant la variabilité de l'estimateur bootstrap autour du paramètre bootstrap θ^* qui est aléatoire sauf si N/n est un entier. En conséquence (2.2.2) va estimer en plus la variance due au changement de θ^* à chaque étape bootstrap. Ainsi, l'estimateur privilégié dans la suite de ce travail sera l'estimateur (2.2.1).

Dans le cas de la variance de \hat{t}_{HT} une estimation bootstrap de cette variance est donnée par :

$$E_{u^*} E_{p^*} \left[\left(\hat{t}_{HT}^* - t^* \right)^2 | U^* \right] = \left[\frac{n-1}{n-f} - \frac{1-fk}{N-1} \left(1 - \frac{r}{n} \right) \right] N^2 (1-f) \frac{s_y^2}{n}.$$

2.2.1.2. Méthode de Chao et Lo (1994)

Comme Booth et coll. (1994), Chao et Lo (1994) proposent de constituer une pseudo-population de taille N , identique à la taille de la population originale. En revanche, Chao et Lo (1994) prennent un échantillon aléatoire simple **avec remise** pour compléter la pseudo-population. Ils construisent leur méthode à partir de principes fondamentaux, en utilisant des idées à partir de la méthode des moments et du maximum de la vraisemblance pour montrer que lorsque N/n est un nombre entier, la seule chose naturelle à faire est de répéter l'échantillon original k fois. Lorsque

N/n n'est pas un entier, ils complètent la pseudo-population avec un échantillon aléatoire simple avec remise de l'échantillon original.

Pour cette méthode l'estimation bootstrap de la variance de \hat{t}_{HT} selon (2.2.1) est

$$E_{u^*} E_{p^*} \left[(\hat{t}_{HT}^* - t^*)^2 | U^* \right] = \left[\frac{n-1}{n-f} - \frac{1-fk}{N-1} \left(1 - \frac{1}{n} \right) \right] N^2 (1-f) \frac{s_y^2}{n}.$$

2.2.1.3. Méthode de Bickel et Freedman (1984) et Chao et Lo (1985)

Bickel et Freedman (1984) et Chao et Lo (1985) proposent de choisir dans un premier temps entre deux pseudo-populations, U_1^* avec une probabilité α_b et U_2^* avec une probabilité $1 - \alpha_b$. Il faut noter que dans l'article de Mashreghi et coll. (2016) la variante proposée par Bickel et Freedman (1984) et celle proposée par Chao et Lo (1985) ont été différenciées selon la probabilité de randomisation α , or il s'agit exactement de la même probabilité qui était exprimée différemment. Nous retenons la probabilité de randomisation présentée par Bickel et Freedman (1984) car elle est plus intuitive et simple à retenir. La pseudo-population U_1^* est constituée en dupliquant $k = \lfloor N/n \rfloor$ fois chaque unité de s alors que U_2^* est constituée en les dupliquant $k + 1$ fois de telle sorte que les tailles des deux pseudo-populations ainsi constituées sont respectivement $N_1 = kn$ et $N_2 = (k + 1)n$. Ensuite on prélève un rééchantillon s^* dans la pseudo-population choisie U_i^* ; $i = 1, 2$ selon un EASSR de taille n . La probabilité de choisir U_1^* (respectivement U_2^*) est

$$\alpha_b = \left(1 - \frac{r}{n} \right) \frac{N_1 - 1}{N - 1}, \left(\text{respectivement } 1 - \alpha_b = \frac{r}{n} \frac{N_2 - 1}{N - 1} \right).$$

Pour la variante proposée par Bickel et Freedman (1984) et Chao et Lo (1985), les deux estimations bootstrap calculées en se basant sur (2.2.2) ou bien (2.2.1) sont équivalentes dans le cas où les paramètres d'intérêt sont la moyenne ou le total. Soit

\bar{y}^* la moyenne de la variable y sur l'échantillon s^* , la variance selon (2.2.2) est

$$\begin{aligned}
\tilde{V}^*(\bar{y}^*) &= E_{u^*} V_{p^*}(\bar{y}^* | U^*) + V_{u^*} E_{p^*}(\bar{y}^* | U^*) \\
&= E_{u^*} V_{p^*}(\bar{y}^* | U^*) + V_{u^*}(\bar{y}) \\
&= E_{u^*} V_{p^*}(\bar{y}^* | U^*) \\
&= E_{u^*} E_{p^*} [(\bar{y}^* - E_{p^*}(\bar{y}^*))^2 | U^*] \\
&= E_{u^*} E_{p^*} [(\bar{y}^* - \bar{y})^2 | U^*] \\
&= E_{u^*} E_{p^*} [(\bar{y}^* - \mu_{y^*})^2 | U^*] = V^*(\bar{y}^*),
\end{aligned}$$

la variance sous (2.2.2) où μ_{y^*} est la moyenne de la pseudo-population U^* qui est égale à \bar{y} pour cette méthode, chose qui n'était pas satisfaite pour les deux variantes proposées par Booth et coll. (1994) et Chao et Lo (1994). Nous obtenons le même résultat pour le total à condition de prendre comme équivalent bootstrap de \hat{t}_{HT} , $\hat{t}_{HT}^* = N\bar{y}^*$ et le total de la pseudo-population U^* , $t^* = N\mu_{y^*}$.

Une estimation bootstrap de la variance de \hat{t}_{HT} est

$$\begin{aligned}
V^*(\hat{t}_{HT}^*) &= E_{u^*} E_{p^*} [(\hat{t}_{HT}^* - t^*)^2 | U^*] \\
&= E_{u^*} V_{p^*}(\hat{t}_{HT}^* | U^*) \\
&= \alpha_b V_{p^*}(\hat{t}_{HT}^* | U_1^*) + (1 - \alpha_b) V_{p^*}(\hat{t}_{HT}^* | U_2^*) \\
&= \frac{n-1}{n-f} N^2 (1-f) \frac{s_y^2}{n}.
\end{aligned}$$

2.2.1.4. Méthode de Sitter (1992a)

Pour la méthode de Sitter (1992a), le nombre de duplications de l'échantillon d'origine s et la taille de l'échantillon bootstrap sont différents de ceux des autres méthodes. Pour cette méthode, la randomisation se fait entre deux paires du nombre

de duplications et de la taille de l'échantillon bootstrap, c'est-à-dire entre deux paires

$$(k_1, n_1) = \left(\left\lceil \frac{N}{n} \left(1 - \frac{1-f}{n} \right) \right\rceil, n-1 \right)$$

et

$$(k_2, n_2) = \left(\left\lfloor \frac{N}{n} \left(1 - \frac{1-f}{n} \right) \right\rfloor, n \right).$$

La randomisation proposée par Sitter (1992a) permet de restituer l'estimateur sans biais habituel de la variance d'une moyenne. On choisit (k_1, n_1) avec une probabilité

$$\alpha_s = \frac{\frac{1-f}{n(n-1)} - a_2}{a_1 - a_2}$$

et (k_2, n_2) avec une probabilité $1 - \alpha_s$ où,

$$a_j = \frac{k_j(1-f_j)}{n_j(N_j-1)},$$

avec $N_j = k_j n$ est la taille de la pseudo population U_j^* constituée en dupliquant l'échantillon s k_j fois et $f_j = n_j/N_j$ est la fraction d'échantillonnage associée au plan de rééchantillonnage p_j^* selon lequel on choisit un rééchantillon s^* de taille n_j dans U_j^* , $j = 1, 2$.

Pour les deux estimations bootstrap calculées en se basant sur (2.2.2) ou bien (2.2.1) sont équivalentes pour l'estimation de la variance de \bar{y} et \hat{t}_{HT} .

$$\begin{aligned} V^*(\hat{t}_{HT}^*) &= E_{u^*} V_{p^*}(\hat{t}_{HT}^* | U^*) \\ &= \alpha_s V_{p_1^*}(\hat{t}_{HT}^* | U_1^*) + (1 - \alpha_s) V_{p_2^*}(\hat{t}_{HT}^* | U_2^*) \\ &= N^2(1-f) \frac{s_y^2}{n}. \end{aligned}$$

Tab. 2.1. Méthodes bootstrap par pseudo-population pour un plan EASSR

Méthode	k	U^{c*}	n'
Booth et coll. (1994)	$\lfloor N/n \rfloor$	EASSR de taille $N - kn$ dans s	n
Chao et Lo (1994)	$\lfloor N/n \rfloor$	EASAR de taille $N - kn$ dans s	n
Bickel et Freedman (1984)	$\lfloor N/n \rfloor$	(*) $\begin{cases} \emptyset, & \text{avec prob. } \alpha_b \\ s & \text{avec prob. } 1 - \alpha_b \end{cases}$	—
Sitter (1992a)	$\lfloor \frac{N}{n} (1 - \frac{1-f}{n}) \rfloor$	Comme (*) avec prob. $(1 - \alpha_s)$	$n - I(U^{c*} = s)$

2.2.1.5. Algorithme général pour les méthodes bootstrap par pseudo-population

Les méthodes bootstrap présentées fonctionnent de la même manière que la méthode de Booth et coll. (1994). La différence apparaît au niveau du plan utilisé pour compléter la pseudo-population.

De ce constat, Mashreghi et coll. (2016) ont proposé un algorithme général (l'algorithme 2.1) pour construire une pseudo-population et pour sélectionner un échantillon bootstrap pour toutes les méthodes existantes. Les éléments en **gras** de l'algorithme doivent être spécifiés pour chaque méthode et seront présentés dans le Tableau 2.1. Ce dernier rapportera le nombre de répétitions k de l'échantillon s , le plan utilisé pour compléter la construction d'une pseudo-population et la taille de l'échantillon bootstrap n' pour les différentes méthodes bootstrap.

Il faut noter que dans l'article de Mashreghi et coll. (2016) la méthode de Sitter (1992a) a été mal décrite en fonction des éléments de l'algorithme proposé dans ledit article. Nous proposons une description de cette méthode en fonction des éléments de cet algorithme.

Algorithme 2.1. Méthode bootstrap par pseudo-population unifiée pour un plan EASSR

- (1) Dupliquer chaque unité de l'échantillon original s , k fois pour constituer, U^f , la partie fixe de la pseudo-population.
- (2) Prélever U^{c*} dans s pour compléter la pseudo-population U^* . En conséquence $U^* = U^f \cup U^{c*}$. Soit θ^* l'équivalent de θ calculé sur la base des données de la pseudo-population U^* ainsi construite.
- (3) Prélever un échantillon s^* dans U^* selon un plan EASSR de taille n' .
- (4) Calculer $\hat{\theta}^*$ l'équivalent bootstrap de $\hat{\theta}$ calculé sur s^* .
- (5) Répéter les étapes 2 à 4 un grand nombre de fois (disons B), pour disposer d'un ensemble de paramètres bootstrap et d'estimations bootstrap

$$\theta_1^*, \dots, \theta_B^* \text{ et } \hat{\theta}_1^*, \dots, \hat{\theta}_B^*.$$

On prend comme estimation de la variance de $\hat{\theta}$:

$$V_B^* = \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_b^* - \theta_b^* \right)^2.$$

2.2.2. Le bootstrap direct

Les méthodes bootstrap appartenant à cette classe partagent l'idée de base du bootstrap d'Efron (1979) c'est-à-dire qu'elles prennent des échantillons bootstrap i.i.d. tirés directement de l'ensemble de données d'origine au lieu de prendre des échantillons bootstrap qui respectent le plan d'échantillonnage d'origine comme les méthodes bootstrap par pseudo-population. En revanche, des ajustements doivent être apportés afin que la variabilité induite par ce type de bootstrap reflète la variabilité du plan d'échantillonnage d'origine.

2.2.2.1. *Le bootstrap avec remise de McCarthy et Snowden (1985)*

La méthode bootstrap proposée par McCarthy et Snowden (1985) est une transposition directe de la méthode classique d'Efron (1979) au cas d'un plan d'échantillonnage sans remise en population finie. Un échantillon bootstrap s^* de taille n' est sélectionné dans s selon un plan d'échantillonnage aléatoire simple **avec remise**. On note par \bar{y}^* la moyenne de la variable y calculée sur l'échantillon bootstrap s^* . En conséquence l'équivalent bootstrap du \hat{t}_{HT} est donné par $\hat{t}_{HT}^* = N\bar{y}^*$. McCarthy et Snowden (1985) montrent que :

$$E_{p^*}(\hat{t}_{HT}^*) = \hat{t}_{HT}$$
$$V_{p^*}(\hat{t}_{HT}^*) = N^2 \left(1 - \frac{1}{n}\right) \frac{s_y^2}{n'}$$

Si $n' = n$, on retrouve la méthode de bootstrap classique du tirage avec remise. Si de plus $f = 1/n$ de telle sorte que $N = n^2$, cette méthode permet de capter la correction de la population finie. Un tel cas est bien sûr rare en pratique.

Pour capter la correction de la population finie, McCarthy et Snowden (1985) suggèrent de choisir $n' = (n - 1)/(1 - f)$. En pratique $(n - 1)/(1 - f)$ est rarement un entier, ils proposent de choisir l'entier le plus proche.

2.2.2.2. *Le bootstrap rééchantillonné de Rao et Wu (1988)*

La méthode du bootstrap rééchantillonné proposée par Rao et Wu (1988), est la plus connue de cette classe. Comme la méthode précédente elle consiste à sélectionner un échantillon bootstrap selon un plan d'échantillonnage aléatoire simple avec remise. Par contre afin de refléter la variabilité du plan d'origine il faut rééchantillonner les valeurs prises par la variable y sur l'échantillon d'origine s avant de procéder au ré-échantillonnage.

Soit n' la taille de l'échantillon bootstrap et $y'_i = \bar{y} + C(y_i - \bar{y})$ est la valeur ré-échelonnée de la i^{me} unité de s , où

$$C = \sqrt{\frac{n'(1-f)}{n-1}}.$$

L'échantillon bootstrap $s^* = \{y_i^*, i = 1 \cdots n'\}$ de taille n' est sélectionné avec remise dans $s' = \{y'_i, i = 1 \cdots n\}$. Dans le cas d'un total, on a

$$\begin{aligned} V_{p^*}(\hat{t}_{HT}^*) &= V_{p^*}\left(\frac{N}{n'} \sum_{j \in s^*} y_j^*\right) \\ &= \frac{N^2}{n'} \frac{1}{n} \sum_{i=1}^n \left(y'_i - n^{-1} \sum_{j=1}^n y'_j\right)^2 \\ &= \frac{N^2 C^2}{nn'} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= N^2(1-f) \frac{s_y^2}{n}. \end{aligned}$$

Rao et Wu (1988) ont montré que le bootstrap ré-échelonné fournit une estimation valide de la variance d'un estimateur qui peut s'écrire comme une fonction lisse de moyennes. Ils ont signalé également qu'un choix inapproprié de n' peut conduire à des valeurs impossibles pour la statistique bootstrappée. En particulier, une statistique strictement positive peut avoir des répliques bootstrap strictement négatives. Pour se prémunir d'une telle possibilité, ils ont proposé de choisir $n' \leq (n-1)/(1-f)$.

2.2.2.3. *Le bootstrap à concordance-miroir de Sitter (1992b)*

Sitter (1992b) a proposé une méthode où l'idée consiste à échantillonner de façon répétée dans l'échantillon s , selon un plan d'échantillonnage aléatoire simple sans remise de taille n'' , et à concaténer les sous-échantillons obtenus jusqu'à obtenir un

rééchantillon s^* de taille $n' = k'n''$, où k' désigne le nombre de sous-échantillonnages réalisés dans s . Soient $s_1^*, \dots, s_{k'}^*$ k' sous-échantillons de taille n'' sélectionnés de façon indépendante dans s selon un plan EASSR. Le rééchantillon s^* est obtenu en concaténant les k' sous-échantillons. Dans le cas où le paramètre d'intérêt est un total, l'estimateur bootstrap obtenu à partir de l'échantillon bootstrap s^* s'écrit ainsi:

$$\hat{t}_{HT}^* = N\bar{y}^* = N/k' \sum_{j=1}^{k'} \bar{y}_j^*,$$

où \bar{y}_j^* est la moyenne de y sur le sous-échantillon s_j^* . On a

$$E^*(\bar{y}_j^*) = \bar{y} \text{ et } V^*(\bar{y}_j^*) = \left(1 - \frac{n''}{n}\right) \frac{s_y^2}{n''},$$

en conséquence

$$E^*(\hat{t}_{HT}^*) = N\bar{y} = \hat{t}_{HT} \text{ et } V^*(\hat{t}_{HT}^*) = \frac{N^2}{k'} \left(1 - \frac{n''}{n}\right) \frac{s_y^2}{n''}.$$

Pour restituer une estimation bootstrap sans biais de $V(\hat{t}_{HT})$ un choix intuitif de k' est $n(1 - f^*)/n''(1 - f)$ où $f^* = n''/n$. Sitter (1992b) propose de choisir $n'' = fn$ pour respecter le plan d'origine en terme de la fraction d'échantillonnage et restituer l'estimateur sans biais du moment d'ordre 3. Or fn et k' sont rarement entiers, dans ce cas il faut passer par une randomisation. La randomisation consiste à choisir entre les quatre paires suivants (n_i'', k_{ij}) ; $i = 1, 2$, $j = 1, 2$ où $n_1'' = \lfloor fn \rfloor$, $n_2'' = \lceil fn \rceil$, $f_i^* = n_i''/n$, $k_{i1} = \lfloor k_i \rfloor$ et $k_{i2} = \lceil k_i \rceil$ avec

$$k_i = \frac{n(1 - f_i^*)}{n_i''(1 - f)}.$$

La randomisation est décrite par la variable aléatoire R comme suit:

$$P(R = (n'', k')) = \begin{cases} \alpha\beta & \text{si } (n'', k') = (n''_1, k_{11}) \\ \alpha(1 - \beta) & \text{si } (n'', k') = (n''_1, k_{12}) \\ (1 - \alpha)\gamma & \text{si } (n'', k') = (n''_2, k_{21}) \\ (1 - \alpha)(1 - \gamma) & \text{si } (n'', k') = (n''_2, k_{22}), \end{cases} \quad (2.2.3)$$

où $\alpha = n''_2 - fn$, $\beta = (k_{11}^{-1} - k_{12}^{-1})/(k_{11}^{-1} - k_{12}^{-1})$ et $\gamma = (k_{21}^{-1} - k_{22}^{-1})/(k_{21}^{-1} - k_{22}^{-1})$. Il faut noter que le choix de n''_1 ou n''_2 peut conduire à une valeur de k_1 ou k_2 inférieure à 1. Ceci peut se produire si $n''_2 > n/2$. Une telle situation ne se produira pas dans notre étude par simulation. Sitter (1992b) a proposé également une façon de procéder plus simple, en choisissant arbitrairement $n'' \leq n/(1 - f)$ et en faisant la randomisation seulement sur k' de la manière suivante; $k' = \lfloor \frac{n(1-f'')}{n''(1-f)} \rfloor + I_q$ où, I_q est une variable aléatoire Bernoulli de paramètre q avec $q = (\lfloor k \rfloor^{-1} - k^{-1})/(\lfloor k \rfloor^{-1} - \lceil k \rceil^{-1})$, $k = n(1 - f'')/(n''(1 - f))$ et $f'' = n''/n$.

2.2.2.4. Algorithme général pour les méthodes bootstrap direct

Mashreghi et coll. (2016) ont présenté un algorithme bootstrap unifié pour ces méthodes. Pour ce faire ils ont identifié trois variantes qui permettent de distinguer une méthode de l'autre en terme de constitution d'un échantillon bootstrap et du calcul de la statistique bootstrap $\hat{\theta}^*$. Pour constituer un échantillon bootstrap la méthode Sitter (1992b) réunit k' sous-échantillons de taille n'' pigés dans s selon un plan aléatoire simple sans remise ainsi la taille n' de s^* est égale à $k'n''$. Pour les méthodes impliquant un seul échantillon i.i.d. de taille n' , il suffit de prendre $n'' = 1$ et $k' = n'$. En d'autres termes, le réglage de $n'' = 1$ dans l'algorithme décrit ci-dessous équivaut à la sélection des échantillons bootstrap avec remise dans s . Pour calculer la statistique bootstrap $\hat{\theta}^*$, la méthode Rao et Wu (1988) utilise les valeurs rééchelonnées, C désigne le facteur de ré-échelonnement. Il faut noter que pour des statistiques comme le coefficient de corrélation (1.1.5), la statistique étant basée sur les totaux estimés de 5 variables, le

Algorithme 2.2. Méthode bootstrap direct unifiée pour un plan EASSR

- 1 Ré-échelonner les valeurs prises par les variables d'intérêt (y_1, \dots, y_J) sur l'échantillon s de cette façon; $\forall j = 1, \dots, J$ et $i \in s$ $y'_{ji} = \bar{y}_j + \mathbf{C}(y_i - \bar{y}_j)$. On construit ainsi l'échantillon des observations ré-échelonnées $s' = \{y'_{ji}, j = 1 \dots J\}_{i=1}^n$.
- 2 Sélectionner un sous-échantillon sans remise de taille n'' dans s' .
- 3 Répéter (2) k' fois indépendamment. Réunir les sous-échantillons obtenus pour constituer un échantillon bootstrap s^* de taille $n' = k'n''$.
- 4 Calculer la statistique bootstrap $\hat{\theta}^*$ en utilisant les observations de s^* .
- 5 Répéter les étapes 2 à 4 un grand nombre de fois (disons B), pour disposer d'un ensemble d'estimateurs bootstrap

$$\hat{\theta}_1^*, \dots, \hat{\theta}_B^*.$$

- 6 Estimer la variance de $\hat{\theta}$ par

$$\hat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^* \right)^2,$$

$$\text{où } \hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*.$$

ré-échelonnement est appliqué à chacune de celles-ci séparément. C'est pourquoi dans l'algorithme 2.2, nous considérons J variables d'intérêt plutôt qu'une seule. Pour les autres méthodes qui utilisent les valeurs de y sur s^* il suffit d'égaliser C à 1.

2.2.3. Le bootstrap pondéré

Comme nous l'avons souligné dans la section 1.1.3 du premier chapitre, un estimateur $\hat{\theta}$ dans un contexte d'échantillonnage s'exprime en fonction des observations et des poids d'échantillonnage, ou d'estimation si de l'information auxiliaire est introduite dans la phase d'estimation. La philosophie du bootstrap pondéré consiste à conserver les observations d'origine et de créer un système de poids aléatoires au lieu de constituer un échantillon

Tab. 2.2. Méthodes bootstrap direct pour un plan EASSR

Méthode	C	(n'', k')	$E_p [V^*(\hat{t}_{HT}^*)] / V(\hat{t}_{HT})$
EFRON	1	$(1, n)$	$\frac{n-1}{n(1-f)}$
McCarthy et Snowden (1985)	1	$(1, \frac{n-1}{1-f}^a)$	1^b
Rao et Wu (1988)	$\sqrt{\frac{n'(1-f)}{n-1}}$	$(1, n')$	1
Sitter (1992b)	1	Selon (2.2.3)	1

^aCe nombre n' est pas nécessairement entier. Si c'est le cas on arrondit à l'entier le plus proche.

^bSeulement si $\frac{n-1}{1-f}$ est un entier.

bootstrap d'observations pour calculer un estimateur bootstrap $\hat{\theta}^*$. Cette manière de procéder est attrayante pour les utilisateurs de données d'enquêtes publiques diffusées par des organismes statistiques tels que Statistique Canada. Ces organismes fournissent des fichiers de micro-données comprenant des colonnes avec les observations originales, une colonne avec les poids d'enquête originaux et les B colonnes de poids bootstrap. En conséquence, les agences n'ont pas besoin de fournir certains détails sur le plan d'échantillonnage utilisé qui pourraient révéler de l'information confidentielle. Nous présentons par la suite la version pondérée de la méthode de Rao et Wu (1988), Rao et coll. (1992), la méthode de Chipperfield et Preston (2007), le bootstrap généralisé appliqué dans un contexte d'échantillonnage de Bertail et Combris (1997) et Beaumont et Patak (2012), et les deux méthodes proposées par Antal et Tillé (2011a) et Antal et Tillé (2014).

Pour une présentation unifiée de ces méthodes, nous écrivons les poids bootstrap sous une forme générale

$$w_i^* = a_i^* w_i, \quad (2.2.4)$$

où les a_i^* sont les ajustements bootstrap des poids d'origine pour refléter la variabilité du plan d'échantillonnage d'origine et $w_i = \pi_i^{-1}$.

2.2.3.1. *Le bootstrap pondéré de Rao et coll. (1992)*

Rao et coll. (1992) ont proposé une version bootstrap pondéré de leur méthode Rao et Wu (1988). Soit s^* un échantillon bootstrap de taille n' tiré dans s' avec remise. On définit les variables indicatrices suivantes

$$I_{ji}^* = \begin{cases} 1 & \text{si l'individu } i \text{ est sélectionné au } j^{\text{ème}} \text{ tirage } j = 1, \dots, n'; i = 1, \dots, n \\ 0 & \text{sinon.} \end{cases}$$

On définit $m_i^* = \sum_{j=1}^{n'} I_{ji}^*$ comme le nombre de fois où l'individu $i \in s$ apparaît dans s^* et par construction des I_{ji}^* on a $\sum_{i=1}^n m_i^* = n'$. L'estimateur bootstrap du total obtenu à partir de la méthode Rao et Wu (1988) s'écrit ainsi:

$$\begin{aligned} \hat{t}_{HT}^* &= \frac{N}{n'} \sum_{j \in s^*} y_j^* \\ &= \frac{N}{n'} \sum_{j \in s^*} \sum_{i \in s} I_{ji}^* (\bar{y} + C(y_i - \bar{y})) \\ &= \frac{N}{n'} \sum_{i \in s} m_i^* (\bar{y} + C(y_i - \bar{y})) \\ &= N\bar{y} + \frac{N}{n'} \sum_{i \in s} m_i^* C y_i - NC\bar{y} \\ &= \sum_{i \in s} \left(1 + C \left(\frac{m_i^* n}{n'} - 1 \right) \right) \frac{N}{n} y_i \\ &= \sum_{i \in s} \left(1 + C \left(\frac{m_i^* n}{n'} - 1 \right) \right) w_i y_i \end{aligned}$$

où $w_i = N/n$ est le poids d'échantillonnage associé à l'individu i et $C = \sqrt{\frac{n'(1-f)}{n-1}}$.

De ce constat Rao et coll. (1992) ont suggéré de conserver les observations originales et de créer des poids bootstrap dont la forme est (2.2.4) avec

$$a_i^* = \left(1 + C \left(\frac{m_i^* n}{n'} - 1 \right) \right).$$

2.2.3.2. Chipperfield et Preston (2007)

La méthode Chipperfield et Preston (2007) a introduit un nouveau jeu de poids bootstrap rééchelonné en terme du plan de rééchantillonnage; un échantillon bootstrap s^* de taille $[n/2]$ est sélectionné sans remise dans s . En conséquence

$$a_i^* = \left(1 + C \left(\frac{m_i^* n}{[n/2]} - 1 \right) \right)$$

où $C = \sqrt{[n/2](1-f)/(n-[n/2])}$ et m_i^* est égal à 1 si l'individu i apparaît dans s^* et 0 sinon.

2.2.3.3. Le bootstrap généralisé de Bertail et Combris (1997) et Beaumont et Patak (2012)

Bertail et Combris (1997) ont proposé d'appliquer dans le cadre d'une population finie la méthode du bootstrap généralisé introduite par Lo (1991) puis généralisée par Mason et Newton (1992). L'idée consiste à affecter directement aux individus des poids aléatoires, sans passer par une procédure de rééchantillonnage d'observations. Les premiers moments du vecteur des ajustements bootstrap $\mathbf{a}^* = (a_1^*, \dots, a_n^*)$ sont choisis de telle sorte que l'espérance et la variance de la statistique bootstrap \hat{t}_{HT}^* sous ce mécanisme aléatoire correspondent à l'estimateur sans biais habituel du total, \hat{t}_{HT} , et à l'estimateur Horvitz-Thompson ou de Sen-Yates-Grundy de la variance de l'estimateur du total, $\hat{V}(\hat{t}_{HT})$. Bertail et Combris (1997) et Beaumont et Patak (2012) indiquent que si n'importe quelle distribution est utilisée pour générer le vecteur \mathbf{a}^* de sorte que:

$$E^*(\mathbf{a}^*) = \mathbf{1}_n, \tag{2.2.5}$$

$$V^*(\mathbf{a}^*) = \mathbf{\Sigma}, \text{ où } \Sigma_{ii} = (1-f) \text{ et } \Sigma_{ij} = -\frac{1-f}{n-1} \forall i \neq j, \tag{2.2.6}$$

alors

$$E^* (\hat{t}_{HT}^*) = E^* \left(\sum_{i \in s} \frac{N}{n} a_i^* y_i \right) = \hat{t}_{HT}$$

et

$$V^* (\hat{t}_{HT}^*) = N^2 (1 - f) \frac{s_y^2}{n}.$$

De nombreux choix pour la distribution du vecteur \mathbf{a}^* peuvent être considérés. Par exemple, Bertail et Combris (1997) ont généré le vecteur \mathbf{a}^* comme

$$\mathbf{a}^* = \mathbf{1}_n + \Sigma^{\frac{1}{2}} \tilde{\mathbf{a}}^*, \quad (2.2.7)$$

où $\tilde{\mathbf{a}}^*$ est un n -vecteur contenant des variables aléatoires mutuellement indépendantes, \tilde{a}_i^* , qui ont toutes une espérance nulle et une variance égale à 1 et $\Sigma^{\frac{1}{2}}$ est obtenu à partir la décomposition spectrale de Σ , i.e., $\Sigma^{\frac{1}{2}} = \Gamma \Lambda^{1/2} \Gamma$, où Λ est la matrice diagonale dont les coefficients diagonaux sont les valeurs propres et Γ est la matrice orthogonale dont les colonnes sont les vecteurs propres. Pour obtenir $\Lambda^{1/2}$ il faut que toutes les valeurs propres de Σ soient positives, ce qui équivaut à vérifier si Γ est semi-définie positive. Un choix simple consiste à générer \tilde{a}_i^* à partir de la distribution normale; le vecteur \mathbf{a}^* suit une distribution normale multivariée $N(\mathbf{1}_n, \Sigma)$. Un autre choix proposé par Beaumont et Patak (2012) consiste à utiliser la distribution

$$\tilde{a}_i^* = \begin{cases} -\epsilon & \text{avec } P(\tilde{a}_i^* = -\epsilon) = 1/(1 + \epsilon^2) \\ 1/\epsilon & \text{avec } P(\tilde{a}_i^* = 1/\epsilon) = \epsilon^2/(1 + \epsilon^2) \end{cases} \quad (2.2.8)$$

où ϵ est une constante strictement positive choisie par le statisticien.

Il faut noter que le choix d'une distribution pour générer les ajustements bootstrap peut conduire à l'occurrence de valeurs négatives de ces ajustements ce que certains spécialistes de l'échantillonnage considèrent être un problème parce que le poids associé à une observation peut être considéré comme le nombre d'unités dans la population représentées par cette unité de l'échantillon. Pour détourner ce problème Beaumont et Patak (2012) ont

suggéré de ré-échelonner les ajustements bootstrap comme suit

$$\check{a}_i^* = \frac{a_i^* + \tau - 1}{\tau}, \quad (2.2.9)$$

où $\tau \geq 1$ est un petit nombre mais assez grand pour que les ajustements bootstrap échelonnés \check{a}_i^* ne soient plus négatifs. Notons que $E^*(\check{a}_i^*) = 1$ et $E^*[(\check{a}_i^* - 1)(\check{a}_j^* - 1)] = \tau^{-2}E^*[(a_i^* - 1)(a_j^* - 1)]$. En conséquence, pour avoir une estimation bootstrap valide il faut multiplier l'estimation bootstrap calculée à partir des ajustements bootstrap échelonnés \check{a}_i^* par τ^2 . Donc, cette valeur doit être fournie à un éventuel utilisateur.

2.2.3.4. Les méthodes d'Antal et Tillé (2011a) et Antal et Tillé (2014)

Antal et Tillé (2011a) et Antal et Tillé (2014) ont proposé deux algorithmes bootstrap dont l'idée consiste à ajuster les poids d'origine par un vecteur $m^* = (m_1^*, \dots, m_n^*) \in \mathbf{N}^n$ avec $\sum_{i=1}^n m_i^* = n$. Ainsi, ces deux méthodes forcent l'ajustement des poids par la multiplication par un entier, ce qui n'est pas le cas de la méthode précédente si la distribution est continue. Ce vecteur m^* doit satisfaire aux conditions (2.2.5) et (2.2.6) pour restituer les estimateurs sans biais habituels des premiers moments dans le cas linéaire. Pour bien comprendre le fonctionnement des deux méthodes bootstrap, nous présentons un nouveau type de plan d'échantillonnage qui porte le nom de "One-One". En effet les deux méthodes ont utilisé deux plans de type "One-One" pour construire une partie du vecteur m^* . Un échantillon \tilde{s} de taille n tiré de s selon un plan de type "One-One" possède les deux propriétés suivantes

$$E_{\tilde{p}}(\tilde{m}_i) = V_{\tilde{p}}(\tilde{m}_i) = 1 \quad \text{et} \quad \sum_{i \in s} \tilde{m}_i = n,$$

où, \tilde{m}_i désigne le nombre de fois où l'unité $i \in s$ apparaît dans l'échantillon \tilde{s} et \tilde{p} désigne un plan de type "One-One".

Pour la méthode de Antal et Tillé (2011a), le rééchantillon s^* est généré selon un mécanisme aléatoire qui mixe deux plans d'échantillonnage: le plan aléatoire simple sans remise et un plan de type "One-One". Ainsi m_i^* sera le nombre de fois où l'unité $i \in s$ apparaîtra dans s^* . En fait, le rééchantillon s^* est constitué en deux étapes. Dans

la première étape, on sélectionne un échantillon s_A^* de s selon un plan d'échantillonnage aléatoire simple sans remise de taille n_A et dans la deuxième étape, parmi les unités non sélectionnées, on sélectionne un échantillon s_B^* avec remise de taille $n_B = n - n_A$ selon un plan d'échantillonnage "One-One". Ce dernier est mis en œuvre en combinant deux plans d'échantillonnage: le plan aléatoire simple avec remise et le plan d'échantillonnage avec sur-remplacement (over-replacement). Pour une description détaillée (incluant la valeur de n_A) voir l'algorithme 2 et l'algorithme 3 dans l'article de Antal et Tillé (2011a).

Dans la même optique que la méthode précédente Antal et Tillé (2014) ont proposé un algorithme bootstrap similaire qui utilise un mécanisme aléatoire qui mixe deux plans, le plan de Bernoulli et un plan de type "One-One". Toutefois, la nouvelle méthode se veut plus allégée en terme de calcul car elle fait appel à un plan de type "One-One" facile à mettre en œuvre. En effet, le ré-échantillon s^* est constitué aussi en deux étapes. Dans la première étape, un échantillon s_A^* est sélectionné dans s selon le plan d'échantillonnage de Bernoulli de paramètre $\pi = n/N$ et de taille observée aléatoire n_A et dans la deuxième étape, si $n - n_A \geq 2$, un échantillon s_B^* de taille $n_B = n - n_A$ est sélectionné dans l'ensemble des unités non pigées à la première étape selon un plan "One-One" qui porte le nom de "doubled half sampling". Ainsi $m_i^* = m_{Ai}^* + (1 - m_{Ai}^*)m_{Bi}^*$, où, m_{Ai}^* est le nombre de fois où l'unité i apparaît dans s_A^* et m_{Bi}^* est le nombre de fois où l'unité i apparaît dans s_B^* . En revanche, si $n - n_A = 1$ c'est-à-dire qu'il existe seulement une unité non sélectionnée selon le premier plan, (disons l'unité l) alors cette unité est dupliquée m_{Bl}^* fois selon la distribution suivante:

$$m_{Bl}^* = \begin{cases} 0 & \text{avec probabilité } 1/4 \\ 1 & \text{avec probabilité } 1/2 \\ 2 & \text{avec probabilité } 1/4. \end{cases}$$

Ensuite parmi les unités sélectionnées selon la première étape (plan de Bernoulli), une unité est sélectionnée avec probabilité égale (disons l'unité k) cette unité sera dupliquée

$m_k^* = 2 - m_{Bl}^*$. On définit

$$m_i^* = \begin{cases} m_{Ai}^* & \text{si } i \neq l, k \\ m_{Bl}^* & \text{si } i = l \\ 2 - m_{Bl}^* & \text{si } i = k. \end{cases}$$

Si $n - n_A = 0$ alors $m^* = m_A^*$.

2.3. Les intervalles de confiance bootstrap

Dans un contexte d'enquête par échantillonnage, l'utilisation du bootstrap pour construire les intervalles de confiance a reçu moins d'attention qu'en statistique classique. Ceci peut être expliqué par le fait que les organismes de statistiques officielles utilisent généralement le coefficient de variation comme mesure de la précision d'une estimation.

Les méthodes bootstrap présentées dans ce chapitre ont des spécificités qui auront des répercussions sur la construction des intervalles de confiance bootstrap. Ces spécificités sont identifiées en répondant, pour chaque méthode bootstrap, à ces deux questions principales:

- (Q1) Est-ce que le paramètre bootstrap θ^* est égal à $\hat{\theta}(s)$ où $\hat{\theta}(s)$ désigne la valeur de $\hat{\theta}$ calculée sur la base de l'échantillon d'observations s ?
- (Q2) Comment estimer la variance V^* dans un t -bootstrap?

La première question a été évoquée par Mashreghi et coll. (2016). Ils ont signalé que seulement pour certaines méthodes bootstrap par pseudo-population, le paramètre bootstrap θ^* est aléatoire et par conséquent différent de $\hat{\theta}(s)$. Le paramètre bootstrap pour ces méthodes correspond au paramètre de la pseudo-population. Dans le cas d'un plan d'échantillonnage aléatoire simple sans remise il s'agit des méthodes de Booth et coll. (1994) et Chao et Lo (1994). De ce constat, ils ont proposé de centrer $\hat{L}_{n;*}$ et $\hat{K}_{n;*}$ par rapport à θ^* au lieu de $\hat{\theta}$ et de calculer les intervalles de confiance bootstrap de base et t -bootstrap en utilisant respectivement les fonctions quantiles bootstrap $\hat{L}_{n;*}^{-1}$ et $\hat{K}_{n;*}^{-1}$. Autrement dit, pour ces méthodes bootstrap $\hat{L}_{n;*}$ et $\hat{K}_{n;*}$ sont définies ainsi

$$\hat{L}_{n;*}(x) = Prob^* \left(\hat{\theta}^* - \theta^* \leq x \right)$$

et

$$\hat{K}_{n,*}(x) = Prob^* \left(\frac{\hat{\theta}^* - \theta^*}{\sqrt{\hat{V}^*}} \leq x \right).$$

Et pour l'intervalle de confiance percentile il faut le calculer à partir la formule (1.2.10) au lieu (1.2.9) qui ne sont pas équivalentes pour ces méthodes bootstrap.

Mashreghi et coll. (2016) ont signalé que ce cas de figure ne se produit que pour certaines méthodes bootstrap par pseudo-population. Ainsi, pour les autres méthodes on aurait toujours $\theta^* = \hat{\theta}(s)$. Or, ceci n'est pas vrai également pour la méthode Rao et Wu (1988) dans certains cas. En effet cette méthode est différente des autres parce qu'elle manipule les données de l'échantillon des observations ré-échelonnées et non pas celles de l'échantillon des observations d'origine. Pour bien illustrer le problème lié à cette méthode nous prenons l'exemple de l'échantillon d'observations suivant $s = ((y_i, z_i), i = 1 \dots 10)$, pour chaque unité échantillonnée on mesure deux variables d'intérêt y et z .

i	1	2	3	4	5	6	7	8	9	10
y_i	0,39	2,09	0,24	0,48	1,64	0,57	0,23	0,00	0,88	0,25
z_i	1,35	3,39	1,65	1,76	2,91	2,66	1,66	2,70	2,04	2,74

On suppose que la fraction de sondage $f = 0,4$ et on choisit, n' la taille des échantillons bootstrap, de cette manière $n' = n - 3$. Donc l'échantillon des observations rééchelonnées, $s' = ((y'_i, z'_i), i = 1 \dots 10)$ qui sera l'objet du ré-échantillonnage dans ce cas est le suivant

i	1	2	3	4	5	6	7	8	9	10
y'_i	0,48	1,64	0,38	0,54	1,33	0,60	0,37	0,21	0,82	0,39
z'_i	1,65	3,04	1,85	1,93	2,71	2,54	1,86	2,57	2,12	2,60.

Le tableau suivant rapporte les valeurs de $\hat{\theta}$ l'estimateur de θ , le paramètre d'intérêt. Ces valeurs sont calculées sur la base de l'échantillon des observations d'origine s et l'échantillon des observations ré-échelonnées s' .

θ	$\hat{\theta}(s)$	$\hat{\theta}(s')$
μ_y	0,677	0,677
μ_z	2,286	2,286
$r_{y,z}$	0,296	0,296
$\rho_{y,z}$	0,611	0,611
m_y	0,435	0,510
G_y	0,486	0,332.

Nous voulons montrer à travers cet exemple que dans le cas des paramètres linéaires et lisses de totaux, nous avons toujours $\hat{\theta}(s) = \hat{\theta}(s')$. Par contre dans le cas des paramètres non lisses comme la médiane et l'indice de Gini cette égalité n'est pas satisfaite. Donc, dans le cas de la méthode bootstrap de Rao et Wu (1988), le paramètre bootstrap θ^* est égal à $\hat{\theta}(s')$ et non pas à $\hat{\theta}(s)$. Ainsi, les intervalles de confiance bootstrap de base et t -bootstrap pour cette méthode doivent être calculés en centrant $\hat{L}_{n;*}$ et $\hat{K}_{n;*}$ par rapport à $\hat{\theta}(s')$ au lieu de $\hat{\theta}(s)$. En conséquence, l'intervalle de confiance percentile pour cette méthode doit être calculé également selon la formule (1.2.10) au lieu (1.2.9). Il faut noter que la méthode de Rao et coll. (1992) donne exactement les mêmes résultats que la méthode de Rao et Wu (1988) dans le cas des paramètres linéaires et lisses de totaux. Toutefois, dans le cas des paramètres non lisses comme la médiane et l'indice de Gini, les deux méthodes sont différentes. En effet, l'estimateur bootstrap de la médiane dans le cas de la méthode Rao et Wu (1988) est la médiane de l'échantillon bootstrap des observations ré-échelonnées, or pour la méthode de Rao et coll. (1992) il s'agit de la médiane pondérée des observations d'origine. Donc, ce problème ne se pose pas pour la méthode de Rao et coll. (1992).

L'autre grande question est comment estimer la variance V^* dans un t -bootstrap. Ce problème a été soulevé par Chauvet (2007) en avançant que l'absence d'un principe de plug-in pour certaines méthodes bootstrap rend l'estimation de la variance V^* délicate. En fait, l'estimation de la variance V^* ne présente aucune difficulté pour les méthodes bootstrap par pseudo-population car ces méthodes ont l'avantage d'obéir comme le bootstrap d'Efron à un véritable principe de plug-in comme nous l'avons mentionné lors de leur introduction. En

fait, pour les méthodes bootstrap par pseudo-population, aucun ajustement n'est nécessaire pour réaliser un second bootstrap. Il suffit, de créer une pseudo-population à partir de l'échantillon bootstrap s^* et de mettre en œuvre le plan d'échantillonnage d'origine.

