

2026

00

Bases
statistiques
et généralités

Neuchâtel 2026

Applications du bootstrap à l'OFS

Estimation de variance à l'aide de méthodes de rééchantillonnage

Rapports de méthodes de la section Méthodes statistiques

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site Internet de l'OFS.

Applications du bootstrap à l'OFS

Estimation de variance à l'aide de méthodes de rééchantillonnage

Rédaction Djalel-Eddine Meskaldji, OFS
Rémy Mariétan, OFS
Éditeur Office fédéral de la statistique (OFS)

Neuchâtel 2026

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Djalel-Eddine Meskaldji, tél. +41 58 463 02 38
djalel-eddine.meskaldji@bfs.admin.ch

Rédaction: Djalel-Eddine Meskaldji, OFS; Rémy Mariétan, OFS

Contenu: Section méthodes statistiques, OFS

Série: Statistique de la Suisse

Domaine: 00 Bases statistiques et généralités

Langue du texte original: français

Mise en page: Publishing et diffusion PUB, OFS

En ligne: www.statistique.ch

Imprimés: www.statistique.ch
Office fédéral de la statistique, CH-2010 Neuchâtel,
order@bfs.admin.ch, tél. +41 58 463 60 60
Impression réalisée en Suisse

Copyright: OFS, Neuchâtel 2026
La reproduction est autorisée si la source est mentionnée.

Numéro OFS: 338-0095

ISBN: 978-3-303-00787-7

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Le Bootstrap	8
2.1 Introduction générale	8
2.2 Variantes du bootstrap	8
2.2.1 Bootstrap non-paramétrique	9
2.2.2 Bootstrap paramétrique	9
2.3 Estimation de la variance	9
2.4 Comparaison avec la méthode de linéarisation	9
3 Base théorique pour les applications	10
3.1 Adaptation générale du bootstrap pour les enquêtes par échantillonnage	10
3.1.1 Contraintes liées aux plans complexes	11
3.1.2 Grandes familles de méthodes	11
3.1.3 Diffusion de poids	12
3.2 Bootstrap non-paramétrique par pseudo-populations	12
3.2.1 Principe général	13
3.2.2 Construction de la pseudo-population	13
3.2.3 Réplication	14
3.2.4 Validité asymptotique	14
3.2.5 Limitations	14
3.3 Parametric Design Bootstrap (PDB)	15
3.3.1 Principe général	15
3.3.2 Construction de la pseudo-population paramétrique	15
3.3.3 Motivation de la correction PDB	16
3.3.4 Estimateur PDB de la MSE	16
3.3.5 Interprétation	16
3.3.6 Propriétés et intérêt pratique	17
3.3.7 Limitations	17
3.4 Méthodes de réplication par poids	17
3.4.1 Méthodes structurées	17
3.4.2 Méthodes bootstrap par génération de poids	18
3.4.3 Estimation de la variance	19
3.4.4 Propriétés et avantages	19
3.4.5 Limitations	20
3.5 Intervalles de confiance	20
3.5.1 IC par approximation normale	20
3.5.2 IC par quantiles empiriques (percentile)	20
3.6 Différents aspects concernant la mise en œuvre	21
3.6.1 Principe général	21
3.6.2 Prise en compte du calage	21
3.6.3 Prise en compte de la non-réponse	22
3.6.4 Prise en compte de l'imputation	22
3.6.5 Diagnostics des répliqués bootstrap	22
3.6.6 Recommandations opérationnelles à l'OFS	22
3.6.7 Reproductibilité	23
4 Application du bootstrap à l'OFS	23
4.1 Projet SAE-SE	24
4.2 Projet SAE-BESTA	25
4.3 Projet STATENT	26
4.4 Projet PIAAC	27

5 Conclusions	28
Annexes	29
A Notations	29
Bibliographie	31
Rapports de méthodes de la section méthodes statistiques de l'OFS	33

Préambule

Le présent rapport de méthodes vise à documenter les approches mises en œuvre à l'Office fédéral de la statistique (OFS) pour l'estimation de la variance à l'aide de méthodes de ré-échantillonnage. Il s'inscrit dans une volonté de formalisation et d'harmonisation des pratiques bootstrap utilisées dans différentes productions statistiques.

Ce rapport repose sur des travaux réalisés dans le cadre de soutiens méthodologiques des sections de production de l'OFS par la section Méthodes statistiques (METH). Il a pour ambition de fournir à la fois une synthèse des pratiques existantes, de capitaliser l'expérience acquise au sein de l'OFS et de servir de référence pour de futurs développements méthodologiques.

Les auteurs remercient chaleureusement les relecteurs de METH pour leurs commentaires précieux et leurs suggestions pertinentes qui ont permis d'améliorer la clarté et la rigueur de ce document : Anne Massiani, Lionel Qualité, Daniel Assoulin, Daniel Kilchmann et Jean-Pierre Renfer.

Résumé

Ce rapport de méthodes présente les principales approches de rééchantillonnage mises en œuvre à l'Office fédéral de la statistique (OFS) pour l'estimation de la variance dans le cadre d'enquêtes par échantillonnage.

L'accent est mis sur l'utilisation du bootstrap pour estimer la variance des estimateurs lorsque les développements analytiques sont difficiles à obtenir.

Après un rappel des principes du bootstrap et de ses variantes (paramétrique ou non-paramétrique), les spécificités de son application dans un contexte d'enquêtes basées sur un échantillon aléatoire sont décrites, notamment en tenant compte des poids d'échantillonnage, de la non-réponse et du calage sur des totaux connus de la population.