Toutefois, pour les méthodes qui sont motivées par un ajustement sur les estimateurs naturels des premiers moments dans le cas linéaire, l'estimation de la variance V^* est compliquée. Il s'agit des méthodes bootstrap direct et bootstrap pondéré. En effet, la complication réside d'une part, sur la difficulté de l'identification de l'échantillon bootstrap qui fera l'objet du ré-échantillonnage, et d'autre part, sur la façon d'appliquer l'algorithme Monte Carlo à cet échantillon bootstrap pour calculer une estimation de la variance V^* . Autrement dit, est-ce qu'on va garder les mêmes paramètres de réglage ou bien un nouveau réglage est-il nécessaire pour avoir une estimation valide dans le cas linéaire. En effet, on souhaite avoir dans le cas d'une moyenne μ_y

$$E(V(\bar{y}^{**} | s^*) | s) = (1 - f) \frac{s_y^2}{n},$$

où \bar{y}^{**} est la moyenne des observations d'un échantillon tiré de l'échantillon bootstrap s^* . Le tableau 2.3 donne pour les méthodes bootstrap direct et pondéré, une description de l'étape de ré-échantillonnage à réaliser dans s pour estimer la variance et celle à réaliser dans s^* pour calculer un intervalle de confiance t -bootstrap. Cette description en parallèle permettra de bien repérer les similitudes et les différences entre le premier et le second ré-échantillonnage.

Tab. 2.3. Description du premier et second ré-échantillonnage pour chaque méthode bootstrap direct ou pondéré pour calculer un intervalle de confiance t -bootstrap

Méthode	Premier ré-échantillonnage à réaliser dans s	Second ré-échantillonnage à réaliser dans s^*
McCarthy et Snowden (1985)	Consiste à sélectionner dans s avec remise un échantillon bootstrap s^* de taille $n' = (n - 1)/(1 - f)$.	Consiste à sélectionner dans s^* avec remise un échantillon bootstrap de taille $n' - 1$.
Rao et Wu (1988)	Consiste à ré-échelonner les observations de la variable y sur s de la manière suivante $\forall i \in s, y'_i = \bar{y} + C(y_i - \bar{y})$ où, $C = (\frac{n'(1-f)}{n-1})^{1/2}$ et n' est choisi arbitrairement. Puis on sélectionne dans l'échantillon des observations ré-échelonnées $s' = (y'_i, i \in s)$ avec remise un échantillon bootstrap s^* de taille n' .	On ré-échelonne les observations de s^* de la même manière que le premier ré-échantillonnage mais en utilisant le facteur de ré-échelonnement $C = \sqrt{n''/(n' - 1)}$, le choix de n'' est arbitraire. Puis, on sélectionne dans l'échantillon bootstrap des observations ré-échelonnées, avec remise, un ré-échantillon de taille n'' . Si on choisit $n'' = n' - 1$ alors $C = 1$, ceci veut dire qu'on n'a pas besoin de ré-échelonner de nouveau les observations de s^* .
Sitter (1992b)	Consiste à sélectionner dans s , sans remise, $k' = \lfloor \frac{n(1-f'')}{n''(1-f)} \rfloor + I_q^a$, sous-échantillons de taille arbitraire, $n'' \leq \frac{n}{2-f}$ et de les concaténer pour constituer un échantillon bootstrap s^* . La taille de l'échantillon bootstrap $n' = k'n''$.	On sélectionne dans s^* sans remise $k'' = \lfloor \frac{n(1-f''')}{n'''(1-f)} \rfloor + I_{q'}^b$, sous-échantillons de taille arbitraire $n''' \leq \frac{nn'}{n'(1-f)+n}$ puis on les concatène.

^a I_q Bernoulli(q) avec $q = (\lfloor k \rfloor^{-1} - k^{-1})/(\lfloor k \rfloor^{-1} - \lceil k \rceil^{-1})$, $k = n(1 - f'')/(n''(1 - f))$ et $f'' = n''/n$.

Méthode	Premier ré-échantillonnage à réaliser dans s	Second ré-échantillonnage à réaliser dans s^*
Rao et coll. (1992)	Identique à Rao et Wu (1988) en terme du plan de ré-échantillonnage. En conséquence l'identification de l'échantillon bootstrap sera identique à Rao et Wu (1988). Autrement dit l'échantillon bootstrap est obtenu en échantillonnant dans l'échantillon des observations ré-échelonnées s' avec remise un échantillon s^* de taille n' .	Identique à Rao et Wu (1988) On manipule les observations de s^* de la même façon que Rao et Wu (1988) dans le second bootstrap.
Chipperfield et Preston (2007)	Similaire à Rao et Wu (1988). La différence apparait au niveau du facteur ré-échelonnement $C = \left(\frac{\lfloor n/2 \rfloor (1-f)}{n - \lfloor n/2 \rfloor}\right)^{1/2}$ et le plan de ré-échantillonnage au lieu d'échantillonner avec remise on échantillonne sans remise dans s' l'échantillon des observations ré-échelonnées, un échantillon s^* de taille $n' = \lfloor n/2 \rfloor$.	Similaire à Rao et Wu (1988). On ré-échelonne les observation de s^* en utilisant le facteur $C = \sqrt{\lfloor n'/2 \rfloor / n'}$, puis on échantillonne sans remise un échantillon de taille $\lfloor n'/2 \rfloor$.
Beaumont et Patak (2012)	Consiste à générer un système de poids bootstrap en utilisant des lois de probabilités. L'absence d'un mécanisme de ré-échantillonnage ne permet pas d'identifier un échantillon bootstrap.	On ne peut pas identifier l'échantillon bootstrap.

Méthode	Premier ré-échantillonnage à réaliser dans s	Second ré-échantillonnage à réaliser dans s^*
Antal et Tillé (2011a)	On échantillonne dans s un échantillon bootstrap s^* de taille n selon un plan de ré-échantillonnage qui mixe deux plans, le plan aléatoire simple sans remise et un plan de type "One-One".	On échantillonne de la même façon que le premier ré-échantillonnage mais dans s^* cette fois-ci.
Antal et Tillé (2014)	Similaire à Antal et Tillé (2011a), la seule différence est le plan d'échantillonnage utilisé. Le plan utilisé cette fois-ci mixe le plan de Bernoulli et un autre plan de type "One-One"	Similaire à Antal et Tillé (2011a).

Chapitre 3

Bootstrap pour un plan d'échantillonnage à probabilités inégales sans remise

Dans ce chapitre, nous étudions les méthodes bootstrap pour un plan d'échantillonnage à probabilités inégales sans remise. Un plan est dit à probabilités inégales s'il existe au moins deux unités de la population qui ont des probabilités d'inclusion différentes. Ce type de plan est souvent utilisé lorsqu'on dispose sur la base de sondage d'une variable auxiliaire corrélée positivement avec la variable d'intérêt. En conséquence il serait plus efficace d'échantillonner les unités à probabilités proportionnelles à cette variable auxiliaire, plutôt que d'échantillonner à probabilités égales.

Pour imiter la variabilité d'un plan d'échantillonnage à probabilités inégales il y a deux approches qui ont guidé les auteurs des méthodes bootstrap que nous présentons dans ce chapitre. La première approche est une extension de l'approche bootstrap par pseudo-population que nous avons décrit dans le deuxième chapitre et plus précisément la variante de Booth et coll. (1994). La seconde approche est une adaptation du bootstrap pondéré au cas d'un échantillonnage en population finie. Nous rappelons que l'idée du bootstrap pondéré consiste à générer un système de poids bootstrap selon un mécanisme aléatoire de telle sorte que les premiers moments de la statistique bootstrap sous ce mécanisme aléatoire restituent les estimateurs habituels des premiers moments dans le cas linéaire sous le plan d'échantillonnage. En conséquence pour se caler sur la variance de Horvitz

Thompson (1.2.2) ou bien celle de Sen-Yates-Grundy (1.2.4) il faut connaître les probabilités d'inclusion d'ordre deux. Or en dehors du plan de Poisson, du plan de Poisson conditionnel et de quelques autres plans (par exemple Rao-Sampford), elles sont difficiles à calculer. Par conséquent les méthodes bootstrap pondéré pour les plans dont on ne peut pas calculer exactement les probabilités d'inclusion d'ordre deux vont se caler sur des formules simplifiées de la variance de l'estimateur de Horvitz Thompson d'un total.

Ce chapitre est organisé en trois sections. En section 1, nous présentons quelques rappels sur l'échantillonnage à probabilités inégales, nous introduisons la notion de l'entropie dans un contexte d'échantillonnage et nous donnons quelques formules de variance simplifiées. En section 2, nous présentons et discutons les différentes méthodes bootstrap par pseudo-population pour un plan d'échantillonnage à probabilités inégales. Ensuite, à la section 3, nous présentons les méthodes bootstrap pondéré pour le plan de Poisson et pour les plans de grande entropie.

3.1. Rappels sur l'échantillonnage à probabilités inégales

En général les probabilités d'inclusion sont définies proportionnellement à une variable auxiliaire disponible pour toute la population. Soit x une variable auxiliaire (souvent appelée variable de taille) disponible sur toute la population. Un plan d'échantillonnage est dit proportionnel à la taille si les probabilités d'inclusion π_i sont définies proportionnellement à x

$$\pi_i = n \frac{x_i}{t_x}, \forall i \in U, \quad (3.1.1)$$

où $t_x = \sum_{i \in U} x_i$ et n désigne la taille espérée (la taille moyenne) de l'échantillon aléatoire. Quand les π_i sont définis par (3.1.1), il est possible que certains π_i soient supérieurs à un. Dans ce cas, nous sélectionnons avec probabilité 1 les unités pour lesquelles $\pi_i > 1$. Ensuite, nous recalculons les probabilités d'inclusion comme dans (3.1.1) pour les unités non sélectionnées, où n et t_x vont changer pour tenir compte des unités sélectionnées. Encore une fois, certains π_i peuvent être supérieurs à un, dans ce cas on choisit avec probabilité 1 les unités pour lesquelles les nouveaux $\pi_i > 1$ et nous continuons ce processus itératif jusqu'à

ce que tous les π_i soient inférieurs ou égaux à 1. Donc quelques probabilités d'inclusion π_i seront égales à 1 et d'autres proportionnelles à x_i . On suppose donc sans perte de généralité que $0 < \pi_i < 1, \forall i \in U$. Il existe de nombreuses procédures d'échantillonnage à probabilités proportionnelles à la taille permettant de respecter (ou respecter approximativement) les probabilités d'inclusion fixées a priori. Nous présentons quelques procédures de tirage que nous utiliserons dans notre étude par simulation. Pour un inventaire exhaustif de ces procédures, voir Hanif et Brewer (1980) et Tillé (2006).

3.1.1. Plan de Poisson

Le tirage poissonien consiste à échantillonner les individus de U indépendamment les uns des autres. Soient $\pi_1, \pi_2, \dots, \pi_N$ les probabilités d'inclusion d'ordre un qui peuvent être définies selon (3.1.1). Pour chaque unité i de la population, on effectue une expérience de Bernoulli avec probabilité π_i . En conséquent la probabilité de tirer l'échantillon s est

$$p(s) = \prod_{i \in s} \pi_i \prod_{i \in U \setminus s} 1 - \pi_i.$$

Du fait de l'indépendance entre les tirages, les probabilités d'inclusion d'ordre 2 peuvent être calculées exactement :

$$\pi_{ij} = \pi_i \pi_j \quad \forall i \neq j \in U.$$

La taille n_s de l'échantillon s est aléatoire. Son espérance est

$$E_p(n_s) = \sum_{i \in U} \pi_i.$$

Cette espérance est égale à n si les probabilités π_i ont été choisies selon (3.1.1). Sa variance est

$$V_p(n_s) = \sum_{i \in U} \pi_i(1 - \pi_i).$$

La variance de \hat{t}_{HT} s'obtient à l'aide de la formule de variance de Horvitz-Thompson (1.2.6)

$$V_p(\hat{t}_{HT}) = \sum_{i \in U} \left(\frac{y_i}{\pi_i} \right)^2 \pi_i(1 - \pi_i).$$

Cette variance apparaît comme un simple total et peut être estimée sans biais à l'aide d'un estimateur de Horvitz Thompson

$$\hat{V}(\hat{t}_{HT}) = \sum_{i \in s} \left(\frac{y_i}{\pi_i} \right)^2 (1 - \pi_i).$$

Malgré sa simplicité cette variance reste élevée en raison de l'indépendance entre les tirages des unités et la variabilité de la taille de l'échantillon.

3.1.2. Plan de Poisson conditionnel

Comme nous l'avons mentionné, l'indépendance dans le tirage poissonien des différentes unités engendre une perte d'efficacité. Une solution consiste à rechercher le plan qui maximise l'entropie parmi les plans à taille fixe pour des probabilités d'inclusion données, où l'entropie d'un plan d'échantillonnage p est définie par

$$L(p) = \sum_{s \subset U} p(s) \ln(p(s)) \text{ avec } 0 \ln(0) = 0.$$

On peut montrer que ce plan correspond à un plan de Poisson avec une taille d'échantillon fixée n d'où le nom de plan de Poisson conditionnel. Une façon de l'obtenir consiste à sélectionner un échantillon selon un plan de Poisson de probabilités d'inclusion p_i , $i \in U$ avec $\sum_{i \in U} p_i = n$ et en rejetant l'échantillon tant qu'il n'est pas de taille n . Les probabilités d'inclusion d'ordre un finales, π_i , ne sont pas égales aux p_i . Les π_i ainsi que les π_{ij} sont calculées de manière itérative à partir des p_i . Notons que grâce à un algorithme développé par Deville (2000), il est possible de fixer à l'avance les π_i et de déterminer les p_i à utiliser à l'étape d'échantillonnage de Poisson. Si les π_i sont calculées selon (3.1.1), le plan de Poisson conditionnel permet un échantillonnage proportionnel à la taille sans remise de taille fixe. Le plan de Poisson conditionnel est par construction à taille fixe. Donc l'estimateur de Sen-Yates-Grundy de la variance de \hat{t}_{HT} est aussi sans biais. La formule d'estimation de variance (1.2.3) peut s'écrire comme

$$\hat{V}_{SYG} = \sum_{i \in s} \sum_{j \in s} \frac{y_i y_j}{\pi_i \pi_j} D_{ij}, \quad (3.1.2)$$

où

$$D_{ij} = \begin{cases} - \sum_{\substack{k \in s \\ k \neq i}} \frac{\Delta_{ik}}{\pi_{ik}} & \text{si } i = j \\ \frac{\Delta_{ij}}{\pi_{ij}} & \text{si } i \neq j, \end{cases}$$

et $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$ pour $i, j \in s$.

3.1.3. Formules simplifiées d'estimation de variance pour les plans à forte entropie

Les deux formules d'estimation de variance (1.2.1) et (1.2.3), que nous avons présentées dans le premier chapitre présentent deux défis principaux. En effet, pour les mettre en oeuvre, il faut calculer les probabilités d'inclusion d'ordre deux, chose qui peut être complexe à réaliser. Le second défi est l'évaluation d'une double somme, qui peut être problématique lorsque la taille de l'échantillon est grande. On présente donc quelques formules simplifiées d'estimation de la variance valides pour les plans à forte entropie, comme le plan de Poisson conditionnel.

Plusieurs estimateurs de variance simplifiés ont été proposés dans la littérature. Ils ont tous la même forme que l'estimateur de Sen-Yates-Grundy de la variance (3.1.2) avec des valeurs différentes pour D_{ij} . Autrement dit les D_{ij} sont substitués par \tilde{D}_{ij} (que nous présentons sous peu) dans la formule (3.1.2). Matei et Tillé (2005) ont discuté les performances d'une famille d'estimateurs simplifiés de variance basée sur les approximations de D_{ij} suivantes:

$$\tilde{D}_{ij} = \begin{cases} c_i - \frac{c_i^2}{\sum_{k \in s} c_k} & \text{si } i = j \\ \frac{-c_i c_j}{\sum_{k \in s} c_k} & \text{si } i \neq j, \end{cases}$$

où c_i sont des valeurs proposées par les différents auteurs qui ont abordé le sujet. Dans leur étude par simulations, Matei et Tillé (2005) ont signalé que le choix de c_i proposé par Hájek (1981) produit un estimateur très efficace et légèrement biaisé :

$$c_i = \frac{n}{n-1}(1 - \pi_i). \quad (3.1.3)$$

Deville et Tillé (2005) ont proposé de choisir c_i telle sorte que

$$c_i - \frac{c_i^2}{\sum_{k \in s} c_k} = 1 - \pi_i. \quad (3.1.4)$$

Dans ce cas on a $\tilde{D}_{ii} = 1 - \pi_i$. Un autre choix possible avancé par Antal et Tillé (2011a) consiste à résoudre un système d'équations non linéaires

$$c_i - \frac{c_i^2}{\sum_{k \in s} c_k} = - \sum_{\substack{k \in s \\ k \neq i}} \frac{\Delta_{ik}}{\pi_{ik}}. \quad (3.1.5)$$

Dans ce cas on a $\tilde{D}_{ii} = D_{ii}$.

3.2. Méthodes bootstrap par pseudo-population

3.2.1. Holmberg (1998)

Holmberg (1998) a généralisé la méthode de Booth et coll. (1994) dans le cas d'un plan d'échantillonnage avec probabilités proportionnelles à la taille, p , en utilisant l'inverse des probabilités d'inclusion $\pi_i = nx_i/t_x$ pour construire U^* . En fait, Holmberg (1998) propose de construire U^* en dupliquant chaque unité i de s , $d_i = \lfloor \pi_i^{-1} \rfloor + \epsilon_i$ fois, où les ϵ_i sont les réalisations de n épreuves de Bernoulli indépendants dont la probabilité de succès de chacun d'eux est $r_i = \pi_i^{-1} - \lfloor \pi_i^{-1} \rfloor$. Pour estimer la variance de $V_p(\hat{\theta})$, Holmberg (1998) a utilisé l'algorithme suivant:

- (1) Construire $U^* = \{1^*, \dots, i^*, \dots, N^*\}$ en dupliquant chaque unité i de s , d_i fois incluant la variable de taille x_i . Ainsi, la taille de la pseudo-population U^* et le total de la variable auxiliaire sur celle-ci sont donnés par $N^* = \sum_{i \in s} d_i$ et $t_x^* = \sum_{i \in s} d_i x_i$,
- (2) Sélectionner un échantillon bootstrap s^* en utilisant le même plan d'échantillonnage p avec les probabilités d'inclusion $\pi_{i^*} = nx_{i^*}/t_x^*$,
- (3) Calculer l'estimateur bootstrap $\hat{\theta}^*$ sur s^* ,

- (4) Répéter les étapes 1 à 3 un grand nombre de fois (disons B), pour disposer d'un ensemble d'estimateurs bootstrap $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$. On prend comme estimateur de la variance de $\hat{\theta}$:

$$\hat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^* \right)^2,$$

$$\text{où } \hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*.$$

Holmberg (1998) a ignoré l'impact de la variabilité due à une partie aléatoire de U^* sur l'estimation de $V_p(\hat{\theta})$. En effet, et comme nous avons mentionné dans le chapitre 2, lors de l'introduction de la méthode de Booth et coll. (1994), l'estimation bootstrap de la variance obtenue par l'algorithme proposé est une approximation Monte Carlo de $\tilde{V}^*(\hat{\theta}^*) = E_{u^*p^*} \left[\hat{\theta}^* - E_{u^*p^*}(\hat{\theta}^*) \right]^2$. En conséquence une variance parasite s'ajoute de plus à la variabilité du plan de rééchantillonnage p^* .

3.2.2. Chauvet (2007)

Chauvet (2007) a proposé un algorithme bootstrap général pour un plan d'échantillonnage à probabilités inégales p . L'algorithme bootstrap proposé peut sembler compliqué en terme de calculs puisqu'il repose sur l'obtention de plusieurs échantillons bootstrap pour chacune de plusieurs pseudo-populations. En revanche il permet de supprimer la variance parasite générée lors de l'étape de la construction de la pseudo-population. En effet, Chauvet (2007) a substitué $V_p(\hat{\theta})$ par $E_{u^*} V_{p^*}(\hat{\theta}^*)$ et l'algorithme Monte Carlo proposé qui permet d'approximer ce dernier consiste à :

- (1) Construire $U^* = \{1^*, \dots, i^*, \dots, N^*\}$ en dupliquant chaque unité i de s , $\lfloor \pi_i^{-1} \rfloor$ fois et de compléter les unités ainsi obtenues d'un échantillon, sélectionné dans s selon le plan d'échantillonnage d'origine p avec des probabilités d'inclusion $r_i = \pi_i^{-1} - \lfloor \pi_i^{-1} \rfloor$.
- (2) Sélectionner un échantillon bootstrap s^* en utilisant le même algorithme de tirage p et avec les mêmes probabilités d'inclusion π_{i^*} ,
- (3) Calculer l'estimateur bootstrap $\hat{\theta}^*$ sur s^* ,

- (4) Répéter les étapes 2 à 3 un grand nombre de fois (disons B), pour disposer d'un ensemble d'estimateurs bootstrap $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$. Soit

$$\hat{V}_B^* = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta}_{(\cdot)}^* \right)^2,$$

où $\hat{\theta}_{(\cdot)}^* = B^{-1} \sum_{b=1}^B \hat{\theta}_b^*$.

- (5) Répéter les étapes 1 à 4 un grand nombre de fois (disons D), pour disposer d'un ensemble d'estimateurs bootstrap de variance $\hat{V}_{1B}^*, \dots, \hat{V}_{dB}^*$. On prend comme estimation de la variance de $\hat{\theta}$:

$$\frac{1}{D} \sum_{b=1}^D \hat{V}_{dB}^*.$$

Il faut noter que si les probabilités d'inclusion sont calculées proportionnellement à une variable de taille, Chauvet (2007) propose de recalculer la probabilité d'inclusion de chaque unité de U^* de la même manière que nous avons indiqué dans l'algorithme de Holmberg (1998) c'est-à-dire $\pi_{i^*} = nx_{i^*}/t_x^*$.

Le principe de plug-in utilisé par Chauvet (2007) permet de restituer l'estimateur habituel de la variance d'un estimateur de Horvitz Thompson dans le cas d'un plan de Poisson.

Chauvet (2007) a montré que son algorithme est consistant dans le cas d'estimateur de Horvitz-Thompson et l'estimateur par substitution pour les plans de grande entropie comme le plan de Poisson et le plan de Poisson conditionnel ou ceux proches de l'entropie maximale comme le plan Rao-Sampford et le plan systématique ordonné. Chauvet (2007) a également proposé un algorithme simplifié qui consiste à créer une seule U^* en dupliquant chaque unité i de $s \lceil 1/\pi_i \rceil$ fois, où $\lceil \cdot \rceil$ désigne l'entier le plus proche. Le reste de l'algorithme reste inchangé. Il faut noter que Mashreghi et coll. (2016) ont proposé un principe de plug-in qui consiste à substituer $V_p(\hat{\theta})$ par $E_{u^*} E_{p^*} (\hat{\theta}^* - \theta^*)^2$, où θ^* est le paramètre bootstrap de U^* . Ce principe de plug-in permet de supprimer la variance parasite générée par la construction de la pseudo-population et l'algorithme Monte Carlo associé à celui-ci n'implique pas le double bootstrap comme dans le cas de Chauvet (2007). Autrement dit, il suffit de répéter les trois premières étapes de l'algorithme général de Chauvet (2007), un grand nombre de

fois (disons B) et de prendre comme estimation bootstrap

$$\hat{V}_B^* = \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_b^* - \theta^* \right)^2. \quad (3.2.1)$$

Mashreghi et coll. (2016) ont omis de mentionner que cette approche ne fonctionne que dans le cas où l'estimateur est sans biais. Dans le cas contraire, cette approche devient un estimateur de l'erreur quadratique moyenne. Pour estimer la variance dans un tel cas, il faut procéder par l'approche décrite précédemment. Puisque cette méthode repose sur la moyenne d'estimateurs de variance conditionnelle, il est possible d'utiliser la même quantité de calculs en prenant par exemple $D = 10$ pseudo-populations à partir desquels $B = 100$ échantillons bootstrap chacun seront tirés plutôt que de créer 1000 pseudo-populations à partir desquels un seul échantillon bootstrap sera sélectionné.

3.2.3. Barbiero et Mecatti (2010)

Dans la même optique que l'algorithme simplifié de Chauvet (2007), Barbiero et Mecatti (2010) ont proposé trois algorithmes simplifiés dans le cas d'un plan d'échantillonnage proportionnel à la taille. Le premier algorithme est identique à la version simplifiée de Chauvet (2007). Les deux autres algorithmes proposés consistent à construire une seule pseudo-population équilibrée par rapport à la variable auxiliaire x utilisée pour calculer les probabilités d'inclusion relatives au plan d'échantillonnage d'origine. Autrement dit dans les deux algorithmes l'unique pseudo-population est construite de telle sorte que le total de la variable auxiliaire x sur U^* , t_x^* , soit le plus proche possible de celui sur U , t_x , qui doit être connu. Pour y arriver la pseudo-population est obtenue d'une façon itérative. La pseudo-population initiale est constituée en dupliquant chaque unité de s , $\lfloor \pi_i^{-1} \rfloor$ fois et à chaque itération k on ajoute l'unité k de l'échantillon s ordonné par ordre décroissant selon la variables $r = \pi^{-1} - \lfloor \pi^{-1} \rfloor$ ou bien selon la variable $q = \pi^{-1} / (\lfloor \pi^{-1} \rfloor + 1)$. Après, on évalue si le critère suivant, $|t_{k,x}^* - t_x| < |t_{k-1,x}^* - t_x|$ est satisfait pour itérer la même procédure sur la pseudo-population constituée lors de l'itération k . Sinon la pseudo-population équilibrée est

celle construite à l'itération k . L'étape de ré-échantillonnage et l'estimation de la variance des trois algorithmes sont similaires à celles de l'algorithme de Holmberg (1998).

3.3. Méthodes bootstrap pondéré

3.3.1. Bertail et Combris (1997) et Beaumont et Patak (2012)

Comme nous avons mentionné au début de ce chapitre. Il serait possible, si les probabilités d'inclusion d'ordre deux étaient calculables, de générer un système de poids bootstrap ($w_i^* = w_i a_i^*, i \in s$) selon un mécanisme aléatoire de telle sorte que les premiers moments sous ce mécanisme restituent les estimateurs habituels des premiers moments dans le cas linéaire sous le plan. Ce qui revient dans le cas d'un plan de Poisson où la probabilité d'inclusion de deux unités distinctes est égale au produit de leurs probabilités d'inclusion d'ordre un, de satisfaire les trois conditions suivantes:

- (1) $E^*(a_i^*|s) = 1$,
- (2) $V^*(a_i^*|s) = 1 - \pi_i$,
- (3) $i \neq j$ alors $Cov(a_i^*, a_j^*|s) = 0$.

Bertail et Combris (1997) et Beaumont et Patak (2012) ont expérimenté plusieurs distributions aléatoires pour générer le vecteur \mathbf{a}^* . En effet, ce dernier est généré comme $\mathbf{a}^* = \mathbf{1}_n + \boldsymbol{\Sigma}^{1/2} \tilde{\mathbf{a}}^*$, où $\tilde{\mathbf{a}}^*$ est un n -vecteur contenant des variables aléatoires mutuellement indépendantes, \tilde{a}_i^* , qui ont toutes une espérance nulle et une variance égale à 1 et $\boldsymbol{\Sigma}^{1/2}$ est une matrice diagonale dont les éléments sont $(\sqrt{1 - \pi_1}, \dots, \sqrt{1 - \pi_n})$.

3.3.2. Antal et Tillé (2011a) et Antal et Tillé (2014)

Antal et Tillé (2011a) et Antal et Tillé (2014) ont proposé deux méthodes bootstrap pour un plan de Poisson. Les deux méthodes font appel à des lois de probabilités discrètes pour générer le vecteur aléatoire \mathbf{m}^* . Pour la méthode de Antal et Tillé (2011a) le vecteur \mathbf{m}^* est généré comme suit

$$\mathbf{m}^* = \mathbf{m}_1^* + (\mathbf{1} - \mathbf{m}_1^*) \cdot \mathbf{m}_2^*,$$

où \cdot désigne le produit de Hadamard, \mathbf{m}_1^* est un vecteur composé de variables aléatoires indépendantes de Bernoulli, m_{1i}^* , $i \in s$, de paramètres π_i et \mathbf{m}_2^* est un vecteur composé de variables aléatoires indépendantes de Poisson, m_{2i}^* , $i \in s$, de paramètre 1.

Pour la méthode Antal et Tillé (2014) le vecteur \mathbf{m}^* est généré comme suit

$$\mathbf{m}^* = \mathbf{m}_1^* + 2(\mathbf{1} - \mathbf{m}_1^*) \cdot \mathbf{m}_2^*,$$

où \mathbf{m}_1^* est obtenu de la même façon que la méthode précédente et \mathbf{m}_2^* est un vecteur composé de variables aléatoires indépendantes de Bernoulli, m_{2i}^* , $i \in s$, de paramètre 1/2.

Antal et Tillé (2011a) et Antal et Tillé (2014) ont proposé également deux méthodes bootstrap valides pour un plan d'échantillonnage à probabilités inégales à forte entropie. Pour sélectionner un ré-échantillon s^* , Antal et Tillé (2011a) utilisent un mécanisme aléatoire qui mixe deux plans d'échantillonnage, le plan d'échantillonnage d'origine mis en oeuvre avec les probabilités d'inclusion ϕ_i , $i \in s$ et un plan de type "One-One". Les probabilités d'inclusion ϕ_i sont choisies de telle sorte que $\forall i \in s$, $1 - \phi_i = \tilde{D}_{ii}$, où \tilde{D}_{ii} sont les approximations données en (3.1.3), (3.1.4) et (3.1.5). En effet, le ré-échantillon s^* est construit en deux étapes. Dans un premier temps on applique dans s le même algorithme de tirage d'origine mais avec les probabilités d'inclusion ϕ_i , $i \in s$ pour sélectionner un échantillon s_1^* et dans un second temps, pour avoir un ré-échantillon s^* de la même taille que s , on sélectionne parmi les unités non sélectionnées dans la première étape, c'est-à-dire $i \in s \setminus s_1^*$, un échantillon s_2^* selon un plan de type "One-One". Finalement, on définit m_i^* comme étant le nombre de fois où l'unité $i \in s$ apparaît dans s^* . Pour une description détaillée de l'algorithme bootstrap proposé par Antal et Tillé (2011a), voir les algorithmes 4 et 5 dans le papier de Antal et Tillé (2011a).

Antal et Tillé (2014) ont généralisé leur méthode bootstrap proposée pour un plan aléatoire simple sans remise à un plan d'échantillonnage à probabilités inégales. En effet, l'algorithme bootstrap proposé est similaire à celui dans le cas d'un plan aléatoire simple sans remise. Dans un premier temps, il faut mettre en oeuvre dans s un plan de Poisson avec les mêmes probabilités d'inclusion que le plan d'origine. À cette étape on se trouve

avec un échantillon s_1^* de taille n_1 ensuite parmi les unités non sélectionnées, on sélectionne un échantillon s_2^* de taille $n - n_1$ selon un plan de type “One-One”. Il faut noter que si $n_1 = n - 1$, c’est-à-dire il reste seulement une unité qui n’était pas sélectionnée lors de l’étape 1, alors on ne peut pas mettre en œuvre un plan de type “One-One”. Antal et Tillé (2014) propose de choisir avec probabilité $1/2$ l’unité restante ou bien mettre en œuvre dans s le même plan d’échantillonnage d’origine avec les probabilités d’inclusion ψ_i , qui sont construites de telle sorte que $\sum_{i \in s} \psi_i = n - 2$ et de sélectionner les deux unités restantes selon un plan de type “One-One”. Pour une description détaillée de la méthode incluant la construction des probabilités d’inclusion, ψ_i , $i \in s$, voir Antal et Tillé (2014).

Chapitre 4

Étude par simulation

4.1. Introduction

Dans le présent chapitre, nous effectuons des études par simulation afin de comparer la performance des méthodes bootstrap décrites aux chapitres 2 et 3. La comparaison se basera sur la précision et la stabilité des estimateurs de variance et les propriétés empiriques des intervalles de confiance. Les plans d'échantillonnage considérés sont le plan aléatoire simple sans remise, le plan de Poisson et le plan de Poisson conditionnel. Pour chacun de ces plans, on s'intéresse d'une part à l'estimation bootstrap de la variance des estimateurs de paramètres linéaires, lisses de totaux et non linéaires. D'autre part, nous nous intéressons aux taux de couverture unilatérale et bilatérale et à la largeur moyenne des intervalles de confiance bootstrap. Toutes les simulations présentées dans ce chapitre ont été effectuées à l'aide du logiciel R, notamment le package "sampling" développé par Tillé et Matei (2016) qu'on a utilisé pour mettre en œuvre les algorithmes de tirage de Poisson et de Poisson conditionnel. Le chapitre est organisé de la façon suivante. En section 4.2, nous décrivons la façon dont nous avons généré les populations ainsi que les échantillons simulés. En section 4.3, nous présentons les indices de mesure des critères de comparaison. En sections 4.4, 4.5 et 4.6, nous présentons et discutons les résultats relatifs à chaque plan d'échantillonnage.

4.2. Populations et échantillons simulés

Nous utilisons une population fictive U de 2175 unités. Cette population contient trois variables notées x , y et z . Elles sont générées comme suit. Pour $i = 1, \dots, 2175$, $z_i = |v_i|$, où les v_i sont *i.i.d.* selon une loi normale de moyenne 0 et d'écart-type 7, $y_i = (\beta_0 + \beta_1 z_i^{1,2} + \sigma \epsilon_i)^2 + c$ où $\beta_0 = 12,5$, $\beta_1 = 3$, $c = 4000$ et les ϵ_i sont *i.i.d.* selon une loi normale de moyenne 0 et d'écart-type 1 et $x_i = y_i^{0,2} p_i$ où les p_i sont *i.i.d.* selon une log-normale de paramètre 0 et 0,25 qui sont, respectivement, la moyenne et l'écart type du logarithme de la variable p .

La variable x est utilisée comme mesure de la taille de la population. Ainsi, les probabilités d'inclusion pour le plan de Poisson et Poisson conditionnel ont été calculées proportionnellement à cette variable. La corrélation entre la variable auxiliaire x et la variable d'intérêt y dans la population est de 0,3. Il faut noter que le modèle utilisé pour générer les trois variables est identique à celui considéré dans Antal et Tillé (2011b). Par contre, dans le papier de Antal et Tillé (2011b) on avance une corrélation entre x et y d'environ 0,5. Une telle valeur de la corrélation n'est pas impossible, mais lorsqu'on a simulé 100 000 populations selon ce modèle, seulement cinq d'entre elles avaient une corrélation dépassant 0,45.

De la population U nous avons extrait quatre sous-population U_1 , U_2 , U_3 et U_4 respectivement de taille $N_1 = 2175$, $N_2 = 725$, $N_3 = 465$ et $N_4 = 155$; la sous-population U_i , $i = 1, \dots, 4$ est constituée par les N_i premiers éléments de U . Nous avons créé quatre populations de taille différente afin de pouvoir avoir quatre scénarios à partir de deux tailles d'échantillon (50 ou 150) ainsi que deux fractions de sondage (7% ou 32%). Dans les populations U_1 et U_3 nous avons mis en œuvre trois plans d'échantillonnage: plan aléatoire simple sans remise, plan de Poisson et plan de Poisson conditionnel de taille $n = 150$. Dans les populations U_2 et U_4 nous avons fait la même chose mais avec une taille $n = 50$. Le tableau 4.1 résume les quatre scénarios de simulations et donne les coefficients de variation CV pour les deux variables d'intérêt y et z .

Tab. 4.1. Populations et scénarios de simulation

Scénario	U_i	N_i	n_i	$f_i = \frac{n_i}{N_i}$	$CV_y(\%)$	$CV_z(\%)$
Scénario 1	U_1	2175	150	0,07	3,63	5,98
Scénario 2	U_2	725	50	0,07	6,36	10,17
Scénario 3	U_3	465	150	0,32	3,01	4,84
Scénario 4	U_4	155	50	0,32	4,64	8,67

Les figures 4.1 et 4.2 donnent la répartition des unités de chacune des quatre populations U_1, U_2, U_3 et U_4 selon les deux variables d'intérêt y et z .

Fig. 4.1. Répartition des quatre populations selon la variable y

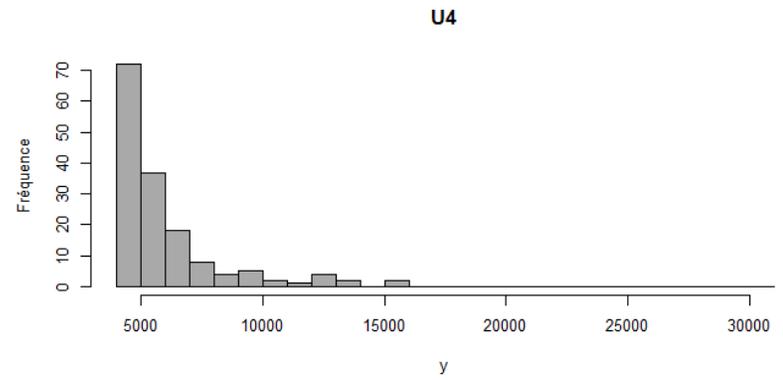
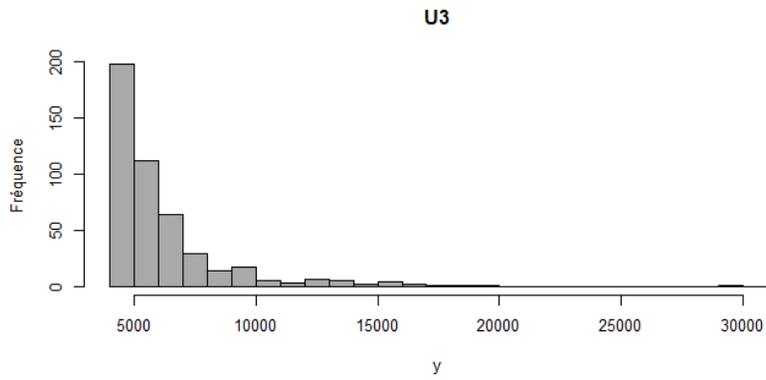
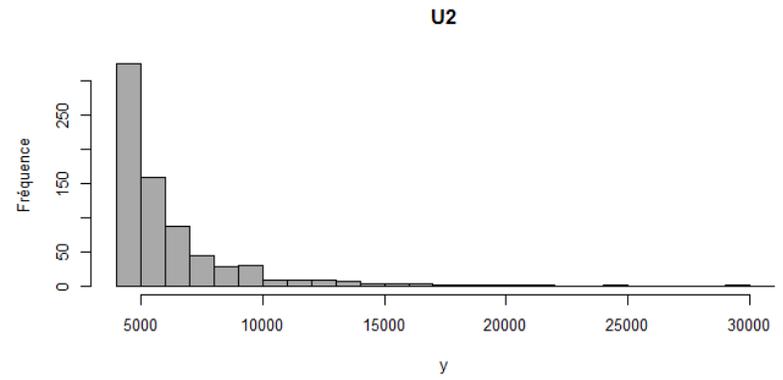
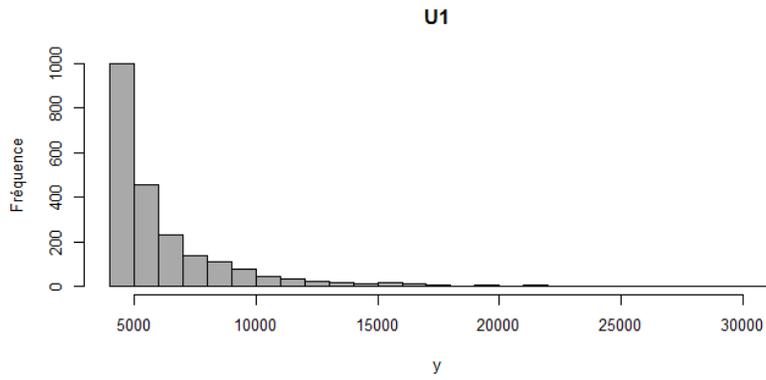
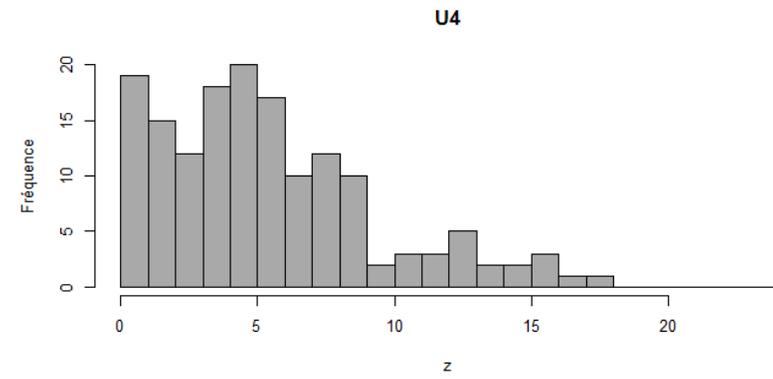
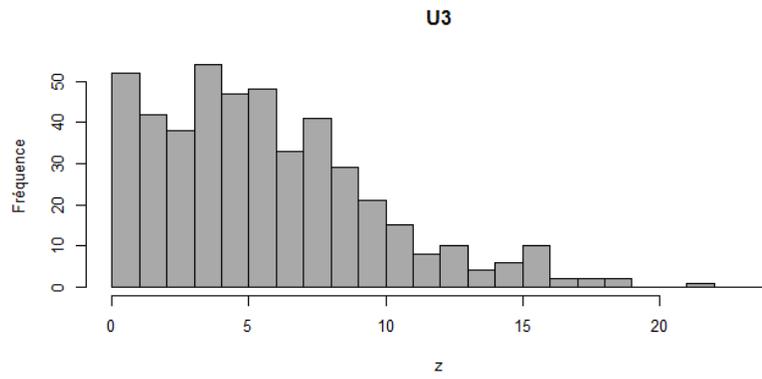
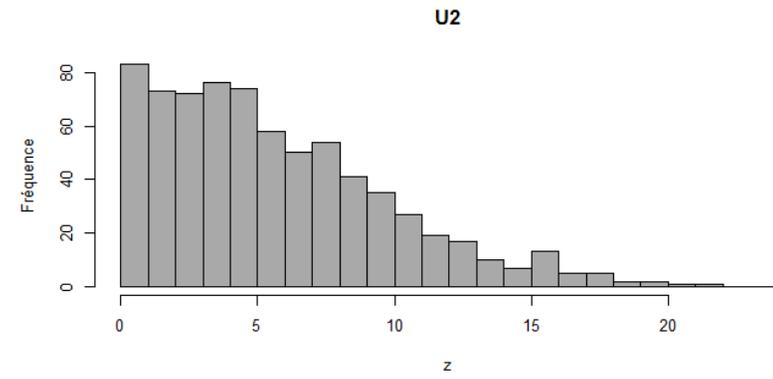
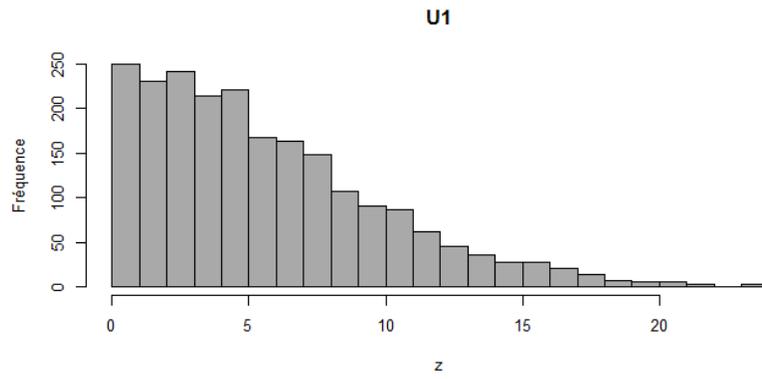


Fig. 4.2. Répartition des quatre populations selon la variable z



Dans notre étude on s'intéresse à l'estimation bootstrap de la variance des estimateurs des paramètres de la population suivants: les totaux des variables y et z , t_y et t_z , le ratio du total de y et du total de z , $r_{y,z}$, le coefficient de corrélation entre les variables y et z , $\rho_{y,z}$, la médiane de la variable y , m_y et l'indice de Gini de la variable y , G_y . Les estimateurs de ces paramètres sont calculés respectivement selon les formules (1.1.2), (1.1.4), (1.1.5), (1.1.6) et (1.1.7). La variance des estimateurs des totaux est calculée de façon exacte à l'aide des formules usuelles pour un estimateur d'Horvitz-Thompson pour le plan aléatoire simple sans remise et le plan de Poisson. Toutefois, pour ces estimateurs dans le cas du plan de Poisson conditionnel ainsi que pour tous les autres estimateurs et dans tous les plans, la variance est calculée par simulation. Soit $\hat{\theta}$ un estimateur d'un paramètre θ , on réalise un grand nombre (noté D) de simulations pour obtenir D échantillons indépendants s_1, \dots, s_D sur lesquels les estimateurs $\hat{\theta}_1, \dots, \hat{\theta}_D$ sont calculés. En conséquence de la loi des grands nombres,

$$V_{sim}(\hat{\theta}) = \frac{1}{D-1} \sum_{d=1}^D (\hat{\theta}_d - \hat{\theta}_{(\cdot)})^2,$$

converge vers la vraie valeur $V(\theta)$ quand D tend vers l'infini, avec $\hat{\theta}_{(\cdot)} = D^{-1} \sum_{d=1}^D \hat{\theta}_d$. On prendra $V_{sim}(\hat{\theta})$ comme approximation de la vraie variance de $\hat{\theta}$, avec $D = 10\,000$. Pour les méthodes bootstrap, l'estimation de la variance et les intervalles de confiance asymptotique, percentile, de base et t -bootstrap sont produits à l'aide de $A = 1\,000$ échantillons indépendants (s_1, \dots, s_A) , pour chacun desquels $B = 1\,000$ échantillons bootstrap sont sélectionnés. On note que les intervalles de confiance de type t -bootstrap pour les paramètres autres que les totaux nécessitent de répéter la procédure de rééchantillonnage une deuxième fois sur chaque échantillon bootstrap. Pour ne pas alourdir davantage le calcul on se limite pour chaque échantillon bootstrap de sélectionner seulement $C = 50$ rééchantillons bootstrap pour estimer la variance de l'estimateur calculé sur l'échantillon bootstrap.