Des applications concrètes sont ensuite présentées dans le cadre de différentes statistiques de l'OFS. Pour chacune de ces statistiques, les choix méthodologiques effectués, les résultats obtenus, ainsi que les défis rencontrés sont illustrés.

L'ensemble du rapport vise à documenter les pratiques en matière d'estimation de variance à l'aide de méthodes de rééchantillonnage dans les statistiques de l'OFS, en tenant compte des contraintes opérationnelles et des exigences de qualité tout en capitalisant l'expérience acquise dans plusieurs projets méthodologiques de l'OFS.

1 Introduction

Dans les enquêtes basées sur un échantillon aléatoire, l'estimation de caractéristiques ou de paramètres d'intérêt s'accompagne inévitablement d'une incertitude due à la variabilité de l'échantillonnage. Parmi les approches qui permettent d'évaluer cette incertitude, le bootstrap constitue une alternative aux méthodes analytiques classiques, en particulier lorsque celles-ci sont difficiles à développer pour des plans d'échantillonnage ou des estimateurs complexes.

Le bootstrap repose sur le principe du rééchantillonnage : à partir d'un échantillon observé, un grand nombre de répliqués simulés sont générés selon différentes variantes permettant ainsi d'approximer empiriquement les principales caractéristiques de la distribution de l'estimateur, notamment sa variance et, dans certains cas, sa distribution. Cette approche est particulièrement utile pour des estimateurs complexes (par exemple non-linéaires ou dépendant explicitement d'un modèle statistique), pour lesquels il n'existe pas de formule explicite de variance ou quand celle-ci est difficile à obtenir.

Après une présentation théorique des principales variantes du bootstrap – non-paramétrique et paramétrique – nous discutons de leur adaptation au contexte des plans d'échantillonnage, notamment en présence de stratification, de pondération, de non-réponse et de calage.

Ce rapport ne se limite pas aux méthodes actuellement mises en œuvre à l'OFS. Il présente également plusieurs approches proposées dans la littérature, dont certaines pourraient constituer des alternatives intéressantes dans de futurs projets. Une attention particulière est accordée à leurs conditions d'application, à leurs propriétés théoriques et à leur pertinence dans le contexte des statistiques officielles.

L'OFS a eu recours à plusieurs méthodes de rééchantillonnage dans plusieurs projets liés aux enquêtes par échantillonnage, en particulier dans des contextes nécessitant une modélisation avancée, une estimation sur petits domaines (Small area estimation - SAE), ou une protection des données via la méthode de réplification de poids. Une synthèse de l'application des méthodes de rééchantillonnage dans les projets suivants est fournie à titre d'illustration :

- SAE-SE : estimation sur petits domaines dans le cadre du relevé structurel du système de recensement fédéral de la population (SE : Strukturhebung);
- SAE-BESTA : estimation sur petits domaines d'indicateurs trimestriels de la statistique de l'emploi (BESTA : Beschäftigungsstatistik);
- STATENT : estimation à l'aide de modèles du total des équivalents plein temps (EPT) dans la statistique structurelle des entreprises;
- PIAAC : estimations au sein du programme international pour l'évaluation des compétences des adultes (PIAAC : Programme for the International Assessment of Adult Competencies).

Le projet SAE-BESTA illustre la mise en œuvre d'une estimation de variance pour des estimations trimestrielles d'indicateurs sur petits domaines. Le projet STATENT illustre l'utilisation du bootstrap non-paramétrique pour estimer la variance du total des EPT. L'enquête PIAAC présente quant à elle une application de la méthode de réplification de poids visant à garantir une estimation fiable de la variance et à éviter de diffuser directement des variables auxiliaires sensibles nécessaires à une estimation de variance analytique. Le projet SAE-SE illustre l'utilisation du bootstrap paramétrique pour l'estimation de l'erreur quadratique moyenne (Mean Squared Error, MSE) sous le plan de sondage (design-MSE) dans le cadre de modèles de type Empirical Best Linear Unbiased Prediction (EBLUP).

Le reste du rapport est organisé comme suit. La section 2 présente les principes généraux du bootstrap et les notions nécessaires à la compréhension des méthodes de rééchantillonnage. La section 3 décrit les principales adaptations du bootstrap aux enquêtes par échantillonnage, en présentant les différentes familles de méthodes ainsi que les aspects liés à la pondération, à la non-réponse et à la construction d'intervalles de confiance. La section 4 illustre ces approches au

travers de plusieurs applications à l'OFS, avant qu'une synthèse des principaux enseignements ne soit présentée dans la conclusion.

2 Le Bootstrap

2.1 Introduction générale

Le bootstrap est une méthode de rééchantillonnage introduite par [Efron \[1979\]](#) dans le but d'estimer, de manière empirique, la précision d'un estimateur statistique. Son principe repose sur l'exploitation directe de l'échantillon observé. Au lieu de supposer une distribution théorique pour l'estimateur, on génère un grand nombre de répliqués d'échantillons par tirage avec remise à partir des données disponibles. La distribution empirique des répliqués de l'estimateur permet ensuite d'estimer sa variance ou de construire des intervalles de confiance. Cette approche s'inscrit dans une perspective d'inférence fondée sur des méthodes de simulation plutôt que sur des développements analytiques explicites [[Efron & Tibshirani, 1993](#), p. xiv].

Le bootstrap présente plusieurs intérêts :

- Il est conceptuellement simple et flexible.
- Il s'applique aussi bien à des estimateurs linéaires que non-linéaires (médianes, ratios, quantiles, estimateurs modélisés).
- Il ne nécessite pas le développement explicite de formules analytiques pour la variance, ce qui est particulièrement utile lorsque ces formules sont difficiles à obtenir. En effet, cette approche est privilégiée lorsque la dérivation d'une expression théorique explicite est nettement plus complexe que la simple démonstration de son existence, condition généralement suffisante pour établir la validité du bootstrap [[Shao, 2003](#), Section 2, pp. 191–192].

Cependant, l'usage du bootstrap classique repose sur certaines hypothèses. Sa validité asymptotique a été largement étudiée dans la littérature (voir par exemple [Davison & Hinkley \[1997\]](#)). La validité théorique du bootstrap est généralement établie en montrant que la distribution des répliqués bootstrap approxime celle de l'estimateur d'intérêt. Dans de nombreuses applications, l'objectif est toutefois plus limité et consiste principalement à reproduire correctement certaines de ces caractéristiques, notamment sa variance [[Shao, 2003](#), p. 191, Section 2]. Dans un cadre asymptotique, cette approximation est généralement justifiée pour des estimateurs suffisamment réguliers, c'est-à-dire pouvant être exprimés comme des fonctions régulières des observations indépendantes et identiquement distribuées (i.i.d.).

Le nombre de répliqués B utilisé dans la pratique influe sur la qualité des estimations obtenues. Des valeurs typiques comprises entre 500 et 2 000 sont couramment utilisées. Le choix de B est un compromis entre les contraintes de calcul et la stabilité des résultats, la variabilité de simulation pouvant être non négligeable lorsque le nombre de répliqués est trop faible [[Beaumont et Patak, 2012](#), p. 127].

Enfin, plusieurs variantes du bootstrap ont été développées pour répondre à des contextes spécifiques, notamment en présence de plans d'échantillonnage en populations finies, simples ou complexes, pour lesquels des adaptations du bootstrap classique sont nécessaires [[Mashreghi et al., 2016](#), pp. 2–3].

2.2 Variantes du bootstrap

Le terme « bootstrap » regroupe plusieurs méthodes de rééchantillonnage qui ont pour objectif commun d'estimer la distribution d'un estimateur ou, plus modestement, ses deux premiers moments (espérance et variance), mais qui diffèrent dans la manière de générer les répliqués.

On distingue principalement deux grandes familles de méthodes : les approches paramétriques et les approches non-paramétriques.

2.2.1 Bootstrap non-paramétrique

Le bootstrap non-paramétrique consiste à rééchantillonner dans l'échantillon observé avec remise. Cette approche ne repose sur aucune hypothèse de distribution et s'applique directement aux données. Elle est facilement adaptable au cas d'échantillon aléatoire simple, mais nécessite des ajustements plus complexes pour d'autres plans d'échantillonnage comme nous allons le voir dans la section 3.

2.2.2 Bootstrap paramétrique

Le bootstrap paramétrique repose sur l'ajustement d'un modèle probabiliste aux données. Une fois les paramètres estimés, de nouvelles données sont simulées sur la base de ce modèle, et l'estimateur est recalculé sur chaque simulation. Cette approche est adaptée lorsque le modèle fournit une représentation suffisamment réaliste du mécanisme ayant généré les données observées [Shao, 2003, Section 2, pp. 191–192].

2.3 Estimation de la variance

Soit $\hat{\theta}$ l'estimateur calculé sur l'échantillon observé et $\hat{\theta}_b^*$, $b = 1, \dots, B$, les répliqués de $\hat{\theta}$ obtenus à partir des B échantillons bootstrap. On note

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*$$

la moyenne des estimations bootstrap. L'estimateur usuel de la variance bootstrap est (voir par exemple [Efron & Tibshirani, 1993, Chapitre 6], et [Davison & Hinkley, 1997, Chapitre 2]) :

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2.$$

Dans la littérature on trouve aussi la variante suivante :

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta})^2.$$

Cette seconde écriture est proche de la précédente lorsque B est grand et que $\bar{\theta}^*$ est proche de $\hat{\theta}$. La différence entre les deux formulations devient alors généralement négligeable en pratique, et dépend essentiellement du nombre de répliqués utilisés.

2.4 Comparaison avec la méthode de linéarisation

La méthode de linéarisation repose sur une approximation asymptotique de l'estimateur par une forme linéaire, obtenue à partir d'un développement de type différentiel. Cette approche permet de dériver des formules analytiques de variance pour des estimateurs suffisamment réguliers [Davison & Sardy, 2007, p. 8]. Elle est largement utilisée en pratique, notamment en raison de son efficacité computationnelle, puisqu'elle ne nécessite pas de procédures de rééchantillonnage. Les méthodes bootstrap adoptent une approche différente en remplaçant les développements analytiques par une approximation empirique de la distribution de l'estimateur à partir de répliqués

simulés. Cette stratégie permet d'éviter le calcul explicite de formules de variance, ce qui peut s'avérer particulièrement avantageux lorsque ces formules sont difficiles à obtenir [Shao, 2003, pp. 191–192].

Dans le contexte des enquêtes par échantillonnage, les deux approches présentent des caractéristiques complémentaires. La linéarisation est généralement rapide à mettre en œuvre une fois les expressions analytiques établies, mais peut devenir délicate à appliquer pour des estimateurs complexes ou non différentiables. Le bootstrap, de son côté, offre une grande flexibilité et une applicabilité large, notamment dans des situations complexes ou pour des estimateurs non lisses (voir [Davison & Sardy, 2007, p. 26] et [Shao, 2003, p. 193]). En contrepartie, il peut entraîner un coût de calcul plus élevé et nécessiter des adaptations spécifiques dans certains plans d'échantillonnage [Davison & Sardy, 2007, p. 28].