4.3. Critères de comparaison des estimateurs de variance

Les performances des différentes méthodes bootstrap sont mesurées par six indices, le premier et le deuxième indice sont liés respectivement à la précision et la stabilité des estimateurs de variance tandis que les autres sont liés aux intervalles de confiance :

- Le biais relatif en pourcentage (*BR*)

$$BR = 100 \frac{E_p(V^*(\hat{\theta}^*)) - V_{sim}(\hat{\theta})}{V_{sim}(\hat{\theta})},$$

où $E_p(V^*(\hat{\theta}^*))$ est approximé par $A^{-1} \sum_{i=1}^A \hat{V}_{iB}^*$ où \hat{V}_{iB}^* est l'estimateur bootstrap de la variance.

- La racine carrée de l'erreur quadratique moyenne relative (*REQMR*)

$$REQMR = 100 \times \frac{\sqrt{E_p \left(V^*(\hat{\theta}^*) - V_{sim}(\hat{\theta}) \right)^2}}{V_{sim}(\hat{\theta})},$$

où $E_p \left(V^*(\hat{\theta}^*) - V_{sim}(\hat{\theta}) \right)^2$ est approximé par $A^{-1} \sum_{i=1}^A (\hat{V}_{iB}^* - V_{sim}(\hat{\theta}))^2$.

- La proportion des intervalles de confiance de niveau 95% pour lesquels le paramètre θ se situe en-dessous des intervalles de confiance calculés pour les différents types d'intervalles de confiance

$$L = \frac{1}{A} \sum_{i=1}^A I \left(\hat{\theta}_{i,inf} > \theta \right),$$

où $\hat{\theta}_{i,inf}$ est la borne inférieure de l'intervalle de confiance i . La proportion théorique est 2,5%.

- La proportion des intervalles de confiance de niveau 95% pour lesquels le paramètre θ se situe au-dessus des intervalles de confiance calculés pour les différents types d'intervalles de confiance. La proportion théorique est 2,5%.

$$U = \frac{1}{A} \sum_{i=1}^A I \left(\hat{\theta}_{i,sup} < \theta \right),$$

où $\hat{\theta}_{i,sup}$ est la borne supérieure de l'intervalle de confiance i .

- La proportion des intervalles de confiance de niveau 95% pour lesquels le paramètre se situe en-dehors des intervalles de confiance calculés pour les différents types d'intervalles de confiance

$$L + U = 1 - \frac{1}{A} \sum_{i=1}^A I(\hat{\theta}_{i,inf} < \theta < \hat{\theta}_{i,sup}).$$

La proportion théorique est 5%.

- La longueur moyenne des intervalles de confiance

$$lg = \frac{1}{A} \sum_{i=1}^A (\hat{\theta}_{i,sup} - \hat{\theta}_{i,inf}).$$

Les deux proportions L et U seront comparées à la proportion théorique (2,5%) via un test de proportion de niveau de signification 95%. Ainsi, les deux proportions seront statistiquement différentes de 2,5% si elles sont hors de l'intervalle $[1,5 ; 3,5]$. Pour la proportion $L + U$ elle sera comparée à 5% via le même test. En conséquence cette proportion sera statistiquement différente de 5% si elle se trouve hors de l'intervalle $[3,6 ; 6,4]$.

4.4. Étude 1: Méthodes bootstrap pour le plan aléatoire simple sans remise

Dans le cadre de cette première étude par simulation, on cherche à comparer les méthodes bootstrap présentées au chapitre 2 pour le plan aléatoire simple sans remise. En effet, on s'attend à ce que toutes les méthodes aient un comportement voisin dans le cas du total, du ratio et du coefficient de corrélation mais pas nécessairement dans le cas des paramètres non lisses comme la médiane et l'indice de Gini. Il faut noter que pour mettre en oeuvre certaines méthodes, il faut spécifier certains paramètres de réglage. Il s'agit des méthodes de Rao et Wu (1988), de Rao et coll. (1992) et de Beaumont et Patak (2012). Donc, les méthodes de Rao et Wu (1988) et de Rao et coll. (1992) ont été mises en oeuvre avec deux valeurs pour la taille d'échantillon bootstrap, $n' = n - 3$ et $n' = \lceil n/2 \rceil$. Quant à la méthode de Beaumont et Patak (2012), elle a été expérimentée en utilisant trois distributions, soit

deux distributions symétriques — la distribution normale et la distribution (2.2.8) avec $\epsilon = 1$, — et une distribution asymétrique générée à partir d'une exponentielle. Notons que certains ajustements bootstrap négatifs ont été générés avec ces distributions. Lorsque ces ajustements négatifs sont survenus nous les avons gardés sans aucun traitement dans le cas de tous les paramètres étudiés sauf pour le coefficient de corrélation. Dans ce dernier cas, l'occurrence de certains ajustements négatifs peut causer deux problèmes. Le premier problème est la production de valeurs manquantes pour l'estimateur bootstrap $\hat{\rho}^*$ car les deux termes suivants $s_y^{*2} = \sum_{i \in s} w_i^* y_i^2 - N^{-1} (\sum_{i \in s} w_i^* y_i)^2$ ou $s_z^{*2} = \sum_{i \in s} w_i^* z_i^2 - N^{-1} (\sum_{i \in s} w_i^* z_i)^2$ peuvent être négatifs; voir (1.1.5). Le second problème est que l'estimateur bootstrap peut prendre des valeurs hors de l'intervalle $[-1, 1]$. Ceci est dû principalement à l'occurrence de certains ajustements bootstrap négatifs extrêmes. Donc pour remédier à ces deux problèmes nous avons ré-échelonné les ajustements bootstrap comme indiqué dans la formule (2.2.9). Soit $\{a_{ib}^*, i = 1, \dots, n, b = 1, \dots, B\}$ un jeu d'ajustements bootstrap généré selon le modèle (2.2.7), où a_{ib}^* désigne l'ajustement bootstrap pour le poids de la i ème observation du b ème échantillon bootstrap. Le facteur de ré-échelonnement τ est choisi ainsi

$$\tau = \begin{cases} 1 & \text{si } \min^* \geq 0 \\ |\min^*| + 1 & \text{sinon,} \end{cases}$$

où $\min^* = \min\{a_{ib}^*, i = 1, \dots, n, b = 1, \dots, B\}$. On construit donc, un nouveau jeu d'ajustements bootstrap rééchelonnés $\check{a}_{ib}^* = \frac{a_{ib}^* + \tau - 1}{\tau}$ non négatifs. A partir de ce nouveau jeu, on calcule B statistiques bootstrap, $(\hat{\theta}_b^{r*}, b = 1, \dots, B)$ en substituant dans la formule de $\hat{\theta}$ les poids d'échantillonnage $(w_i, i = 1, \dots, n)$ par des poids bootstrap ré-échelonnés $(w_i^{r*} = \check{a}_{ib}^* w_i, i = 1, \dots, n)$. Il faut noter que Beaumont et Patak (2012) n'ont pas abordé le sujet de la construction des intervalles de confiance bootstrap si un système de poids bootstrap ré-échelonnés a été utilisé pour calculer les statistiques bootstrap. Autrement dit, si on fournit à un utilisateur un jeu de poids bootstrap ré-échelonnés et la valeur du facteur ré-échelonnement τ , comment faut-il procéder pour calculer un intervalle de confiance percentile par exemple? En effet, pour calculer les intervalles de confiance bootstrap de base et percentile, il ne faut pas substituer la distribution de $\hat{\theta}$ par celle de $\hat{\theta}^{r*}$ où

$\hat{\theta}^{r*}$ est l'estimateur bootstrap calculé à partir des poids bootstrap ré-échelonnés, car dans le cas d'un paramètre linéaire $V^*(\hat{\theta}^{r*}) = \tau^{-2}\hat{V}(\hat{\theta}) \neq \hat{V}(\hat{\theta})$. Pour détourner ce problème nous proposons de substituer la distribution de $\hat{\theta} - \theta$ par celle de $\tau(\hat{\theta}^{r*} - \hat{\theta})$ et de calculer les intervalles de confiance bootstrap de base et percentile respectivement selon les formules (1.2.8) et (1.2.9) où $\hat{L}_{n,*}(x) = Prob^*(\tau(\hat{\theta}^{r*} - \hat{\theta}) \leq x)$.

Notons que dans le cas du coefficient de corrélation $\rho_{y,z}$ l'estimation de la variance et des intervalles de confiance bootstrap de base et percentile ont été basés sur un jeu de poids bootstrap ré-échelonnés. En conséquence, il faut d'une part multiplier par τ^2 l'estimation de la variance et d'autre part calculer les intervalles de confiance bootstrap de base et percentile comme nous venons d'indiquer.

Rappelons que les intervalles de confiance t -bootstrap n'ont pas été calculés pour la méthode de Beaumont et Patak (2012).

4.4.1. Résultats: cas de paramètres linéaires

Le tableau 4.2 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne relative des estimateurs bootstrap de la variance étudiés pour les estimateurs Horvitz-Thompson des totaux t_y et t_z . Les tableaux 4.3 à 4.6 et 4.8 à 4.11 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap. Les deux tableaux 4.7 et 4.12 rapportent pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Booth et coll. (1994).

Tab. 4.2. Étude 1: Précision et stabilité des méthodes bootstrap (paramètres linéaires t_y et t_z)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
t_y								
Booth et coll. (1994)	-0,36	32,13	-3,64	50,38	-2,33	28,31	-2,18	30,50
Chao et Lo (1994)	-0,35	32,67	-3,73	50,56	-2,10	28,35	-2,07	30,84
Bickel et Freedman (1984)	-0,44	32,50	-3,48	50,74	-2,36	28,37	-2,05	30,91
Sitter (1992a)	0,60	32,92	-1,56	51,42	-1,78	28,34	-0,44	31,36
McCarthy et Snowden (1985)	0,21	32,64	-0,52	51,90	-1,46	28,64	-0,19	31,28
Sitter (1992b)	0,59	32,55	-1,40	51,96	-1,85	28,21	-0,42	31,46
Rao et Wu (1988)	0,28	32,91	-1,71	51,48	-1,66	28,52	-0,71	30,94
Rao et Wu (1988) ($n' = n/2$)	0,38	32,72	-1,32	51,60	-1,25	28,06	-0,97	31,02
Rao et coll. (1992)	0,40	32,55	-1,77	51,79	-1,60	28,35	-0,38	31,12
Rao et coll. (1992) ($n' = n/2$)	0,25	32,67	-1,75	51,57	-1,54	28,82	-0,68	31,24
Chipperfield et Preston (2007)	0,15	32,25	-1,45	51,78	-1,76	28,37	-0,74	31,01
Antal et Tillé (2014)	0,34	32,73	-1,93	51,25	-1,65	28,69	-0,60	31,00
Antal et Tillé (2011a)	0,26	32,53	-1,84	51,14	-1,47	28,60	-0,67	31,30
Beaumont et Patak (2012) Norm	0,27	32,68	-1,87	51,32	-1,49	28,61	-0,72	31,09
Beaumont et Patak (2012) Exp	0,06	32,79	-1,87	51,89	-1,74	28,60	-0,43	31,49
Beaumont et Patak (2012) $\epsilon = 1$	0,39	32,71	-1,92	51,50	-1,88	28,55	-0,77	31,01
t_z								
Booth et coll. (1994)	-0,99	13,65	-2,59	21,17	-0,94	11,65	-0,62	18,98
Chao et Lo (1994)	-1,22	13,64	-2,85	21,10	-0,82	11,47	-0,62	19,38
Bickel et Freedman (1984)	-1,10	13,63	-2,67	20,92	-0,82	11,46	-0,33	19,23
Sitter (1992a)	-0,29	13,71	-0,62	21,40	-0,31	11,64	1,08	19,50
McCarthy et Snowden (1985)	-0,27	13,83	0,66	21,76	-0,02	11,63	1,25	19,36
Sitter (1992b)	-0,38	13,64	-0,54	21,69	-0,47	11,62	0,95	19,44
Rao et Wu (1988)	-0,51	13,72	-0,88	21,42	-0,38	11,46	0,57	19,55
Rao et Wu (1988) ($n' = n/2$)	-0,44	13,59	-0,76	21,42	-0,39	11,46	0,67	19,22
Rao et coll. (1992)	-0,54	13,77	-1,02	21,37	-0,27	11,46	0,58	19,37
Rao et coll. (1992) ($n' = n/2$)	-0,18	14,11	-0,90	21,50	-0,68	11,29	0,73	19,40
Chipperfield et Preston (2007)	-0,49	13,83	-0,69	21,50	-0,35	11,62	0,47	19,26
Antal et Tillé (2014)	-0,43	13,87	-0,84	21,35	-0,46	11,45	0,68	19,38
Antal et Tillé (2011a)	-0,21	13,66	-0,86	21,60	-0,24	11,42	0,80	19,53
Beaumont et Patak (2012) Norm	-0,45	13,65	-0,77	21,24	-0,50	11,72	0,70	19,34
Beaumont et Patak (2012) Exp	-0,19	13,93	-0,81	21,64	-0,38	11,69	0,88	19,77
Beaumont et Patak (2012) $\epsilon = 1$	-0,67	13,70	-0,43	21,35	-0,20	11,52	0,91	19,40

Tab. 4.3. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (t_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,4	6,8	7,2	1,000	0,4	7,8	8,2	1,000	0,5	5,5	6,0	1,000	0,7	5,9	6,6	1,000
Chao et Lo (1994)	0,4	6,7	7,1	1,000	0,5	8,0	8,5	0,999	0,6	5,2	5,8	1,001	0,7	6,1	6,8	1,000
Bickel et Freedman (1984)	0,5	6,5	7,0	0,999	0,4	7,8	8,2	1,000	0,6	5,6	6,2	1,000	0,7	6,5	7,2	1,000
Sitter (1992a)	0,4	6,5	6,9	1,005	0,4	7,7	8,1	1,010	0,6	5,5	6,1	1,003	0,6	6,2	6,8	1,009
McCarthy et Snowden (1985)	0,5	6,6	7,1	1,003	0,4	7,5	7,9	1,016	0,5	5,4	5,9	1,004	0,7	5,9	6,6	1,010
Sitter (1992b)	0,4	6,5	6,9	1,005	0,4	7,8	8,2	1,011	0,6	5,3	5,9	1,003	0,7	6,0	6,7	1,009
Rao et Wu (1988)	0,4	6,6	7,0	1,002	0,4	7,6	8,0	1,009	0,5	5,2	5,7	1,002	0,6	6,2	6,8	1,006
Rao et Wu (1988) ($n' = n/2$)	0,4	6,6	7,0	1,004	0,4	7,9	8,3	1,010	0,6	5,5	6,1	1,003	0,8	6,2	7,0	1,007
Rao et coll. (1992)	0,5	6,4	6,9	1,002	0,4	7,9	8,3	1,008	0,6	5,2	5,8	1,003	0,7	5,9	6,6	1,005
Rao et coll. (1992) ($n' = n/2$)	0,4	6,7	7,1	1,003	0,4	7,7	8,1	1,009	0,5	5,5	6,0	1,001	0,7	6,3	7,0	1,007
Chipperfield et Preston (2007)	0,5	6,4	6,9	1,003	0,4	7,6	8,0	1,011	0,7	5,5	6,2	1,003	0,7	5,9	6,6	1,007
Antal et Tillé (2014)	0,4	6,4	6,8	1,004	0,4	7,8	8,2	1,009	0,5	5,5	6,0	1,002	0,7	6,0	6,7	1,007
Antal et Tillé (2011a)	0,4	6,8	7,2	1,003	0,4	8,0	8,4	1,010	0,6	5,4	6,0	1,003	0,7	5,8	6,5	1,008
Beaumont et Patak (2012) Norm	0,4	6,8	7,2	1,003	0,4	7,8	8,2	1,010	0,6	5,3	5,9	1,003	0,7	6,0	6,7	1,008
Beaumont et Patak (2012) Exp	0,4	6,8	7,2	1,003	0,4	7,8	8,2	1,009	0,4	5,6	6,0	1,002	0,7	6,0	6,7	1,007
Beaumont et Patak (2012) $\epsilon = 1$	0,4	6,6	7,0	1,003	0,5	7,8	8,3	1,009	0,6	5,5	6,1	1,004	0,8	6,1	6,9	1,007

Tab. 4.4. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (t_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,9	6,4	7,3	1,000	0,7	7,5	8,2	1,000	0,8	5,2	6,0	1,000	0,9	5,7	6,6	1,000
Chao et Lo (1994)	0,7	6,4	7,1	1,001	0,8	7,5	8,3	0,999	0,9	5,2	6,1	1,000	0,9	6,0	6,9	1,000
Bickel et Freedman (1984)	0,8	6,3	7,1	1,001	0,9	7,4	8,3	1,001	0,7	5,8	6,5	1,000	1,4	5,9	7,3	1,000
Sitter (1992a)	0,7	6,2	6,9	1,004	0,9	7,1	8,0	1,011	0,8	5,4	6,2	1,003	1,0	6,0	7,0	1,008
McCarthy et Snowden (1985)	0,8	6,0	6,8	1,003	0,8	7,1	7,9	1,016	1,1	4,9	6,0	1,007	1,0	5,3	6,3	1,012
Sitter (1992b)	0,7	6,0	6,7	1,005	0,8	7,3	8,1	1,012	0,7	5,1	5,8	1,004	1,0	6,1	7,1	1,011
Rao et Wu (1988)	0,8	6,1	6,9	1,002	0,9	6,7	7,6	1,009	0,9	4,7	5,6	1,004	1,4	5,4	6,8	1,007
Rao et Wu (1988) ($n' = n/2$)	1,2	5,7	6,9	1,004	1,1	6,7	7,8	1,005	1,4	4,2	5,6	1,002	1,8	5,1	6,9	1,003
Rao et coll. (1992)	0,8	5,9	6,7	1,002	0,9	7,2	8,1	1,010	0,9	4,6	5,5	1,004	1,0	5,3	6,3	1,005
Rao et coll. (1992) ($n' = n/2$)	1,3	6,1	7,4	1,000	1,1	7,0	8,1	1,003	1,3	4,3	5,6	1,001	1,4	5,3	6,7	1,003
Chipperfield et Preston (2007)	0,6	6,5	7,1	0,996	0,4	7,9	8,3	0,986	0,7	5,7	6,4	0,993	1,0	6,5	7,5	0,997
Antal et Tillé (2014)	0,5	6,7	7,2	0,996	0,5	7,9	8,4	0,987	0,6	5,8	6,4	0,998	0,8	6,4	7,2	1,001
Antal et Tillé (2011a)	0,8	6,1	6,9	1,004	0,8	6,8	7,6	1,014	1,0	4,4	5,4	1,010	1,3	5,1	6,4	1,018
Beaumont et Patak (2012) Norm	0,6	6,6	7,2	1,005	0,5	7,9	8,4	1,018	0,6	5,6	6,2	1,006	0,8	5,7	6,5	1,014
Beaumont et Patak (2012) Exp	1,4	5,5	6,9	1,007	1,4	6,2	7,6	1,020	1,5	4,0	5,5	1,008	1,7	4,8	6,5	1,015
Beaumont et Patak (2012) $\epsilon = 1$	0,6	6,8	7,4	0,996	0,5	8,1	8,6	0,985	0,5	5,8	6,3	0,995	0,7	6,6	7,3	0,997

Tab. 4.5. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (t_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,2	7,6	7,8	1,000	0,4	8,8	9,2	1,000	0,5	5,9	6,4	1,000	0,7	6,6	7,3	1,000
Chao et Lo (1994)	0,2	7,2	7,4	1,001	0,2	8,9	9,1	0,999	0,6	6,0	6,6	1,000	0,8	7,1	7,9	1,000
Bickel et Freedman (1984)	0,4	7,3	7,7	1,001	0,2	8,8	9,0	1,001	0,4	6,1	6,5	1,000	0,6	6,9	7,5	1,000
Sitter (1992a)	0,4	7,4	7,8	1,004	0,4	8,7	9,1	1,011	0,5	6,1	6,6	1,003	0,6	6,8	7,4	1,008
McCarthy et Snowden (1985)	0,4	7,6	8,0	1,003	0,2	8,9	9,1	1,016	0,5	6,7	7,2	1,007	0,6	7,4	8,0	1,012
Sitter (1992b)	0,2	7,2	7,4	1,005	0,2	8,5	8,7	1,012	0,6	5,6	6,2	1,004	0,7	7,1	7,8	1,011
Rao et Wu (1988)	0,3	7,5	7,8	1,002	0,2	9,3	9,5	1,009	0,4	6,3	6,7	1,004	0,6	7,8	8,4	1,007
Rao et Wu (1988) ($n' = n/2$)	0,1	7,8	7,9	1,004	0,0	9,5	9,5	1,005	0,3	7,0	7,3	1,002	0,7	8,9	9,6	1,003
Rao et coll. (1992)	0,1	7,2	7,3	1,002	0,3	9,5	9,8	1,010	0,6	6,1	6,7	1,004	0,6	7,7	8,3	1,005
Rao et coll. (1992) ($n' = n/2$)	0,3	7,7	8,0	1,000	0,1	9,6	9,7	1,003	0,3	6,7	7,0	1,001	0,6	8,4	9,0	1,003
Chipperfield et Preston (2007)	0,6	6,6	7,2	0,996	0,4	7,9	8,3	0,986	0,6	5,6	6,2	0,993	0,9	6,7	7,6	0,997
Antal et Tillé (2014)	0,5	6,8	7,3	0,996	0,4	7,9	8,3	0,987	0,6	5,7	6,3	0,998	0,7	6,4	7,1	1,001
Antal et Tillé (2011a)	0,2	7,6	7,8	1,004	0,3	9,3	9,6	1,014	0,5	6,4	6,9	1,010	0,7	7,7	8,4	1,018
Beaumont et Patak (2012) Norm	0,5	6,8	7,3	1,005	0,5	7,8	8,3	1,018	0,4	5,4	5,8	1,006	0,7	6,2	6,9	1,014
Beaumont et Patak (2012) Exp	0,2	8,1	8,3	1,007	0,0	9,6	9,6	1,020	0,3	7,0	7,3	1,008	0,3	8,5	8,8	1,015
Beaumont et Patak (2012) $\epsilon = 1$	0,5	6,7	7,2	0,996	0,5	8,1	8,6	0,985	0,8	5,8	6,6	0,995	1,0	6,5	7,5	0,997

Tab. 4.6. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (t_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	2,1	4,2	6,3	1,000	1,6	3,6	5,2	1,000	2,1	3,1	5,2	1,000	2,4	3,0	5,4	1,000
Chao et Lo (1994)	1,9	4,4	6,3	1,001	1,7	4,0	5,7	0,999	1,9	3,0	4,9	1,000	2,1	3,3	5,4	1,002
Bickel et Freedman (1984)	2,0	3,8	5,8	1,001	1,7	4,0	5,7	1,003	1,9	2,6	4,5	0,999	2,0	3,4	5,4	1,000
Sitter (1992a)	1,7	4,2	5,9	1,009	1,3	3,2	4,5	1,026	2,1	2,7	4,8	1,004	1,8	3,2	5,0	1,019
McCarthy et Snowden (1985)	2,1	3,9	6,0	1,002	1,4	3,2	4,6	1,030	1,8	3,4	5,2	0,993	1,8	3,6	5,4	0,996
Sitter (1992b)	2,2	4,0	6,2	1,010	1,7	3,4	5,1	1,023	2,2	2,8	5,0	1,006	1,9	3,1	5,0	1,021
Rao et Wu (1988)	1,6	4,0	5,6	1,005	1,6	3,7	5,3	1,023	2,6	2,6	5,2	1,008	2,1	3,0	5,1	1,036
Rao et Wu (1988)($n' = n/2$)	2,6	2,4	5,0	1,061	1,9	2,5	4,4	1,161	3,1	1,9	5,0	1,061	3,1	1,9	5,0	1,199
Rao et coll. (1992)	1,8	3,9	5,7	1,009	1,1	3,4	4,5	1,031	1,8	2,6	4,4	1,011	2,0	2,9	4,9	1,049
Rao et coll. (1992)($n' = n/2$)	2,7	2,7	5,4	1,063	1,9	2,5	4,4	1,169	2,7	1,8	4,5	1,068	2,5	1,9	4,4	1,211
Chipperfield et Preston (2007)	2,5	3,2	5,7	1,051	2,3	2,9	5,2	1,133	2,5	1,8	4,3	1,052	2,1	1,9	4,0	1,137
Antal et Tillé (2014)	2,3	3,0	5,3	1,046	1,6	2,6	4,2	1,113	2,0	2,1	4,1	1,021	2,0	2,2	4,2	1,045
Antal et Tillé (2011a)	1,9	4,4	6,3	1,000	1,0	3,5	4,5	1,010	1,7	3,8	5,5	0,987	1,6	3,9	5,5	0,979

Tab. 4.7. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (t_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,006	1,065	1,011	1,195	1,007	1,067	1,009	1,121
Chao et Lo (1994)	1,006	1,066	1,010	1,194	1,008	1,067	1,009	1,123
Bickel et Freedman (1984)	1,006	1,066	1,011	1,198	1,007	1,066	1,009	1,120
Sitter (1992a)	1,011	1,074	1,022	1,226	1,010	1,071	1,017	1,142
McCarthy et Snowden (1985)	1,009	1,067	1,027	1,230	1,012	1,059	1,019	1,117
Sitter (1992b)	1,011	1,075	1,022	1,223	1,010	1,073	1,017	1,144
Rao et Wu (1988)	1,009	1,070	1,021	1,222	1,011	1,075	1,016	1,161
Rao et Wu (1988) ($n' = n/2$)	1,010	1,130	1,023	1,388	1,008	1,131	1,015	1,343
Rao et coll. (1992)	1,010	1,072	1,021	1,225	1,010	1,072	1,016	1,158
Rao et coll. (1992) ($n' = n/2$)	1,009	1,128	1,021	1,385	1,011	1,135	1,016	1,349
Chipperfield et Preston (2007)	1,009	1,120	1,020	1,348	1,009	1,121	1,015	1,272
Antal et Tillé (2014)	1,010	1,113	1,020	1,329	1,011	1,089	1,017	1,170
Antal et Tillé (2011a)	1,009	1,064	1,020	1,207	1,012	1,053	1,016	1,097
Beaumont et Patak (2012) Norm	1,009	–	1,020	–	1,012	–	1,016	–
Beaumont et Patak (2012) Exp	1,008	–	1,020	–	1,010	–	1,017	–
Beaumont et Patak (2012) $\epsilon = 1$	1,010	–	1,020	–	1,009	–	1,016	–

Tab. 4.8. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,7	4,0	4,7	1,000	1,2	4,2	5,4	1,000	2,0	2,8	4,8	1,000	1,8	4,6	6,4	1,000
Chao et Lo (1994)	0,7	4,1	4,8	0,999	1,5	4,1	5,6	0,999	2,2	3,1	5,3	1,002	1,8	4,5	6,3	1,000
Bickel et Freedman (1984)	0,8	4,1	4,9	0,999	1,3	4,1	5,4	0,999	2,0	2,5	4,5	1,002	1,5	4,5	6,0	1,000
Sitter (1992a)	0,7	4,0	4,7	1,004	1,1	3,9	5,0	1,010	2,2	3,1	5,3	1,004	1,6	4,5	6,1	1,007
McCarthy et Snowden (1985)	0,7	4,0	4,7	1,004	1,3	3,9	5,2	1,016	2,0	2,8	4,8	1,006	1,7	4,4	6,1	1,007
Sitter (1992b)	0,7	3,9	4,6	1,003	1,2	4,1	5,3	1,010	2,1	2,8	4,9	1,002	1,7	4,5	6,2	1,007
Rao et Wu (1988)	0,7	4,1	4,8	1,002	1,3	4,0	5,3	1,009	1,9	2,7	4,6	1,002	1,6	4,5	6,1	1,004
Rao et Wu (1988) ($n' = n/2$)	0,6	3,9	4,5	1,004	1,1	4,1	5,2	1,009	2,2	2,9	5,1	1,004	1,6	4,4	6,0	1,007
Rao et coll. (1992)	0,7	4,0	4,7	1,002	1,2	4,1	5,3	1,008	1,9	2,9	4,8	1,004	1,7	4,1	5,8	1,004
Rao et coll. (1992) ($n' = n/2$)	0,5	3,9	4,4	1,004	1,3	3,9	5,2	1,009	2,2	2,8	5,0	1,002	1,7	4,2	5,9	1,007
Chipperfield et Preston (2007)	0,8	3,8	4,6	1,002	1,2	4,0	5,2	1,009	2,2	3,0	5,2	1,002	1,8	4,3	6,1	1,004
Antal et Tillé (2014)	0,7	4,0	4,7	1,002	1,3	4,0	5,3	1,009	2,3	2,8	5,1	1,002	1,7	4,5	6,2	1,007
Antal et Tillé (2011a)	0,7	3,9	4,6	1,004	1,3	4,0	5,3	1,009	2,1	2,9	5,0	1,004	1,6	4,4	6,0	1,007
Beaumont et Patak (2012) Norm	0,7	4,0	4,7	1,002	1,2	4,0	5,2	1,009	2,0	3,0	5,0	1,002	1,8	4,3	6,1	1,007
Beaumont et Patak (2012) Exp	0,7	3,8	4,5	1,001	1,4	4,0	5,4	1,009	2,0	2,8	4,8	1,002	1,5	4,3	5,8	1,007
Beaumont et Patak (2012) $\epsilon = 1$	0,7	4,0	4,7	1,003	1,2	4,0	5,2	1,009	2,0	2,6	4,6	1,004	1,6	4,4	6,0	1,007

Tab. 4.9. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,8	3,9	4,7	1,000	1,4	3,6	5,0	1,000	2,4	2,8	5,2	1,000	1,9	4,3	6,2	1,000
Chao et Lo (1994)	0,9	4,0	4,9	0,999	1,9	3,6	5,5	0,999	2,3	3,0	5,3	1,000	2,0	4,5	6,5	1,000
Bickel et Freedman (1984)	0,9	3,8	4,7	1,000	1,8	3,8	5,6	1,000	2,0	2,9	4,9	1,000	2,0	4,1	6,1	1,000
Sitter (1992a)	0,8	4,0	4,8	1,003	1,4	3,6	5,0	1,009	2,3	2,8	5,1	1,002	1,8	4,0	5,8	1,007
McCarthy et Snowden (1985)	0,9	4,0	4,9	1,004	1,5	3,4	4,9	1,017	2,1	2,5	4,6	1,004	2,0	4,0	6,0	1,011
Sitter (1992b)	1,0	3,8	4,8	1,003	1,6	3,6	5,2	1,011	2,3	2,6	4,9	1,002	1,8	4,1	5,9	1,007
Rao et Wu (1988)	0,8	4,2	5,0	1,001	1,6	3,6	5,2	1,008	2,4	2,4	4,8	1,002	2,1	3,6	5,7	1,004
Rao et Wu (1988) ($n' = n/2$)	0,9	3,6	4,5	1,005	1,4	3,4	4,8	1,008	2,4	2,3	4,7	1,002	1,9	3,6	5,5	1,007
Rao et coll. (1992)	0,8	3,7	4,5	1,003	1,3	3,6	4,9	1,008	2,4	2,6	5,0	1,004	1,7	3,9	5,6	1,004
Rao et coll. (1992) ($n' = n/2$)	1,0	3,5	4,5	1,004	1,7	3,1	4,8	1,008	2,4	2,7	5,1	1,000	1,8	3,9	5,7	1,007
Chipperfield et Preston (2007)	0,8	3,9	4,7	1,001	1,3	4,1	5,4	1,004	2,4	3,2	5,6	1,002	1,9	4,3	6,2	1,000
Antal et Tillé (2014)	0,9	4,3	5,2	1,001	1,3	4,0	5,3	1,006	2,2	3,1	5,3	1,002	2,0	4,5	6,5	1,004
Antal et Tillé (2011a)	0,8	3,9	4,7	1,005	1,7	3,8	5,5	1,009	2,2	2,3	4,5	1,004	1,7	4,2	5,9	1,011
Beaumont et Patak (2012) Norm	0,8	4,2	5,0	1,002	1,2	4,1	5,3	1,011	2,0	2,9	4,9	1,000	1,7	4,1	5,8	1,007
Beaumont et Patak (2012) Exp	0,9	3,3	4,2	1,006	1,6	2,8	4,4	1,019	2,3	2,1	4,4	1,006	2,0	3,5	5,5	1,018
Beaumont et Patak (2012) $\epsilon = 1$	0,8	4,3	5,1	1,001	1,1	4,2	5,3	1,006	2,1	2,8	4,9	1,002	2,0	4,0	6,0	1,000

Tab. 4.10. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,6	4,0	4,6	1,000	1,4	4,7	6,1	1,000	2,0	3,3	5,3	1,000	1,7	4,6	6,3	1,000
Chao et Lo (1994)	0,7	4,7	5,4	0,999	1,4	4,8	6,2	0,999	1,9	2,9	4,8	1,000	1,8	4,6	6,4	1,000
Bickel et Freedman (1984)	0,8	4,2	5,0	1,000	0,8	4,5	5,3	1,000	2,2	2,7	4,9	1,000	1,6	4,5	6,1	1,000
Sitter (1992a)	0,7	4,0	4,7	1,003	0,9	4,4	5,3	1,009	1,9	3,2	5,1	1,002	1,8	4,3	6,1	1,007
McCarthy et Snowden (1985)	0,6	4,2	4,8	1,004	1,0	4,4	5,4	1,017	1,8	3,1	4,9	1,004	1,6	4,5	6,1	1,011
Sitter (1992b)	0,6	4,4	5,0	1,003	1,3	4,3	5,6	1,011	2,0	3,2	5,2	1,002	1,6	4,4	6,0	1,007
Rao et Wu (1988)	0,6	4,4	5,0	1,001	1,3	4,3	5,6	1,008	2,0	3,0	5,0	1,002	1,5	4,8	6,3	1,004
Rao et Wu (1988) ($n' = n/2$)	0,5	4,5	5,0	1,005	1,2	4,7	5,9	1,008	2,1	3,5	5,6	1,002	1,6	4,8	6,4	1,007
Rao et coll. (1992)	0,6	4,6	5,2	1,003	1,1	5,0	6,1	1,008	2,0	3,3	5,3	1,004	1,7	4,6	6,3	1,004
Rao et coll. (1992) ($n' = n/2$)	0,6	4,4	5,0	1,004	1,1	4,4	5,5	1,008	1,8	3,4	5,2	1,000	1,6	4,6	6,2	1,007
Chipperfield et Preston (2007)	0,8	4,4	5,2	1,001	1,4	4,2	5,6	1,004	2,5	3,0	5,5	1,002	1,6	4,6	6,2	1,000
Antal et Tillé (2014)	0,7	4,2	4,9	1,001	1,4	3,9	5,3	1,006	2,3	2,9	5,2	1,002	1,8	4,5	6,3	1,004
Antal et Tillé (2011a)	0,5	4,2	4,7	1,005	1,1	5,0	6,1	1,009	1,9	3,3	5,2	1,004	1,4	4,7	6,1	1,011
Beaumont et Patak (2012) Norm	0,7	4,1	4,8	1,002	1,3	3,7	5,0	1,011	2,1	3,1	5,2	1,000	1,7	4,4	6,1	1,007
Beaumont et Patak (2012) Exp	0,6	4,3	4,9	1,006	1,1	4,5	5,6	1,019	2,0	3,2	5,2	1,006	1,3	4,8	6,1	1,018
Beaumont et Patak (2012) $\epsilon = 1$	0,8	4,2	5,0	1,001	1,3	4,0	5,3	1,006	2,0	2,7	4,7	1,002	1,8	4,5	6,3	1,000

Tab. 4.11. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	1,2	3,2	4,4	1,000	2,2	2,0	4,2	1,000	2,2	2,1	4,3	1,000	2,6	2,9	5,5	1,000
Chao et Lo (1994)	1,4	3,0	4,4	0,999	1,8	1,8	3,6	0,996	2,5	2,2	4,7	1,002	2,3	2,7	5,0	1,000
Bickel et Freedman (1984)	1,1	2,9	4,0	1,000	1,4	2,0	3,4	1,001	2,6	1,8	4,4	1,000	2,0	2,6	4,6	1,000
Sitter (1992a)	1,2	3,1	4,3	1,008	1,4	1,9	3,3	1,022	2,5	2,3	4,8	1,006	2,2	2,4	4,6	1,017
McCarthy et Snowden (1985)	0,7	3,3	4,0	1,008	1,5	1,8	3,3	1,033	2,2	1,8	4,0	1,008	2,0	2,9	4,9	1,010
Sitter (1992b)	1,1	3,0	4,1	1,007	1,6	1,6	3,2	1,021	2,5	2,0	4,5	1,006	2,2	2,7	4,9	1,017
Rao et Wu (1988)	0,9	3,1	4,0	1,003	1,7	2,1	3,8	1,013	2,4	2,2	4,6	1,004	2,1	2,6	4,7	1,010
Rao et Wu (1988)($n' = n/2$)	1,3	2,7	4,0	1,019	2,0	1,6	3,6	1,059	2,7	1,7	4,4	1,016	2,1	2,1	4,2	1,059
Rao et coll. (1992)	1,0	2,8	3,8	1,006	1,8	1,8	3,6	1,014	2,6	1,7	4,3	1,004	2,1	2,9	5,0	1,014
Rao et coll. (1992)($n' = n/2$)	1,2	2,4	3,6	1,018	1,8	1,4	3,2	1,063	2,6	1,8	4,4	1,020	2,3	2,2	4,5	1,063
Chipperfield et Preston (2007)	1,4	2,6	4,0	1,013	1,9	1,8	3,7	1,045	2,3	1,7	4,0	1,018	2,3	2,2	4,5	1,045
Antal et Tillé (2014)	1,0	2,7	3,7	1,016	1,7	1,5	3,2	1,046	2,6	2,0	4,6	1,008	2,0	2,4	4,4	1,028
Antal et Tillé (2011a)	1,2	3,0	4,2	1,007	1,5	2,2	3,7	1,016	2,3	2,2	4,5	1,000	1,8	2,9	4,7	0,997

Tab. 4.12. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (t_z)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,005	1,014	1,006	1,043	1,004	1,012	1,007	1,044
Chao et Lo (1994)	1,004	1,013	1,004	1,038	1,006	1,014	1,007	1,044
Bickel et Freedman (1984)	1,004	1,014	1,005	1,043	1,006	1,012	1,007	1,044
Sitter (1992a)	1,009	1,022	1,016	1,065	1,008	1,018	1,015	1,062
McCarthy et Snowden (1985)	1,009	1,022	1,022	1,077	1,010	1,021	1,015	1,055
Sitter (1992b)	1,008	1,021	1,016	1,065	1,006	1,018	1,015	1,062
Rao et Wu (1988)	1,007	1,017	1,014	1,056	1,008	1,016	1,015	1,055
Rao et Wu (1988) ($n' = n/2$)	1,008	1,034	1,016	1,105	1,006	1,029	1,015	1,105
Rao et coll. (1992)	1,009	1,020	1,014	1,058	1,006	1,016	1,015	1,058
Rao et coll. (1992) ($n' = n/2$)	1,007	1,033	1,016	1,108	1,008	1,033	1,015	1,109
Chipperfield et Preston (2007)	1,008	1,028	1,014	1,086	1,006	1,029	1,015	1,087
Antal et Tillé (2014)	1,007	1,030	1,014	1,091	1,006	1,021	1,015	1,073
Antal et Tillé (2011a)	1,009	1,021	1,014	1,060	1,008	1,012	1,015	1,040
Beaumont et Patak (2012) Normal	1,008	–	1,014	–	1,006	–	1,015	–
Beaumont et Patak (2012) Exp	1,008	–	1,014	–	1,008	–	1,015	–
Beaumont et Patak (2012) $\epsilon = 1$	1,008	–	1,014	–	1,006	–	1,015	–

Tab. 4.13. Biais relatif théorique de la variance de l'estimateur d'Horvitz-Thompson des estimateurs bootstrap par pseudo-population de la variance, $BRT = 100(E_p(V^*(\hat{t}_{HT}^*) - V_p(\hat{t}_{HT}))/V_p(\hat{t}_{HT}))$

Méthode	<i>BRT</i>			
	$f = 7\%$		$f = 32\%$	
	$n = 150$	$n = 50$	$n = 150$	$n = 50$
Booth et coll. (1994)	-0,62	-1,87	-0,46	-1,38
Chao et Lo (1994)	-0,62	-1,87	-0,46	-1,38
Bickel et Freedman (1984)	-0,62	-1,86	-0,45	-1,36.