De manière générale, ces deux approches constituent des outils importants pour l'estimation de la variance dans les enquêtes complexes, et leur utilisation dépend du type d'estimateur considéré, de la complexité du plan de sondage et des contraintes opérationnelles.

Les sections suivantes présentent les principales adaptations du bootstrap au contexte des enquêtes par échantillonnage et des plans de sondage complexes.

3 Base théorique pour les applications

Les erreurs affectant les estimations issues d'enquêtes par échantillonnage peuvent être décomposées en deux grandes composantes : le biais et la variance. Le biais peut résulter de plusieurs sources, notamment :

- les erreurs de mesure ;
- la non-réponse ;
- l'estimation de fonctions non-linéaires de totaux ;
- le recours à une modélisation, notamment dans les approches d'estimation sur petits domaines (SAE).

Dans ce rapport, l'accent est mis sur l'estimation de la variance, en particulier à l'aide des méthodes de bootstrap.

3.1 Adaptation générale du bootstrap pour les enquêtes par échantillonnage

La section précédente a présenté les principes généraux du bootstrap dans un cadre i.i.d. Dans les enquêtes par échantillonnage, ces principes doivent être adaptés au plan de sondage, où la population est finie et considérée comme fixe, et où l'aléa provient du mécanisme de sélection. Dans ce contexte, la dérivation de formules analytiques de variance peut rapidement devenir complexe, ce qui motive le recours à des méthodes de rééchantillonnage telles que le bootstrap [Davison & Sardy, 2007, p. 8]. La complexité des estimateurs considérés peut aussi plaider en faveur d'un bootstrap. En pratique à l'OFS, le bootstrap a été utilisé dans des cas où la complexité des estimateurs, plus que celle du plan de sondage, constituait une difficulté pour l'obtention de formules.

L'adaptation du bootstrap aux enquêtes par échantillonnage est toutefois délicate. Comme les données d'enquête ne sont en général pas i.i.d., les méthodes bootstrap doivent être modifiées de manière à reproduire correctement le plan de sondage et les structures de dépendance qu'il induit [Mashreghi et al., 2016, p. 2–3]. Plus généralement, la validité théorique du bootstrap est établie en montrant que la distribution empirique des estimateurs bootstrap $\hat{\theta}^*$, approxime la distribution de l'estimateur $\hat{\theta}$. Toutefois, dans de nombreuses applications en sondage, l'objectif est plus restreint et consiste principalement à obtenir une approximation correcte de certaines caractéristiques de cette distribution, notamment la variance [Shao, 2003, p. 191, Section 2]. Sous

certaines conditions, les estimateurs de variance obtenus par rééchantillonnage sont consistants pour une large classe d'estimateurs, incluant des estimateurs lisses et non lisses [Sitter, 2001, Section 4]. Une grande partie des résultats théoriques concernent les estimateurs dits par substitution, c'est-à-dire des estimateurs pouvant être exprimés comme des fonctionnelles de la distribution empirique pondérée de l'échantillon (voir par exemple Efron & Tibshirani, 1993). Cette classe inclut de nombreux estimateurs usuels en statistique d'enquête, tels que les moyennes, totaux, ratios ou quantiles.

Dans les enquêtes par échantillonnage, l'adaptation du bootstrap vise principalement à :

- intégrer correctement les probabilités d'inclusion, notamment lorsqu'elles sont inégales ;
- préserver la structure du plan (stratification, grappes, multi-degrés) ;
- reproduire les traitements appliqués aux poids, en particulier les ajustements pour la non-réponse et le calage, afin que la variance estimée corresponde bien à l'estimateur effectivement publié [Davison & Sardy, 2007, p. 16].

Ces adaptations conditionnent directement la validité de l'estimation de variance sous le plan de sondage. Une attention particulière doit également être portée à la correction de population finie, dont l'absence peut conduire à un biais dans l'estimation de la variance [Davison & Sardy, 2007, p. 17].

3.1.1 Contraintes liées aux plans complexes

Stratification Lorsque le plan est stratifié, le rééchantillonnage doit en général être effectué indépendamment dans chaque strate afin de préserver la structure du plan initial [Davison & Sardy, 2007, p. 17].

Échantillonnage en grappes Dans les plans multi-degrés, le mécanisme bootstrap doit reproduire la structure de sélection à chaque niveau du plan, notamment au niveau des unités primaires d'échantillonnage, afin de préserver les dépendances induites par le plan [Rao & Wu, 1988, Mashreghi et al., 2016].

Plans à probabilités inégales Lorsque les probabilités d'inclusion diffèrent selon les unités, le mécanisme bootstrap doit reproduire cette hétérogénéité. Cela constitue l'une des difficultés majeures de l'adaptation du bootstrap aux enquêtes par échantillonnage.

3.1.2 Grandes familles de méthodes

Dans la littérature sur le bootstrap en sondage, on distingue plusieurs grandes familles de méthodes [Mashreghi et al., 2016, pp. 2–3].

Méthodes par pseudo-population Les méthodes par pseudo-population reconstruisent, à partir de l'échantillon, une population artificielle à laquelle un plan de sondage bootstrap est ensuite appliqué. Elles sont particulièrement naturelles lorsque l'on souhaite reproduire explicitement la logique du plan de sondage et peuvent être adaptées à des contextes complexes, notamment pour des plans à probabilités inégales ou à plusieurs degrés [Chauvet, 2007].

Méthodes directes par rééchantillonnage Les méthodes directes construisent les réplicats à partir de l'échantillon observé, sans reconstruire explicitement de pseudo-population. C'est le cas, par exemple, du bootstrap rééchelonné de Rao & Wu [1988]. Ces méthodes sont souvent plus simples à mettre en œuvre en pratique que la méthode par pseudo-population, mais leur domaine de validité dépend du plan considéré et des hypothèses retenues.

Méthodes de réplification par poids Les méthodes de réplification par poids construisent directement des jeux de poids répliqués, sans sélection explicite de nouveaux échantillons ni reconstruction d’une pseudo-population. Elles constituent une famille particulièrement attractive en statistique officielle, car elles sont généralement plus simples à mettre en œuvre que les approches par pseudo-population et permettent de diffuser directement des poids répliqués aux utilisateurs.

Cette famille regroupe notamment deux grandes catégories de méthodes. La première correspond aux méthodes bootstrap par génération de poids, parmi lesquelles figurent les approches d’Antal et Tillé [Antal & Tillé, 2011] ainsi que le *generalized bootstrap* proposé par Beaumont et al. [2008] et étendu par Beaumont et Patak [2012]. La seconde regroupe les méthodes de réplification structurées, telles que le Balanced Repeated Replication (BRR), la méthode de Fay ou certaines variantes du jackknife, qui construisent les répliqués selon une structure prédéfinie plutôt qu’au moyen d’un mécanisme bootstrap aléatoire [Wolter, 2007].

Bootstrap paramétrique Enfin, une quatrième famille regroupe les méthodes de bootstrap paramétrique. Contrairement aux approches précédentes, elles reposent sur un modèle probabiliste ajusté aux données. Dans le contexte des estimateurs SAE, une variante particulière, le parametric design bootstrap (PDB), est décrite à la section 3.3.

Toutes les méthodes précédentes visent à reproduire la variabilité induite par le plan de sondage. Leur mise en œuvre dépend toutefois du type de l’estimateur considéré.

En toute généralité, un estimateur est une fonction des observations de l’échantillon. Néanmoins, dans certains cas – notamment en estimation sur petits domaines (SAE) – cette fonction fait intervenir, en plus des observations, des informations auxiliaires définies au niveau de la population. Dans ces situations, des approches fondées sur une pseudo-population ou sur une modélisation explicite sont souvent privilégiées, car elles permettent de reproduire de manière cohérente l’ensemble des composantes intervenant dans le calcul de l’estimateur.

Les méthodes directes par rééchantillonnage constituent l’extension la plus naturelle du bootstrap classique aux enquêtes par échantillonnage. Elles sont principalement adaptées à certains plans spécifiques et ne sont, à notre connaissance, pas utilisées actuellement dans les développements méthodologiques de l’OFS. Elles ne seront donc pas détaillées davantage dans ce rapport.

3.1.3 Diffusion de poids

Dans la pratique des statistiques officielles, de nombreuses institutions diffusent des poids bootstrap permettant aux utilisateurs de calculer des estimateurs de variance sans recourir à des formules analytiques explicites [Mashreghi et al., 2016, p. 2]. Deux stratégies principales sont alors utilisées :

- le rééchantillonnage direct des unités;
- la génération directe de poids bootstrap.

La seconde approche est notamment utilisée dans des enquêtes internationales, par exemple PIAAC, où la méthode de Fay permet de construire des poids de rééchantillonnage (poids répliqués) qui intègrent l’information nécessaire à l’estimation de la variance, sans nécessiter l’accès aux variables auxiliaires sensibles, notamment celles utilisées pour le calage [Bianchi et al., 2025, OECD, 2022b].

3.2 Bootstrap non-paramétrique par pseudo-populations

La méthode des pseudo-populations est une approche non-paramétrique adaptée aux plans d’échantillonnage en population finie. Introduite par Gross [1980] et développée notamment par

[Chauvet \[2007\]](#), elle vise à reproduire la variabilité sous le plan en reconstruisant une population artificielle, appelée pseudo-population, à partir de l'échantillon observé, puis en y appliquant un plan de sondage bootstrap.

Dans ce cadre, la pseudo-population est considérée comme fixe et l'aléa provient exclusivement du mécanisme de sélection. Les méthodes par pseudo-population s'inscrivent ainsi naturellement dans une approche d'estimation de variance sous le plan [[Mashreghi et al., 2016](#), p. 3].

3.2.1 Principe général

Le principe consiste à construire une pseudo-population à partir de l'échantillon en dupliquant les unités de l'échantillon, puis à tirer des échantillons bootstrap dans cette population artificielle selon un plan inspiré du plan initial [[Mashreghi et al., 2016](#), Section 4.1, pp. 11–12]. La variabilité observée entre les réplicats vise alors à reproduire la variabilité due au plan de sondage.

3.2.2 Construction de la pseudo-population

Soit un échantillon s de taille n tiré dans une population finie U de taille N , selon des probabilités d'inclusion π_i .

On distingue les différentes étapes de la pondération :

$$w_i^{(0)} = \frac{1}{\pi_i}$$

poids de sondage, reflétant le nombre d'unités de la population que l'unité i représente sous le plan de sondage,

$$w_i^{(nr)}$$

poids ajustés pour la non-réponse, obtenus en corrigeant les poids de sondage afin de compenser la non-réponse totale de certaines unités échantillonnées,

$$w_i^{(cal)}$$

poids finaux après calage, résultant d'un ajustement des poids corrigés pour la non-réponse de manière à respecter des contraintes sur des totaux auxiliaires connus au niveau de la population. Ces différentes étapes peuvent être vues comme une succession d'ajustements appliqués aux poids de sondage, permettant des estimations correctes pour la population.