4.4.2. Discussion: cas de paramètres linéaires

À la lumière des résultats indiqués au tableau 4.2, les trois premières méthodes bootstrap par pseudo-population, à savoir Booth et coll. (1994), Chao et Lo (1994) et Bickel et Freedman (1984) montrent un comportement similaire en terme du signe du BR observé de de la variance de l'estimateur Horvitz-Tompson du total et de sa valeur. Ce résultat n'est pas surprenant car d'une part $E_p(V^*(\hat{t}_{HT}^*)) < V_p(\hat{t}_{HT})$ et d'autre part parce que les BRT de chacune de ces trois méthodes bootstrap évalués dans les différents scénarios sont semblables. Le tableau 4.13 donne les valeurs du BRT de chacune des trois méthodes en fonction des valeurs de la fraction de sondage f et de la taille de l'échantillon n . Pour ce qui est des autres méthodes bootstrap, qui se calent explicitement sur l'estimateur de variance dans le cas linéaire, à savoir la méthode de Sitter (1992a) et les méthodes bootstrap direct ou pondéré, elles montrent un comportement similaire en terme du BR . En effet, les BR associés à ces méthodes bootstrap sont inférieurs à 2% en valeur absolue dans les quatre scénarios tandis que les BR associés aux trois premières méthodes bootstrap par pseudo-population dépassent 2% en valeur absolue dans les trois derniers scénarios. En ce qui concerne la stabilité mesurée ici par la racine carrée de l'erreur quadratique moyenne relative $REQMR$, toutes les méthodes ont des performances équivalentes. Les différentes méthodes bootstrap sont plus stables dans le cas d'un coefficient de variation petit; voir le

tableau 4.1. Dans le cas de t_y , il s'agit des scénarios 1, 3 et 4 alors qu'il s'agit des scénarios 1 et 3 dans le cas de t_z .

Les tableaux 4.3 à 4.11 donnent les performances des méthodes en terme des erreurs de couverture et de la longueur moyenne des intervalles de confiance. Toutes les méthodes ont des comportements relativement voisins en terme des propriétés des intervalles de confiance, notamment les intervalles de confiance asymptotique, percentile et bootstrap de base. Ceci se manifeste clairement dans les quatre scénarios par des valeurs proches pour les erreurs de couverture bilatérale et pour la longueur moyenne des intervalles de confiance. Statistiquement parlant, l'erreur globale de couverture théorique (5%) est respectée par les intervalles de confiance asymptotique, percentile et bootstrap de base dans le cas du total t_z (tableau 4.8, tableau 4.9 et tableau 4.10). En revanche, elle n'est respectée que par les intervalles de confiance asymptotique et percentile dans le troisième scénario dans le cas du total t_y et pour certaines méthodes dans le dernier scénario. Toutefois, les taux de couverture bilatérale observés pour les intervalles de confiance asymptotique, percentile et bootstrap de base restent proches de 95%. Ils sont autour de 93% dans les scénarios 1 et 4 et autour de 92% dans le scénario 2 (tableau 4.3, tableau 4.4 et tableau 4.5). On peut également constater que les taux d'erreur inférieure et supérieure L et U de couverture unilatérale des intervalles de confiance asymptotique, percentile et bootstrap de base de t_y sont statistiquement différents de 2,5%. Ce constat ne s'applique pas à la lettre aux intervalles de confiance asymptotique, percentile et bootstrap de base de t_z . En fait, les deux taux d'erreur de couverture unilatérale observés L et U ne sont pas statistiquement différents de 2,5% dans le scénario 3 et seulement pour L dans le scénario 4.

Les intervalles de confiance t -bootstrap du t_y produits par les méthodes bootstrap ont pour la plupart des taux d'erreur de couverture unilatérale proches de la valeur théorique et respectent bien le taux de couverture bilatérale (tableau 4.6).

Pour le total t_z le t -bootstrap a amélioré les taux d'erreur L et U comparativement aux autres type d'intervalle de confiance, toutefois le taux de couverture global s'est trouvé affecté par ce gain de symétrie dans le scénario 2 et pour certaines méthodes bootstrap particulières (tableau 4.11). Il s'agit des méthodes suivantes: Bickel et Freedman (1984),

Sitter (1992b), McCarthy et Snowden (1985) Sitter (1992a), Rao et coll. (1992) mise en œuvre avec $n' = \lceil \frac{n}{2} \rceil$ et Antal et Tillé (2014). En effet, le taux de couverture bilatérale observé dans le deuxième scénario pour ces méthodes avoisine 97%.

Le gain apporté par le t -bootstrap en terme des taux d'erreur de couverture unilatérale vient du fait que les intervalles de confiance t -bootstrap sont plus larges que les intervalles percentiles ou asymptotiques (voir tableaux 4.7 et 4.12). Toutefois, il faut signaler que le choix de la méthode pour mettre en œuvre le t -bootstrap semble avoir un effet sur la longueur dudit intervalle de confiance. En effet, les intervalles de confiance t -bootstrap produits par les méthodes de Rao et Wu (1988) et Rao et coll. (1992) mises en œuvre avec $n' = \lceil \frac{n}{2} \rceil$, ainsi que ceux produits par Chipperfield et Preston (2007) et Antal et Tillé (2014) sont plus larges comparativement à ceux produits par les autres méthodes bootstrap. Il faut noter que cette remarque ne s'applique à la méthode de Antal et Tillé (2014) que dans les deux premiers scénarios. Ce constat peut être expliqué par le fait que ces méthodes manipulent seulement la moitié des données lors du second bootstrap pour estimer la variance de l'estimateur bootstrap. En effet, pour les méthodes de Rao et Wu (1988) et Rao et coll. (1992) elles sont mises en œuvre avec $n' = \lceil \frac{n}{2} \rceil$. Pour la méthode de Chipperfield et Preston (2007) l'échantillon bootstrap qui sera utilisé pour estimer la variance de l'estimateur bootstrap est aussi de taille $n' = \lceil \frac{n}{2} \rceil$. Pour la méthode de Antal et Tillé (2014), dans les deux premiers scénarios elle se comporte comme la méthode de Chipperfield et Preston (2007). En fait, les deux premiers scénarios se caractérisent par un taux de sondage relativement faible ($f = 0,07$). Pour simplifier les choses, supposons que $f = 0$ de telle sorte que les ajustements bootstrap pour chacune des deux méthodes sont égaux à $a_i^* = 2m_i$ où les $m_i, i = 1, \dots, n$ sont générés selon un mécanisme qui consiste à échantillonner sans remise dans s un échantillon de taille $n/2$. Pour vérifier cela il suffit de remplacer dans la formule relative à la construction des ajustements bootstrap de la méthode Chipperfield et Preston (2007) f par 0 et $\lceil \frac{n}{2} \rceil$ par $n/2$ qui est entier. Rappelons que la méthode de Antal et Tillé (2014) consiste à générer des ajustements bootstrap selon un mécanisme qui mixe deux plans d'échantillonnage, le plan de Bernoulli de paramètre f et un plan de type "One-One" qui consiste à sélectionner la moitié des unités selon le plan

aléatoire simple sans remise puis les doubler. Donc si $f = 0$ les ajustements bootstrap seront générés seulement selon le second plan et par conséquent $a_i^* = 2m_i$. D'où la ressemblance entre les deux méthodes dans les deux premiers scénarios qui se caractérisent par un taux de sondage proche de 0.

4.4.3. Résultats: cas des paramètres lisses de totaux

Le tableau 4.14 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne relative des estimateurs bootstrap de la variance de l'estimateur par substitution du ratio et du coefficient de corrélation. Les tableaux 4.15 à 4.18 et 4.20 à 4.23 donnent, pour les différents scénarios, les performances des méthodes bootstrap dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap du ratio et de coefficient de corrélation. Les deux tableaux 4.19 et 4.24 rapportent pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Booth et coll. (1994).

Tab. 4.14. Étude 1: Biais relatif et stabilité des méthodes bootstrap (paramètres lisses de totaux $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
$r_{y,z}$								
Booth et coll. (1994)	1,75	20,34	5,53	40,42	1,71	17,53	3,71	33,12
Chao et Lo (1994)	1,79	20,51	5,33	40,43	1,67	17,72	3,70	33,64
Bickel et Freedman (1984)	1,95	20,38	5,35	40,75	2,08	17,80	3,76	33,04
Sitter (1992a)	2,49	20,58	7,56	41,46	2,41	17,80	5,28	33,73
McCarthy et Snowden (1985)	2,27	20,56	8,52	41,99	2,23	17,84	4,47	33,26
Sitter (1992b)	2,04	20,46	6,96	41,10	2,00	17,73	4,41	32,71
Rao et Wu (1988)	2,20	20,70	6,63	40,85	2,08	17,68	3,41	32,42
Rao et Wu (1988) ($n' = n/2$)	1,99	20,69	5,86	39,72	2,00	17,80	2,87	31,95
Rao et coll. (1992)	2,01	20,45	6,51	40,70	2,00	17,65	3,59	32,97
Rao et coll. (1992) ($n' = n/2$)	2,06	20,43	5,94	40,09	1,70	17,47	3,21	32,92
Chipperfield et Preston (2007)	2,43	21,02	7,55	41,95	2,35	17,73	4,69	33,75
Antal et Tillé (2014)	2,54	20,81	7,81	41,89	2,12	17,53	4,91	33,89
Antal et Tillé (2011a)	2,34	20,34	6,49	40,33	2,05	17,80	3,33	32,57
Beaumont et Patak (2012) Norm	2,58	21,06	7,87	42,30	2,17	17,88	5,26	33,98
Beaumont et Patak (2012) Exp	1,83	20,43	6,35	40,03	1,99	17,68	2,88	32,20
Beaumont et Patak (2012) $\epsilon = 1$	2,36	20,66	7,96	42,31	2,30	17,61	5,08	34,04
$\rho_{y,z}$								
Booth et coll. (1994)	3,12	36,90	26,88	89,71	3,17	29,00	6,48	61,14
Chao et Lo (1994)	2,96	36,60	26,41	88,47	3,20	28,73	5,88	59,32
Bickel et Freedman (1984)	2,97	37,03	26,75	88,77	2,76	28,69	6,40	60,23
Sitter (1992a)	3,83	37,26	29,69	92,14	3,53	28,92	8,22	62,31
McCarthy et Snowden (1985)	1,89	36,61	27,56	88,52	0,60	27,93	0,08	53,37
Sitter (1992b)	2,49	36,80	26,90	89,32	1,80	29,20	6,85	59,43
Rao et Wu (1988)	1,23	36,54	24,32	84,59	-1,28	28,08	-4,88	50,49
Rao et Wu (1988) ($n' = n/2$)	-2,43	35,00	14,39	71,16	-2,83	27,93	-12,17	46,07
Rao et coll. (1992)	1,33	36,44	24,31	85,10	-0,89	28,54	-5,21	49,84
Rao et coll. (1992) ($n' = n/2$)	-2,53	34,88	14,25	70,82	-2,81	27,76	-12,30	46,06
Chipperfield et Preston (2007)	7,57	41,89	49,39	116,01	2,14	30,55	10,37	62,36
Antal et Tillé (2014)	7,77	41,24	48,86	116,75	3,32	29,75	13,23	67,02
Antal et Tillé (2011a)	1,44	36,28	24,21	86,34	-0,91	27,46	-4,17	51,78
Beaumont et Patak (2012) Norm	-3,01	35,51	9,48	72,42	-3,73	27,63	-11,72	47,81
Beaumont et Patak (2012) Exp	-6,63	33,52	1,06	60,39	-5,19	27,01	-18,55	45,19
Beaumont et Patak (2012) $\epsilon = 1$	3,26	38,61	23,79	85,67	1,66	30,23	1,30	54,98

Tab. 4.15. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	2,2	1,9	4,1	1,000	1,7	3,3	5,0	1,000	1,6	3,0	4,6	1,000	1,5	3,2	4,7	1,000
Chao et Lo (1994)	2,3	2,3	4,6	1,000	1,7	3,2	4,9	1,000	1,7	3,3	5,0	1,000	1,5	3,2	4,7	1,000
Bickel et Freedman (1984)	2,2	2,3	4,5	1,000	1,6	3,2	4,8	1,000	1,5	2,8	4,3	1,000	1,7	3,0	4,7	1,000
Sitter (1992a)	2,2	1,8	4,0	1,006	1,6	3,1	4,7	1,007	1,5	2,9	4,4	1,000	1,4	3,0	4,4	1,008
McCarthy et Snowden (1985)	2,1	2,1	4,2	1,000	1,6	3,0	4,6	1,014	1,5	3,1	4,6	1,000	1,4	3,0	4,4	1,004
Sitter (1992b)	2,3	2,3	4,6	1,000	1,6	3,1	4,7	1,007	1,5	3,0	4,5	1,000	1,5	2,9	4,4	1,004
Rao et Wu (1988)	2,0	2,1	4,1	1,000	1,6	3,1	4,7	1,007	1,6	2,8	4,4	1,000	1,5	2,9	4,4	1,000
Rao et Wu (1988) ($n' = n/2$)	2,0	2,2	4,2	1,000	1,7	3,1	4,8	1,007	1,6	3,0	4,6	1,000	1,5	2,9	4,4	1,000
Rao et coll. (1992)	2,4	2,1	4,5	1,000	1,6	3,4	5,0	1,007	1,4	3,1	4,5	1,000	1,4	2,9	4,3	1,000
Rao et coll. (1992) ($n' = n/2$)	2,2	2,2	4,4	1,000	1,7	3,2	4,9	1,000	1,7	3,1	4,8	1,000	1,2	3,1	4,3	1,000
Chipperfield et Preston (2007)	2,0	2,2	4,2	1,006	1,6	3,1	4,7	1,007	1,6	3,0	4,6	1,000	1,5	2,9	4,4	1,008
Antal et Tillé (2014)	2,3	2,3	4,6	1,006	1,6	3,1	4,7	1,011	1,7	3,0	4,7	1,000	1,4	2,9	4,3	1,008
Antal et Tillé (2011a)	2,1	2,2	4,3	1,000	1,8	3,0	4,8	1,007	1,7	2,9	4,6	1,000	1,5	3,3	4,8	1,000
Beaumont et Patak (2012) Norm	2,0	2,2	4,2	1,006	1,6	3,1	4,7	1,014	1,5	3,1	4,6	1,000	1,5	3,1	4,6	1,008
Beaumont et Patak (2012) Exp	2,3	2,3	4,6	1,000	1,6	3,5	5,1	1,007	1,5	2,7	4,2	1,000	1,5	3,0	4,5	1,000
Beaumont et Patak (2012) $\epsilon = 1$	2,2	2,2	4,4	1,006	1,5	3,3	4,8	1,007	1,8	3,3	5,1	1,000	1,5	2,9	4,4	1,008

Tab. 4.16. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	3,3	1,5	4,8	1,000	3,4	2,3	5,7	1,000	2,6	2,2	4,8	1,000	3,4	2,4	5,8	1,000
Chao et Lo (1994)	3,3	1,6	4,9	1,000	3,5	2,3	5,8	0,996	2,9	2,3	5,2	1,000	3,3	2,2	5,5	0,996
Bickel et Freedman (1984)	3,3	1,4	4,7	1,006	3,6	2,3	5,9	0,996	2,8	2,2	5,0	1,000	3,1	2,1	5,2	1,000
Sitter (1992a)	3,3	1,4	4,7	1,006	3,5	2,3	5,8	1,007	2,9	1,9	4,8	1,000	3,0	1,9	4,9	1,008
McCarthy et Snowden (1985)	3,3	1,4	4,7	1,006	3,5	2,2	5,7	1,014	2,8	2,4	5,2	1,000	3,0	2,0	5,0	1,008
Sitter (1992b)	3,2	1,3	4,5	1,006	3,6	2,4	6,0	1,011	2,9	1,9	4,8	1,000	3,1	2,4	5,5	1,008
Rao et Wu (1988)	3,4	2,0	5,4	1,006	3,6	2,5	6,1	1,007	2,4	1,8	4,2	1,000	3,0	2,1	5,1	1,004
Rao et Wu (1988) ($n' = n/2$)	3,1	1,8	4,9	1,006	3,4	2,3	5,7	1,007	2,4	2,4	4,8	1,000	2,8	2,3	5,1	1,004
Rao et coll. (1992)	3,2	1,2	4,4	1,006	3,3	2,6	5,9	1,004	2,5	2,3	4,8	1,000	3,0	2,3	5,3	1,004
Rao et coll. (1992) ($n' = n/2$)	3,2	1,5	4,7	1,006	3,8	2,3	6,1	1,004	2,6	2,3	4,9	1,000	2,9	2,2	5,1	1,004
Chipperfield et Preston (2007)	3,2	1,3	4,5	1,006	3,7	2,4	6,1	1,007	2,8	2,4	5,2	1,000	3,4	2,1	5,5	1,004
Antal et Tillé (2014)	3,3	1,4	4,7	1,006	3,6	2,3	5,9	1,011	2,4	2,0	4,4	0,992	3,4	2,2	5,6	1,004
Antal et Tillé (2011a)	3,2	1,5	4,7	1,006	3,7	2,2	5,9	1,004	2,6	1,9	4,5	1,000	3,0	2,5	5,5	1,004
Beaumont et Patak (2012) Norm	3,5	1,4	4,9	1,006	3,7	2,3	6,0	1,011	2,5	2,3	4,8	1,000	3,2	1,9	5,1	1,012
Beaumont et Patak (2012) Exp	3,2	1,5	4,7	1,000	3,4	2,3	5,7	1,014	2,6	2,0	4,6	1,000	3,1	2,2	5,3	1,008
Beaumont et Patak (2012) $\epsilon = 1$	3,1	1,5	4,6	1,006	3,8	2,2	6,0	1,007	2,7	2,2	4,9	1,000	3,4	2,0	5,4	1,004

Tab. 4.17. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	1,6	3,2	4,8	1,000	0,8	5,2	6,0	1,000	1,2	3,8	5,0	1,000	0,7	5,4	6,1	1,000
Chao et Lo (1994)	1,2	3,3	4,5	1,000	0,7	5,3	6,0	0,996	0,9	3,9	4,8	1,000	1,0	5,3	6,3	0,996
Bickel et Freedman (1984)	1,1	3,3	4,4	1,006	0,7	5,1	5,8	0,996	1,2	4,2	5,4	1,000	0,6	5,0	5,6	1,000
Sitter (1992a)	1,4	3,4	4,8	1,006	0,4	4,9	5,3	1,007	0,8	4,3	5,1	1,000	0,5	5,4	5,9	1,008
McCarthy et Snowden (1985)	1,5	3,0	4,5	1,006	0,5	5,3	5,8	1,014	1,1	4,0	5,1	1,000	0,5	5,2	5,7	1,008
Sitter (1992b)	1,0	3,4	4,4	1,006	0,5	4,7	5,2	1,011	1,1	4,1	5,2	1,000	0,6	5,1	5,7	1,008
Rao et Wu (1988)	1,1	3,0	4,1	1,006	0,5	5,3	5,8	1,007	1,2	3,9	5,1	1,000	0,5	5,3	5,8	1,004
Rao et Wu (1988) ($n' = n/2$)	1,3	3,2	4,5	1,006	0,7	4,7	5,4	1,007	1,4	4,0	5,4	1,000	0,7	5,0	5,7	1,004
Rao et coll. (1992)	1,4	3,4	4,8	1,006	0,5	4,6	5,1	1,004	0,9	4,2	5,1	1,000	0,6	5,2	5,8	1,004
Rao et coll. (1992) ($n' = n/2$)	1,3	3,2	4,5	1,006	0,4	5,0	5,4	1,004	1,0	4,1	5,1	1,000	0,7	5,3	6,0	1,004
Chipperfield et Preston (2007)	1,3	3,2	4,5	1,006	0,2	5,1	5,3	1,007	1,0	4,2	5,2	1,000	0,7	5,3	6,0	1,004
Antal et Tillé (2014)	0,9	3,6	4,5	1,006	0,6	5,3	5,9	1,011	1,0	4,1	5,1	0,992	0,5	5,4	5,9	1,004
Antal et Tillé (2011a)	1,1	3,2	4,3	1,006	0,6	5,0	5,6	1,004	1,3	4,1	5,4	1,000	0,5	5,0	5,5	1,004
Beaumont et Patak (2012) Norm	1,0	3,3	4,3	1,006	0,4	5,0	5,4	1,011	1,3	4,1	5,4	1,000	0,6	5,1	5,7	1,012
Beaumont et Patak (2012) Exp	1,2	3,0	4,2	1,000	0,5	4,7	5,2	1,014	1,2	3,7	4,9	1,000	0,5	4,8	5,3	1,008
Beaumont et Patak (2012) $\epsilon = 1$	1,4	3,5	4,9	1,006	0,4	5,0	5,4	1,007	1,1	4,0	5,1	1,000	0,4	5,5	5,9	1,004

Tab. 4.18. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) ($r_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	2,8	1,6	4,4	1,000	3,0	2,2	5,2	1,000	1,9	1,7	3,6	1,000	2,4	2,1	4,5	1,000
Chao et Lo (1994)	3,0	1,6	4,6	0,994	2,9	2,1	5,0	1,000	2,1	2,4	4,5	1,000	2,1	2,3	4,4	1,000
Bickel et Freedman (1984)	2,7	1,1	3,8	1,000	3,1	1,9	5,0	1,000	2,2	1,9	4,1	1,007	2,5	2,2	4,7	1,004
Sitter (1992a)	2,8	1,0	3,8	1,000	2,7	2,3	5,0	1,014	2,0	2,1	4,1	1,007	2,2	2,0	4,2	1,012
McCarthy et Snowden (1985)	2,8	1,1	3,9	1,000	3,1	1,9	5,0	1,021	1,7	2,2	3,9	1,007	2,2	1,9	4,1	1,023
Sitter (1992b)	3,0	1,3	4,3	1,000	2,8	1,8	4,6	1,010	2,1	2,4	4,5	1,015	2,4	2,0	4,4	1,020
Rao et Wu (1988)	2,7	1,1	3,8	1,000	2,7	2,1	4,8	1,014	2,2	1,9	4,1	1,007	2,1	1,9	4,0	1,020
Rao et Wu (1988) ($n' = n/2$)	2,2	1,5	3,7	1,012	2,1	1,9	4,0	1,052	2,1	2,1	4,2	1,022	1,6	2,0	3,6	1,059
Rao et coll. (1992)	2,8	1,2	4,0	1,000	3,1	2,0	5,1	1,003	2,0	2,0	4,0	1,000	2,2	2,1	4,3	1,012
Rao et coll. (1992) ($n' = n/2$)	2,6	1,3	3,9	1,012	1,8	1,9	3,7	1,035	2,1	2,4	4,5	1,015	1,9	2,1	4,0	1,039
Chipperfield et Preston (2007)	2,5	1,1	3,6	1,012	2,4	1,8	4,2	1,038	1,9	1,6	3,5	1,015	1,9	2,1	4,0	1,047
Antal et Tillé (2014)	2,7	1,3	4,0	1,006	2,8	1,7	4,5	1,024	2,4	2,4	4,8	1,007	2,3	2,2	4,5	1,012
Antal et Tillé (2011a)	2,6	1,3	3,9	1,000	3,5	2,1	5,6	1,007	2,2	2,0	4,2	1,000	2,6	2,3	4,9	1,000

Tab. 4.19. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (le ratio $r_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,006	1,031	1,007	1,029	1,008	1,023	1,008	1,032
Chao et Lo (1994)	1,006	1,025	1,007	1,029	1,008	1,023	1,008	1,032
Bickel et Freedman (1984)	1,006	1,031	1,007	1,029	1,008	1,031	1,008	1,036
Sitter (1992a)	1,012	1,031	1,014	1,043	1,008	1,031	1,016	1,044
McCarthy et Snowden (1985)	1,006	1,031	1,022	1,050	1,008	1,031	1,012	1,056
Sitter (1992b)	1,006	1,031	1,014	1,039	1,008	1,038	1,012	1,052
Rao et Wu (1988)	1,006	1,031	1,014	1,043	1,008	1,031	1,008	1,052
Rao et Wu (1988) ($n' = n/2$)	1,006	1,043	1,007	1,082	1,008	1,046	1,008	1,093
Rao et coll. (1992)	1,006	1,031	1,014	1,032	1,008	1,023	1,008	1,044
Rao et coll. (1992) ($n' = n/2$)	1,006	1,043	1,014	1,065	1,008	1,038	1,008	1,073
Chipperfield et Preston (2007)	1,012	1,043	1,014	1,068	1,008	1,038	1,016	1,081
Antal et Tillé (2014)	1,012	1,037	1,022	1,054	1,008	1,031	1,016	1,048
Antal et Tillé (2011a)	1,012	1,031	1,014	1,039	1,008	1,023	1,008	1,032
Beaumont et Patak (2012) Normal	1,012	–	1,022	–	1,008	–	1,016	–
Beaumont et Patak (2012) Exp	1,006	–	1,014	–	1,008	–	1,008	–
Beaumont et Patak (2012) $\epsilon = 1$	1,012	–	1,014	–	1,008	–	1,016	–

Tab. 4.20. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	6,4	1,4	7,8	1,000	6,0	0,7	6,7	1,000	4,7	1,4	6,1	1,000	6,4	2,2	8,6	1,000
Chao et Lo (1994)	6,4	1,2	7,6	1,000	6,0	0,7	6,7	1,000	4,5	1,4	5,9	1,000	6,5	2,5	9,0	1,000
Bickel et Freedman (1984)	6,4	1,1	7,5	1,000	5,7	0,6	6,3	1,000	4,6	1,4	6,0	1,000	6,1	2,4	8,5	1,000
Sitter (1992a)	6,4	1,3	7,7	1,009	5,8	0,6	6,4	1,011	4,7	1,4	6,1	1,011	6,1	2,4	8,5	1,000
McCarthy et Snowden (1985)	6,3	1,3	7,6	0,991	6,0	0,8	6,8	1,000	4,6	1,4	6,0	0,989	6,6	2,4	9,0	0,975
Sitter (1992b)	6,2	1,4	7,6	1,000	5,7	0,8	6,5	1,000	4,6	1,5	6,1	0,989	6,2	2,2	8,4	1,000
Rao et Wu (1988)	6,4	1,6	8,0	0,991	5,8	0,8	6,6	0,989	4,8	1,5	6,3	0,979	7,1	2,6	9,7	0,949
Rao et Wu (1988) ($n' = n/2$)	7,0	1,6	8,6	0,973	5,9	1,0	6,9	0,953	5,0	1,4	6,4	0,968	7,4	3,3	10,7	0,911
Rao et coll. (1992)	6,3	1,4	7,7	0,991	6,0	0,8	6,8	0,989	4,9	1,5	6,4	0,979	7,2	2,4	9,6	0,949
Rao et coll. (1992) ($n' = n/2$)	6,7	1,5	8,2	0,973	6,1	0,8	6,9	0,953	5,1	1,3	6,4	0,968	7,4	3,1	10,5	0,911
Chipperfield et Preston (2007)	5,3	1,5	6,8	1,018	4,7	0,4	5,1	1,079	4,6	1,6	6,2	0,989	6,0	2,0	8,0	1,013
Antal et Tillé (2014)	5,5	1,3	6,8	1,027	5,0	0,3	5,3	1,074	4,6	1,4	6,0	1,000	5,6	2,2	7,8	1,025
Antal et Tillé (2011a)	6,7	1,3	8,0	0,991	5,9	0,8	6,7	0,989	4,9	1,4	6,3	0,979	7,1	3,0	10,1	0,949
Beaumont et Patak (2012) Norm	6,8	2,0	8,8	0,973	6,9	1,0	7,9	0,932	5,0	1,6	6,6	0,968	7,6	3,3	10,9	0,911
Beaumont et Patak (2012) Exp	7,3	1,8	9,1	0,955	7,3	1,2	8,5	0,900	5,0	1,4	6,4	0,968	7,9	3,5	11,4	0,873
Beaumont et Patak (2012) $\epsilon = 1$	6,2	1,4	7,6	1,000	5,9	0,9	6,8	0,989	4,5	1,4	5,9	0,989	6,6	2,5	9,1	0,975

Tab. 4.21. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	4,5	1,3	5,8	1,000	4,0	1,7	5,7	1,000	3,5	1,5	5,0	1,000	4,3	3,6	7,9	1,000
Chao et Lo (1994)	4,8	1,0	5,8	1,000	4,4	1,5	5,9	0,995	3,8	1,5	5,3	1,000	4,7	3,6	8,3	1,000
Bickel et Freedman (1984)	4,9	1,3	6,2	1,000	4,2	1,3	5,5	0,995	3,6	1,5	5,1	1,000	4,2	3,6	7,8	1,000
Sitter (1992a)	4,8	1,2	6,0	1,000	3,7	1,3	5,0	1,011	3,8	1,8	5,6	1,000	3,9	3,4	7,3	1,013
McCarthy et Snowden (1985)	5,2	1,1	6,3	1,000	4,0	1,4	5,4	1,011	4,0	1,5	5,5	1,000	4,7	3,5	8,2	0,974
Sitter (1992b)	4,4	1,1	5,5	1,000	4,2	1,3	5,5	1,005	3,9	1,7	5,6	1,000	3,6	3,6	7,2	1,013
Rao et Wu (1988)	5,0	1,6	6,6	1,000	4,0	1,6	5,6	1,000	4,1	1,5	5,6	0,989	4,7	3,8	8,5	0,949
Rao et Wu (1988) ($n' = n/2$)	5,4	1,7	7,1	0,973	4,1	2,3	6,4	0,952	4,1	1,5	5,6	0,978	5,3	3,8	9,1	0,904
Rao et coll. (1992)	4,4	1,2	5,6	1,000	4,1	1,8	5,9	1,000	4,1	1,7	5,8	0,989	5,1	4,1	9,2	0,949
Rao et coll. (1992) ($n' = n/2$)	5,3	1,6	6,9	0,973	4,2	2,0	6,2	0,952	4,0	1,5	5,5	0,978	5,3	4,1	9,4	0,904
Chipperfield et Preston (2007)	4,3	1,1	5,4	1,027	3,4	1,3	4,7	1,080	3,5	1,7	5,2	1,000	3,6	3,8	7,4	1,019
Antal et Tillé (2014)	3,6	1,2	4,8	1,018	4,0	1,5	5,5	1,080	3,2	1,6	4,8	1,000	3,6	3,5	7,1	1,032
Antal et Tillé (2011a)	5,1	1,3	6,4	1,000	4,1	1,8	5,9	1,000	4,0	1,5	5,5	0,989	4,9	3,4	8,3	0,962
Beaumont et Patak (2012) Norm	5,8	2,1	7,9	0,982	6,3	1,3	7,6	0,941	4,3	1,5	5,8	0,978	6,8	4,0	10,8	0,923
Beaumont et Patak (2012) Exp	6,7	1,6	8,3	0,964	5,8	1,4	7,2	0,920	4,5	1,5	6,0	0,968	6,8	3,6	10,4	0,891
Beaumont et Patak (2012) $\epsilon = 1$	4,5	1,5	6,0	1,000	4,3	1,5	5,8	0,984	3,5	1,9	5,4	1,000	4,6	4,4	9,0	0,968

Tab. 4.22. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	8,7	1,9	10,6	1,000	10,7	0,4	11,1	1,000	6,7	1,3	8,0	1,000	10,0	2,4	12,4	1,000
Chao et Lo (1994)	8,7	1,8	10,5	1,000	10,5	0,6	11,1	0,995	6,7	1,5	8,2	1,000	9,6	2,3	11,9	1,000
Bickel et Freedman (1984)	8,3	2,0	10,3	1,000	10,9	0,6	11,5	0,995	6,6	1,3	7,9	1,000	9,6	2,2	11,8	1,000
Sitter (1992a)	8,7	1,9	10,6	1,000	10,4	0,4	10,8	1,011	6,6	1,4	8,0	1,000	9,4	2,3	11,7	1,013
McCarthy et Snowden (1985)	8,9	1,8	10,7	1,000	10,6	0,6	11,2	1,011	6,5	1,4	7,9	1,000	9,7	2,2	11,9	0,974
Sitter (1992b)	8,9	1,8	10,7	1,000	10,4	0,4	10,8	1,005	6,7	1,4	8,1	1,000	9,6	2,4	12,0	1,013
Rao et Wu (1988)	8,5	1,8	10,3	1,000	10,5	0,4	10,9	1,000	6,6	1,6	8,2	0,989	10,1	2,3	12,4	0,949
Rao et Wu (1988) ($n' = n/2$)	8,6	1,8	10,4	0,973	11,2	0,4	11,6	0,952	6,2	1,3	7,5	0,978	10,3	2,2	12,5	0,904
Rao et coll. (1992)	8,4	1,7	10,1	1,000	10,8	0,5	11,3	1,000	6,3	1,3	7,6	0,989	10,0	2,3	12,3	0,949
Rao et coll. (1992) ($n' = n/2$)	8,4	1,9	10,3	0,973	11,1	0,7	11,8	0,952	6,4	1,4	7,8	0,978	10,2	2,4	12,6	0,904
Chipperfield et Preston (2007)	8,6	2,1	10,7	1,027	9,8	0,2	10,0	1,080	6,6	1,1	7,7	1,000	9,8	1,7	11,5	1,019
Antal et Tillé (2014)	8,4	2,0	10,4	1,018	10,2	0,4	10,6	1,080	7,0	1,1	8,1	1,000	9,6	1,9	11,5	1,032
Antal et Tillé (2011a)	8,5	1,9	10,4	1,000	11,0	0,5	11,5	1,000	6,4	1,8	8,2	0,989	9,7	2,7	12,4	0,962
Beaumont et Patak (2012) Norm	7,1	2,1	9,2	0,982	8,2	1,1	9,3	0,941	5,7	1,3	7,0	0,978	8,4	2,8	11,2	0,923
Beaumont et Patak (2012) Exp	7,9	2,2	10,1	0,964	9,3	0,8	10,1	0,920	6,1	2,1	8,2	0,968	9,3	3,1	12,4	0,891
Beaumont et Patak (2012) $\epsilon = 1$	8,7	2,1	10,8	1,000	10,2	0,5	10,7	0,984	6,8	1,0	7,8	1,000	9,8	2,4	12,2	0,968

Tab. 4.23. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) ($\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	3,0	3,4	6,4	1,000	2,3	2,3	4,6	1,000	2,9	2,8	5,7	1,000	2,1	4,2	6,3	1,000
Chao et Lo (1994)	3,3	3,4	6,7	1,000	2,4	2,3	4,7	0,995	2,6	2,0	4,6	1,000	2,5	4,0	6,5	1,000
Bickel et Freedman (1984)	3,6	3,2	6,8	0,992	2,2	2,2	4,4	0,995	2,8	2,9	5,7	1,000	2,2	4,2	6,4	1,000
Sitter (1992a)	2,9	3,4	6,3	1,000	2,4	2,0	4,4	1,014	2,7	3,1	5,8	1,000	2,0	4,0	6,0	1,012
McCarthy et Snowden (1985)	2,8	3,4	6,2	1,000	2,3	1,7	4,0	1,014	2,8	2,6	5,4	0,990	1,9	4,3	6,2	1,006
Sitter (1992b)	3,1	3,3	6,4	1,000	2,4	2,5	4,9	1,005	3,1	2,4	5,5	0,990	2,3	4,1	6,4	1,017
Rao et Wu (1988)	3,2	3,3	6,5	0,992	2,3	1,8	4,1	1,029	2,6	2,3	4,9	0,990	2,6	3,5	6,1	1,023
Rao et Wu (1988) ($n' = n/2$)	3,5	2,6	6,1	1,033	2,0	1,5	3,5	1,148	3,1	1,1	4,2	1,041	3,3	2,2	5,5	1,075
Rao et coll. (1992)	3,2	3,3	6,5	0,992	2,4	2,0	4,4	1,014	3,2	1,9	5,1	0,990	3,2	3,3	6,5	1,006
Rao et coll. (1992) ($n' = n/2$)	3,7	2,6	6,3	1,017	2,6	1,0	3,6	1,120	3,4	1,2	4,6	1,041	3,0	2,5	5,5	1,064
Chipperfield et Preston (2007)	3,0	2,8	5,8	1,058	2,2	1,5	3,7	1,163	2,8	2,1	4,9	1,020	1,9	2,3	4,2	1,173
Antal et Tillé (2014)	3,2	2,9	6,1	1,042	1,8	1,7	3,5	1,110	2,7	2,7	5,4	1,000	2,0	3,7	5,7	1,052
Antal et Tillé (2011a)	3,1	3,2	6,3	0,983	2,2	2,0	4,2	1,000	2,8	3,0	5,8	0,980	3,0	3,5	6,5	0,965

Tab. 4.24. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (Le coefficient de corrélation $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,009	1,081	1,016	1,118	1,011	1,054	1,013	1,109
Chao et Lo (1994)	1,009	1,081	1,016	1,112	1,011	1,054	1,013	1,109
Bickel et Freedman (1984)	1,009	1,072	1,016	1,112	1,011	1,054	1,013	1,109
Sitter (1992a)	1,018	1,081	1,027	1,134	1,022	1,054	1,013	1,122
McCarthy et Snowden (1985)	1,000	1,081	1,016	1,134	1,000	1,043	0,987	1,115
Sitter (1992b)	1,009	1,081	1,016	1,123	1,000	1,043	1,013	1,128
Rao et Wu (1988)	1,000	1,072	1,005	1,150	0,989	1,043	0,962	1,135
Rao et Wu (1988) ($n' = n/2$)	0,982	1,117	0,968	1,283	0,978	1,097	0,923	1,192
Rao et coll. (1992)	1,000	1,072	1,005	1,134	0,989	1,043	0,962	1,115
Rao et coll. (1992) ($n' = n/2$)	0,982	1,099	0,968	1,251	0,978	1,097	0,923	1,179
Chipperfield et Preston (2007)	1,027	1,144	1,091	1,299	1,000	1,075	1,026	1,301
Antal et Tillé (2014)	1,036	1,126	1,091	1,241	1,011	1,054	1,038	1,167
Antal et Tillé (2011a)	1,000	1,063	1,005	1,118	1,000	1,032	0,962	1,071
Beaumont et Patak (2012) Normal	0,982	–	0,947	–	0,978	–	0,923	–
Beaumont et Patak (2012) Exp	0,964	–	0,914	–	0,978	–	0,885	–
Beaumont et Patak (2012) $\epsilon = 1$	1,009	–	1,005	–	1,000	–	0,987	–

4.4.4. Discussion: cas des paramètres lisses de totaux

Le tableau 4.14 met en évidence des faits que nous énumérons dans les points suivants:

- Les différents estimateurs bootstrap de la variance, dans le cas du ratio, sont relativement semblables et plus stables que dans le cas du coefficient de corrélation;
- Les estimateurs bootstrap de la variance sont plus stables dans les scénarios qui se caractérisent par une taille d'échantillon importante, $n = 150$.
- Les trois premières méthodes bootstrap par pseudo-population, à savoir Booth et coll. (1994), Chao et Lo (1994) et Bickel et Freedman (1984) sont semblables et font relativement mieux que la méthode de Sitter (1992a) en terme de la précision et de la stabilité pour les paramètres étudiés;
- La méthode de McCarthy et Snowden (1985) fait mieux que les autres en terme de BR dans le cas du coefficient de corrélation, si la fraction de sondage est importante;
- Le choix de la distribution pour générer les ajustements bootstrap pour la méthode de Beaumont et Patak (2012) et le choix de la taille de l'échantillon bootstrap n' pour les méthodes de Rao et Wu (1988) et de Rao et coll. (1992) semblent affecter la précision et la stabilité des estimateurs bootstrap de la variance dans le cas du coefficient de corrélation, chose qui n'était pas constatée dans le cas des deux totaux et du ratio. Toutefois, il n'y a pas un réglage optimal qui permet d'obtenir des résultats performants dans tous les scénarios;
- La méthode de Rao et coll. (1992) donne presque les mêmes résultats que la méthode de Rao et Wu (1988). Ce résultat n'est pas surprenant et nous aurions obtenu des résultats identiques si les estimateurs bootstrap avaient été calculés sur la base des mêmes échantillons bootstrap;
- La méthode de Chipperfield et Preston (2007) et la méthode de Antal et Tillé (2014) ont un comportement voisin dans les deux premiers scénarios et elles sont moins précises et moins stables que les autres méthodes, notamment dans le cas du coefficient de corrélation.

Les tableaux 4.15 à 4.17 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance du ratio. Toutes les méthodes ont des comportements voisins pour un même type d'intervalles de confiance. Le taux de couverture théorique est bien respecté par les intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap (les taux de couverture observés sont compris entre 94% et 97% et ceci dans tous les scénarios). Le t -bootstrap dans le cas du ratio ne semble pas apporter un gain particulier sur les taux d'erreur unilatérale comparative-ment aux intervalles de confiance asymptotique et percentile malgré que les intervalles de confiance t -bootstrap soient plus larges d'environ 3% (voir tableau 4.19).

En ce qui concerne le coefficient de corrélation et à la lumière des résultats indiqués dans les tableaux 4.20 à 4.22, on constate dans les scénarios 1 et 4 et pour toutes les méthodes bootstrap que les taux de couverture observés des intervalles asymptotiques diffèrent statistiquement du taux théorique (les taux observés sont compris entre 90,9% et 93,2% dans le scénario 1 et entre 89,6% et 92,2% dans le scénario 4). Dans le scénario 3 les taux de couverture observés sont très proches du taux théorique, ils sont compris entre de 93,4% et 94,1%. Dans le scénario 2 et seulement pour certaines méthodes le taux de couverture est bien respecté. Il s'agit des méthodes Bickel et Freedman (1984), Sitter (1992a), Chipperfield et Preston (2007) et Antal et Tillé (2014). Toutefois, cette performance n'est attribuable qu'à la grande surestimation de la variance et en particulier pour les deux dernières méthodes. Les intervalles de confiance percentile font mieux que les intervalles asymptotiques et bootstrap de base en terme de la couverture bilatérale et des taux d'erreur unilatérale. Ce constat est vrai pour toutes les méthodes bootstrap. Il faut noter que la construction d'un intervalle de confiance via la technique des percentiles ou bootstrap de base conduit à un intervalle de confiance identique si la distribution bootstrap est parfaitement symétrique. Or, dans le cas du coefficient de corrélation les taux de couverture bilatérale observés des intervalles de confiance bootstrap de base et ceux des intervalles percentiles sont différents; par exemple dans le premier scénario les taux de couverture observés sont autour de 94% pour les intervalles de confiance percentile par contre ils sont autour de 90% pour les intervalles bootstrap de base.

Les intervalles de confiance t -bootstrap du $\rho_{y,z}$ produits par les méthodes bootstrap ont pour la plupart des taux d'erreur de couverture très proches des valeurs théoriques (tableau 4.23). Toutefois, il faut signaler que les intervalles de confiance t -bootstrap sont plus longs que les intervalles de confiance percentile. En conséquence le choix de la méthode bootstrap pour mettre en œuvre le t -bootstrap doit être basé sur le critère de la longueur. Les résultats indiqués dans le tableau 4.24 montrent que le choix de la méthode Antal et Tillé (2011a) pour mettre en œuvre le t -bootstrap dans le cas du coefficient de corrélation est optimal.

4.4.5. Résultats: cas des paramètres non lisses

Le tableau 4.25 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne relative des estimateurs bootstrap de la variance des estimateurs de la médiane et de l'indice de Gini de la variable y . Les tableaux 4.26 à 4.29 et 4.31 à 4.34 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap de la médiane et de l'indice de Gini de la variable y . Les deux tableaux 4.30 et 4.35 rapportent pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Booth et coll. (1994).

Tab. 4.25. Étude 1: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses de totaux, m_y et G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
m_y								
Booth et coll. (1994)	14,13	56,33	18,97	67,78	7,45	40,58	30,67	80,80
Chao et Lo (1994)	14,15	56,59	18,54	67,29	7,52	40,91	30,35	79,59
Bickel et Freedman (1984)	13,16	55,63	15,97	64,31	6,09	40,03	26,87	75,83
Sitter (1992a)	15,32	57,65	21,97	69,27	6,99	40,38	30,00	78,08
McCarthy et Snowden (1985)	10,79	53,11	14,63	62,30	5,83	39,12	27,00	73,32
Sitter (1992b)	10,71	52,94	14,87	62,24	4,31	38,48	23,72	71,23
Rao et Wu (1988)	11,58	52,48	16,51	62,71	2,69	33,35	25,51	64,72
Rao et Wu (1988) ($n' = n/2$)	10,12	42,14	14,12	56,04	-1,87	26,41	20,10	51,74
Rao et coll. (1992)	17,10	57,57	29,06	73,53	5,19	38,55	33,22	80,33
Rao et coll. (1992) ($n' = n/2$)	12,88	54,78	21,06	67,49	8,38	40,39	36,78	82,26
Chipperfield et Preston (2007)	18,06	57,48	19,82	66,40	4,03	39,04	40,49	88,91
Antal et Tillé (2014)	9,40	52,93	10,71	60,19	4,30	38,51	21,98	71,59
Antal et Tillé (2011a)	9,89	52,26	13,26	61,31	5,07	38,77	25,54	73,78
Beaumont et Patak (2012) Norm	10,15	53,06	13,34	61,92	4,77	38,46	26,61	74,39
Beaumont et Patak (2012) Exp	13,47	54,38	20,96	66,65	6,43	39,11	33,88	78,78
Beaumont et Patak (2012) $\epsilon = 1$	11,51	53,78	17,59	65,19	4,96	38,80	23,84	71,72
G_y								
Booth et coll. (1994)	-4,16	39,96	-11,95	54,95	-5,96	35,67	-6,87	29,52
Chao et Lo (1994)	-3,98	40,54	-12,03	55,00	-5,74	36,09	-6,90	29,09
Bickel et Freedman (1984)	-4,06	40,45	-11,85	55,28	-5,77	35,95	-6,81	29,21
Sitter (1992a)	-3,30	40,56	-10,11	56,07	-5,53	35,80	-5,40	29,16
McCarthy et Snowden (1985)	-5,02	40,05	-13,01	54,94	-6,97	35,54	-10,11	29,63
Sitter (1992b)	-4,22	40,15	-13,24	55,07	-6,12	35,67	-7,61	29,34
Rao et Wu (1988)	-2,22	41,10	-10,89	55,63	3,47	38,98	0,56	30,43
Rao et Wu (1988) ($n' = n/2$)	16,48	50,76	2,48	61,62	17,66	46,88	12,05	36,02
Rao et coll. (1992)	-5,33	39,83	-14,64	54,59	-8,04	35,06	-12,17	29,44
Rao et coll. (1992) ($n' = n/2$)	-6,90	39,15	-18,71	52,46	-9,63	34,57	-14,86	29,78
Chipperfield et Preston (2007)	-0,97	42,09	-4,43	60,87	-4,56	36,09	-4,95	29,86
Antal et Tillé (2014)	-0,51	42,54	-4,74	60,64	-4,68	36,6	-5,14	29,69
Antal et Tillé (2011a)	-5,12	39,85	-14,7	54,41	-8,52	34,98	-13,02	29,61
Beaumont et Patak (2012) Norm	-0,56	42,57	-2,53	63,13	-4,36	36,75	-4,27	30,06
Beaumont et Patak (2012) Exp	-8,60	38,83	-22,46	50,94	-11,17	34,60	-17,89	30,39
Beaumont et Patak (2012) $\epsilon = 1$	-1,11	41,80	-4,36	61,02	-4,69	36,26	-4,93	30,06

Tab. 4.26. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (m_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	2,5	3,7	6,2	1,000	2,3	2,8	5,1	1,000	3,5	2,6	6,1	1,000	1,4	3,4	4,8	1,000
Chao et Lo (1994)	2,6	3,5	6,1	1,000	2,4	3,0	5,4	0,998	4,0	2,5	6,5	1,000	1,7	3,1	4,8	1,000
Bickel et Freedman (1984)	2,5	3,3	5,8	0,997	2,5	3,4	5,9	0,988	3,7	2,6	6,3	0,995	1,6	3,4	5,0	0,987
Sitter (1992a)	2,5	3,2	5,7	1,006	2,3	2,9	5,2	1,013	4,0	2,7	6,7	0,998	1,5	3,2	4,7	1,000
McCarthy et Snowden (1985)	2,6	3,5	6,1	0,988	2,6	3,2	5,8	0,983	4,2	2,8	7,0	0,993	1,9	3,6	5,5	0,989
Sitter (1992b)	2,6	3,5	6,1	0,988	2,6	3,4	6,0	0,985	4,2	2,5	6,7	0,986	1,8	3,8	5,6	0,976
Rao et Wu (1988)	2,7	3,6	6,3	0,991	2,5	3,3	5,8	0,992	4,7	2,5	7,2	0,982	1,8	3,5	5,3	0,991
Rao et Wu (1988) ($n' = n/2$)	2,9	2,9	5,8	0,994	1,8	2,9	4,7	0,988	4,9	2,3	7,2	0,965	1,9	3,4	5,3	0,978
Rao et coll. (1992)	2,0	3,7	5,7	1,014	1,7	3,8	5,5	1,044	3,1	3,7	6,8	0,991	1,2	4,6	5,8	1,012
Rao et coll. (1992) ($n' = n/2$)	2,1	4,1	6,2	0,995	2,0	4,4	6,4	1,011	3,2	3,0	6,2	1,005	1,3	3,9	5,2	1,027
Chipperfield et Preston (2007)	2,2	3,6	5,8	1,020	1,9	4,2	6,1	1,006	3,3	3,6	6,9	0,984	1,0	4,2	5,2	1,037
Antal et Tillé (2014)	2,3	4,5	6,8	0,980	2,0	5,2	7,2	0,966	3,2	3,9	7,1	0,986	1,2	5,8	7,0	0,969
Antal et Tillé (2011a)	2,3	4,5	6,8	0,983	2,0	5,1	7,1	0,977	3,0	3,7	6,7	0,991	1,2	5,5	6,7	0,983
Beaumont et Patak (2012) Norm	2,8	3,5	6,3	0,984	2,5	3,4	5,9	0,975	4,3	2,7	7,0	0,989	1,7	3,5	5,2	0,989
Beaumont et Patak (2012) Exp	2,7	3,3	6,0	0,997	2,6	2,7	5,3	1,013	4,2	2,2	6,4	0,996	1,4	3,2	4,6	1,016
Beaumont et Patak (2012) $\epsilon = 1$	2,6	3,8	6,4	0,991	2,6	3,3	5,9	0,995	4,3	2,7	7,0	0,989	1,9	3,9	5,8	0,978

Tab. 4.27. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (m_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	2,1	1,9	4,0	1,000	2,4	2,0	4,4	1,000	2,6	2,4	5,0	1,000	2,5	2,6	5,1	1,000
Chao et Lo (1994)	2,0	2,3	4,3	0,998	2,8	2,1	4,9	0,996	2,4	2,8	5,2	1,000	2,0	2,4	4,4	1,002
Bickel et Freedman (1984)	2,1	2,0	4,1	0,990	2,6	2,3	4,9	0,982	2,4	2,4	4,8	0,985	2,4	2,7	5,1	0,977
Sitter (1992a)	2,2	2,4	4,6	0,995	2,6	2,3	4,9	0,993	2,4	2,4	4,8	0,993	2,2	2,6	4,8	0,988
McCarthy et Snowden (1985)	2,5	2,3	4,8	0,995	2,5	2,1	4,6	0,996	2,2	2,2	4,4	0,995	2,1	2,7	4,8	0,995
Sitter (1992b)	2,2	2,2	4,4	0,997	2,5	2,3	4,8	0,993	2,5	2,2	4,7	0,991	2,3	2,8	5,1	0,986
Rao et Wu (1988)	2,5	2,7	5,2	0,995	3,1	2,3	5,4	0,986	3,7	2,5	6,2	0,985	3,1	3,2	6,3	0,993
Rao et Wu (1988) ($n' = n/2$)	3,1	2,1	5,2	1,008	2,1	1,7	3,8	0,986	3,9	1,5	5,4	0,969	3,6	1,4	5,0	0,978
Rao et coll. (1992)	2,2	2,0	4,2	0,998	2,3	2,1	4,4	1,041	2,2	2,0	4,2	1,007	1,6	2,0	3,6	1,043
Rao et coll. (1992) ($n' = n/2$)	3,1	2,9	6,0	0,948	2,5	2,4	4,9	0,984	2,6	2,4	5,0	0,969	2,1	2,2	4,3	0,995
Chipperfield et Preston (2007)	2,0	2,1	4,1	1,023	2,5	2,3	4,8	0,994	2,4	1,9	4,3	0,996	5,0	4,6	9,6	0,821
Antal et Tillé (2014)	2,0	3,0	5,0	0,984	2,1	3,4	5,5	0,955	2,0	2,4	4,4	0,989	1,8	3,6	5,4	0,970
Antal et Tillé (2011a)	1,7	2,7	4,4	0,990	1,8	2,9	4,7	0,981	1,9	2,3	4,2	0,993	1,5	3,7	5,2	0,998
Beaumont et Patak (2012) Norm	2,3	2,3	4,6	0,989	2,3	2,7	5,0	0,980	2,4	2,2	4,6	0,989	2,1	2,6	4,7	0,975
Beaumont et Patak (2012) Exp	2,4	2,1	4,5	1,002	2,1	2,1	4,2	1,018	2,3	1,9	4,2	1,002	2,0	2,3	4,3	1,006
Beaumont et Patak (2012) $\epsilon = 1$	2,5	2,2	4,7	0,994	2,3	2,4	4,7	0,987	2,4	2,3	4,7	0,993	2,2	2,8	5,0	0,973

Tab. 4.28. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (m_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	4,9	7,9	12,8	1,000	6,1	9,6	15,7	1,000	9,6	6,5	16,1	1,000	3,4	8,1	11,5	1,000
Chao et Lo (1994)	4,8	7,9	12,7	0,998	6,5	9,8	16,3	0,996	9,4	6,0	15,4	1,000	3,7	8,0	11,7	1,002
Bickel et Freedman (1984)	4,9	7,9	12,8	0,990	6,4	10,0	16,4	0,982	9,9	6,1	16,0	0,985	4,0	8,4	12,4	0,977
Sitter (1992a)	5,2	8,6	13,8	0,995	6,5	9,9	16,4	0,993	9,8	6,3	16,1	0,993	3,7	8,5	12,2	0,988
McCarthy et Snowden (1985)	5,1	8,1	13,2	0,995	6,2	10,1	16,3	0,996	9,5	5,8	15,3	0,995	3,9	8,3	12,2	0,995
Sitter (1992b)	5,1	8,1	13,2	0,997	6,3	9,8	16,1	0,993	9,7	6,2	15,9	0,991	3,8	8,3	12,1	0,986
Rao et Wu (1988)	4,4	6,8	11,2	0,995	4,3	8,7	13,0	0,986	7,8	4,2	12,0	0,985	3,3	7,4	10,7	0,993
Rao et Wu (1988) ($n' = n/2$)	3,8	6,0	9,8	1,008	2,6	8,6	11,2	0,986	7,3	3,7	11,0	0,969	3,0	8,0	11,0	0,978
Rao et coll. (1992)	4,1	10,4	14,5	0,998	3,6	13,6	17,2	1,041	7,9	8,6	16,5	1,007	2,7	13,2	15,9	1,043
Rao et coll. (1992) ($n' = n/2$)	4,5	11,2	15,7	0,948	4,6	14,6	19,2	0,984	8,3	9,4	17,7	0,969	2,8	13,2	16,0	0,995
Chipperfield et Preston (2007)	4,2	10,1	14,3	1,023	4,6	14,5	19,1	0,994	8,3	9,0	17,3	0,996	3,7	15,7	19,4	0,821
Antal et Tillé (2014)	4,3	9,6	13,9	0,984	6,2	13,4	19,6	0,955	9,0	7,1	16,1	0,989	2,9	12,9	15,8	0,970
Antal et Tillé (2011a)	4,3	9,6	13,9	0,990	5,2	13,1	18,3	0,981	8,3	7,0	15,3	0,993	3,0	12,3	15,3	0,998
Beaumont et Patak (2012) Norm	5,7	8,3	14,0	0,989	6,7	9,9	16,6	0,980	9,8	5,2	15,0	0,989	4,0	8,4	12,4	0,975
Beaumont et Patak (2012) Exp	5,0	8,3	13,3	1,002	6,1	9,7	15,8	1,018	9,7	6,1	15,8	1,002	4,0	8,2	12,2	1,006
Beaumont et Patak (2012) $\epsilon = 1$	5,1	8,5	13,6	0,994	6,4	9,9	16,3	0,987	9,4	5,9	15,3	0,993	4,3	8,5	12,8	0,973

Tab. 4.29. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (m_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	3,3	3,8	7,1	1,000	5,3	2,9	8,2	1,000	5,6	4,2	9,8	1,000	2,2	2,8	5,0	1,000
Chao et Lo (1994)	3,6	3,3	6,9	0,999	5,8	3,0	8,8	0,996	5,5	3,9	9,4	0,998	2,1	2,5	4,6	0,999
Bickel et Freedman (1984)	3,3	3,8	7,1	0,996	5,2	3,0	8,2	0,990	5,2	3,8	9,0	0,992	2,3	2,6	4,9	0,992
Sitter (1992a)	3,2	3,5	6,7	1,003	4,4	2,9	7,3	1,013	5,3	4,0	9,3	0,998	2,2	2,6	4,8	1,004
McCarthy et Snowden (1985)	3,3	3,5	6,8	1,018	5,2	2,9	8,1	1,025	5,1	3,6	8,7	1,051	2,0	2,1	4,1	1,102
Sitter (1992b)	3,2	3,5	6,7	1,009	5,0	3,1	8,1	1,014	5,2	4,4	9,6	1,015	2,2	2,8	5,0	1,015
Rao et Wu (1988)	2,4	2,1	4,5	1,026	3,2	1,2	4,4	1,052	3,6	1,9	5,5	1,011	1,4	0,8	2,2	1,054
Rao et Wu (1988) ($n' = n/2$)	2,9	1,4	4,3	1,038	2,7	0,6	3,3	1,069	3,9	0,4	4,3	1,007	2,8	0,4	3,2	1,060
Rao et coll. (1992)	2,5	4,5	7,0	1,050	2,2	4,0	6,2	1,134	4,0	4,5	8,5	1,057	1,3	3,9	5,2	1,135
Rao et coll. (1992) ($n' = n/2$)	3,3	4,2	7,5	1,102	3,9	3,3	7,2	1,232	4,3	3,0	7,3	1,179	1,4	3,0	4,4	1,314
Chipperfield et Preston (2007)	2,5	4,5	7,0	1,059	1,7	4,2	5,9	1,171	4,5	5,3	9,8	1,038	1,0	4,1	5,1	1,222
Antal et Tillé (2014)	3,0	5,0	8,0	0,993	4,6	5,0	9,6	0,986	4,4	4,7	9,1	0,997	1,6	4,8	6,4	0,988
Antal et Tillé (2011a)	2,8	4,4	7,2	1,000	4,2	5,3	9,5	1,010	4,4	4,6	9,0	1,010	1,6	4,6	6,2	1,008

Tab. 4.30. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (La médiane m_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,032	1,100	1,052	1,177	1,033	1,123	1,042	1,154
Chao et Lo (1994)	1,032	1,098	1,050	1,173	1,033	1,121	1,042	1,152
Bickel et Freedman (1984)	1,029	1,095	1,040	1,165	1,027	1,114	1,028	1,144
Sitter (1992a)	1,039	1,103	1,066	1,193	1,031	1,121	1,042	1,158
McCarthy et Snowden (1985)	1,019	1,119	1,034	1,207	1,026	1,179	1,030	1,271
Sitter (1992b)	1,019	1,110	1,036	1,194	1,018	1,139	1,018	1,171
Rao et Wu (1988)	1,026	1,129	1,046	1,239	1,015	1,136	1,033	1,216
Rao et Wu (1988) ($n' = n/2$)	1,026	1,142	1,041	1,258	0,996	1,130	1,018	1,223
Rao et coll. (1992)	1,048	1,155	1,102	1,335	1,024	1,187	1,059	1,309
Rao et coll. (1992) ($n' = n/2$)	1,026	1,213	1,066	1,451	1,040	1,324	1,070	1,516
Chipperfield et Preston (2007)	1,053	1,164	1,058	1,378	1,016	1,165	1,081	1,409
Antal et Tillé (2014)	1,011	1,092	1,015	1,160	1,018	1,119	1,011	1,140
Antal et Tillé (2011a)	1,016	1,100	1,029	1,189	1,024	1,134	1,023	1,163
Beaumont et Patak (2012) Normal	1,016	–	1,027	–	1,022	–	1,030	–
Beaumont et Patak (2012) Exp	1,029	–	1,065	–	1,031	–	1,059	–
Beaumont et Patak (2012) $\epsilon = 1$	1,023	–	1,045	–	1,022	–	1,019	–

Tab. 4.31. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Booth et coll. (1994) (G_y)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,6	8,4	9,0	1,000	1,4	13,6	15,0	1,000	0,5	7,7	8,2	1,000	1,7	7,9	9,6	1,000
Chao et Lo (1994)	0,6	8,5	9,1	1,000	1,5	13,7	15,2	1,000	0,5	7,4	7,9	1,000	1,6	8,3	9,9	1,000
Bickel et Freedman (1984)	0,7	8,6	9,3	1,000	1,4	13,4	14,8	1,000	0,5	7,4	7,9	1,000	1,7	8,1	9,8	1,000
Sitter (1992a)	0,7	8,6	9,3	1,000	1,3	13,5	14,8	1,018	0,4	7,6	8,0	1,000	1,5	7,8	9,3	1,000
McCarthy et Snowden (1985)	0,6	8,8	9,4	1,000	1,5	13,5	15,0	1,000	0,4	7,3	7,7	0,983	1,7	8,4	10,1	0,976
Sitter (1992b)	0,6	8,5	9,1	1,000	1,4	13,8	15,2	1,000	0,4	7,3	7,7	1,000	1,7	8,3	10,0	1,000
Rao et Wu (1988)	0,6	8,5	9,1	1,000	1,1	13,5	14,6	1,018	0,3	6,8	7,1	1,051	1,1	7,8	8,9	1,037
Rao et Wu (1988) ($n' = n/2$)	0,1	6,5	6,6	1,103	0,6	11,6	12,2	1,091	0,2	6,1	6,3	1,119	0,8	7,3	8,1	1,098
Rao et coll. (1992)	0,7	8,6	9,3	1,000	1,6	13,7	15,3	0,991	0,4	7,7	8,1	0,983	1,7	8,6	10,3	0,976
Rao et coll. (1992) ($n' = n/2$)	0,8	8,6	9,4	0,985	1,9	14,2	16,1	0,964	0,5	8,0	8,5	0,983	1,7	8,8	10,5	0,951
Chipperfield et Preston (2007)	0,5	8,6	9,1	1,029	0,9	12,6	13,5	1,045	0,4	7,1	7,5	1,000	1,6	8,0	9,6	1,000
Antal et Tillé (2014)	0,5	8,4	8,9	1,029	0,9	12,6	13,5	1,036	0,4	7,3	7,7	1,000	1,7	8,0	9,7	1,000
Antal et Tillé (2011a)	0,7	8,6	9,3	1,000	1,7	13,7	15,4	0,991	0,4	7,6	8,0	0,983	1,8	8,4	10,2	0,963
Beaumont et Patak (2012) Norm	0,6	8,6	9,2	1,029	1,2	13,1	14,3	1,055	0,4	7,2	7,6	1,000	1,4	8,2	9,6	1,000
Beaumont et Patak (2012) Exp	0,9	8,9	9,8	0,971	1,9	14,8	16,7	0,945	0,6	8,0	8,6	0,966	1,8	9,6	11,4	0,927
Beaumont et Patak (2012) $\epsilon = 1$	0,4	8,4	8,8	1,029	1,1	12,6	13,7	1,045	0,4	7,7	8,1	1,000	1,6	7,8	9,4	1,000

Tab. 4.32. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Booth et coll. (1994) (G_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	0,4	9,6	10,0	1,000	0,5	16,3	16,8	1,000	0,3	9,7	10,0	1,000	0,9	11,4	12,3	1,000
Chao et Lo (1994)	0,3	10,1	10,4	1,000	0,6	16,4	17,0	1,000	0,4	9,6	10,0	1,000	0,5	11,3	11,8	1,000
Bickel et Freedman (1984)	0,3	10,0	10,3	1,000	0,6	16,4	17,0	1,009	0,3	9,3	9,6	1,000	0,8	11,6	12,4	1,000
Sitter (1992a)	0,4	9,7	10,1	1,000	0,5	16,0	16,5	1,009	0,3	9,3	9,6	1,000	0,5	11,0	11,5	1,025
McCarthy et Snowden (1985)	0,4	9,2	9,6	1,000	0,6	16,3	16,9	1,019	0,4	8,4	8,8	1,000	1,0	10,8	11,8	1,013
Sitter (1992b)	0,4	9,7	10,1	1,000	0,5	15,8	16,3	1,019	0,4	9,4	9,8	1,000	0,5	11,3	11,8	1,025
Rao et Wu (1988)	0,6	8,1	8,7	1,015	0,8	13,5	14,3	1,028	0,4	6,8	7,2	1,069	0,9	7,9	8,8	1,063
Rao et Wu (1988) ($n' = n/2$)	0,2	6,1	6,3	1,104	0,4	11,8	12,2	1,075	0,2	5,9	6,1	1,121	0,4	7,3	7,7	1,101
Rao et coll. (1992)	0,3	10,0	10,3	1,015	0,7	15,9	16,6	1,009	0,6	8,4	9,0	1,000	1,0	10,2	11,2	0,987
Rao et coll. (1992) ($n' = n/2$)	1,0	8,7	9,7	0,985	1,1	15,9	17,0	0,962	0,7	7,8	8,5	0,966	1,3	9,3	10,6	0,962
Chipperfield et Preston (2007)	0,1	11,5	11,6	1,000	0,2	19,1	19,3	0,991	0,3	10,5	10,8	1,000	0,4	13,0	13,4	1,000
Antal et Tillé (2014)	0,1	11,3	11,4	1,000	0,3	19,1	19,4	0,991	0,3	10,5	10,8	1,000	0,4	12,1	12,5	1,013
Antal et Tillé (2011a)	0,3	9,5	9,8	1,015	0,6	15,9	16,5	1,009	0,6	7,9	8,5	1,000	1,0	9,3	10,3	1,000
Beaumont et Patak (2012) Norm	0,1	11,1	11,2	1,030	0,1	17,4	17,5	1,085	0,2	10,4	10,6	1,017	0,4	12,0	12,4	1,038
Beaumont et Patak (2012) Exp	0,9	8,0	8,9	1,000	1,0	14,9	15,9	0,981	0,9	7,2	8,1	0,983	1,1	9,1	10,2	0,962
Beaumont et Patak (2012) $\epsilon = 1$	0,1	11,3	11,4	1,000	0,2	19,1	19,3	0,991	0,3	10,7	11,0	0,983	0,5	12,6	13,1	1,000

Tab. 4.33. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Booth et coll. (1994) (G_y)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	1,7	8,0	9,7	1,000	3,0	11,8	14,8	1,000	1,2	6,6	7,8	1,000	3,8	6,9	10,7	1,000
Chao et Lo (1994)	1,4	8,0	9,4	1,000	3,0	12,2	15,2	1,000	0,9	7,0	7,9	1,000	4,1	6,9	11,0	1,000
Bickel et Freedman (1984)	1,6	7,8	9,4	1,000	3,1	11,9	15,0	1,009	1,3	6,8	8,1	1,000	3,8	6,9	10,7	1,000
Sitter (1992a)	1,6	8,2	9,8	1,000	2,8	11,4	14,2	1,009	1,0	6,6	7,6	1,000	4,0	6,9	10,9	1,025
McCarthy et Snowden (1985)	1,4	7,8	9,2	1,000	2,9	11,4	14,3	1,019	0,8	6,9	7,7	1,000	3,7	7,6	11,3	1,013
Sitter (1992b)	1,8	7,8	9,6	1,000	3,1	11,7	14,8	1,019	1,2	6,7	7,9	1,000	4,0	6,8	10,8	1,025
Rao et Wu (1988)	0,9	8,9	9,8	1,015	1,8	13,5	15,1	1,028	0,4	7,4	7,8	1,069	1,7	8,5	10,2	1,063
Rao et Wu (1988) ($n' = n/2$)	0,2	7,6	7,8	1,104	1,1	12,7	13,8	1,075	0,2	6,9	7,1	1,121	1,5	9,1	10,6	1,101
Rao et coll. (1992)	1,9	7,9	9,8	1,015	2,9	12,1	15,0	1,009	0,7	7,3	8,0	1,000	3,6	8,1	11,7	0,987
Rao et coll. (1992) ($n' = n/2$)	1,4	8,9	10,3	0,985	2,6	14,0	16,6	0,962	0,6	8,6	9,2	0,966	3,1	9,9	13,0	0,962
Chipperfield et Preston (2007)	2,4	6,5	8,9	1,000	4,2	10,4	14,6	0,991	2,3	6,2	8,5	1,000	4,4	6,9	11,3	1,000
Antal et Tillé (2014)	2,5	6,5	9,0	1,000	4,3	9,9	14,2	0,991	1,4	6,4	7,8	1,000	4,4	6,6	11,0	1,013
Antal et Tillé (2011a)	1,4	8,1	9,5	1,015	2,9	11,8	14,7	1,009	1,0	7,6	8,6	1,000	3,4	7,8	11,2	1,000
Beaumont et Patak (2012) Norm	2,5	5,5	8,0	1,030	3,5	9,0	12,5	1,085	1,3	5,9	7,2	1,017	4,2	5,7	9,9	1,038
Beaumont et Patak (2012) Exp	1,2	9,7	10,9	1,000	2,3	13,4	15,7	0,981	0,4	9,1	9,5	0,983	2,9	10,1	13,0	0,962
Beaumont et Patak (2012) $\epsilon = 1$	2,4	6,6	9,0	1,000	4,5	10,3	14,8	0,991	1,9	6,4	8,3	0,983	4,2	6,8	11,0	1,000

Tab. 4.34. Étude 1: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Booth et coll. (1994) (G_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
Booth et coll. (1994)	1,7	3,3	5,0	1,000	1,8	5,8	7,6	1,000	1,6	2,9	4,5	1,000	2,1	2,9	5,0	1,000
Chao et Lo (1994)	1,3	3,4	4,7	1,012	1,8	5,9	7,7	1,000	1,1	2,5	3,6	1,000	1,7	2,8	4,5	1,010
Bickel et Freedman (1984)	1,4	3,3	4,7	1,000	1,9	6,3	8,2	1,000	1,3	2,5	3,8	1,015	2,0	2,9	4,9	1,000
Sitter (1992a)	1,4	3,6	5,0	1,012	1,4	6,0	7,4	1,024	1,3	2,7	4,0	1,015	1,8	2,8	4,6	1,010
McCarthy et Snowden (1985)	1,2	3,6	4,8	1,012	1,7	5,9	7,6	1,030	0,9	3,0	3,9	1,000	1,7	3,2	4,9	1,029
Sitter (1992b)	1,4	3,0	4,4	1,012	1,5	5,9	7,4	1,024	1,5	2,5	4,0	1,015	1,5	2,8	4,3	1,010
Rao et Wu (1988)	0,6	3,8	4,4	1,012	0,5	6,7	7,2	1,024	0,5	2,2	2,7	1,074	0,3	2,7	3,0	1,135
Rao et Wu (1988) ($n' = n/2$)	0,1	1,9	2,0	1,235	0,3	3,2	3,5	1,393	0,5	1,0	1,5	1,309	0,2	1,4	1,6	1,683
Rao et coll. (1992)	1,3	3,8	5,1	0,988	2,2	6,1	8,3	1,012	2,3	3,0	5,3	0,956	2,1	2,9	5,0	1,000
Rao et coll. (1992) ($n' = n/2$)	3,2	3,3	6,5	1,000	3,0	3,9	6,9	1,137	3,5	1,9	5,4	1,015	2,7	1,8	4,5	1,298
Chipperfield et Preston (2007)	2,6	2,3	4,9	1,099	3,0	3,6	6,6	1,238	3,3	1,6	4,9	1,044	3,0	1,7	4,7	1,231
Antal et Tillé (2014)	2,0	2,3	4,3	1,086	2,6	3,6	6,2	1,196	1,6	2,0	3,6	1,029	1,9	2,6	4,5	1,058
Antal et Tillé (2011a)	1,2	3,7	4,9	1,000	1,7	6,0	7,7	1,000	0,9	3,6	4,5	0,985	1,5	3,3	4,8	0,952

Tab. 4.35. Étude 1: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Booth et coll. (1994) (Gini G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Booth et coll. (1994)	1,015	1,209	1,038	1,585	1,017	1,172	1,038	1,316
Chao et Lo (1994)	1,015	1,224	1,038	1,585	1,017	1,172	1,038	1,329
Bickel et Freedman (1984)	1,015	1,209	1,038	1,585	1,017	1,190	1,038	1,316
Sitter (1992a)	1,015	1,224	1,057	1,623	1,017	1,190	1,038	1,329
McCarthy et Snowden (1985)	1,015	1,224	1,038	1,632	1,000	1,172	1,013	1,354
Sitter (1992b)	1,015	1,224	1,038	1,623	1,017	1,190	1,038	1,329
Rao et Wu (1988)	1,015	1,224	1,057	1,623	1,069	1,259	1,089	1,494
Rao et Wu (1988) ($n' = n/2$)	1,119	1,493	1,132	2,208	1,138	1,534	1,139	2,215
Rao et coll. (1992)	1,015	1,194	1,028	1,604	1,000	1,121	1,013	1,316
Rao et coll. (1992) ($n' = n/2$)	1,015	1,209	1,000	1,802	1,000	1,190	0,987	1,709
Chipperfield et Preston (2007)	1,030	1,328	1,075	1,962	1,017	1,224	1,038	1,620
Antal et Tillé (2014)	1,045	1,313	1,075	1,896	1,017	1,207	1,038	1,392
Antal et Tillé (2011a)	1,015	1,209	1,028	1,585	1,000	1,155	0,987	1,253
Beaumont et Patak (2012) Norm	1,045	–	1,094	–	1,017	–	1,038	–
Beaumont et Patak (2012) Exp	0,985	–	0,981	–	0,983	–	0,975	–
Beaumont et Patak (2012) $\epsilon = 1$	1,045	–	1,085	–	1,017	–	1,038	–

4.4.6. Discussion: cas des paramètres non lisses

Le tableau 4.25 met en évidence des faits que nous énumérons dans les points suivants:

- Dans le cas de la médiane, on peut dire que les méthodes bootstrap direct font légèrement mieux que les méthode bootstrap par pseudo-population. Ce constat peut s'expliquer par le fait que l'algorithme général 2.1 proposé par Mashreghi et coll. (2016) ne fonctionne que dans le cas d'un estimateur sans biais, c'est-à-dire $E_{p^*}(\hat{\theta}^*) = \theta^*$. Dans le cas contraire comme la médiane, cet algorithme permet d'estimer l'erreur quadratique moyenne;
- Pour la classe de méthodes bootstrap pondéré on peut dire que la méthode de Antal et Tillé (2011a), la méthode de Antal et Tillé (2014) et la méthode de Beaumont et Patak (2012) mise en œuvre avec les deux distributions symétriques font mieux que les autres méthodes dans le cas de la médiane;
- Le choix d'une distribution symétrique pour générer les ajustements bootstrap pour la méthode de Beaumont et Patak (2012) semble approprié dans le contexte des paramètres non lisses comme la médiane et l'indice de Gini;
- Dans le cas des deux paramètres étudiés, on constate que les deux méthodes Rao et Wu (1988) et Rao et coll. (1992) ne donnent pas des résultats proches en terme de BR et $REQMR$ comme c'était le cas des paramètres linéaires ou lisses. Ceci s'explique par le fait que les formules de calcul de l'estimateur bootstrap $\hat{\theta}^*$ ne sont pas équivalentes dans le cas de ces deux paramètres. En particulier, Rao et Wu (1988) fait mieux que Rao et coll. (1992) pour la médiane dans les quatre scénarios, alors que c'est l'inverse pour Gini, du moins pour la $REQMR$;
- Dans le cas de l'indice de Gini on constate que les BR observés dans tous les scénarios sont tous négatifs à l'exception de la méthode de Rao et Wu (1988) mise en œuvre avec une taille d'échantillon bootstrap $n' = \lceil n/2 \rceil$ et la même méthode mise en œuvre avec une taille $n' = n - 3$ mais seulement dans les deux derniers scénarios. Ceci peut s'expliquer par le fait que cette méthode manipule les données ré-échelonnées et pas les données d'origine.

- D'une manière générale et compte tenu de tous les scénarios et des deux paramètres étudiés, on peut dire que la méthode de Antal et Tillé (2014) et la méthode de Beaumont et Patak (2012) mise en œuvre avec les deux distributions symétriques font mieux que les autres méthodes.

Examinons maintenant les performances des différentes méthodes dans la production d'intervalles de confiance de la médiane (tableaux 4.26 à 4.28). On constate que sauf pour Chipperfield et Preston (2007) pour le scénario 4, toutes les probabilités de couverture bilatérale des intervalles percentiles ne sont pas statistiquement différentes de 95%. Dans le scénario 3, plusieurs intervalles asymptotiques ont une couverture bilatérale statistiquement inférieure à 95%, ce qui est également le cas de tous les scénarios pour les méthodes d'Antal et Tillé (2011a) et Antal et Tillé (2014).

On peut également remarquer que les intervalles de confiance percentile font mieux que les intervalles asymptotiques, notamment en terme des erreurs de couverture unilatérale et la longueur des intervalles (voir le tableau 4.30 de la comparaison de la longueur). Cette remarque s'applique à toutes les méthodes et pour tous les scénarios sauf pour la méthode de Chipperfield et Preston (2007) dans le dernier scénario, où les taux d'erreur de couverture unilatérale observés avoisinent le 5%.

Pour toutes les méthodes bootstrap, les intervalles de confiance *t*-bootstrap ne font pas mieux que les intervalles de confiance percentile. En effet, pour la plupart des méthodes, la couverture bilatérale des intervalles *t*-bootstrap est statistiquement inférieure à 95% pour tous les scénarios sauf pour le dernier où la plupart sont corrects (voir tableau 4.29).

En ce qui concerne les performances des différentes méthodes dans la production d'intervalles de confiance de l'indice de Gini (tableaux 4.31 à 4.33), on constate que pour toutes les méthodes le taux de couverture bilatérale théorique n'est pas respecté par les intervalles de confiance asymptotique, percentile et bootstrap de base dans tous les scénarios sauf pour la méthode de Rao et Wu (1988) mise en œuvre avec une taille d'échantillon bootstrap égale à $\lfloor n/2 \rfloor$ dans les scénarios 1 et 3 pour les intervalles de confiance percentile et seulement dans le scénario 3 pour les intervalles de confiance asymptotique. En effet, les taux de couverture bilatérale observés sont autour de 83% dans le scénario 2. Le *t*-bootstrap a amélioré

substantiellement la situation pour toutes les méthodes bootstrap en terme des erreurs de couverture unilatérales et bilatérales sauf pour la méthode de Rao et Wu (1988) dans les scénarios 1, 3 et 4. En effet, le gain que le t -bootstrap apporte par rapport à la méthode percentile en terme du taux de couverture bilatérale est autour de 5 points en pourcentage dans les scénarios 1, 3 et 4 et autour de 10 points en pourcentage dans le scénario 2.

Il faut noter que les intervalles de confiance t -bootstrap dans le cas de l'indice de Gini sont plus larges que ceux de type percentile d'environ 20% dans les scénarios 1, 3 et 4 et d'environ 60% dans le scénario 2 (voir tableau 4.35). Toutefois, il faut rappeler que les estimateurs bootstrap de la variance dans le cas de ce paramètre ont tendance à sous-estimer la variance. En conséquence, le t -bootstrap s'avère une bonne solution pour remédier à ce problème.

4.5. Étude 2: Plan de Poisson

Dans notre deuxième étude par simulation nous comparons les méthodes bootstrap présentées dans le chapitre 3 pour le plan de Poisson. Les probabilités d'inclusion sont calculées proportionnellement à la variable auxiliaire x . Il faut noter que pour mettre en œuvre la méthode de Chauvet (2007), nous avons créé $D = 10$ pseudo-populations à partir desquelles $B = 100$ échantillons bootstrap seront tirés. Nous avons également testé l'algorithme bootstrap proposé par Mashreghi et coll. (2016); voir l'équation (3.2.1). Pour les deux méthodes bootstrap par pseudo-population x -équilibrée, nous les dénotons respectivement $x-equ1$ et $x-equ2$. Pour la méthode proposée par Beaumont et Patak (2012), nous la mettons en œuvre en utilisant les trois distributions décrites dans la première étude. Il faut noter que dans le cas du coefficient de corrélation, le calcul de l'estimateur bootstrap de la variance par la méthode de Beaumont et Patak (2012) se fait sur la base des ajustements bootstrap ré-échelonnés. En conséquence, l'estimation de la variance et la construction des intervalles de confiance percentile et bootstrap de base sont effectués de la même façon que nous avons indiquée dans la première étude.

La construction des intervalles de confiance percentile, bootstrap de base et t -bootstrap pour les méthodes bootstrap par pseudo-population est réalisée d'une façon analogue à l'approche proposée par Mashreghi et coll. (2016) pour la méthode Booth et coll. (1994) dans le cas du plan aléatoire simple sans remise. En conséquence, les intervalles de confiance percentile et bootstrap de base calculés pour les trois premières méthodes, à savoir Holmberg (1998), Chauvet (2007) et Mashreghi et coll. (2016) seront identiques car d'une part l'algorithme de l'estimation de la variance n'intervient pas dans la construction desdits intervalles de confiance et d'autre part la construction de la pseudo-population est identique pour ces méthodes. En conséquence, nous rapportons seulement les résultats empiriques des intervalles de confiance percentile et bootstrap de base relatifs à la méthode de Mashreghi et coll. (2016). En ce qui concerne les intervalles de confiance t -bootstrap pour ces trois méthodes, ils seront différents car l'algorithme de l'estimation de la variance diffère d'une méthode à l'autre. Bien que la mise en œuvre de la technique t -bootstrap est théoriquement faisable pour la méthode de Chauvet (2007) cela génère un temps de calcul très important car l'algorithme de l'estimation de la variance par cette méthode en particulier est complexe. Pour cette raison les intervalles de confiance t -bootstrap ne seront pas calculés pour cette méthode.

Pour les méthodes bootstrap pondéré nous calculons les intervalles de confiance t -bootstrap seulement pour les deux méthodes Antal et Tillé (2011a) et Antal et Tillé (2014). En effet, la réalisation d'un double bootstrap pour ces deux méthodes ne pose aucun problème. Il suffit d'appliquer le même algorithme bootstrap d'estimation de la variance au ré-échantillon s^* , chose qui s'avère compliquée pour la méthode de Beaumont et Patak (2012) car l'identification du ré-échantillon s^* qui fera l'objet d'un double bootstrap est impossible.

4.5.1. Résultats: cas des paramètres linéaires

Le tableau 4.36 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne relative des estimateurs bootstrap de la variance étudiés pour le plan de Poisson, dans le cas des totaux t_y et t_z . Le tableau 4.37 au tableau 4.40 donnent,

Tab. 4.36. Étude 2: Biais relatif et stabilité des méthodes bootstrap (paramètres linéaires, t_y et t_z).

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
t_y								
Holmberg (1998)	0,05	13,04	-0,65	20,74	3,73	11,56	3,05	17,01
Chauvet (2007)	-0,20	12,93	-0,66	20,39	0,48	10,57	0,00	16,63
Mashreghi et coll. (2016)	0,45	12,83	-0,80	20,38	0,73	10,55	-0,14	16,28
Chauvet (2007) simplifié	-0,23	12,86	-0,70	20,67	0,10	10,47	-1,94	16,06
Barbiero et Mecatti (2010) <i>x-equ1</i>	-0,21	11,85	-1,07	18,92	0,46	8,25	0,12	12,18
Barbiero et Mecatti (2010) <i>x-equ2</i>	-0,18	11,44	-1,04	18,90	1,46	8,80	1,02	12,53
Beaumont et Patak (2012) Norm	0,19	12,75	-0,71	20,64	0,73	10,31	0,17	16,17
Beaumont et Patak (2012) Exp	-0,02	12,89	-0,76	20,91	0,81	10,78	0,10	16,44
Beaumont et Patak (2012) $\epsilon = 1$	0,23	12,80	-1,12	20,49	0,68	10,60	-0,06	16,35
Antal et Tillé (2014)	0,06	12,95	-0,69	20,74	0,69	10,66	-0,05	16,35
Antal et Tillé (2011a)	0,13	12,84	-0,63	20,68	0,45	10,68	-0,19	16,20
t_z								
Holmberg (1998)	0,11	14,21	-0,17	22,89	3,73	13,20	4,23	20,41
Chauvet (2007)	0,02	14,24	-0,21	22,47	0,52	12,39	0,60	19,57
Mashreghi et coll. (2016)	0,49	14,12	-0,29	22,45	0,82	12,35	0,62	19,37
Chauvet (2007) simplifié	-0,19	14,21	-0,27	22,70	-0,45	12,24	-1,81	19,25
Barbiero et Mecatti (2010) <i>x-equ1</i>	-0,07	13,15	-0,69	21,13	0,43	10,75	0,73	16,46
Barbiero et Mecatti (2010) <i>x-equ2</i>	-0,23	12,73	-0,56	21,09	1,28	11,33	1,57	16,75
Beaumont et Patak (2012) Norm	0,22	14,07	-0,29	22,61	0,77	12,23	0,83	19,24
Beaumont et Patak (2012) Exp	0,23	14,21	-0,45	22,90	0,86	12,62	0,88	19,49
Beaumont et Patak (2012) $\epsilon = 1$	0,43	14,00	-0,59	22,48	0,73	12,46	0,75	19,34
Antal et Tillé (2014)	0,17	14,27	-0,28	22,69	0,75	12,59	0,74	19,40
Antal et Tillé (2011a)	0,22	14,14	-0,26	22,66	0,50	12,76	0,64	19,46

pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap des totaux t_y et t_z . Le tableau 4.41 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.37. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres linéaires, t_y et t_z).