Dans les expériences décrites dans ce rapport, la pseudo-population est construite à partir des poids corrigés pour la non-réponse mais d'autres variantes sont possibles. Le choix de la version des poids utilisée pour construire la pseudo-population peut avoir un impact sur la variance estimée et sera discuté ultérieurement dans la section des recommandations (voir section 3.6.6). En pratique, la construction de la pseudo-population consiste à répéter les unités de l'échantillon un nombre de fois proche de leur poids, éventuellement complété par une règle d'arrondi ou un mécanisme aléatoire afin d'obtenir une population artificielle de taille proche de N [[Mashreghi et al., 2016](#), Section 4.1.1, p. 12].

Dans les cas simples (par exemple échantillonnage aléatoire simple sans remise), cette construction revient à répéter chaque unité environ N/n fois, avec des ajustements lorsque ce rapport n'est pas entier [[Mashreghi et al., 2016](#), p. 12].

Dans un cas plus général, une construction possible consiste à définir, pour chaque unité $i \in s$, un nombre de copies aléatoire dans la pseudo-population $U^{*(c)}$, $c = 1, \dots, C$, :

$$m_i^{(c)} = \lfloor w_i \rfloor + \Psi_i^{(c)}, \quad \Psi_i^{(c)} \sim \text{Bernoulli}(w_i - \lfloor w_i \rfloor).$$

Cette construction garantit que la taille (en espérance) de la pseudo-population est proche de N . Lorsque les poids utilisés pour construire la pseudo-population sont suffisamment grands, l'effet relatif des arrondis devient généralement faible. Dans ce cas, la génération d'une seule pseudo-population peut déjà fournir une approximation satisfaisante tout en réduisant fortement le coût de calcul.

3.2.3 Réplication

À partir de chaque pseudo-population $U^{*(c)}$, des échantillons bootstrap $s^{*(b,c)}$ sont tirés en imitant le processus de sélection de l'échantillon initial (méthode de tirage, stratification, grappes, tailles d'échantillon, non-réponse). Soit :

$$\hat{\theta}_{b,c}^*$$

l'estimateur calculé sur le répliat (b, c) . La variance estimée à partir de la pseudo-population c est :

$$\widehat{\text{Var}}_c(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_{b,c}^* - \bar{\theta}_c^* \right)^2,$$

avec

$$\bar{\theta}_c^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{b,c}^*.$$

La variance bootstrap globale est obtenue par moyenne sur les pseudo-populations :

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{C} \sum_{c=1}^C \widehat{\text{Var}}_c(\hat{\theta}).$$

3.2.4 Validité asymptotique

Les propriétés asymptotiques des méthodes par pseudo-population ont été étudiées dans la littérature, en particulier pour des plans à forte entropie tels que le tirage de Poisson ou certains plans à probabilités inégales [Chauvet, 2007, chapitre 3]. Dans ces cadres, l'objectif est de montrer que la variance des estimateurs bootstrap reproduit asymptotiquement la variance sous le plan de sondage. Les résultats de Chauvet [2007] indiquent que l'algorithme proposé est consistant pour l'estimation de variance d'estimateurs par substitution dans le cas de plans à forte entropie, et qu'il peut être étendu à certains plans multi-degrés [Chauvet, 2007, p. 15]. Ces résultats reposent sur la capacité de la pseudo-population et du plan bootstrap associé à reproduire, au moins asymptotiquement, les principales composantes de variabilité du plan initial.

Il convient toutefois de souligner que ces propriétés sont établies dans des cadres asymptotiques spécifiques. En pratique, leur qualité d'approximation dépend notamment de la taille d'échantillon, de l'hétérogénéité des poids et de la complexité du plan de sondage.

3.2.5 Limitations

- La construction de la pseudo-population dépend du plan considéré et peut devenir délicate lorsque les poids sont très hétérogènes [Mashreghi et al., 2016, Section 4.1].

- Dans les plans complexes, la reproduction exacte des probabilités d’inclusion d’ordre supérieur peut seulement être approximative. Ceci motive notamment les développements proposés par [Chauvet \[2007\]](#) pour le cas multi-degrés, où la méthode initiale de [Gross \[1980\]](#) n’est pas directement consistante, c’est-à-dire qu’elle ne garantit pas que les estimations de variance obtenues par bootstrap convergent vers la vraie variance lorsque la taille de l’échantillon augmente [[Chauvet, 2007](#), p. 206].
- La construction explicite d’une pseudo-population peut entraîner un coût mémoire et computationnel important lorsque la population reconstruite est de grande taille.

3.3 Parametric Design Bootstrap (PDB)

Dans cette sous-section, nous nous limitons au cadre des estimateurs SAE de type EBLUP basés sur un modèle linéaire mixte (Linear Mixed Model, LMM). Le *parametric design bootstrap* (PDB) est une méthode proposée pour estimer directement la MSE sous le plan d’échantillonnage d’estimateurs SAE fondés sur un modèle, en particulier des EBLUPs [[Molina & Strzalkowska-Kominiak, 2020](#), Section 4.4, pp. 294–295]. L’idée est de conserver la stabilité apportée par le modèle tout en approchant l’espérance sous le plan de sondage par rééchantillonnage répété dans une population bootstrap fixe.