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Holmberg (1998)	1,2	3,4	4,6	1,000	1,3	4,2	5,5	1,000	1,7	3,1	4,8	1,000	1,5	3,8	5,3	1,000
Chauvet (2007)	1,3	3,5	4,8	0,999	1,5	4,2	5,7	1,000	2,0	3,2	5,2	0,984	1,3	3,8	5,1	0,985
Mashreghi et coll. (2016)	1,0	3,4	4,4	1,002	1,3	4,0	5,3	0,999	1,7	3,4	5,1	0,985	1,9	3,9	5,8	0,984
Chauvet (2007) simplifié	1,2	3,3	4,5	0,999	1,5	4,0	5,5	1,000	1,7	3,2	4,9	0,982	1,7	4,1	5,8	0,975
Barbiero et Mecatti (2010) $x-equ1$	1,3	3,1	4,4	0,999	1,4	4,0	5,4	0,999	2,3	2,7	5,0	0,985	2,8	2,7	5,5	0,987
Barbiero et Mecatti (2010) $x-equ2$	1,6	3,1	4,7	0,999	1,6	3,7	5,3	0,999	2,2	2,6	4,8	0,989	2,8	3,1	5,9	0,991
Beaumont et Patak (2012) Norm	1,4	3,3	4,7	1,001	1,1	3,8	4,9	1,000	1,9	3,2	5,1	0,985	1,7	3,9	5,6	0,986
Beaumont et Patak (2012) Exp	1,3	3,4	4,7	1,000	1,1	4,1	5,2	0,999	2,0	3,3	5,3	0,986	1,5	4,1	5,6	0,985
Beaumont et Patak (2012) $\epsilon = 1$	1,3	3,3	4,6	1,001	1,5	4,1	5,6	0,998	1,9	3,3	5,2	0,985	1,6	3,9	5,5	0,985
Antal et Tillé (2014)	1,3	3,2	4,5	1,000	1,4	3,9	5,3	1,000	1,7	3,3	5,0	0,985	1,8	4,0	5,8	0,985
Antal et Tillé (2011a)	1,8	3,3	5,1	1,000	1,4	4,2	5,6	1,000	1,6	3,5	5,1	0,984	1,4	3,9	5,3	0,984
	t_z															
Holmberg (1998)	1,4	3,8	5,2	1,000	1,1	5,3	6,4	1,000	1,3	3,3	4,6	1,000	1,6	4,1	5,7	1,000
Chauvet (2007)	1,4	3,9	5,3	0,999	1,0	5,3	6,3	1,000	1,5	3,5	5,0	0,985	1,8	4,4	6,2	0,983
Mashreghi et coll. (2016)	1,4	3,8	5,2	1,002	1,1	5,2	6,3	0,999	1,5	3,2	4,7	0,987	1,9	4,5	6,4	0,983
Chauvet (2007) simplifié	1,4	3,8	5,2	0,998	1,2	5,2	6,4	0,999	1,4	3,4	4,8	0,980	1,7	4,4	6,1	0,970
Barbiero et Mecatti (2010) $x-equ1$	1,5	3,7	5,2	0,999	1,1	5,1	6,2	0,998	2,1	2,4	4,5	0,984	2,2	3,6	5,8	0,985
Barbiero et Mecatti (2010) $x-equ2$	1,4	3,8	5,2	0,999	1,4	5,0	6,4	0,998	1,8	2,4	4,2	0,989	2,3	3,5	5,8	0,989
Beaumont et Patak (2012) Norm	1,5	3,8	5,3	1,000	1,0	5,2	6,2	0,999	1,5	3,5	5,0	0,986	1,8	4,4	6,2	0,983
Beaumont et Patak (2012) Exp	1,4	4,1	5,5	1,000	1,0	5,2	6,2	0,998	1,5	3,4	4,9	0,987	1,6	4,3	5,9	0,983
Beaumont et Patak (2012) $\epsilon = 1$	1,4	4,2	5,6	1,002	1,1	5,3	6,4	0,998	1,4	3,2	4,6	0,986	1,8	4,3	6,1	0,983
Antal et Tillé (2014)	1,4	3,8	5,2	1,000	1,0	5,2	6,2	0,999	1,4	3,4	4,8	0,986	1,5	4,3	5,8	0,983
Antal et Tillé (2011a)	1,5	4,0	5,5	1,000	1,0	5,2	6,2	0,999	1,4	3,5	4,9	0,984	1,6	4,1	5,7	0,983

Tab. 4.38. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires, t_y et t_z).

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Mashreghi et coll. (2016)	1,6	3,0	4,6	1,000	1,8	3,5	5,3	1,000	1,7	3,1	4,8	1,000	2,0	3,8	5,8	1,000
Chauvet (2007) simplifié	1,5	3,1	4,6	0,999	1,7	3,4	5,1	1,000	1,8	3,0	4,8	0,998	2,1	4,0	6,1	0,989
Barbiero et Mecatti (2010) <i>x-equ1</i>	1,5	3,0	4,5	0,999	2,0	3,4	5,4	1,000	2,8	2,7	5,5	1,000	3,2	2,9	6,1	1,001
Barbiero et Mecatti (2010) <i>x-equ2</i>	2,0	2,7	4,7	1,000	2,0	3,6	5,6	1,000	3,0	2,9	5,9	1,004	3,2	3,0	6,2	1,006
Beaumont et Patak (2012) Norm	1,7	3,6	5,3	1,000	1,3	4,2	5,5	1,001	1,9	3,0	4,9	1,001	1,9	3,9	5,8	1,001
Beaumont et Patak (2012) Exp	2,2	2,8	5,0	0,999	2,1	3,1	5,2	0,999	2,3	2,9	5,2	1,002	2,9	3,3	6,2	0,998
Beaumont et Patak (2012) $\epsilon = 1$	1,5	3,4	4,9	0,999	1,4	4,5	5,9	0,995	1,7	3,3	5,0	1,000	1,8	3,8	5,6	0,997
Antal et Tillé (2014)	1,2	3,4	4,6	0,999	1,3	4,1	5,4	1,001	1,6	3,1	4,7	1,000	2,0	4,0	6,0	0,998
Antal et Tillé (2011a)	2,0	3,0	5,0	1,000	1,8	3,9	5,7	1,001	2,2	3,1	5,3	1,001	2,2	3,5	5,7	0,998
	t_z															
Mashreghi et coll. (2016)	1,5	3,9	5,4	1,000	1,3	5,0	6,3	1,000	1,7	3,4	5,1	1,000	1,9	3,9	5,8	1,000
Chauvet (2007) simplifié	1,5	3,7	5,2	0,998	1,4	5,0	6,4	1,000	1,5	3,1	4,6	0,996	1,7	4,6	6,3	0,985
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,1	3,3	5,4	0,999	1,5	4,9	6,4	0,999	2,1	2,3	4,4	1,002	2,2	3,4	5,6	0,998
Barbiero et Mecatti (2010) <i>x-equ2</i>	1,7	3,7	5,4	0,999	1,5	4,5	6,0	0,999	2,2	2,4	4,6	1,004	2,7	3,4	6,1	1,002
Beaumont et Patak (2012) Norm	1,6	3,8	5,4	1,001	1,1	5,3	6,4	1,001	1,6	3,2	4,8	1,001	1,6	4,3	5,9	1,000
Beaumont et Patak (2012) Exp	2,0	3,6	5,6	1,002	2,1	4,0	6,1	0,998	2,1	2,5	4,6	1,001	2,5	3,6	6,1	0,998
Beaumont et Patak (2012) $\epsilon = 1$	1,5	4,1	5,6	1,001	1,3	5,1	6,4	0,996	1,4	3,5	4,9	1,002	1,9	4,3	6,2	0,993
Antal et Tillé (2014)	1,7	4,2	5,9	0,999	1,0	5,4	6,4	0,998	1,4	3,4	4,8	1,000	1,6	4,5	6,1	0,996
Antal et Tillé (2011a)	2,0	3,7	5,7	1,000	1,6	4,8	6,4	0,999	1,8	2,9	4,7	1,000	2,0	3,5	5,5	0,998

Tab. 4.39. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires, t_y et t_z).

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Mashreghi et coll. (2016)	1,2	3,8	5,0	1,000	1,4	4,9	6,3	1,000	2,0	3,6	5,6	1,000	1,7	4,3	6,0	1,000
Chauvet (2007) simplifié	1,1	3,8	4,9	0,999	1,1	4,8	5,9	1,000	1,9	3,4	5,3	0,998	1,6	4,4	6,0	0,989
Barbiero et Mecatti (2010) $x-equ1$	1,1	4,0	5,1	0,999	1,3	4,9	6,2	1,000	2,3	2,7	5,0	1,000	2,6	3,7	6,3	1,001
Barbiero et Mecatti (2010) $x-equ2$	1,5	3,3	4,8	1,000	1,3	4,7	6,0	1,000	2,1	2,7	4,8	1,004	2,7	3,4	6,1	1,006
Beaumont et Patak (2012) Norm	1,3	3,6	4,9	1,000	1,2	4,1	5,3	1,001	1,9	3,3	5,2	1,001	1,8	3,7	5,5	1,001
Beaumont et Patak (2012) Exp	0,7	4,0	4,7	0,999	0,8	5,6	6,4	0,999	1,5	3,9	5,4	1,002	0,9	5,2	6,1	0,998
Beaumont et Patak (2012) $\epsilon = 1$	1,5	3,5	5,0	0,999	1,5	4,4	5,9	0,995	1,9	3,4	5,3	1,000	1,9	4,1	6,0	0,997
Antal et Tillé (2014)	1,5	3,5	5,0	0,999	1,5	3,9	5,4	1,001	1,8	3,4	5,2	1,000	2,0	4,0	6,0	0,998
Antal et Tillé (2011a)	1,3	3,7	5,0	1,000	1,1	4,7	5,8	1,001	1,5	3,3	4,8	1,001	1,2	4,9	6,1	0,998
	t_z															
Mashreghi et coll. (2016)	1,4	4,3	5,7	1,000	0,9	5,5	6,4	1,000	1,2	3,7	4,9	1,000	1,7	4,5	6,2	1,000
Chauvet (2007) simplifié	1,5	4,5	6,0	0,998	1,0	5,4	6,4	1,000	1,3	3,5	4,8	0,996	1,5	4,5	6,0	0,985
Barbiero et Mecatti (2010) $x-equ1$	1,5	4,1	5,6	0,999	0,9	5,4	6,3	0,999	2,2	2,5	4,7	1,002	2,3	3,9	6,2	0,998
Barbiero et Mecatti (2010) $x-equ2$	1,3	3,8	5,1	0,999	1,0	5,4	6,4	0,999	1,8	2,6	4,4	1,004	2,2	3,7	5,9	1,002
Beaumont et Patak (2012) Norm	1,3	4,1	5,4	1,001	1,2	5,2	6,4	1,001	1,5	3,3	4,8	1,001	2,0	4,3	6,3	1,000
Beaumont et Patak (2012) Exp	1,4	5,0	6,4	1,002	0,4	6,0	6,4	0,998	1,3	4,2	5,5	1,001	0,8	5,6	6,4	0,998
Beaumont et Patak (2012) $\epsilon = 1$	1,4	4,1	5,5	1,001	1,2	5,2	6,4	0,996	1,7	3,2	4,9	1,002	1,8	4,6	6,4	0,993
Antal et Tillé (2014)	1,6	3,7	5,3	0,999	1,3	5,0	6,3	0,998	1,8	3,3	5,1	1,000	1,8	4,4	6,2	0,996
Antal et Tillé (2011a)	1,4	4,7	6,1	1,000	0,7	5,7	6,4	0,999	1,1	3,9	5,0	1,000	0,9	5,3	6,2	0,998

Tab. 4.40. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres linéaires, t_y et t_z).

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Holmberg (1998)	2,4	2,4	4,8	1,000	2,3	2,7	5,0	1,000	2,6	2,7	5,3	1,000	3,0	1,9	4,9	1,000
Mashreghi et coll. (2016)	2,9	2,6	5,5	1,003	2,5	2,8	5,3	0,998	2,6	2,5	5,1	0,987	3,1	2,1	5,2	0,982
Chauvet (2007) simplifié	2,6	2,7	5,3	0,999	2,6	2,5	5,1	1,000	2,8	2,6	5,4	0,983	3,2	1,9	5,1	0,973
Barbiero et Mecatti (2010) $x-equ1$	2,6	2,5	5,1	1,000	2,4	2,3	4,7	0,999	3,2	2,2	5,4	0,986	4,3	1,4	5,7	0,985
Barbiero et Mecatti (2010) $x-equ2$	3,3	2,4	5,7	1,000	2,7	2,4	5,1	0,998	3,6	2,2	5,8	0,990	4,2	1,7	5,9	0,991
Antal et Tillé (2014)	2,8	2,3	5,1	1,031	2,1	2,1	4,2	1,044	2,6	2,1	4,7	1,014	3,0	1,7	4,7	1,014
Antal et Tillé (2011a)	2,5	2,2	4,7	1,025	2,2	2,4	4,6	1,025	2,0	2,5	4,5	1,006	2,4	2,5	4,9	0,997
	t_z															
Holmberg (1998)	2,5	2,8	5,3	1,000	2,0	2,6	4,6	1,000	2,2	2,3	4,5	1,000	2,6	2,2	4,8	1,000
Mashreghi et coll. (2016)	2,5	2,8	5,3	1,005	2,1	2,6	4,7	0,999	2,2	1,8	4,0	0,986	2,4	2,4	4,8	0,977
Chauvet (2007) simplifié	2,7	3,0	5,7	0,997	2,4	2,2	4,6	1,000	2,3	2,2	4,5	0,980	3,2	2,4	5,6	0,967
Barbiero et Mecatti (2010) $x-equ1$	2,7	2,6	5,3	1,001	2,6	2,4	5,0	0,997	3,4	1,8	5,2	0,986	3,4	1,7	5,1	0,979
Barbiero et Mecatti (2010) $x-equ2$	2,9	2,7	5,6	0,999	2,5	2,4	4,9	0,999	3,6	1,8	5,4	0,988	3,7	1,9	5,6	0,983
Antal et Tillé (2014)	2,6	2,3	4,9	1,035	2,2	1,8	4,0	1,050	2,3	1,8	4,1	1,014	2,7	1,7	4,4	1,013
Antal et Tillé (2011a)	2,1	2,3	4,4	1,026	2,3	2,1	4,4	1,022	1,8	2,3	4,1	1,004	2,1	2,3	4,4	0,990

Tab. 4.41. Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (Totaux t_y et t_z)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
	Holmberg (1998)	1,004	1,011	1,006	1,034	1,021	1,025	1,019
Chauvet (2007)	1,003	–	1,006	–	1,005	–	1,004	–
Mashreghi et coll. (2016)	1,006	1,014	1,006	1,032	1,006	1,011	1,003	1,024
Chauvet (2007) simplifié	1,003	1,010	1,006	1,034	1,003	1,008	0,994	1,015
Barbiero et Mecatti (2010) $x-equ1$	1,003	1,011	1,005	1,034	1,005	1,011	1,006	1,027
Barbiero et Mecatti (2010) $x-equ2$	1,004	1,011	1,005	1,033	1,010	1,014	1,011	1,034
Beaumont et Patak (2012) Norm	1,004	–	1,005	–	1,006	–	1,003	–
Beaumont et Patak (2012) Exp	1,025	–	1,027	–	1,098	–	1,105	–
Beaumont et Patak (2012) $\epsilon = 1$	1,024	–	1,026	–	1,097	–	1,104	–
Antal et Tillé (2014)	1,004	1,042	1,006	1,080	1,006	1,040	1,004	1,058
Antal et Tillé (2011a)	1,005	1,036	1,006	1,060	1,005	1,031	1,003	1,041
t_z								
Holmberg (1998)	1,005	1,014	1,006	1,046	1,021	1,029	1,020	1,060
Chauvet (2007)	1,004	–	1,006	–	1,005	–	1,004	–
Mashreghi et coll. (2016)	1,007	1,020	1,006	1,045	1,007	1,015	1,002	1,035
Chauvet (2007) simplifié	1,004	1,012	1,006	1,046	1,000	1,009	0,989	1,024
Barbiero et Mecatti (2010) $x-equ1$	1,005	1,015	1,005	1,043	1,005	1,015	1,004	1,038
Barbiero et Mecatti (2010) $x-equ2$	1,004	1,013	1,005	1,045	1,010	1,017	1,009	1,042
Beaumont et Patak (2012) Norm	1,006	–	1,006	–	1,006	–	1,002	–
Beaumont et Patak (2012) Exp	1,027	–	1,026	–	1,099	–	1,104	–
Beaumont et Patak (2012) $\epsilon = 1$	1,025	–	1,025	–	1,098	–	1,102	–
Antal et Tillé (2014)	1,006	1,050	1,006	1,098	1,006	1,044	1,002	1,073
Antal et Tillé (2011a)	1,006	1,041	1,006	1,069	1,005	1,033	1,002	1,049

4.5.2. Discussion: cas des paramètres linéaires

À la lumière des résultats indiqués dans le tableau 4.36, on constate que les estimateurs bootstrap de la variance dans le cas linéaire exhibent des biais relatifs faibles (inférieur à 2% en valeur absolue) dans tous les scénarios sauf pour la méthode de Holmberg (1998) qui a produit un biais relatif avoisinant 4% dans les deux derniers scénarios. En ce qui concerne la stabilité des estimateurs de variance, on note que les méthodes bootstrap par pseudo-population x -équilibré proposées par Barbiero et Mecatti (2010) sont relativement plus stables.

Examinons maintenant les performances des différentes méthodes dans la production d'intervalles de confiance dans le cas des paramètres linéaires t_y et t_z (tableaux 4.37 à 4.40). Toutes les méthodes ont des comportements relativement voisins, le taux de couverture théorique bilatérale est bien respecté par tous les intervalles de confiance et dans tous les scénarios. Il faut signaler également que les taux d'erreur unilatérale sont, à une exception près, bien respectés par les intervalles de confiance t -bootstrap pour toutes les méthodes bootstrap et dans les quatre scénarios. Toutefois il faut signaler que ce gain en terme des taux d'erreur de couverture unilatérale est accompagné par des intervalles de confiance plus larges que les intervalles percentiles d'environ 1% à 5% pour les méthodes bootstrap par pseudo-population et de 4% à 10% pour les deux méthodes Antal et Tillé (2011a) et Antal et Tillé (2014) (voir tableau 4.41). Pour certaines méthodes bootstrap par pseudo-population on peut également remarquer que les intervalles de confiance asymptotique et percentile font mieux que les intervalles de confiance bootstrap de base en terme d'erreur de couverture unilatérale.

4.5.3. Résultats: cas des paramètres lisses de totaux

Le tableau 4.42 présente le biais relatif (en %) ainsi que la racine de l'erreur quadratique moyenne relative des estimateurs bootstrap de la variance étudiés pour le plan de Poisson, dans le cas du ratio et du coefficient de corrélation. Les tableaux 4.43 à 4.46 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production

Tab. 4.42. Étude 2: Biais relatif et stabilité des méthodes bootstrap, (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
$r_{y,z}$								
Holmberg (1998)	2,24	23,53	6,43	46,16	3,47	21,03	7,01	41,01
Chauvet (2007)	1,96	23,59	6,30	45,89	1,50	20,64	4,33	39,54
Mashreghi et coll. (2016)	2,29	23,57	7,20	46,45	1,59	20,49	4,85	40,35
Chauvet (2007) simplifié	2,09	23,52	6,47	45,94	3,17	21,02	8,82	42,75
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,12	22,85	5,84	44,68	1,99	20,53	2,16	35,37
Barbiero et Mecatti (2010) <i>x-equ2</i>	2,27	22,58	5,92	44,28	4,32	21,39	3,93	36,20
Beaumont et Patak (2012) Norm	4,01	24,81	13,45	53,00	1,68	20,65	6,66	41,66
Beaumont et Patak (2012) Exp	0,37	23,05	0,32	42,12	-0,86	20,08	-2,82	36,30
Beaumont et Patak (2012) $\epsilon = 1$	3,88	24,45	12,21	51,06	1,83	20,63	6,07	40,68
Antal et Tillé (2014)	3,92	24,39	12,04	50,69	1,89	20,46	6,30	41,25
Antal et Tillé (2011a)	1,98	24,09	5,77	45,77	0,25	20,21	0,69	37,86
$\rho_{y,z}$								
Holmberg (1998)	2,34	36,23	21,23	91,31	4,72	25,69	12,66	69,25
Chauvet (2007)	1,85	36,04	21,09	92,26	0,57	24,81	8,91	65,30
Mashreghi et coll. (2016)	3,11	36,21	22,86	92,04	0,57	24,72	9,97	67,98
Chauvet (2007) simplifié	1,76	35,77	20,56	94,34	5,81	26,44	12,57	67,41
Barbiero et Mecatti (2010) <i>x-equ1</i>	1,69	34,65	20,74	88,90	2,05	24,75	6,39	61,73
Barbiero et Mecatti (2010) <i>x-equ2</i>	2,45	35,30	21,17	90,08	7,88	27,28	11,68	64,85
Beaumont et Patak (2012) Norm	-2,74	34,88	2,55	72,49	-4,74	23,51	-9,65	51,31
Beaumont et Patak (2012) Exp	-5,91	32,65	-1,47	64,89	-8,31	23,31	-17,03	47,96
Beaumont et Patak (2012) $\epsilon = 1$	7,18	39,81	39,55	116,66	0,10	24,78	5,06	61,31
Antal et Tillé (2014)	8,17	40,38	44,37	124,99	1,88	25,41	14,21	72,34
Antal et Tillé (2011a)	0,84	34,89	18,54	89,70	-4,26	23,90	-4,02	55,23

d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap du ratio et de coefficient de corrélation. Le tableau 4.47 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.43. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Holmberg (1998)	1,6	3,2	4,8	1,000	0,6	4,5	5,1	1,000	2,0	3,5	5,5	1,000	1,5	4,3	5,8	1,000
Chauvet (2007)	1,7	3,1	4,8	0,995	1,0	4,6	5,6	1,000	2,0	3,3	5,3	0,988	1,6	4,5	6,1	0,987
Mashreghi et coll. (2016)	1,5	3,2	4,7	1,000	0,8	4,6	5,4	1,003	2,1	3,7	5,8	0,986	1,5	4,5	6,0	0,988
Chauvet (2007) simplifié	1,7	3,1	4,8	1,000	0,7	4,6	5,3	1,000	1,9	3,6	5,5	1,000	1,4	4,4	5,8	1,008
Barbiero et Mecatti (2010) $x-equ1$	1,6	3,3	4,9	1,000	0,8	4,6	5,4	0,997	1,8	3,4	5,2	0,993	1,4	4,8	6,2	0,977
Barbiero et Mecatti (2010) $x-equ2$	1,7	3,0	4,7	1,000	0,8	4,6	5,4	1,000	1,7	3,5	5,2	1,000	1,4	4,6	6,0	0,984
Beaumont et Patak (2012) Norm	1,5	3,0	4,5	1,000	0,5	4,6	5,1	1,031	2,0	3,6	5,6	0,986	1,3	4,4	5,7	0,996
Beaumont et Patak (2012) Exp	1,6	3,2	4,8	0,988	1,0	4,7	5,7	0,972	2,1	3,9	6,0	0,978	1,5	5,0	6,5	0,953
Beaumont et Patak (2012) $\epsilon = 1$	1,5	2,9	4,4	1,000	0,4	4,6	5,0	1,024	2,0	3,7	5,7	0,986	1,5	4,4	5,9	0,992
Antal et Tillé (2014)	1,4	3,2	4,6	1,000	0,3	4,5	4,8	1,024	2,0	3,5	5,5	0,986	1,6	4,6	6,2	0,996
Antal et Tillé (2011a)	1,5	3,0	4,5	1,000	0,8	4,6	5,4	0,997	2,0	3,5	5,5	0,986	1,6	4,7	6,3	0,969
	$\rho_{y,z}$															
Holmberg (1998)	5,5	1,3	6,8	1,000	7,0	1,1	8,1	1,000	3,8	2,0	5,8	1,000	4,5	1,4	5,9	1,000
Chauvet (2007)	5,3	1,2	6,5	0,991	7,2	0,9	8,1	0,998	4,2	2,0	6,2	0,982	4,6	1,3	5,9	0,986
Mashreghi et coll. (2016)	5,4	1,2	6,6	1,000	7,1	0,8	7,9	1,006	4,0	2,2	6,2	0,977	4,7	1,4	6,1	0,986
Chauvet (2007) simplifié	5,8	1,3	7,1	0,981	7,2	1,0	8,2	0,994	3,8	1,6	5,4	1,011	4,6	1,5	6,1	1,000
Barbiero et Mecatti (2010) $x-equ1$	5,5	1,3	6,8	0,991	7,5	0,9	8,4	1,000	4,3	2,1	6,4	0,989	5,5	1,7	7,2	0,973
Barbiero et Mecatti (2010) $x-equ2$	5,2	1,2	6,4	1,000	6,9	1,0	7,9	1,000	3,3	1,7	5,0	1,011	5,1	1,5	6,6	1,000
Beaumont et Patak (2012) Norm	5,6	1,8	7,4	0,962	8,1	1,9	10,0	0,922	4,4	2,1	6,5	0,955	6,0	2,5	8,5	0,899
Beaumont et Patak (2012) Exp	6,1	1,8	7,9	0,962	8,4	1,7	10,1	0,906	4,5	2,6	7,1	0,932	7,1	2,8	9,9	0,858
Beaumont et Patak (2012) $\epsilon = 1$	4,5	1,1	5,6	1,019	6,5	0,8	7,3	1,067	4,0	2,1	6,1	0,977	4,9	1,7	6,6	0,973
Antal et Tillé (2014)	4,7	0,8	5,5	1,019	6,2	0,5	6,7	1,083	4,1	2,1	6,2	0,989	4,7	1,4	6,1	1,007
Antal et Tillé (2011a)	5,7	1,4	7,1	0,981	7,4	1,1	8,5	0,989	4,2	2,4	6,6	0,955	5,9	1,9	7,8	0,926

Tab. 4.44. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Mashreghi et coll. (2016)	3,1	2,4	5,5	1,000	2,9	3,9	6,8	1,000	3,3	3,0	6,3	1,000	2,6	3,7	6,3	1,000
Chauvet (2007) simplifié	2,8	2,7	5,5	1,000	2,8	3,3	6,1	1,000	3,3	2,7	6,0	1,007	2,5	3,4	5,9	1,020
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,7	2,6	5,3	0,994	3,5	3,5	7,0	0,997	3,4	2,8	6,2	1,000	3,0	3,8	6,8	0,992
Barbiero et Mecatti (2010) <i>x-equ2</i>	3,1	2,5	5,6	1,000	3,1	3,9	7,0	0,997	3,0	2,5	5,5	1,015	3,1	3,7	6,8	0,996
Beaumont et Patak (2012) Norm	2,8	2,3	5,1	1,006	2,9	3,0	5,9	1,031	3,2	2,8	6,0	1,000	3,0	3,4	6,4	1,012
Beaumont et Patak (2012) Exp	2,8	2,9	5,7	0,988	3,2	4,2	7,4	0,972	3,3	2,9	6,2	0,985	2,9	4,1	7,0	0,964
Beaumont et Patak (2012) $\epsilon = 1$	2,6	2,5	5,1	1,006	2,9	3,3	6,2	1,028	3,2	3,1	6,3	1,000	2,8	3,4	6,2	1,008
Antal et Tillé (2014)	3,1	2,5	5,6	1,006	2,8	3,3	6,1	1,028	3,0	3,0	6,0	1,000	2,8	3,3	6,1	1,008
Antal et Tillé (2011a)	3,1	2,7	5,8	0,994	3,2	3,9	7,1	0,997	3,3	2,8	6,1	0,993	2,9	3,7	6,6	0,980
	$\rho_{y,z}$															
Mashreghi et coll. (2016)	4,1	1,2	5,3	1,000	6,0	2,1	8,1	1,000	2,8	2,8	5,6	1,000	3,5	1,7	5,2	1,000
Chauvet (2007) simplifié	3,9	1,1	5,0	1,000	5,8	2,1	7,9	1,000	2,5	2,5	5,0	1,023	3,3	1,7	5,0	1,021
Barbiero et Mecatti (2010) <i>x-equ1</i>	4,0	1,0	5,0	1,000	5,6	2,1	7,7	1,000	3,2	2,7	5,9	1,012	4,3	2,2	6,5	0,986
Barbiero et Mecatti (2010) <i>x-equ2</i>	3,9	1,0	4,9	1,000	5,9	2,0	7,9	1,000	2,9	2,5	5,4	1,035	3,4	2,2	5,6	1,014
Beaumont et Patak (2012) Norm	5,9	1,7	7,6	0,971	7,7	2,2	9,9	0,927	4,2	2,3	6,5	0,977	5,2	2,5	7,7	0,911
Beaumont et Patak (2012) Exp	5,0	1,5	6,5	0,962	6,8	2,4	9,2	0,904	4,1	2,6	6,7	0,953	5,7	3,0	8,7	0,870
Beaumont et Patak (2012) $\epsilon = 1$	3,6	1,1	4,7	1,029	5,2	1,7	6,9	1,079	3,1	2,9	6,0	1,000	3,9	2,1	6,0	0,979
Antal et Tillé (2014)	3,7	1,0	4,7	1,029	5,2	1,6	6,8	1,084	2,8	2,9	5,7	1,012	3,3	1,5	4,8	1,021
Antal et Tillé (2011a)	4,4	1,2	5,6	0,990	6,0	2,1	8,1	0,994	3,5	2,7	6,2	0,977	3,8	2,2	6,0	0,945

Tab. 4.45. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Mashreghi et coll. (2016)	1,2	4,0	5,2	1,000	0,3	6,2	6,5	1,000	1,3	5,1	6,4	1,000	0,6	6,7	7,3	1,000
Chauvet (2007) simplifié	1,1	4,2	5,3	1,000	0,1	6,2	6,3	1,000	1,4	4,9	6,3	1,007	0,3	6,2	6,5	1,020
Barbiero et Mecatti (2010) <i>x-equ1</i>	1,1	3,7	4,8	0,994	0,2	6,4	6,6	0,997	1,4	5,0	6,4	1,000	0,7	6,6	7,3	0,992
Barbiero et Mecatti (2010) <i>x-equ2</i>	0,9	4,1	5,0	1,000	0,2	6,5	6,7	0,997	1,5	4,8	6,3	1,015	0,7	6,5	7,2	0,996
Beaumont et Patak (2012) Norm	1,0	4,0	5,0	1,006	0,2	5,7	5,9	1,031	1,3	5,1	6,4	1,000	0,6	6,4	7,0	1,012
Beaumont et Patak (2012) Exp	1,3	4,0	5,3	0,988	0,4	6,6	7,0	0,972	1,8	5,1	6,9	0,985	1,0	7,0	8,0	0,964
Beaumont et Patak (2012) $\epsilon = 1$	0,9	3,7	4,6	1,006	0,1	5,5	5,6	1,028	1,4	5,3	6,7	1,000	0,7	6,3	7,0	1,008
Antal et Tillé (2014)	1,1	3,7	4,8	1,006	0,2	5,6	5,8	1,028	1,5	5,1	6,6	1,000	0,6	6,5	7,1	1,008
Antal et Tillé (2011a)	1,1	4,3	5,4	0,994	0,2	6,2	6,4	0,997	1,3	5,1	6,4	0,993	0,7	6,3	7,0	0,980
	$\rho_{y,z}$															
Mashreghi et coll. (2016)	7,1	1,3	8,4	1,000	11,3	0,8	12,1	1,000	5,6	1,9	7,5	1,000	7,9	1,4	9,3	1,000
Chauvet (2007) simplifié	7,4	1,7	9,1	1,000	11,0	0,9	11,9	1,000	5,0	1,4	6,4	1,023	7,9	1,3	9,2	1,021
Barbiero et Mecatti (2010) <i>x-equ1</i>	7,1	1,5	8,6	1,000	11,1	0,9	12,0	1,000	5,5	1,6	7,1	1,012	7,8	1,5	9,3	0,986
Barbiero et Mecatti (2010) <i>x-equ2</i>	7,3	1,4	8,7	1,000	11,2	0,8	12,0	1,000	4,9	1,3	6,2	1,035	7,4	1,0	8,4	1,014
Beaumont et Patak (2012) Norm	5,9	1,6	7,5	0,971	9,4	1,4	10,8	0,927	4,9	2,1	7,0	0,977	6,9	2,1	9,0	0,911
Beaumont et Patak (2012) Exp	7,8	1,8	9,6	0,962	11,8	1,1	12,9	0,904	5,4	2,6	8,0	0,953	8,9	2,4	11,3	0,870
Beaumont et Patak (2012) $\epsilon = 1$	6,7	1,5	8,2	1,029	10,5	0,7	11,2	1,079	5,2	1,4	6,6	1,000	7,9	1,4	9,3	0,979
Antal et Tillé (2014)	6,7	1,3	8,0	1,029	10,0	0,7	10,7	1,084	5,7	1,5	7,2	1,012	7,4	1,6	9,0	1,021
Antal et Tillé (2011a)	7,4	1,7	9,1	0,990	11,2	1,2	12,4	0,994	5,5	1,9	7,4	0,977	8,4	1,6	10,0	0,945

Tab. 4.46. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Holmberg (1998)	2,3	2,3	4,6	1,000	2,5	3,4	5,9	1,000	3,3	2,7	6,0	1,000	2,4	3,2	5,6	1,000
Mashreghi et coll. (2016)	2,4	2,4	4,8	1,000	2,7	2,9	5,6	1,003	3,0	2,6	5,6	1,000	2,4	3,5	5,9	1,000
Chauvet (2007) simplifié	2,4	2,7	5,1	1,000	2,6	3,4	6,0	1,000	2,4	2,6	5,0	1,014	2,2	3,3	5,5	1,008
Barbiero et Mecatti (2010) $x-equ1$	2,3	2,3	4,6	1,000	2,1	3,3	5,4	1,003	2,7	2,3	5,0	1,007	1,9	3,4	5,3	1,004
Barbiero et Mecatti (2010) $x-equ2$	2,7	2,0	4,7	1,000	2,6	3,0	5,6	1,000	2,9	2,8	5,7	1,007	2,4	3,6	6,0	1,008
Antal et Tillé (2014)	2,1	2,2	4,3	1,012	2,2	2,9	5,1	1,037	3,2	2,5	5,7	1,014	1,9	3,4	5,3	1,015
Antal et Tillé (2011a)	2,0	2,8	4,8	1,000	2,5	3,3	5,8	1,000	2,7	2,5	5,2	1,000	2,4	3,4	5,8	0,981
	$\rho_{y,z}$															
Holmberg (1998)	2,0	2,5	4,5	1,000	3,5	3,6	7,1	1,000	2,4	2,7	5,1	1,000	1,9	3,3	5,2	1,000
Mashreghi et coll. (2016)	2,2	2,6	4,8	1,000	3,1	3,5	6,6	1,005	2,4	2,8	5,2	1,000	2,5	3,2	5,7	0,994
Chauvet (2007) simplifié	1,9	2,5	4,4	1,000	3,3	3,8	7,1	1,000	2,3	2,4	4,7	1,022	1,8	2,8	4,6	1,031
Barbiero et Mecatti (2010) $x-equ1$	2,0	2,8	4,8	1,000	3,9	3,6	7,5	1,005	2,3	2,6	4,9	1,011	2,4	3,1	5,5	1,012
Barbiero et Mecatti (2010) $x-equ2$	2,2	2,8	5,0	1,000	3,4	4,1	7,5	1,005	2,2	2,8	5,0	1,022	2,1	2,9	5,0	1,000
Antal et Tillé (2014)	1,8	2,4	4,2	1,054	2,5	3,1	5,6	1,121	2,5	2,5	5,0	1,011	1,9	3,2	5,1	1,043
Antal et Tillé (2011a)	1,9	2,7	4,6	0,991	3,3	4,1	7,4	0,995	2,4	3,2	5,6	0,978	2,3	3,5	5,8	0,951

Tab. 4.47. Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) ($r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
$r_{y,z}$								
Holmberg (1998)	1,006	1,024	1,003	1,028	1,015	1,022	1,016	1,032
Chauvet (2007)	1,001	–	1,003	–	1,003	–	1,003	–
Mashreghi et coll. (2016)	1,006	1,024	1,007	1,031	1,000	1,022	1,004	1,032
Chauvet (2007) simplifié	1,006	1,024	1,003	1,028	1,015	1,037	1,024	1,040
Barbiero et Mecatti (2010) $x-equ1$	1,006	1,024	1,000	1,031	1,007	1,029	0,992	1,036
Barbiero et Mecatti (2010) $x-equ2$	1,006	1,024	1,003	1,028	1,015	1,029	1,000	1,040
Beaumont et Patak (2012) Norm	0,994		0,969		0,993		0,968	
Beaumont et Patak (2012) Exp	1,018		0,997		1,088		1,063	
Beaumont et Patak (2012) $\epsilon = 1$	1,030		1,049		1,096		1,111	
Antal et Tillé (2014)	1,006	1,036	1,028	1,066	1,000	1,037	1,012	1,048
Antal et Tillé (2011a)	1,006	1,024	1,000	1,028	1,000	1,022	0,984	1,012
$\rho_{y,z}$								
Holmberg (1998)	1,010	1,067	1,011	1,112	1,023	1,047	1,014	1,110
Chauvet (2007)	1,001	–	1,009	–	1,005	–	1,000	–
Mashreghi et coll. (2016)	1,010	1,067	1,017	1,118	1,000	1,047	1,000	1,103
Chauvet (2007) simplifié	0,990	1,067	1,006	1,112	1,035	1,070	1,014	1,144
Barbiero et Mecatti (2010) $x-equ1$	1,000	1,067	1,011	1,118	1,012	1,058	0,986	1,123
Barbiero et Mecatti (2010) $x-equ2$	1,010	1,067	1,011	1,118	1,035	1,070	1,014	1,110
Beaumont et Patak (2012) Norm	0,971	–	0,933	–	0,977	–	0,918	–
Beaumont et Patak (2012) Exp	0,971	–	0,938	–	1,047	–	0,966	–
Beaumont et Patak (2012) $\epsilon = 1$	1,048	–	1,112	–	1,105	–	1,110	–
Antal et Tillé (2014)	1,029	1,124	1,096	1,247	1,012	1,058	1,021	1,158
Antal et Tillé (2011a)	0,990	1,057	1,000	1,107	0,977	1,023	0,938	1,055

4.5.4. Discussion: cas des paramètres lisses de totaux

Nous énumérons les conclusions tirées du tableau 4.36 dans les points suivants:

- Les estimateurs bootstrap de la variance dans le cas du ratio sont plus précis et plus stables que dans le cas du coefficient de corrélation;
- Les méthodes bootstrap par pseudo-population sont relativement semblables dans les deux premiers scénarios et pour les deux paramètres étudiés. Cependant, dans les deux derniers scénarios, il semble que la méthode de Chauvet (2007), la méthode de Mashreghi et coll. (2016) et la première version de la méthode de Barbiero et Mecatti (2010) x -équilibré font mieux que les autres;
- La méthode de Chauvet (2007) et la méthode de Mashreghi et coll. (2016) donnent des résultats très proches en terme du BR et $REQMR$ avec un léger avantage en faveur de la méthode de Chauvet (2007);
- Pour la méthode de Beaumont et Patak (2012), il semble que le choix de la distribution pour générer les ajustements bootstrap a un effet sur la précision et la stabilité de l'estimateur bootstrap de la variance par cette méthode. En effet, il semble que le choix de la distribution exponentielle soit plus approprié dans le cas du ratio mais pas nécessairement dans le cas du coefficient de corrélation;
- L'estimation bootstrap de la variance par la méthode de Beaumont et Patak (2012) mise en œuvre avec une distribution exponentielle est inférieure à celle issue des autres méthodes bootstrap. Cette caractéristique lui a permis de faire toujours mieux que les autres méthodes bootstrap quand ces dernières sur-estiment considérablement la vraie variance (scénario 2 dans le cas du coefficient de corrélation);
- La méthode de Antal et Tillé (2014) et la méthode de Beaumont et Patak (2012) mise en œuvre avec la distribution de l'équation (2.2.8) avec $\epsilon = 1$ ont produit les plus grands BR et $REQMR$ dans les deux premiers scénarios et ceci pour les deux paramètres étudiés.

A la lumière des résultats indiqués dans le tableau 4.43 qui rapporte les propriétés empiriques d'intervalles de confiance asymptotique, on constate que le taux de couverture

théorique bilatérale est bien respecté dans le cas du ratio; les taux de couverture bilatérale observés dans les quatre scénarios sont entre 93,5% et 96,5%. Les taux d'erreur unilatérale sont aussi respectés par la plupart des méthodes dans les scénarios 1 et 3. En ce qui concerne le coefficient de corrélation, il semble que le taux de couverture bilatérale des intervalles de confiance asymptotique est affecté par la précision et la stabilité des estimateurs bootstrap de la variance, notamment la stabilité. Prenons les deux cas extrêmes en terme de la stabilité, le scénario 2 (le plus instable) et le scénario 3 (le plus stable). Dans ce dernier scénario, le taux de couverture bilatérale est bien respecté par les intervalles de confiance asymptotique pour toutes les méthodes à l'exception de la méthode de Beaumont et Patlak (2012) mise en œuvre avec une distribution exponentielle dont le taux de couverture bilatérale observé était de 93% qui reste quand même proche de 95%. En revanche, dans le scénario 2 les estimateurs bootstrap de la variance sont moins stables ce qui a affecté le taux de couverture bilatérale pour la plupart des méthodes, notamment celles qui ont produit un BR faible.

Pour les intervalles de confiance percentile (tableau 4.44) et les intervalles de confiance bootstrap de base (tableau 4.45), on constate pour les deux paramètres étudiés, que les méthodes bootstrap par pseudo-population ont des performances proches en terme des propriétés empiriques de ces deux types d'intervalles de confiance. Notons également que les intervalles de confiance percentile font mieux que les intervalles de confiance asymptotique et les intervalles bootstrap de base en terme des erreurs unilatérales et ceci pour les différentes méthodes bootstrap. La technique bootstrap de base a produit les résultats les moins performants en terme d'erreur de couverture bilatérale dans le cas du coefficient de corrélation.

Les intervalles de confiance t -bootstrap associés aux différentes méthodes (tableau 4.46) respectent bien les taux d'erreur de couverture bilatérale et unilatérale dans tous les scénarios et pour les deux paramètres étudiés sauf pour la borne supérieure dans le deuxième scénario dans le cas du coefficient de corrélation où seulement la méthode de Antal et Tillé (2014) a une erreur unilatérale qui n'est pas statistiquement différente de 2,5%. Le tableau 4.47 nous renseigne que le choix de la méthode de Antal et Tillé (2011a) pour mettre en

œuvre le t -bootstrap est optimal en terme de la largeur dans le cas des deux paramètres étudiés. En effet, compte tenu des quatre scénarios, la largeur relative ne dépasse guère le 3% dans le cas du ratio et entre 2% et 11% dans le cas du coefficient de corrélation.

4.5.5. Résultats: cas des paramètres non lisses de totaux

Le tableau 4.48 présente les biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne des estimateurs bootstrap de la variance étudiés, dans le cas de la médiane et de l'indice de Gini de la variable y . Les tableaux 4.49 à 4.52 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile et t -bootstrap de la médiane et de l'indice de Gini de la variable y . Le tableau 4.53 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.48. Étude 2: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses m_y et G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
m_y								
Holmberg (1998)	11,61	58,41	11,42	66,66	6,72	42,35	23,41	70,67
Chauvet (2007)	11,30	57,48	10,88	65,91	4,94	41,52	20,82	68,35
Mashreghi et coll. (2016)	16,20	62,23	19,27	76,05	9,06	44,82	28,83	78,24
Chauvet (2007) simplifié	11,47	58,04	11,34	65,77	3,50	41,10	24,78	70,97
Barbiero et Mecatti (2010) <i>x-equ1</i>	11,41	56,84	10,99	66,16	5,12	40,89	20,37	67,34
Barbiero et Mecatti (2010) <i>x-equ2</i>	11,12	57,04	10,27	65,19	3,80	40,53	16,49	65,57
Beaumont et Patak (2012) Norm	10,93	57,90	10,37	65,39	4,40	41,79	20,08	68,52
Beaumont et Patak (2012) Exp	10,51	57,55	9,03	65,13	3,91	41,26	17,83	67,52
Beaumont et Patak (2012) $\epsilon = 1$	12,23	57,89	13,01	66,99	4,69	41,45	21,69	69,93
Antal et Tillé (2014)	12,07	58,06	13,37	68,54	4,98	41,54	21,97	70,15
Antal et Tillé (2011a)	11,53	58,52	11,33	66,41	3,99	41,12	19,24	68,00
G_y								
Holmberg (1998)	-5,08	38,16	-11,54	45,28	6,71	20,99	-1,01	31,48
Chauvet (2007)	-5,27	38,23	-11,97	45,07	-0,22	19,47	-5,18	31,11
Mashreghi et coll. (2016)	-4,12	38,76	-9,14	45,76	0,58	19,58	-2,88	31,68
Chauvet (2007) simplifié	-4,75	38,33	-11,59	44,97	-0,59	20,47	-5,53	31,26
Barbiero et Mecatti (2010) <i>x-equ1</i>	-5,21	37,77	-12,17	44,89	-0,86	18,77	-6,52	29,46
Barbiero et Mecatti (2010) <i>x-equ2</i>	-5,27	37,47	-12,12	44,92	-0,43	19,24	-5,20	29,67
Beaumont et Patak (2012) Norm	-1,01	40,73	0,99	54,26	0,55	19,67	-0,76	33,41
Beaumont et Patak (2012) Exp	-9,79	36,74	-22,68	43,23	-5,26	18,95	-17,82	30,99
Beaumont et Patak (2012) $\epsilon = 1$	-1,32	40,26	-2,51	50,27	0,60	19,86	-2,38	32,35
Antal et Tillé (2014)	-1,22	40,19	-2,56	50,46	0,60	19,82	-2,07	32,31
Antal et Tillé (2011a)	-5,90	38,15	-13,58	44,84	-3,11	19,05	-11,73	30,59

Tab. 4.49. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Holmberg (1998)	2,2	3,1	5,3	1,000	2,2	3,3	5,5	1,000	4,0	2,8	6,8	1,000	3,0	2,8	5,8	1,000
Chauvet (2007)	2,2	3,1	5,3	0,998	2,2	3,3	5,5	0,998	4,2	3,2	7,4	0,990	3,3	2,7	6,0	0,990
Mashreghi et coll. (2016)	2,0	2,8	4,8	1,020	2,1	2,7	4,8	1,033	3,7	2,9	6,6	1,010	2,9	2,5	5,4	1,019
Chauvet (2007) simplifié	2,5	3,0	5,5	0,998	2,2	3,2	5,4	1,001	4,4	2,9	7,3	0,983	3,3	2,8	6,1	1,006
Barbiero et Mecatti (2010) $x-equ1$	2,3	3,1	5,4	0,998	2,4	3,1	5,5	0,998	4,5	2,8	7,3	0,990	3,5	3,0	6,5	0,987
Barbiero et Mecatti (2010) $x-equ2$	2,3	3,1	5,4	0,998	2,2	3,2	5,4	0,995	4,4	3,5	7,9	0,985	3,5	2,9	6,4	0,970
Beaumont et Patak (2012) Norm	2,2	3,0	5,2	0,997	2,2	3,4	5,6	0,996	4,2	3,2	7,4	0,987	3,2	2,8	6,0	0,985
Beaumont et Patak (2012) Exp	2,4	3,1	5,5	0,995	2,3	3,3	5,6	0,989	4,3	3,3	7,6	0,987	3,3	2,9	6,2	0,977
Beaumont et Patak (2012) $\epsilon = 1$	2,3	2,9	5,2	1,003	2,1	3,3	5,4	1,008	4,2	3,2	7,4	0,990	3,2	2,8	6,0	0,992
Antal et Tillé (2014)	2,2	3,1	5,3	1,002	2,2	3,2	5,4	1,008	4,1	3,2	7,3	0,990	3,1	2,7	5,8	0,994
Antal et Tillé (2011a)	2,2	2,9	5,1	0,998	2,2	3,2	5,4	1,000	4,6	2,9	7,5	0,987	3,0	2,8	5,8	0,983
	G_y															
Holmberg (1998)	1,2	8,5	9,7	1,000	1,1	12,6	13,7	1,000	2,1	4,4	6,5	1,000	1,0	7,1	8,1	1,000
Chauvet (2007)	1,2	8,3	9,5	0,990	0,9	12,6	13,5	0,992	2,3	5,4	7,7	0,955	1,2	7,6	8,8	0,987
Mashreghi et coll. (2016)	1,3	8,3	9,6	1,000	0,9	12,4	13,3	1,010	2,2	5,3	7,5	0,959	1,0	7,3	8,3	1,000
Chauvet (2007) simplifié	1,2	8,6	9,8	1,000	1,0	12,7	13,7	1,000	2,3	5,5	7,8	0,959	1,0	7,4	8,4	0,987
Barbiero et Mecatti (2010) $x-equ1$	1,3	8,6	9,9	1,000	1,0	12,8	13,8	0,990	2,5	5,3	7,8	0,959	1,3	7,3	8,6	0,974
Barbiero et Mecatti (2010) $x-equ2$	1,3	8,2	9,5	1,000	1,2	12,8	14,0	0,990	2,3	5,4	7,7	0,959	1,2	7,5	8,7	0,987
Beaumont et Patak (2012) Norm	1,3	8,3	9,6	1,016	0,6	11,0	11,6	1,058	2,4	5,3	7,7	0,959	1,0	7,0	8,0	1,000
Beaumont et Patak (2012) Exp	1,3	8,9	10,2	0,968	1,5	14,0	15,5	0,932	2,5	5,7	8,2	0,918	1,4	8,9	10,3	0,921
Beaumont et Patak (2012) $\epsilon = 1$	1,3	8,0	9,3	1,000	0,8	11,6	12,4	1,039	2,3	5,4	7,7	0,959	0,9	7,3	8,2	1,000
Antal et Tillé (2014)	1,2	8,2	9,4	1,000	0,8	11,7	12,5	1,039	2,0	5,2	7,2	0,959	0,9	7,2	8,1	1,000
Antal et Tillé (2011a)	1,3	8,5	9,8	0,984	1,1	12,7	13,8	0,981	2,5	5,5	8,0	0,939	1,2	7,9	9,1	0,947

Tab. 4.50. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Mashreghi et coll. (2016)	2,6	2,3	4,9	1,000	2,5	2,3	4,8	1,000	2,4	2,5	4,9	1,000	1,9	1,6	3,5	1,000
Chauvet (2007) simplifié	3,0	2,5	5,5	0,997	2,7	2,0	4,7	0,994	3,3	2,5	5,8	0,979	2,8	2,1	4,9	0,986
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,8	2,4	5,2	0,994	2,8	1,9	4,7	0,991	2,5	2,7	5,2	0,988	3,3	2,2	5,5	0,971
Barbiero et Mecatti (2010) <i>x-equ2</i>	3,2	2,6	5,8	0,992	2,3	1,8	4,1	0,991	2,3	3,0	5,3	0,983	3,8	2,7	6,5	0,950
Beaumont et Patak (2012) Norm	3,0	2,6	5,6	0,992	2,6	2,1	4,7	0,994	2,6	2,6	5,2	0,983	2,5	1,9	4,4	0,968
Beaumont et Patak (2012) Exp	2,9	2,5	5,4	0,991	2,6	2,4	5,0	0,981	2,4	2,6	5,0	0,985	3,1	1,8	4,9	0,961
Beaumont et Patak (2012) $\epsilon = 1$	3,0	2,1	5,1	0,998	2,7	2,0	4,7	1,002	2,9	2,6	5,5	0,986	2,5	1,8	4,3	0,973
Antal et Tillé (2014)	2,8	2,3	5,1	0,998	2,6	2,0	4,6	1,003	2,7	2,6	5,3	0,986	2,7	1,8	4,5	0,975
Antal et Tillé (2011a)	3,0	2,3	5,3	0,992	2,5	1,9	4,4	0,994	3,0	2,7	5,7	0,981	2,7	1,7	4,4	0,965
	G_y															
Mashreghi et coll. (2016)	1,2	10,1	11,3	1,000	0,1	14,8	14,9	1,000	1,3	6,1	7,4	1,000	0,2	9,6	9,8	1,000
Chauvet (2007) simplifié	1,1	9,9	11,0	1,000	0,2	14,9	15,1	1,010	1,4	6,5	7,9	0,979	0,4	9,9	10,3	0,987
Barbiero et Mecatti (2010) <i>x-equ1</i>	1,0	10,1	11,1	1,000	0,2	14,9	15,1	1,000	1,5	6,3	7,8	1,000	0,3	9,6	9,9	0,987
Barbiero et Mecatti (2010) <i>x-equ2</i>	1,1	9,8	10,9	1,000	0,0	14,7	14,7	1,010	1,6	6,8	8,4	0,979	0,5	9,9	10,4	0,987
Beaumont et Patak (2012) Norm	0,9	11,1	12,0	1,016	0,0	16,5	16,5	1,090	1,3	6,2	7,5	1,000	0,2	9,8	10,0	1,013
Beaumont et Patak (2012) Exp	1,3	9,0	10,3	0,984	1,1	13,9	15,0	0,950	2,9	5,1	8,0	0,979	1,0	8,2	9,2	0,933
Beaumont et Patak (2012) $\epsilon = 1$	0,8	11,5	12,3	1,000	0,0	16,7	16,7	1,020	1,7	6,4	8,1	1,000	0,3	10,5	10,8	0,987
Antal et Tillé (2014)	0,8	11,5	12,3	1,000	0,0	16,9	16,9	1,030	1,6	6,4	8,0	1,000	0,1	10,0	10,1	1,000
Antal et Tillé (2011a)	1,2	9,4	10,6	1,000	0,2	14,9	15,1	1,000	2,0	5,3	7,3	0,979	0,8	8,3	9,1	0,973

Tab. 4.51. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Mashreghi et coll. (2016)	3,6	6,7	10,3	1,000	5,0	10,1	15,1	1,000	8,9	5,9	14,8	1,000	5,4	8,2	13,6	1,000
Chauvet (2007) simplifié	3,7	6,3	10,0	0,997	5,9	10,7	16,6	0,994	9,5	5,4	14,9	0,979	6,7	9,6	16,3	0,986
Barbiero et Mecatti (2010) $x-equ1$	3,4	6,2	9,6	0,994	6,0	10,6	16,6	0,991	9,5	6,0	15,5	0,988	6,6	9,4	16,0	0,971
Barbiero et Mecatti (2010) $x-equ2$	3,6	6,7	10,3	0,992	6,6	10,5	17,1	0,991	9,0	6,1	15,1	0,983	5,8	9,3	15,1	0,950
Beaumont et Patak (2012) Norm	7,8	1,8	9,6	0,992	11,8	1,1	12,9	0,994	5,4	2,6	8,0	0,983	8,9	2,4	11,3	0,968
Beaumont et Patak (2012) Exp	3,6	6,3	9,9	0,991	6,3	11,2	17,5	0,981	9,3	6,0	15,3	0,985	6,5	9,8	16,3	0,961
Beaumont et Patak (2012) $\epsilon = 1$	3,4	6,4	9,8	0,998	5,8	11,0	16,8	1,002	9,0	6,5	15,5	0,986	6,3	9,8	16,1	0,973
Antal et Tillé (2014)	3,3	7,1	10,4	0,998	5,8	10,6	16,4	1,003	9,1	6,1	15,2	0,986	6,3	9,8	16,1	0,975
Antal et Tillé (2011a)	3,7	6,5	10,2	0,992	5,9	10,7	16,6	0,994	9,1	6,0	15,1	0,981	6,3	10,0	16,3	0,965
	G_y															
Mashreghi et coll. (2016)	1,5	7,2	8,7	1,000	2,5	11,0	13,5	1,000	3,5	4,3	7,8	1,000	3,7	6,0	9,7	1,000
Chauvet (2007) simplifié	1,6	7,8	9,4	1,000	2,3	10,8	13,1	1,010	3,3	4,3	7,6	0,979	3,2	6,4	9,6	0,987
Barbiero et Mecatti (2010) $x-equ1$	1,8	7,7	9,5	1,000	2,8	10,5	13,3	1,000	3,7	4,5	8,2	1,000	3,8	6,1	9,9	0,987
Barbiero et Mecatti (2010) $x-equ2$	1,4	7,4	8,8	1,000	2,3	10,4	12,7	1,010	3,4	4,8	8,2	0,979	3,5	6,5	10,0	0,987
Beaumont et Patak (2012) Norm	6,7	1,5	8,2	1,016	10,5	0,7	11,2	1,090	5,2	1,4	6,6	1,000	7,9	1,4	9,3	1,013
Beaumont et Patak (2012) Exp	1,3	9,4	10,7	0,984	2,4	14,0	16,4	0,950	2,7	6,3	9,0	0,979	3,1	9,4	12,5	0,933
Beaumont et Patak (2012) $\epsilon = 1$	2,6	6,1	8,7	1,000	3,2	8,3	11,5	1,020	3,6	4,3	7,9	1,000	3,9	6,0	9,9	0,987
Antal et Tillé (2014)	2,7	6,4	9,1	1,000	3,2	8,7	11,9	1,030	3,8	4,1	7,9	1,000	4,0	5,6	9,6	1,000
Antal et Tillé (2011a)	1,5	7,9	9,4	1,000	2,3	10,8	13,1	1,000	2,9	5,5	8,4	0,979	3,4	7,4	10,8	0,973

Tab. 4.52. Étude 2: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Holmberg (1998) (paramètres non lisses m_y et G_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Holmberg (1998)	2,8	3,3	6,1	1,000	4,3	3,6	7,9	1,000	4,8	3,1	7,9	1,000	3,0	2,0	5,0	1,000
Mashreghi et coll. (2016)	2,9	3,0	5,9	0,987	5,4	3,4	8,8	0,986	4,3	3,2	7,5	0,992	3,6	2,1	5,7	0,983
Chauvet (2007) simplifié	3,1	2,7	5,8	0,996	5,0	3,3	8,3	0,999	4,9	3,1	8,0	1,003	3,7	2,7	6,4	1,027
Barbiero et Mecatti (2010) $x-equ1$	2,6	2,8	5,4	0,996	5,1	3,4	8,5	1,002	4,9	2,9	7,8	1,008	4,3	3,1	7,4	1,000
Barbiero et Mecatti (2010) $x-equ2$	2,9	3,1	6,0	0,997	4,8	3,6	8,4	0,997	5,3	3,3	8,6	1,006	3,7	3,2	6,9	1,002
Antal et Tillé (2014)	3,0	3,0	6,0	0,992	5,3	3,5	8,8	0,989	4,9	3,6	8,5	0,997	4,0	3,5	7,5	0,984
Antal et Tillé (2011a)	3,1	2,9	6,0	0,997	5,1	3,4	8,5	1,002	5,4	3,7	9,1	0,994	4,4	3,0	7,4	0,992
	G_y															
Holmberg (1998)	2,0	4,2	6,2	1,000	1,3	5,2	6,5	1,000	3,3	2,2	5,5	1,000	1,3	2,8	4,1	1,000
Mashreghi et coll. (2016)	1,6	4,1	5,7	1,000	1,4	5,1	6,5	0,986	2,9	2,1	5,0	1,020	1,2	2,8	4,0	1,000
Chauvet (2007) simplifié	1,8	3,7	5,5	1,000	1,3	5,2	6,5	1,000	3,2	2,0	5,2	1,020	1,3	2,9	4,2	1,011
Barbiero et Mecatti (2010) $x-equ1$	1,9	4,4	6,3	1,000	1,3	5,0	6,3	1,000	3,2	2,0	5,2	1,020	1,3	2,7	4,0	1,011
Barbiero et Mecatti (2010) $x-equ2$	1,7	4,3	6,0	1,000	1,4	5,0	6,4	1,007	3,2	2,0	5,2	1,020	1,6	3,4	5,0	1,022
Antal et Tillé (2014)	2,8	2,8	5,6	1,070	1,8	4,0	5,8	1,148	3,3	1,8	5,1	1,020	1,7	2,7	4,4	1,043
Antal et Tillé (2011a)	2,0	4,0	6,0	0,986	1,2	5,2	6,4	0,979	2,6	2,5	5,1	0,980	0,9	3,9	4,8	0,957

Tab. 4.53. Étude 2: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (m_y et G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
m_y								
Holmberg (1998)	1,014	1,100	1,051	1,195	1,031	1,139	1,026	1,160
Chauvet (2007)	1,012	–	1,049	–	1,021	–	1,016	–
Mashreghi et coll. (2016)	1,034	1,086	1,085	1,179	1,041	1,131	1,046	1,140
Chauvet (2007) simplifié	1,012	1,095	1,052	1,194	1,014	1,143	1,033	1,192
Barbiero et Mecatti (2010) $x-equ1$	1,012	1,095	1,049	1,197	1,021	1,148	1,013	1,160
Barbiero et Mecatti (2010) $x-equ2$	1,012	1,097	1,045	1,191	1,015	1,146	0,996	1,162
Beaumont et Patak (2012) Norm	1,054	–	1,187	–	1,110	–	1,110	–
Beaumont et Patak (2012) Exp	1,028	–	1,061	–	1,110	–	1,103	–
Beaumont et Patak (2012) $\epsilon = 1$	1,037	–	1,081	–	1,113	–	1,120	–
Antal et Tillé (2014)	1,015	1,090	1,059	1,181	1,021	1,136	1,020	1,141
Antal et Tillé (2011a)	1,012	1,097	1,051	1,197	1,017	1,132	1,009	1,151
G_y								
Holmberg (1998)	1,016	1,164	1,030	1,420	1,043	1,064	1,013	1,227
Chauvet (2007)	1,006	–	1,022	–	0,996	–	1,000	–
Mashreghi et coll. (2016)	1,016	1,164	1,040	1,400	1,000	1,085	1,013	1,227
Chauvet (2007) simplifié	1,016	1,164	1,030	1,420	1,000	1,085	1,000	1,240
Barbiero et Mecatti (2010) $x-equ1$	1,016	1,164	1,020	1,420	1,000	1,085	0,987	1,240
Barbiero et Mecatti (2010) $x-equ2$	1,016	1,164	1,020	1,430	1,000	1,085	1,000	1,253
Beaumont et Patak (2012) Norm	1,016	–	1,020	–	1,000	–	0,987	–
Beaumont et Patak (2012) Exp	1,000	–	0,980	–	1,043	–	1,027	–
Beaumont et Patak (2012) $\epsilon = 1$	1,049	–	1,090	–	1,085	–	1,120	–
Antal et Tillé (2014)	1,016	1,246	1,070	1,630	1,000	1,085	1,013	1,280
Antal et Tillé (2011a)	1,000	1,148	1,010	1,390	0,979	1,043	0,960	1,173

4.5.6. Discussion: cas des paramètres non lisses de totaux

Il ressort des résultats indiqués dans le tableau 4.48 les conclusions suivantes:

- Toutes les méthodes bootstrap par pseudo-population à l'exception de la méthode de Mashreghi et coll. (2016) ont des comportements relativement voisins en terme de la précision et de la stabilité dans le cas de la médiane, avec un léger avantage en faveur de la deuxième version x -équilibrée de la méthode de Barbiero et Mecatti (2010). Le comportement de la méthode de Mashreghi et coll. (2016) s'explique par le fait que cette méthode approche dans le cas d'un estimateur biaisé, $E(\hat{\theta}^*) \neq \theta^*$, l'erreur quadratique moyenne et non pas la variance;
- Dans le cas de l'indice de Gini, on constate que la méthode de Beaumont et Patak (2012) mise en œuvre avec une distribution symétrique et la méthode d'Antal et Tillé (2014) ont produit des biais relatifs négligeables dans tous les scénarios, inférieurs à 3% en valeur absolue;
- Dans le cas de l'indice de Gini, on constate que la méthode Beaumont et Patak (2012) mise en œuvre avec une distribution exponentielle a produit des biais relativement importants en comparaison avec les autres méthodes et ceci dans tous les scénarios.

Pour les propriétés empiriques d'intervalles de confiance asymptotique de m_y et G_y , indiquées dans le tableau 4.49, on constate que toutes les méthodes bootstrap ont un comportement relativement voisin en terme des erreurs de couverture. En effet, le taux de couverture bilatérale et le taux des erreurs unilatérales sont bien respectés pour m_y dans tous les scénarios sauf dans le scénario 3. Toutefois, pour G_y ils ne sont pas respectés dans aucun des scénarios.

En ce qui concerne les propriétés empiriques d'intervalles percentiles de m_y et G_y indiqués dans le tableau 4.50, on remarque que le taux de couverture bilatérale et le taux des erreurs unilatérales sont bien respectés dans le cas de m_y mais pas dans le cas de G_y et ceci dans tous les scénarios. On remarque également que globalement les taux d'erreurs

bilatérales dans le cas de Gini sont plus grands pour les intervalles percentiles que pour les intervalles asymptotiques.

Les propriétés empiriques d'intervalles bootstrap de base de m_y et G_y sont présentées dans le tableau 4.51. On constate que le taux de couverture bilatérale et le taux des erreurs unilatérales ne sont pas respectés dans le cas des deux paramètres. Toutefois, on peut dire que les taux d'erreurs bilatérales des intervalles de confiance bootstrap de base se situent généralement entre ceux des intervalles percentiles et ceux des intervalles asymptotiques dans le cas de l'indice de Gini.

D'après les résultats indiqués dans le tableau 4.52 on constate que la technique t -bootstrap a amélioré substantiellement la couverture bilatérale et a diminué les erreurs de couverture unilatérale dans le cas de G_y comparativement avec la technique percentile, dans tous les scénarios et pour toutes les méthodes bootstrap. En fait, le gain apporté par le t -bootstrap dans le scénario 2 en terme de la couverture bilatérale par rapport à la technique percentile est d'environ 9 points en pourcentage. Toutefois, il faut signaler que ce gain en terme de couverture bilatérale s'accompagne d'une augmentation de la largeur d'environ 40% par rapport à celle obtenue via la technique percentile (voir le tableau 4.53). En revanche, dans le cas de m_y , la technique percentile a fait toujours mieux que le t -bootstrap.

4.6. Étude 3: Plan de Poisson conditionnel

Dans notre dernière étude par simulation nous comparons les méthodes bootstrap par pseudo-population et les deux méthodes proposées par Antal et Tillé (2011a) et Antal et Tillé (2014) pour un plan de Poisson conditionnel. Les probabilités d'inclusion sont toujours calculées proportionnellement à la variable auxiliaire x . Pour les intervalles de confiance percentile et bootstrap de base relatifs aux méthodes bootstrap Holmberg (1998), Chauvet (2007) et Mashreghi et coll. (2016), nous rapportons seulement les résultats de ceux de la méthode de Mashreghi et coll. (2016), pour les mêmes raisons que nous avons annoncées lors de l'introduction de la deuxième étude. Il faut noter que l'algorithme de

l'estimation de la variance par les méthodes bootstrap par pseudo-population est très pesant en terme de calcul car pour le même jeu de probabilités d'inclusion, le temps nécessaire pour sélectionner un seul ré-échantillon de taille $n = 150$ selon le plan de Poisson conditionnel à partir d'une pseudo-population de taille $N = 2175$ est plus grand d'environ 567 fois que celui nécessaire pour tirer un ré-échantillon selon le plan de Poisson. Pour cette raison, le t -bootstrap n'est pas mis en œuvre pour ces méthodes car ceci va générer un volume de calcul prohibitif supplémentaire. Ce problème n'est pas posé pour les deux méthodes d'Antal et Tillé (2011a) et d'Antal et Tillé (2014), notamment la dernière car elle ne fait pas appel au plan de Poisson conditionnel lors du ré-échantillonnage.

4.6.1. Résultats: cas des paramètres linéaires

Le tableau 4.54 présente le biais relatif (en %) ainsi que la racine de l'erreur quadratique moyenne des estimateurs bootstrap de la variance pour le plan de Poisson conditionnel, dans le cas des totaux t_y et t_z . Les tableaux 4.55 à 4.58 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap des totaux t_y et t_z . Le tableau 4.59 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.54. Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres linéaires t_y et t_z)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
t_y								
Holmberg (1998)	0,94	27,76	-3,86	40,18	20,01	28,20	27,67	34,74
Chauvet (2007)	0,24	28,83	-4,04	39,71	-2,33	19,02	1,19	28,92
Mashreghi et coll. (2016)	0,40	27,94	-4,50	40,16	-2,51	18,85	1,23	27,54
Chauvet (2007) simplifié	-0,09	27,85	-4,73	40,20	-3,23	19,30	-3,47	27,15
Barbiero et Mecatti (2010) <i>x-equ1</i>	0,34	28,02	-4,08	41,04	-2,88	19,34	4,31	27,68
Barbiero et Mecatti (2010) <i>x-equ2</i>	0,14	27,82	-4,30	40,85	-2,34	19,30	1,93	27,78
Antal et Tillé (2014)	0,44	27,70	-2,66	41,19	-1,84	19,12	2,83	27,80
Antal et Tillé (2011a)	0,80	27,83	-2,29	41,50	-1,24	19,42	2,81	28,01
t_z								
Holmberg (1998)	-1,71	14,43	-2,59	22,08	8,24	15,74	8,73	17,97
Chauvet (2007)	-1,99	13,99	-2,53	22,08	-1,21	13,31	-1,61	17,64
Mashreghi et coll. (2016)	-1,73	14,84	-2,90	22,04	-1,10	13,35	-1,37	17,86
Chauvet (2007) simplifié	-2,43	14,73	-2,79	22,74	-1,75	12,94	-4,02	19,03
Barbiero et Mecatti (2010) <i>x-equ1</i>	-1,86	15,01	-2,41	23,06	-1,53	13,22	0,49	17,82
Barbiero et Mecatti (2010) <i>x-equ2</i>	-1,49	14,49	-2,55	22,72	-0,33	13,72	-0,35	18,10
Antal et Tillé (2014)	-1,41	14,81	-1,14	23,03	-0,69	13,23	-0,39	17,69
Antal et Tillé (2011a)	-1,38	14,84	-0,60	23,51	-0,51	13,34	-0,13	17,43

Tab. 4.55. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres linéaires t_y et t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Holmberg (1998)	1,1	3,6	4,7	1,000	0,7	5,9	6,6	1,000	1,0	2,4	3,4	1,000	1,1	4,4	5,5	1,000
Chauvet (2007)	1,2	4,2	5,4	0,996	1,1	5,9	7,0	0,999	1,5	4,6	6,1	0,901	1,5	5,4	6,9	0,888
Mashreghi et coll. (2016)	1,1	4,0	5,1	0,997	0,7	6,0	6,7	0,996	1,5	4,6	6,1	0,900	1,6	5,7	7,3	0,888
Chauvet (2007) simplifié	1,3	4,3	5,6	0,995	0,9	6,1	7,0	0,995	1,5	4,8	6,3	0,897	1,9	5,8	7,7	0,867
Barbiero et Mecatti (2010) $x-equ1$	1,2	4,2	5,4	0,997	0,7	6,2	6,9	0,998	1,6	4,9	6,5	0,898	1,5	5,6	7,1	0,901
Barbiero et Mecatti (2010) $x-equ2$	1,2	3,9	5,1	0,996	0,7	6,2	6,9	0,997	1,5	4,6	6,1	0,901	1,6	5,8	7,4	0,891
Antal et Tillé (2014)	1,2	4,0	5,2	0,997	0,7	6,0	6,7	1,006	1,5	4,8	6,3	0,903	1,5	5,5	7,0	0,895
Antal et Tillé (2011a)	1,1	4,0	5,1	0,999	0,9	6,1	7,0	1,008	1,4	4,8	6,2	0,905	1,6	5,4	7,0	0,895
	t_z															
Holmberg (1998)	1,8	3,3	5,1	1,000	2,4	3,7	6,1	1,000	1,6	3,1	4,7	1,000	1,1	3,3	4,4	1,000
Chauvet (2007)	2,0	3,5	5,5	0,999	2,1	3,8	5,9	1,001	2,3	4,2	6,5	0,954	1,6	3,8	5,4	0,949
Mashreghi et coll. (2016)	2,0	3,4	5,4	1,000	2,4	3,7	6,1	0,999	2,3	4,0	6,3	0,955	1,7	4,0	5,7	0,949
Chauvet (2007) simplifié	2,0	3,3	5,3	0,996	2,2	3,7	5,9	0,999	2,2	4,2	6,4	0,951	1,7	3,6	5,3	0,938
Barbiero et Mecatti (2010) $x-equ1$	1,9	3,3	5,2	0,999	2,0	3,7	5,7	1,001	2,2	4,0	6,2	0,953	1,6	3,3	4,9	0,960
Barbiero et Mecatti (2010) $x-equ2$	1,9	3,3	5,2	1,001	2,2	3,7	5,9	1,000	2,2	4,0	6,2	0,959	1,5	3,6	5,1	0,956
Antal et Tillé (2014)	1,8	3,5	5,3	1,001	1,9	3,6	5,5	1,008	2,1	4,1	6,2	0,955	1,7	3,7	5,4	0,956
Antal et Tillé (2011a)	1,7	3,4	5,1	1,001	1,9	3,6	5,5	1,011	2,1	4,0	6,1	0,956	1,7	3,4	5,1	0,956

Tab. 4.56. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires t_y et t_z)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Mashreghi et coll. (2016)	1,7	3,8	5,5	1,000	1,6	5,6	7,2	1,000	1,9	4,4	6,3	1,000	2,0	5,2	7,2	1,000
Chauvet (2007) simplifié	1,8	4,2	6,0	0,999	1,8	5,5	7,3	0,999	1,9	4,4	6,3	0,999	2,5	5,8	8,3	0,977
Barbiero et Mecatti (2010) $x-equ1$	1,7	4,0	5,7	1,000	1,1	6,1	7,2	1,000	1,8	4,8	6,6	1,000	1,7	5,4	7,1	1,015
Barbiero et Mecatti (2010) $x-equ2$	1,9	3,7	5,6	0,999	1,5	5,3	6,8	1,000	1,8	4,4	6,2	1,002	1,8	5,5	7,3	1,003
Antal et Tillé (2014)	1,8	4,0	5,8	0,994	0,9	6,2	7,1	0,994	1,7	4,8	6,5	1,003	1,7	5,7	7,4	1,002
Antal et Tillé (2011a)	1,8	3,9	5,7	1,003	1,4	5,2	6,6	1,012	1,7	3,8	5,5	1,014	2,0	5,2	7,2	1,015
	t_z															
Mashreghi et coll. (2016)	2,0	3,4	5,4	1,000	2,6	3,6	6,2	1,000	2,3	4,0	6,3	1,000	1,7	3,8	5,5	1,000
Chauvet (2007) simplifié	2,0	3,2	5,2	0,999	2,3	3,5	5,8	0,999	2,4	3,6	6,0	1,000	1,8	3,5	5,3	0,984
Barbiero et Mecatti (2010) $x-equ1$	2,1	2,9	5,0	1,000	2,3	3,2	5,5	1,003	2,5	3,9	6,4	1,000	1,6	3,5	5,1	1,008
Barbiero et Mecatti (2010) $x-equ2$	1,9	2,9	4,8	1,005	2,3	3,8	6,1	0,999	2,3	3,7	6,0	1,009	1,7	3,4	5,1	1,004
Antal et Tillé (2014)	1,9	3,4	5,3	1,002	2,2	3,8	6,0	1,004	2,2	4,2	6,4	1,004	1,7	3,8	5,5	1,000
Antal et Tillé (2011a)	2,0	2,9	4,9	1,003	2,2	3,4	5,6	1,011	2,2	3,2	5,4	1,015	1,7	3,3	5,0	1,008

Tab. 4.57. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres linéaires t_y et t_z)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Mashreghi et coll. (2016)	1,0	4,6	5,6	1,000	0,7	6,9	7,6	1,000	1,3	5,4	6,7	1,000	1,4	5,9	7,3	1,000
Chauvet (2007) simplifié	1,1	4,5	5,6	0,999	0,4	7,1	7,5	0,999	1,6	4,9	6,5	0,999	1,9	5,9	7,8	0,977
Barbiero et Mecatti (2010) $x-equ1$	0,9	4,5	5,4	1,000	0,5	7,1	7,6	1,000	1,6	5,2	6,8	1,000	1,4	5,7	7,1	1,015
Barbiero et Mecatti (2010) $x-equ2$	0,9	4,4	5,3	0,999	0,5	7,4	7,9	1,000	1,3	5,4	6,7	1,002	1,4	6,0	7,4	1,003
Antal et Tillé (2014)	1,2	4,2	5,4	0,994	1,1	6,2	7,3	0,994	1,4	4,7	6,1	1,003	1,7	5,6	7,3	1,002
Antal et Tillé (2011a)	0,9	4,6	5,5	1,003	0,7	7,0	7,7	1,012	1,0	5,4	6,4	1,014	1,1	5,9	7,0	1,015
	t_z															
Mashreghi et coll. (2016)	1,9	3,6	5,5	1,000	2,2	4,3	6,5	1,000	2,3	3,8	6,1	1,000	1,5	4,2	5,7	1,000
Chauvet (2007) simplifié	1,9	3,5	5,4	0,999	2,4	4,4	6,8	0,999	2,0	4,2	6,2	0,998	1,8	3,8	5,6	0,988
Barbiero et Mecatti (2010) $x-equ1$	1,5	3,7	5,2	1,001	2,0	4,3	6,3	1,003	2,2	4,5	6,7	0,998	1,7	3,9	5,6	1,012
Barbiero et Mecatti (2010) $x-equ2$	2,0	3,7	5,7	1,005	2,1	4,5	6,6	0,999	2,2	4,3	6,5	1,007	1,5	4,1	5,6	1,008
Antal et Tillé (2014)	1,9	3,6	5,5	1,003	2,2	3,7	5,9	1,004	2,5	3,9	6,4	1,002	1,6	3,9	5,5	1,008
Antal et Tillé (2011a)	1,8	3,7	5,5	1,004	1,8	4,2	6,0	1,011	2,5	4,0	6,5	1,015	1,3	3,8	5,1	1,012

Tab. 4.58. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres linéaires t_y et t_z)

Méthode	Scénario 1 $f = 0,07$ $n = 150$				Scénario 2 $f = 0,07$ $n = 50$				Scénario 3 $f = 0,3$ $n = 150$				Scénario 4 $f = 0,3$ $n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	t_y															
Antal et Tillé (2014)	2,5	2,5	5,0	1,000	2,1	2,6	4,7	1,000	2,4	2,3	4,7	1,000	3,0	3,5	6,5	1,000
Antal et Tillé (2011a)	2,2	3,0	5,2	0,974	2,1	3,0	5,1	0,948	2,4	2,5	4,9	0,987	2,2	4,1	6,3	0,967
	t_z															
Antal et Tillé (2014)	2,0	1,7	3,7	1,000	1,9	1,8	3,7	1,000	2,8	2,6	5,4	1,000	1,8	2,2	4,0	1,000
Antal et Tillé (2011a)	2,0	1,8	3,8	0,990	2,0	2,1	4,1	0,982	2,8	2,7	5,5	0,987	1,4	2,1	3,5	0,986

Tab. 4.59. Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (Totaux t_y et t_z)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
Holmberg (1998)	1,009	–	1,010	–	1,119	–	1,135	–
Chauvet (2007)	1,005	–	1,009	–	1,008	–	1,007	–
Mashreghi et coll. (2016)	1,006	–	1,007	–	1,007	–	1,008	–
Chauvet (2007) simplifié	1,003	–	1,006	–	1,003	–	0,984	–
Barbiero et Mecatti (2010) $x-equ1$	1,005	–	1,009	–	1,005	–	1,023	–
Barbiero et Mecatti (2010) $x-equ2$	1,004	–	1,008	–	1,008	–	1,011	–
Antal et Tillé (2014)	1,006	1,095	1,016	1,202	1,011	1,068	1,015	1,125
Antal et Tillé (2011a)	1,008	1,067	1,018	1,140	1,010	1,050	1,015	1,088
t_z								
Holmberg (1998)	1,007	–	1,006	–	1,057	–	1,058	–
Chauvet (2007)	1,006	–	1,006	–	1,008	–	1,004	–
Mashreghi et coll. (2016)	1,007	–	1,005	–	1,009	–	1,004	–
Chauvet (2007) simplifié	1,003	–	1,005	–	1,004	–	0,992	–
Barbiero et Mecatti (2010) $x-equ1$	1,006	–	1,007	–	1,007	–	1,016	–
Barbiero et Mecatti (2010) $x-equ2$	1,008	–	1,006	–	1,013	–	1,012	–
Antal et Tillé (2014)	1,009	1,055	1,014	1,106	1,009	1,050	1,012	1,074
Antal et Tillé (2011a)	1,009	1,045	1,017	1,087	1,011	1,041	1,012	1,058

4.6.2. Discussion: cas des paramètres linéaires

À la lumière des résultats indiqués dans le tableau 4.54, on constate que toutes les méthodes bootstrap ont un comportement relativement voisin en terme du biais relatif à l'exception de la méthode de Holmberg (1998) qui a produit des biais relatifs importants dans les deux derniers scénarios; ils sont supérieurs à 20% dans le cas du total, t_y , et supérieurs à 8% dans le cas du total, t_z . Les taux d'erreur de couverture bilatérale et unilatérale des intervalles de confiance asymptotique produits par les différentes méthodes bootstrap sont similaires, à une exception près, à ceux produits par la méthode de Holmberg (1998) dans les deux derniers scénarios, où la couverture bilatérale et unilatérale droite étaient meilleures que celles des autres méthodes bootstrap (voir tableau 4.55). Toutefois, cette performance est due principalement à la sur-estimation de la variance par cette méthode en particulier. En effet, à l'exception de la méthode Holmberg (1998) dans les deux derniers scénarios, les taux d'erreur de couverture unilatérale gauche, L , pour t_y (resp t_z) sont autour de 1% dans les deux premiers scénarios et 1,5% dans les deux derniers scénarios (resp autour de 2% dans tous les scénarios) et les taux d'erreur de couverture droite, U , pour t_y (resp pour t_z) sont autour de 5% (resp autour de 4%) et les $L + U$ pour t_y (resp pour t_z) sont autour de 6% dans les scénarios 1 et 3 et 7% dans les scénarios 2 et 4 (resp autour de 6% dans tous les scénarios).

La technique des percentiles semble faire mieux que la technique asymptotique au moins en terme de la couverture unilatérale gauche et la largeur des intervalles de confiance. En effet à l'exception de Barbiero et Mecatti (2010) *x-equ1*, Antal et Tillé (2014) et Antal et Tillé (2011a) dans le scénario 2, les taux d'erreur de couverture unilatérale gauche des intervalles de confiance percentile sont autour de 2% (voir tableau 4.56). En outre, les intervalles de confiance asymptotique sont relativement plus larges que les intervalles percentiles d'environ 1% pour toutes les méthodes à l'exception de la méthode Holmberg (1998) dans les deux derniers scénarios, où la différence entre les deux techniques en terme de la largeur d'environ 12% est notable (voir tableau 4.59). La technique bootstrap de base donne des résultats similaires à la technique des percentiles dans le cas de t_z mais pas dans le cas

de t_y où les résultats sont proches de ceux obtenus par la technique asymptotique (voir tableau 4.57). Rappelons-nous que le t -bootstrap a été mis en œuvre seulement pour les deux méthodes Antal et Tillé (2011a) et Antal et Tillé (2014) pour les raisons que nous avons avancées lors de l'introduction de l'étude 3. Le t -bootstrap fait mieux que les autres techniques au moins dans le cas de t_y en terme de la couverture unilatérale et bilatérale (voir le tableau 4.58). Toutefois il faut noter que les intervalles t -bootstrap produits par les deux méthodes sont plus larges d'environ 5% à 20%. On note également que les intervalles de confiance t -bootstrap produits par Antal et Tillé (2014) sont plus larges que ceux produits par Antal et Tillé (2011a). Il faut noter aussi que la méthode de Antal et Tillé (2014) est extrêmement plus rapide que les autres méthodes bootstrap, notamment les méthodes bootstrap par pseudo-population. Cette caractéristique a rendu la mise en œuvre du t -bootstrap par cette méthode moins pesante en terme de calcul.