Contrairement au bootstrap paramétrique classique, qui génère une population à chaque réplication et cible la MSE sous le modèle, le PDB repose sur la génération d’une seule pseudo-population paramétrique à partir du modèle ajusté. On répète ensuite le plan d’échantillonnage dans cette population fixe afin d’approximer, par simulation de Monte Carlo, la variabilité sous le plan [[Molina & Strzalkowska-Kominiak, 2020](#), Section 4.4, pp. 294–295].

3.3.1 Principe général

Soit $\hat{\theta}_d$ un estimateur pour le domaine d , obtenu à partir de l’échantillon observé et de variables auxiliaires connues pour toute la population. La procédure PDB comporte les étapes suivantes :

1. ajuster le modèle sur les données observées ;
2. générer une pseudo-population paramétrique fixe U^* à partir du modèle ajusté ;
3. pour $b = 1, \dots, B$, tirer un échantillon bootstrap $s_d^{*(b)}$ dans U^* selon le plan d’échantillonnage initial ;
4. recalculer l’estimateur $\hat{\theta}_d^{*(b)}$ sur chaque échantillon bootstrap.

Ainsi, la variabilité bootstrap provient exclusivement de la répétition du sondage dans une pseudo-population fixée une seule fois.

3.3.2 Construction de la pseudo-population paramétrique

Dans le cadre d’un modèle linéaire mixte, nous avons :

$$Y_{di} = x'_{di}\beta + u_d + e_{di}, \quad i = 1, \dots, N_d, \quad d = 1, \dots, D,$$

où Y_{di} désigne la variable d’intérêt pour l’unité i du domaine d , x_{di} le vecteur des variables auxiliaires associées, β le vecteur des coefficients de régression, u_d l’effet aléatoire propre au domaine d et e_{di} le terme d’erreur individuel. On suppose généralement que

$$u_d \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_u^2), \quad e_{di} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e^2),$$

avec indépendance entre les effets aléatoires de domaine u_d et les erreurs individuelles e_{di} . Les paramètres σ_u^2 et σ_e^2 représentent respectivement la variance entre domaines et la variance

résiduelle au niveau des unités. Le PDB proposé par [Molina & Strzalkowska-Kominiak \[2020, Section 4.4, p. 295\]](#) consiste à ajuster d'abord le modèle sur l'échantillon observé, puis à générer une pseudo-population fixe conditionnellement aux effets de domaine estimés. Plus précisément, on génère

$$Y_{di}^* \sim \mathcal{N}(x'_{di}\hat{\beta} + \hat{u}_d, \hat{\sigma}_e^2), \quad i = 1, \dots, N_d, \quad d = 1, \dots, D.$$

Cette construction utilise l'ensemble des valeurs auxiliaires disponibles dans la population, contrairement aux méthodes non paramétriques par réplication de l'échantillon. Contrairement à un bootstrap paramétrique classique sous modèle mixte, les effets de domaine ne sont pas régénérés. La pseudo-population est construite conditionnellement aux effets estimés $\hat{u}_d$, qui sont considérés comme fixes dans la procédure PDB.

3.3.3 Motivation de la correction PDB

Sous l'hypothèse du modèle (en supposant que le modèle est vrai), on pourrait comparer directement $\hat{\theta}_d^{*(b)}$ à la valeur du paramètre dans la pseudo-population bootstrap. En pratique, la pseudo-population générée à partir du modèle ne coïncide pas exactement avec la population réelle. Pour éviter une dépendance trop forte au modèle, le PDB remplace donc la vraie valeur inconnue par une cible de comparaison qui dépend de la taille d'échantillon du domaine [[Molina & Strzalkowska-Kominiak, 2020, Section 4.4, p. 295](#)].

Plus précisément, pour les domaines de petite taille d'échantillon, la comparaison se fait avec l'estimateur SAE initial $\hat{\theta}_d$, plus stable. Pour les domaines avec de plus grands échantillons, elle se fait avec un estimateur plus direct, noté ici $\hat{\theta}_{d,\text{FIX}}$, obtenu en remplaçant l'effet aléatoire de domaine par un effet fixe, selon la construction retenue dans la méthode PDB [[Molina & Strzalkowska-Kominiak, 2020, Section 4.4, p. 295](#)].

3.3.4 Estimateur PDB de la MSE

L'estimateur PDB de la MSE sous le plan de $\hat{\theta}_d$ s'écrit alors

$$\widehat{\text{MSE}}_{\text{PDB}}(\hat{\theta}_d) = \hat{\gamma}_d \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_d^{*(b)} - \hat{\theta}_{d,\text{FIX}} \right)^2 + (1 - \hat{\gamma}_d) \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}_d^{*(b)} - \hat{\theta}_d \right)^2,$$

où $\hat{\gamma}_d$ est le coefficient de combinaison associé au domaine d .

Dans le cas particulier de l'EBLUP sous modèle linéaire mixte, $\hat{\gamma}_d$ est le même coefficient que celui intervenant dans l'EBLUP, à savoir

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_d}.$$

Ainsi, un compromis entre deux références est utilisé :

- une référence plus directe pour les domaines ayant une grande taille d'échantillon ;
- une référence plus stable, fondée sur l'estimateur SAE initial, pour les petits domaines.