4.6.3. Résultats: cas des paramètres lisses de totaux

Le tableau 4.60 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne des estimateurs bootstrap de la variance étudiés pour le plan de Poisson conditionnel, dans le cas du ratio, $r_{y,z}$, et du coefficient de corrélation, $\rho_{y,z}$. Les tableaux 4.61 à 4.64 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap du ratio, $r_{y,z}$, et du coefficient de corrélation, $\rho_{y,z}$. Le tableau 4.65 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.60. Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
$r_{y,z}$								
Holmberg (1998)	2,47	19,98	5,65	41,95	3,07	22,45	4,02	32,72
Chauvet (2007)	2,72	21,29	5,36	39,61	0,67	18,64	1,14	32,73
Mashreghi et coll. (2016)	2,91	20,26	6,26	42,53	0,99	21,06	1,94	32,20
Chauvet (2007) simplifié	2,40	20,04	5,50	42,00	3,05	22,38	4,60	33,61
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,73	20,16	5,58	42,07	2,48	22,12	4,57	34,34
Barbiero et Mecatti (2010) <i>x-equ2</i>	3,61	20,60	6,30	42,50	4,95	23,67	6,32	34,74
Antal et Tillé (2014)	3,63	20,07	8,10	44,22	1,42	21,68	2,42	31,97
Antal et Tillé (2011a)	3,04	19,81	7,47	42,92	0,62	21,02	1,84	32,42
$\rho_{y,z}$								
Holmberg (1998)	-0,01	33,06	12,77	70,64	4,40	29,58	12,66	63,89
Chauvet (2007)	-0,02	34,09	12,63	70,30	0,54	25,13	8,79	61,10
Mashreghi et coll. (2016)	0,83	33,28	14,29	71,88	0,35	27,52	9,72	61,79
Chauvet (2007) simplifié	-0,29	33,40	11,81	69,23	5,82	31,06	11,11	62,12
Barbiero et Mecatti (2010) <i>x-equ1</i>	-0,46	33,23	11,42	68,78	5,05	30,28	7,11	60,80
Barbiero et Mecatti (2010) <i>x-equ2</i>	1,06	33,59	14,50	71,98	11,44	35,13	21,91	72,02
Antal et Tillé (2014)	4,69	35,90	29,86	91,36	1,65	28,04	13,53	65,36
Antal et Tillé (2011a)	-0,50	32,87	12,03	68,73	0,55	27,03	-0,48	54,34

Tab. 4.61. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Holmberg (1998)	1,6	2,7	4,3	1,000	0,8	4,3	5,1	1,000	1,9	2,5	4,4	1,000	0,6	3,0	3,6	1,000
Chauvet (2007)	1,6	2,7	4,3	1,002	1,0	4,3	5,3	0,999	1,8	2,9	4,7	0,988	0,7	3,1	3,8	0,984
Mashreghi et coll. (2016)	1,6	2,7	4,3	1,000	0,8	4,3	5,1	1,004	1,9	3,1	5,0	0,993	0,9	3,1	4,0	0,984
Chauvet (2007) simplifié	1,5	2,8	4,3	1,000	0,8	4,1	4,9	1,000	1,7	2,8	4,5	1,000	0,5	3,2	3,7	1,000
Barbiero et Mecatti (2010) $x-equ1$	1,4	2,6	4,0	1,000	0,9	4,4	5,3	1,000	1,7	2,7	4,4	0,993	1,0	3,4	4,4	1,000
Barbiero et Mecatti (2010) $x-equ2$	1,4	2,6	4,0	1,006	0,8	4,3	5,1	1,004	1,8	2,7	4,5	1,007	0,5	3,1	3,6	1,008
Antal et Tillé (2014)	1,7	2,6	4,3	1,006	0,9	4,3	5,2	1,011	1,8	2,7	4,5	0,993	0,9	3,2	4,1	0,992
Antal et Tillé (2011a)	1,3	2,7	4,0	1,000	0,9	4,2	5,1	1,011	1,9	2,6	4,5	0,993	0,7	3,2	3,9	0,984
	$\rho_{y,z}$															
Holmberg (1998)	5,0	0,7	5,7	1,000	7,4	1,0	8,4	1,000	2,8	2,0	4,8	1,000	5,3	1,0	6,3	1,000
Chauvet (2007)	4,7	0,6	5,3	1,001	7,3	0,9	8,2	1,000	2,9	2,1	5,0	0,980	5,7	1,2	6,9	0,984
Mashreghi et coll. (2016)	4,7	0,5	5,2	1,000	7,2	0,7	7,9	1,011	2,6	2,3	4,9	0,978	5,5	1,1	6,6	0,986
Chauvet (2007) simplifié	5,0	0,5	5,5	1,000	7,5	0,9	8,4	1,000	2,6	2,0	4,6	1,011	5,1	1,0	6,1	1,000
Barbiero et Mecatti (2010) $x-equ1$	5,0	0,6	5,6	1,000	8,2	0,9	9,1	0,994	2,5	2,0	4,5	1,000	5,9	1,2	7,1	0,973
Barbiero et Mecatti (2010) $x-equ2$	4,9	0,5	5,4	1,000	7,0	0,8	7,8	1,011	2,4	2,0	4,4	1,034	5,1	0,8	5,9	1,041
Antal et Tillé (2014)	4,5	0,6	5,1	1,019	6,1	0,6	6,7	1,068	2,5	2,1	4,6	0,989	5,4	1,2	6,6	1,000
Antal et Tillé (2011a)	5,1	0,6	5,7	1,000	7,6	1,0	8,6	1,000	2,6	2,2	4,8	1,000	6,4	1,2	7,6	0,946

Tab. 4.62. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Mashreghi et coll. (2016)	2,7	2,1	4,8	1,000	3,1	3,0	6,1	1,000	2,9	2,2	5,1	1,000	3,2	2,2	5,4	1,000
Chauvet (2007) simplifié	2,7	2,0	4,7	0,994	2,6	2,9	5,5	1,000	3,0	2,2	5,2	1,015	3,0	2,4	5,4	1,020
Barbiero et Mecatti (2010) <i>x-equ1</i>	2,3	2,1	4,4	1,006	2,5	2,7	5,2	1,000	2,7	2,1	4,8	1,015	3,0	2,3	5,3	1,020
Barbiero et Mecatti (2010) <i>x-equ2</i>	2,4	2,0	4,4	1,006	3,0	2,9	5,9	1,004	2,6	1,9	4,5	1,030	2,7	2,3	5,0	1,028
Antal et Tillé (2014)	3,0	2,2	5,2	1,012	2,7	2,6	5,3	1,007	3,1	2,1	5,2	1,007	3,0	2,2	5,2	1,004
Antal et Tillé (2011a)	2,4	2,3	4,7	1,006	3,1	2,3	5,4	1,007	2,6	2,4	5,0	1,004	3,0	2,4	5,4	1,008
	$\rho_{y,z}$															
Mashreghi et coll. (2016)	3,5	1,4	4,9	1,000	5,5	1,6	7,1	1,000	2,0	3,1	5,1	1,000	4,0	1,6	5,6	1,000
Chauvet (2007) simplifié	3,5	0,9	4,4	1,000	5,7	1,3	7,0	1,000	1,9	2,7	4,6	1,023	4,0	1,7	5,7	1,014
Barbiero et Mecatti (2010) <i>x-equ1</i>	3,3	0,9	4,2	1,010	6,0	1,2	7,2	1,000	2,1	2,7	4,8	1,011	4,0	1,7	5,7	0,986
Barbiero et Mecatti (2010) <i>x-equ2</i>	3,2	1,0	4,2	1,010	5,4	1,5	6,9	1,011	1,7	2,6	4,3	1,057	3,7	1,3	5,0	1,056
Antal et Tillé (2014)	2,8	1,2	4,0	1,029	4,7	0,8	5,5	1,069	2,2	2,9	5,1	1,011	4,2	1,5	5,7	1,021
Antal et Tillé (2011a)	3,5	1,2	4,7	1,010	5,6	1,2	6,8	1,006	2,5	2,8	5,3	1,009	4,6	1,7	6,3	0,965

Tab. 4.63. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres lisses des totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Holmberg (1998)	0,9	3,9	4,8	1,000	0,1	7,1	7,2	1,000	1,6	3,7	5,3	1,000	0,2	4,5	4,7	1,000
Chauvet (2007)	0,6	3,7	4,3	1,000	0,1	7,1	7,2	1,000	1,5	3,6	5,1	1,007	0,2	4,5	4,7	1,000
Chauvet (2007) simplifié	0,9	3,3	4,2	1,000	0,0	7,2	7,2	1,000	1,5	3,6	5,1	1,015	0,2	4,7	4,9	1,020
Barbiero et Mecatti (2010) $x-equ1$	1,0	3,4	4,4	1,000	0,1	6,9	7,0	1,004	1,4	3,4	4,8	1,015	0,1	4,5	4,6	1,020
Barbiero et Mecatti (2010) $x-equ2$	1,1	3,9	5,0	1,006	0,0	7,0	7,0	1,007	1,4	3,3	4,7	1,030	0,0	4,5	4,5	1,028
Antal et Tillé (2014)	0,7	3,8	4,5	1,006	0,0	6,9	6,9	1,011	1,5	3,9	5,4	1,015	0,1	4,8	4,9	1,004
Antal et Tillé (2011a)	0,9	3,5	4,4	1,006	0,0	7,1	7,1	1,011	1,6	3,6	5,2	1,004	0,1	4,9	5,0	1,008
	$\rho_{y,z}$															
Holmberg (1998)	6,5	0,9	7,4	1,000	11,7	1,0	12,7	1,000	4,5	1,6	6,1	1,000	8,8	0,8	9,6	1,000
Chauvet (2007)	6,4	0,6	7,0	1,000	11,7	1,0	12,7	1,000	4,7	1,7	6,4	1,000	8,8	0,8	9,6	1,000
Chauvet (2007) simplifié	6,7	0,6	7,3	1,000	11,6	1,3	12,9	1,000	3,6	1,7	5,3	1,023	8,6	0,8	9,4	1,007
Barbiero et Mecatti (2010) $x-equ1$	6,5	0,5	7,0	1,000	11,5	0,9	12,4	0,994	4,1	1,4	5,5	1,023	9,2	0,7	9,9	0,986
Barbiero et Mecatti (2010) $x-equ2$	7,0	0,7	7,7	1,010	11,3	1,0	12,3	1,006	3,5	1,6	5,1	1,046	8,3	0,5	8,8	1,048
Antal et Tillé (2014)	6,8	0,9	7,7	1,029	10,6	1,2	11,8	1,063	4,2	1,7	5,9	1,000	8,1	0,7	8,8	1,014
Antal et Tillé (2011a)	7,1	0,8	7,9	1,010	11,1	1,0	12,1	1,000	4,1	1,6	5,7	1,009	8,6	1,1	9,7	0,959

Tab. 4.64. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres lisses de totaux, $r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 $f = 0,07$ $n = 150$				Scénario 2 $f = 0,07$ $n = 50$				Scénario 3 $f = 0,3$ $n = 150$				Scénario 4 $f = 0,3$ $n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	$r_{y,z}$															
Antal et Tillé (2014)	2,0	1,6	3,6	1,000	1,4	1,9	3,3	1,000	2,0	2,0	4,0	1,000	2,2	2,4	4,6	1,000
Antal et Tillé (2011a)	2,1	2,0	4,1	0,994	2,4	2,0	4,4	0,983	2,2	2,2	4,4	0,985	2,0	2,1	4,1	0,989
	$\rho_{y,z}$															
Antal et Tillé (2014)	2,5	2,3	4,8	1,000	1,7	3,0	4,7	1,000	2,3	2,4	4,7	1,000	2,0	1,7	3,7	1,000
Antal et Tillé (2011a)	1,9	3,5	5,4	0,935	1,1	4,6	5,7	0,863	2,5	2,4	4,9	0,957	1,0	2,8	3,8	0,937

Tab. 4.65. Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) ($r_{y,z}$ et $\rho_{y,z}$)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
$r_{y,z}$								
Holmberg (1998)	1,006	–	1,004	–	1,022	–	1,024	–
Chauvet (2007)	1,008	–	1,003	–	1,010	–	1,007	–
Mashreghi et coll. (2016)	1,006	–	1,007	–	1,015	–	1,008	–
Chauvet (2007) simplifié	1,006	–	1,004	–	1,022	–	1,024	–
Barbiero et Mecatti (2010) $x-equ1$	1,006	–	1,004	–	1,015	–	1,024	–
Barbiero et Mecatti (2010) $x-equ2$	1,012	–	1,007	–	1,030	–	1,032	–
Antal et Tillé (2014)	1,012	1,042	1,014	1,067	1,015	1,052	1,016	1,056
Antal et Tillé (2011a)	1,006	1,036	1,014	1,049	1,015	1,037	1,008	1,044
$\rho_{y,z}$								
Holmberg (1998)	1,010	–	1,011	–	1,023	–	1,021	–
Chauvet (2007)	1,010	–	1,012	–	1,002	–	1,005	–
Mashreghi et coll. (2016)	1,010	–	1,023	–	1,000	–	1,007	–
Chauvet (2007) simplifié	1,010	–	1,011	–	1,034	–	1,021	–
Barbiero et Mecatti (2010) $x-equ1$	1,010	–	1,006	–	1,023	–	0,993	–
Barbiero et Mecatti (2010) $x-equ2$	1,010	–	1,023	–	1,057	–	1,062	–
Antal et Tillé (2014)	1,029	1,125	1,080	1,247	1,011	1,057	1,021	1,174
Antal et Tillé (2011a)	1,010	1,077	1,011	1,121	0,989	1,034	0,965	1,083

4.6.4. Discussion: cas des paramètres lisses de totaux

Compte tenu des résultats des différents scénarios pour les deux paramètres étudiés, à savoir le ratio, $r_{y,z}$ et le coefficient de corrélation, $\rho_{y,z}$, il nous semble que la méthode de Chauvet (2007) suivie de la méthode de Mashreghi et coll. (2016) font relativement mieux que les autres méthodes bootstrap par pseudo-population dans les deux derniers scénarios (voir tableau 4.60). Il apparait également que la méthode de Antal et Tillé (2011a) fait toujours mieux que la méthode de Antal et Tillé (2014) et elle fait mieux que les autres méthodes dans le scénario 4 pour $\rho_{y,z}$ (voir tableau 4.60). Les résultats indiqués dans les tableaux 4.61 à 4.63 montrent un comportement relativement voisin des différentes méthodes bootstrap en terme du taux de couverture unilatérale et bilatérale, notamment dans le cas du ratio. En effet, pour la technique asymptotique, les taux d'erreur de couverture unilatérale gauche, L , pour $r_{y,z}$ sont entre 1,3% et 1,9% dans les scénarios 1 et 3 et entre 0,5% et 1% dans les scénarios 2 et 3, les taux d'erreur de couverture droite, U , sont autour de 3% dans les scénarios 1 et 2 et autour 4% dans les scénarios 2 et 4 et les $L+U$ sont entre 3,6% et 5,3%. Pour les intervalles de confiance asymptotique du coefficient de corrélation, on constate que le taux de couverture bilatérale est bien respecté par toutes les méthodes bootstrap dans les scénarios 1 et 3. Dans le scénario 4, il est respecté que par certaines méthodes bootstrap, pour les autres méthodes, les taux de la couverture bilatérale observés sont autour de 7%. Pour la couverture unilatérale, elle n'est respectée que dans le scénario 3 et ceci pour toutes les méthodes (voir tableau 4.61).

La technique des percentiles pour les deux paramètres fait mieux que la technique asymptotique en terme de la couverture bilatérale et unilatérale et la largeur (voir les tableaux 4.62 et 4.65). La technique bootstrap de base fait moins bien que la technique des percentiles et asymptotique au moins en terme de la couverture unilatérale, notamment dans le cas du coefficient de corrélation.

Le t -bootstrap ne semble pas apporter un gain particulier en terme de la couverture unilatérale dans le cas du ratio comparativement à la technique des percentiles malgré que les intervalles t -bootstrap soient plus larges d'environ 5%. Dans le cas du coefficient de

corrélation, le t -bootstrap mis en œuvre par la méthode de Antal et Tillé (2014) respecte mieux la couverture unilatérale et bilatérale que la technique asymptotique, des percentiles et bootstrap de base. Toutefois il faut signaler que les intervalles de confiance t -bootstrap sont plus larges, notamment dans le deuxième scénario où la largeur relative est environ 25% plus grande (voir le tableau 4.65).

4.6.5. Résultats: cas des paramètres non lisses de totaux

Le tableau 4.66 présente le biais relatif (en %) ainsi que la racine carrée de l'erreur quadratique moyenne des estimateurs bootstrap de la variance étudiés pour le plan de Poisson conditionnel, dans le cas de la médiane et de l'indice de Gini de la variable y . Les tableaux 4.67 à 4.70 donnent, pour les différents scénarios, les performances des différentes méthodes dans la production d'intervalles de confiance asymptotique, percentile, bootstrap de base et t -bootstrap de la médiane et de l'indice de Gini de la variable y . Le tableau 4.71 rapporte pour chaque méthode bootstrap la longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile produit par la méthode de Mashreghi et coll. (2016).

Tab. 4.66. Étude 3: Biais relatif et stabilité des méthodes bootstrap (paramètres non lisses, m_y et G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>	<i>BR</i>	<i>REQMR</i>
m_y								
Holmberg (1998)	10,68	52,31	15,36	65,53	8,47	43,18	21,54	66,90
Chauvet (2007)	10,53	52,18	15,61	67,46	6,32	40,56	18,86	67,28
Mashreghi et coll. (2016)	15,35	55,96	23,89	74,87	10,20	44,41	26,91	73,71
Chauvet (2007) simplifié	10,64	51,75	15,56	65,60	5,94	41,56	21,05	65,70
Barbiero et Mecatti (2010) <i>x-equ1</i>	10,80	52,28	15,56	65,76	7,11	41,93	23,02	67,60
Barbiero et Mecatti (2010) <i>x-equ2</i>	10,03	51,94	13,99	64,77	5,62	41,72	15,84	63,24
Antal et Tillé (2014)	10,68	52,63	15,20	65,77	6,39	42,01	19,52	65,85
Antal et Tillé (2011a)	11,40	52,43	17,58	67,49	6,59	42,55	22,07	66,41
G_y								
Holmberg (1998)	-3,98	39,39	-12,19	44,32	3,26	18,82	1,05	28,07
Chauvet (2007)	-3,93	37,62	-12,43	44,50	-3,52	18,86	-3,22	28,37
Mashreghi et coll. (2016)	-3,19	39,32	-9,55	44,45	-3,24	18,55	-0,95	28,16
Chauvet (2007) simplifié	-3,90	39,35	-12,46	43,95	-3,50	19,84	-4,30	28,41
Barbiero et Mecatti (2010) <i>x-equ1</i>	-4,01	39,13	-12,38	43,83	-3,51	19,65	-2,13	27,63
Barbiero et Mecatti (2010) <i>x-equ2</i>	-4,11	39,45	-12,97	44,01	-2,25	20,09	-1,14	28,58
Antal et Tillé (2014)	-0,83	41,04	-4,56	47,66	-2,51	18,65	0,30	28,84
Antal et Tillé (2011a)	-4,08	39,20	-12,29	44,04	-3,11	18,23	-7,06	28,26

Tab. 4.67. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance asymptotique par rapport à la méthode Holmberg (1998) (paramètres non lisses, m_y et G_y)

Méthode	Scénario 1				Scénario 2				Scénario 3				Scénario 4			
	$f = 0,07 \ n = 150$				$f = 0,07 \ n = 50$				$f = 0,3 \ n = 150$				$f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Holmberg (1998)	2,2	3,9	6,1	1,000	2,8	3,7	6,5	1,000	4,3	3,3	7,6	1,000	2,5	3,6	6,1	1,000
Chauvet (2007)	2,1	4,0	6,1	0,999	3,0	3,9	6,9	1,001	4,5	3,2	7,7	0,99	2,5	3,5	6,0	0,989
Mashreghi et coll. (2016)	2,1	4,1	6,2	1,021	2,6	3,0	5,6	1,034	4,0	2,8	6,8	1,008	2,0	3,3	5,3	1,020
Chauvet (2007) simplifié	2,3	3,7	6,0	1,000	3,0	3,7	6,7	1,001	4,8	2,6	7,4	0,988	2,8	3,8	6,6	0,999
Barbiero et Mecatti (2010) $x-equ1$	2,4	3,9	6,3	1,000	2,9	3,6	6,5	1,001	4,6	3,0	7,6	0,993	2,4	3,7	6,1	1,006
Barbiero et Mecatti (2010) $x-equ2$	2,4	4,1	6,5	0,997	2,7	4,1	6,8	0,994	4,1	3,5	7,6	0,987	2,6	4,2	6,8	0,976
Antal et Tillé (2014)	2,4	3,7	6,1	1,000	3,0	3,8	6,8	0,999	4,4	3,4	7,8	0,990	2,6	3,5	6,1	0,991
Antal et Tillé (2011a)	2,3	3,9	6,2	1,003	3,0	3,5	6,5	1,010	4,0	3,4	7,4	0,998	2,3	3,5	5,8	1,003
	G_y															
Holmberg (1998)	1,1	6,8	7,9	1,000	1,1	10,1	11,2	1,000	1,5	5,5	7,0	1,000	1,2	7,2	8,4	1,000
Chauvet (2007)	1,1	6,9	8,0	1,005	1,0	10,1	11,1	1,007	1,8	6,1	7,9	0,949	1,2	7,2	8,4	0,975
Mashreghi et coll. (2016)	1,1	6,9	8,0	1,016	1,0	9,8	10,8	1,020	1,8	6,2	8,0	0,959	1,2	7,2	8,4	0,987
Chauvet (2007) simplifié	1,1	6,7	7,8	1,016	0,8	10,0	10,8	1,000	1,9	6,4	8,3	0,959	1,2	7,8	9,0	0,974
Barbiero et Mecatti (2010) $x-equ1$	1,1	6,8	7,9	1,000	1,0	10,0	11,0	1,000	1,8	6,4	8,2	0,959	1,1	7,3	8,4	0,974
Barbiero et Mecatti (2010) $x-equ2$	1,1	6,9	8,0	1,000	1,0	9,9	10,9	1,000	1,7	6,1	7,8	0,959	1,2	7,4	8,6	0,987
Antal et Tillé (2014)	1,1	6,7	7,8	1,016	0,7	9,7	10,4	1,050	1,9	6,2	8,1	0,959	1,2	7,2	8,4	1,000
Antal et Tillé (2011a)	1,2	7,2	8,4	1,000	0,9	10,1	11,0	1,000	2,0	6,1	8,1	0,959	1,3	7,5	8,8	0,948

Tab. 4.68. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance percentile par rapport à la méthode Mashreghi et coll. (2016) (paramètres non lisses, m_y et G_y)

Méthode	Scénario 1 $f = 0,07 \ n = 150$				Scénario 2 $f = 0,07 \ n = 50$				Scénario 3 $f = 0,3 \ n = 150$				Scénario 4 $f = 0,3 \ n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Mashreghi et coll. (2016)	2,0	3,2	5,2	1,000	2,2	2,3	4,5	1,000	2,7	2,6	5,3	1,000	2,3	2,3	4,6	1,000
Chauvet (2007) simplifié	2,0	3,3	5,3	0,995	2,7	2,2	4,9	0,992	3,5	2,5	6,0	0,983	2,0	2,1	4,1	0,983
Barbiero et Mecatti (2010) $x-equ1$	2,1	3,2	5,3	1,000	2,7	2,1	4,8	0,991	3,5	2,9	6,4	0,986	2,3	2,4	4,7	0,987
Barbiero et Mecatti (2010) $x-equ2$	1,9	3,1	5,0	0,997	2,7	2,0	4,7	0,989	3,5	2,9	6,4	0,985	3,0	3,0	6,0	0,956
Antal et Tillé (2014)	1,9	3,3	5,2	0,995	2,6	2,1	4,7	0,984	3,0	2,7	5,7	0,985	2,0	1,9	3,9	0,973
Antal et Tillé (2011a)	1,8	3,3	5,1	1,002	2,1	2,2	4,3	1,004	3,0	2,6	5,6	0,987	1,9	1,9	3,8	0,991
	G_y															
Mashreghi et coll. (2016)	0,9	8,2	9,1	1,000	0,1	11,9	12,0	1,000	1,2	7,7	8,9	1,000	0,4	9,9	10,3	1,000
Chauvet (2007) simplifié	0,9	8,8	9,7	1,016	0,1	11,9	12,0	1,000	1,1	7,7	8,8	1,000	0,2	9,7	9,9	1,000
Barbiero et Mecatti (2010) $x-equ1$	0,8	7,9	8,7	1,016	0,1	11,5	11,6	1,000	1,1	8,0	9,1	1,000	0,2	9,6	9,8	1,000
Barbiero et Mecatti (2010) $x-equ2$	1,1	8,4	9,5	1,016	0,1	12,0	12,1	1,000	1,1	8,1	9,2	1,000	0,2	9,9	10,1	1,014
Antal et Tillé (2014)	0,6	10,0	10,6	1,016	0,1	14,3	14,4	1,010	1,1	7,6	8,7	1,022	0,0	10,7	10,7	1,000
Antal et Tillé (2011a)	1,1	7,8	8,9	1,016	0,2	11,5	11,7	1,010	1,2	7,2	8,4	1,012	0,5	8,0	8,5	1,000

Tab. 4.69. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance bootstrap de base par rapport à la méthode Mashreghi et coll. (2016) (paramètres non lisses, m_y et G_y)

Méthode	Scénario 1 $f = 0,07$ $n = 150$				Scénario 2 $f = 0,07$ $n = 50$				Scénario 3 $f = 0,3$ $n = 150$				Scénario 4 $f = 0,3$ $n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Mashreghi et coll. (2016)	3,8	7,5	11,3	1,000	7,4	10,7	18,1	1,000	7,5	5,5	13,0	1,000	5,4	9,1	14,5	1,000
Chauvet (2007) simplifié	4,3	7,8	12,1	0,995	7,4	11,3	18,7	0,991	8,7	6,4	15,1	0,983	6,6	10,1	16,7	0,983
Barbiero et Mecatti (2010) $x-equ1$	4,0	7,6	11,6	1,000	7,3	11,6	18,9	0,991	8,7	6,4	15,1	0,986	6,7	9,9	16,6	0,987
Barbiero et Mecatti (2010) $x-equ2$	3,9	7,8	11,7	0,997	7,5	11,5	19,0	0,989	7,2	6,1	13,3	0,985	6,4	11,0	17,4	0,956
Antal et Tillé (2014)	4,1	8,2	12,3	0,995	7,7	11,5	19,2	0,985	8,4	7,0	15,4	0,985	7,1	10,3	17,4	0,973
Antal et Tillé (2011a)	4,1	7,6	11,7	1,002	6,9	10,7	17,6	1,003	8,2	7,0	15,2	0,985	6,6	10,0	16,6	0,992
	G_y															
Mashreghi et coll. (2016)	1,6	6,0	7,6	1,000	2,6	9,1	11,7	1,000	2,7	5,2	7,9	1,000	3,4	6,1	9,5	1,000
Chauvet (2007) simplifié	1,6	6,2	7,8	1,000	2,6	9,5	12,1	1,000	2,6	5,3	7,9	1,000	3,9	6,2	10,1	0,986
Barbiero et Mecatti (2010) $x-equ1$	1,8	6,1	7,9	1,000	2,5	9,0	11,5	1,000	2,4	5,3	7,7	1,000	4,0	6,0	10,0	1,000
Barbiero et Mecatti (2010) $x-equ2$	1,7	6,5	8,2	1,000	2,5	9,1	11,6	1,000	2,6	5,6	8,2	1,022	3,3	6,5	9,8	1,014
Antal et Tillé (2014)	2,6	5,0	7,6	0,984	3,4	7,3	10,7	1,000	2,8	5,1	7,9	1,000	4,4	5,7	10,1	1,000
Antal et Tillé (2011a)	1,5	6,5	8,0	1,000	2,2	8,9	11,1	1,010	2,5	5,0	7,5	1,012	2,9	7,3	10,2	1,000

Tab. 4.70. Étude 3: Taux de couverture et longueur moyenne des intervalles de confiance t -bootstrap par rapport à la méthode Antal et Tillé (2014) (paramètres non lisses, m_y et G_y)

Méthode	Scénario 1 $f = 0,07$ $n = 150$				Scénario 2 $f = 0,07$ $n = 50$				Scénario 3 $f = 0,3$ $n = 150$				Scénario 4 $f = 0,3$ $n = 50$			
	L	U	T	lg	L	U	T	lg	L	U	T	lg	L	U	T	lg
	m_y															
Antal et Tillé (2014)	3,2	4,1	7,3	1,000	5,8	4,3	10,1	1,000	5,1	3,9	9,0	1,000	3,6	4,2	7,8	1,000
Antal et Tillé (2011a)	3,1	3,8	6,9	1,010	5,3	4,1	9,4	1,016	5,0	3,8	8,8	1,010	3,5	4,0	7,5	1,023
	G_y															
Antal et Tillé (2014)	2,5	2,3	4,8	1,000	1,7	3,0	4,7	1,000	2,3	2,4	4,7	1,000	2,0	1,7	3,7	1,000
Antal et Tillé (2011a)	1,9	3,5	5,4	0,935	1,1	4,6	5,7	0,863	2,5	2,6	5,1	0,965	1,0	2,8	3,8	0,937

Tab. 4.71. Étude 3: La longueur des intervalles de confiance asymptotique et t -bootstrap par rapport à celle de l'intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016) (m_y et G_y)

Méthode	Scénario 1 ($f = 0,07 ; n = 150$)		Scénario 2 ($f = 0,07 ; n = 50$)		Scénario 3 ($f = 0,3 ; n = 150$)		Scénario 4 ($f = 0,3 ; n = 50$)	
	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap	Asympt	t -bootstrap
m_y								
Holmberg (1998)	1,028	–	1,052	–	1,034	–	1,025	–
Chauvet (2007)	1,027	–	1,053	–	1,024	–	1,014	–
Mashreghi et coll. (2016)	1,049	–	1,088	–	1,043	–	1,046	–
Chauvet (2007) simplifié	1,028	–	1,053	–	1,022	–	1,024	–
Barbiero et Mecatti (2010) $x-equ1$	1,028	–	1,053	–	1,027	–	1,032	–
Barbiero et Mecatti (2010) $x-equ2$	1,025	–	1,046	–	1,021	–	1,001	–
Antal et Tillé (2014)	1,028	1,114	1,051	1,212	1,024	1,149	1,017	1,159
Antal et Tillé (2011a)	1,031	1,125	1,062	1,232	1,027	1,154	1,029	1,186
G_y								
Holmberg (1998)	1,016	–	1,010	–	1,065	–	1,041	–
Chauvet (2007)	1,021	–	1,017	–	1,011	–	1,015	–
Mashreghi et coll. (2016)	1,033	–	1,030	–	1,022	–	1,027	–
Chauvet (2007) simplifié	1,033	–	1,010	–	1,022	–	1,014	–
Barbiero et Mecatti (2010) $x-equ1$	1,016	–	1,010	–	1,022	–	1,014	–
Barbiero et Mecatti (2010) $x-equ2$	1,016	–	1,010	–	1,022	–	1,027	–
Antal et Tillé (2014)	1,033	1,262	1,061	1,616	1,022	1,130	1,041	1,284
Antal et Tillé (2011a)	1,016	1,180	1,010	1,394	1,022	1,087	0,986	1,203

4.6.6. Discussion: cas des paramètres non lisses de totaux

Il ressort des résultats indiqués dans le tableau 4.66 que la méthode bootstrap par pseudo-population de Barbiero et Mecatti (2010) *x-equ2* fait mieux que les autres méthodes en terme du *BR* et *REQMR* dans le cas de la médiane. Au contraire, la méthode de Mashreghi et coll. (2016) a fait moins bien que les autres méthodes. On constate également que la méthode de Antal et Tillé (2014) fait mieux que les autres méthodes dans le cas de l'indice de Gini. Pour ce qui est des propriétés empiriques d'intervalles de confiance asymptotique, percentile et bootstrap de base, on constate que toutes les méthodes bootstrap ont un comportement voisin en terme des erreurs de couverture bilatérale et unilatérale (voir les tableaux 4.67 à 4.69). Dans le cas de la médiane, on remarque que la technique percentile fait mieux que les autres techniques notamment le bootstrap de base et le *t*-bootstrap. En effet, les taux d'erreurs de couverture théoriques bilatérales et unilatérales sont bien respectés par les intervalles de confiance percentile dans tous les scénarios (voir tableau 4.68). Dans le cas de l'indice de Gini, on constate que le taux d'erreur de couverture bilatérale n'est respecté que par les intervalles de confiance *t*-bootstrap. On constate également que les intervalles de confiance *t*-bootstrap par la méthode de Antal et Tillé (2014) respectent bien les taux d'erreur de couverture bilatérale et unilatérale, alors que les intervalles de confiance *t*-bootstrap de la méthode d' Antal et Tillé (2011a) ne respectent pas toujours les taux d'erreur de couverture unilatérale (voir tableau 4.70). En plus la méthode d'(Antal et Tillé, 2014) permet de calculer un intervalle de confiance *t*-bootstrap dans un temps record.

Conclusion

Dans ce travail nous avons comparé plusieurs méthodes bootstrap de l'estimation de la variance pour trois plans d'échantillonnage, à savoir le plan aléatoire simple sans remise, le plan de Poisson et le plan de Poisson conditionnel. Nous avons proposé pour les méthodes qui n'utilisent pas un principe de plug-in, à l'exception de la méthode de Beaumont et Patak (2012), une façon de réaliser un double bootstrap à partir du rééchantillon pour estimer la variance V^* afin de calculer un intervalle de confiance t -bootstrap. En effet, la méthode de Beaumont et Patak (2012) est la seule qui génère les ajustements bootstrap sans utiliser un mécanisme de ré-échantillonnage ce qui a rendu l'identification du rééchantillon impossible.

Nous avons également proposé une façon de calculer les intervalles de confiance percentile par la méthode de Beaumont et Patak (2012) si les ajustements bootstrap ont été ré-échelonnés et nous l'avons testée dans le cas du coefficient de corrélation pour le plan aléatoire simple sans remise et le plan de Poisson. La méthode de Rao et Wu (1988) est différente des autres car elle manipule les données de l'échantillon des observations ré-échelonnés et non pas celles de l'échantillon des observations d'origine. Cette particularité ne cause aucun problème pour évaluer la précision d'une estimation par un intervalle de confiance bootstrap pour les paramètres linéaires ou lisses car la valeur du paramètre bootstrap étant donné un échantillon est la même soit en utilisant les observations d'origine ou bien les observations ré-échelonnées. Toutefois, dans le cas des paramètres non lisses comme la médiane, cette égalité n'est plus satisfaite. Ainsi, nous avons proposé une façon de calculer les

intervalles de confiance bootstrap par cette méthode pour ces paramètres et nous l'avons testée dans nos simulations, ce qui a donné des résultats satisfaisants.

Dans notre première étude par simulation nous avons comparé plusieurs méthodes bootstrap de l'estimation de la variance pour le plan aléatoire simple sans remise dans quatre scénarios. Les principales recommandations tirées de cette étude sont énumérées dans les points suivants:

- 1- Pour les paramètres linéaires ainsi que le ratio, il semble que toutes les méthodes bootstrap de l'estimation de la variance ont un comportement voisin. En conséquence, les méthodes bootstrap ayant l'avantage d'être plus simple et rapide, et permettant d'accoler au fichier d'enquête des variables de poids bootstrap sont recommandées, comme la méthode de Rao et coll. (1992) et la méthode Antal et Tillé (2014);
- 2- Pour évaluer la précision dans le cas du coefficient de corrélation, nous recommandons l'utilisation de la technique t -bootstrap car la couverture bilatérale et unilatérale est mieux respectée comparativement aux autres techniques de calcul des intervalles de confiance. Nous recommandons la méthode Antal et Tillé (2011a) pour mettre en œuvre le t -bootstrap car les intervalles de confiance t -bootstrap sont relativement moins larges que ceux produits par les autres méthodes;
- 3- Dans le cas de la médiane nous recommandons la technique des percentiles pour évaluer la précision car la couverture unilatérale et bilatérale est mieux respectée et les intervalles de confiance percentile sont moins larges que les intervalles de confiance asymptotique et t -bootstrap ;
- 4- Nous recommandons également la technique t -bootstrap pour évaluer la précision dans le cas de l'indice de Gini car le gain apporté par cette technique est très notable en terme de la couverture unilatérale et bilatérale. Le choix de la méthode d'Antal et Tillé (2014) pour mettre en œuvre le t -bootstrap nous semble approprié dans le cas de ce paramètre en terme de la couverture unilatérale et bilatérale, sa simplicité et sa rapidité comparativement aux autres méthodes bootstrap;

- 5- Pour les méthodes où il faut spécifier certains paramètres de réglage, il nous semble que pour s'épargner d'un comportement extrême de ces méthodes, il faille choisir pour les méthodes Rao et Wu (1988) et Rao et coll. (1992) une taille du ré-échantillon proche de celle de la taille de l'échantillon d'origine alors que le choix d'une distribution symétrique soit plus approprié pour la méthode Beaumont et Patak (2012).

Dans notre deuxième étude nous avons comparé 9 méthodes bootstrap pour le plan de Poisson dans 4 scénarios différents. Les principales conclusions issues de cette étude sont présentées dans les points suivants:

- 1- Pour les paramètres linéaires ainsi que le ratio, il semble que toutes les méthodes bootstrap de l'estimation de la variance ont un comportement voisin. Toutefois, nous pouvons recommander les deux versions x -équilibrée de Barbiero et Mecatti (2010) pour estimer la variance dans le cas de ces paramètres car elles sont relativement plus stables que les autres méthodes bootstrap. Nous recommandons également la technique des percentiles pour évaluer la précision pour ces paramètres;
- 2- Pour évaluer la précision dans le cas du coefficient de corrélation, nous recommandons l'utilisation du t -bootstrap par la méthode de Antal et Tillé (2014) car la couverture unilatérale et bilatérale est bien respectée. Toutefois, il faut s'attendre à ce qu'un intervalle de confiance t -bootstrap soit environ 6% à 20% plus large qu'un intervalle de confiance percentile;
- 3- Pour évaluer la précision dans le cas de la médiane nous recommandons l'utilisation de la technique percentile;
- 4- Dans le cas de l'indice de Gini, nous recommandons l'utilisation de la technique t -bootstrap par la méthode de Antal et Tillé (2014) car la couverture unilatérale et bilatérale est notablement mieux respectée. Toutefois, il faut s'attendre à un intervalle de confiance plus large.

Dans notre dernière étude nous avons comparé 8 méthodes bootstrap de l'estimation de la variance pour le plan de Poisson conditionnel. Cette étude a nécessité un temps énorme pour avoir des résultats parce que l'algorithme de tirage d'un échantillon selon un plan de

Poisson conditionnel est pesant en terme de calcul. Les recommandations tirées de cette étude sont:

- 1- Dans le cas d'un paramètre linéaire ainsi que le ratio, toutes les méthodes bootstrap ont un comportement voisin sauf la méthode de Holmberg (1998) dans le cas d'une fraction de sondage importante. Nous recommandons la méthode de Antal et Tillé (2014) car elle est simple et plus rapide en comparaison avec les autres méthodes, notamment les méthodes bootstrap par pseudo-population;
- 2- Nous recommandons également la technique des percentiles pour évaluer la précision dans le cas des paramètres linéaires, le ratio et la médiane;
- 3- Pour évaluer la précision dans le cas du coefficient de corrélation et l'indice de Gini nous recommandons la technique *t*-bootstrap par la méthode Antal et Tillé (2014) car d'une part la couverture unilatérale et bilatérale est bien respectée et d'autre part la réalisation d'un double bootstrap par cette méthode est extrêmement rapide comparativement aux autres méthodes bootstrap. En effet, le temps nécessaire pour calculer un intervalle de confiance *t*-bootstrap par cette méthode est moins long d'environ 1482 fois de celui nécessaire pour calculer un intervalle de confiance percentile par la méthode de Mashreghi et coll. (2016).

De manière générale toutes les méthodes bootstrap avaient un comportement relativement voisin au moins dans le cas d'un paramètre linéaire à l'exception de la méthode de Holmberg (1998) et plus précisément si le taux de sondage est important. Toutefois, il nous semble que la méthode d'Antal et Tillé (2014) est la plus prometteuse pour l'utilisation dans les enquêtes par échantillonnage à probabilités inégales car les ajustements bootstrap de poids sont des entiers générés directement à partir d'un plan de ré-échantillonnage simple et rapide sans avoir besoin de créer une population artificielle comme c'est le cas des méthodes bootstrap par pseudo-population. En effet les ajustements négatifs peuvent conduire à des valeurs impossibles pour la statistique calculée à partir des poids bootstrap comme nous l'avons évoquées pour la méthode de Beaumont et Patak (2012) dans le cas du coefficient de corrélation. Le ré-échelonnement des ajustements bootstrap proposé par Beaumont et Patak (2012) résout le problème pour les paramètres lisses mais il rend

l'estimation de la précision plus compliquée. La rapidité et la simplicité de la méthode d'Antal et Tillé (2014) constituent ses points forts et rendent la mise en œuvre d'une technique avancée comme le t -bootstrap abordable et faisable en pratique. En ce qui concerne l'évaluation de la précision par un intervalle de confiance il semble d'après notre étude que la technique bootstrap de base a fait toujours moins bien que les autres techniques de construction d'intervalles de confiance.

Bibliographie

- Antal, E., Langel, M. et Tillé, Y. (2011). Variance estimation of inequality indices in complex sampling designs. In *Proceedings of the 58th World Statistical Congress*.
- Antal, E. et Tillé, Y. (2011b). Simple random sampling with over-replacement. *Journal of Statistical Planning and Inference*, **141**, 597–601.
- Antal, E. et Tillé, Y. (2014). A new resampling method for sampling designs without replacement: the doubled half bootstrap. *Computational Statistics*, **29**, 1345–1363.
- Antal, E. et Tillé, Y. (2011a). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, **106**, 534–543.
- Barbiero, A. et Mecatti, F. (2010). Bootstrap algorithms for variance estimation in π ps sampling. In *Complex Data Modeling and Computationally Intensive Statistical Methods*, 57–69. Springer.
- Beaumont, J.-F. et Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review*, **80**, 127–148.
- Bertail, P. et Combris, P. (1997). Bootstrap généralisé d'un sondage. *Annales d'Economie et de Statistique*, **46**, 49–83.
- Bickel, P. J. et Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, **12**, 470–482.
- Booth, J. G., Butler, R. W. et Hall, P. (1994). Bootstrap methods for finite populations. *Journal of the American Statistical Association*, **89**, 1282–1289.

- Brewer, K. et Donadio, M. E. (2003). La variance sous grande entropie de l'estimateur de Horvitz-Thompson. *Techniques d'enquête*, **29**, 213–220.
- Chao, M. T. et Lo, S. H. (1985). A bootstrap method for finite population. *Sankhyā: The Indian Journal of Statistics, Series A*, **47**, 399–405.
- Chao, M. T. et Lo, S. H. (1994). Maximum likelihood summary and the bootstrap method in structured finite populations. *Statistica Sinica*, **4**, 389–406.
- Chauvet, G. (2007). *Méthodes de bootstrap en population finie*. Ph.D. thesis, ENSAI.
- Chipperfield, J. et Preston, J. (2007). Efficient bootstrap for business surveys. *Survey Methodology*, **33**, 167–172.
- Deville, J. C. (1993). Estimation de la variance pour les enquêtes en deux phases. *Manuscrit. INSEE, Paris*.
- Deville, J. C. (2000). Note sur l'algorithme de Chen, Dempster et Liu. *Rapport technique, CREST-ENSAI, Rennes*.
- Deville, J. C. et Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, **128**, 569–591.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.
- Efron, B. et Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall New York.
- Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons, New York.
- Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, 181–184.
- Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, **35**, 1491–1523.
- Hájek, J. (1981). *Sampling from a Finite Population*. New York Marcel Dekker.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion* Springer-Verlag. New York.
- Hanif, M. et Brewer, K. (1980). Sampling with unequal probabilities without replacement: a review. *International Statistical Review/Revue Internationale de Statistique*, **48**, 317–335.

- Holmberg, A. (1998). A bootstrap approach to probability proportional-to-size sampling. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 378–383.
- Horvitz, D. G. et Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, **47**, 663–685.
- Lo, A. Y. (1991). Bayesian bootstrap clones and a biometry function. *Sankhyā: The Indian Journal of Statistics, Series A*, **53**, 320–333.
- Mashreghi, Z., Haziza, D. et Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, **10**, 1–52.
- Mason, D. M. et Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *The Annals of Statistics*, **20**, 1611–1624.
- Matei, A. et Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, **21**, 543–570.
- McCarthy, P. J. et Snowden, C. B. (1985). The bootstrap and finite population sampling. *Vital and Health Statistics, Series 2, No. 95*. DHHS Publication No. (PHS) 85–1369. Public Health Service. Washington. U.S. Government Printing Office.
- Rao, J. N. K. et Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, **83**, 231–241.
- Rao, J. N. K., Wu, C. F. J. et Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, **18**, 209–217.
- Särndal, C.-E., Swensson, B. et Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer Verlag, New York.
- Sitter, R. R. (1992a). Comparing three bootstrap methods for survey data. *Canadian Journal of Statistics*, **20**, 135–154.
- Sitter, R. R. (1992b). A resampling procedure for complex survey data. *Journal of the American Statistical Association*, **87**, 755–765.
- Tillé, Y. (2001). *Théorie des sondages: échantillonnage et estimation en population finie: cours et exercices avec solutions*. Dunod, Paris.

Tillé, Y. et Matei (2016). The sampling package, <https://cran.r-project.org/package=sampling>.