3.3.5 Interprétation

Le PDB emprunte de la force entre domaines de manière analogue à l'EBLUP lui-même, mais cette fois pour l'estimation de la MSE. Lorsque n_d est grand, $\hat{\gamma}_d$ est proche de 1 et la composante plus directe domine. Lorsque n_d est petit, $\hat{\gamma}_d$ est proche de 0 et l'on privilégie la composante plus stable fondée sur $\hat{\theta}_d$ [[Molina & Strzalkowska-Kominiak, 2020, Section 4.4, p. 295](#)].

3.3.6 Propriétés et intérêt pratique

Les résultats théoriques présentés [Molina & Strzalkowska-Kominiak, 2020, matériel supplémentaire] montrent de bonnes propriétés du PDB dans le cadre d'un modèle linéaire mixte sous tirage aléatoire simple sans remise à l'intérieur des domaines. En particulier, les auteurs montrent que, sous les hypothèses du modèle, les MSE sous le modèle et sous le plan présentent un comportement similaire lorsqu'elles sont moyennées sur un grand nombre de domaines. Le PDB hérite ainsi d'une partie des propriétés favorables des approches fondées sur le modèle tout en ciblant directement la MSE sous le plan. Dans les études de simulation, le PDB reproduit mieux les pics de la MSE sous le plan que le bootstrap paramétrique classique, tout en restant plus stable que le bootstrap non paramétrique dans les petits domaines [Molina & Strzalkowska-Kominiak, 2020, Section 4.4, pp. 294–295].

3.3.7 Limitations

Le PDB reste toutefois dépendant du modèle utilisé pour générer la pseudo-population initiale. Par ailleurs, les études de simulation réalisées dans le cadre du projet SAE-SE ont montré que cette procédure fournit des résultats satisfaisants essentiellement pour des domaines dont la taille d'échantillon est d'au moins de l'ordre de 100 observations. Cette dépendance à une taille d'échantillon relativement élevée constitue une limitation importante pour les applications portant sur de très petits domaines.

3.4 Méthodes de réplcation par poids

Les méthodes de réplcation par poids constituent une alternative aux approches fondées sur le rééchantillonnage explicite ou sur la construction d'une pseudo-population. Elles reposent sur la génération de jeux de poids réplqués appliqués aux unités de l'échantillon observé. Les estimateurs sont ensuite recalculés à partir de ces poids afin d'évaluer leur variabilité.

Ces approches sont particulièrement attractives en statistique officielle lorsqu'il est souhaitable de diffuser des poids réplqués aux utilisateurs sans transmettre certaines informations auxiliaires utilisées lors de la pondération ou du calage.

3.4.1 Méthodes structurées

Certaines méthodes de réplcation reposent sur une construction déterministe ou semi-déterministe des poids réplqués.

Balanced Repeated Replication (BRR) La méthode du *Balanced Repeated Replication* (BRR) repose sur la construction de demi-échantillons équilibrés. Elle a été initialement développée pour des plans stratifiés comportant deux unités primaires d'échantillonnage par strate. Elle peut se généraliser pour d'autres plans grâce à la création de *Variance strata* et *Variance unit*, chaque *Variance strata* comportant deux *Variance unit* qui sont des regroupements d'unités. À chaque réplcation, une *Variance unit* est sélectionnée dans chaque strate selon un schéma équilibré, généralement fondé sur des matrices de Hadamard. Les poids des unités retenues sont ajustés afin de préserver les propriétés de l'estimateur.

Méthode de Fay La méthode de Fay constitue une variante du BRR. Au lieu de supprimer complètement certaines unités primaires dans chaque réplcat, elle applique un facteur d'atténuation aux poids, ce qui permet d'obtenir des estimations de variance plus stables.

Jackknife La méthode jackknife repose sur la suppression successive d'unités primaires ou de groupes d'unités. À chaque réplication, une partie de l'échantillon est retirée et les poids des unités restantes sont ajustés afin de préserver les propriétés de l'estimateur. Le jackknife est particulièrement adapté aux plans complexes et constitue une méthode standard dans de nombreux logiciels statistiques.

3.4.2 Méthodes bootstrap par génération de poids

Les méthodes bootstrap par génération de poids reposent sur la construction de facteurs aléatoires appliqués aux poids de l'échantillon observé.

Soit w_i le poids final associé à l'unité $i \in s$. Les poids bootstrap sont définis par

$$w_i^{*(b)} = a_i^{*(b)} w_i, \quad b = 1, \dots, B,$$

où les facteurs $a_i^{*(b)}$ sont des variables aléatoires construites de manière à reproduire certaines propriétés du plan d'échantillonnage.

Dans le cadre du *generalized bootstrap*, ces facteurs satisfont les conditions

$$E^*(a_i^*) = 1, \quad (1)$$

et

$$E^*[(a_i^* - 1)(a_j^* - 1)] = \frac{\Delta_{ij}}{\pi_{ij}}, \quad \Delta_{ij} = \pi_{ij} - \pi_i \pi_j. \quad (2)$$

Ces conditions permettent, pour des estimateurs linéaires tels que le total d'Horvitz–Thompson, une cohérence avec les estimateurs de variance classiques sous le plan [Rao & Wu, 1988, Rao, Wu & Yue, 1992, Beaumont et Patak, 2012].

Exemples de génération des facteurs bootstrap Les facteurs a_i^* peuvent être générés selon différentes constructions satisfaisant les conditions (1) et (2) citées plus haut.

Distribution Gamma : Dans le cas du sondage de Poisson, une construction simple consiste à générer indépendamment les facteurs bootstrap selon

$$a_i^* \stackrel{ind}{\sim} \Gamma\left(\frac{1}{1 - \pi_i}, 1 - \pi_i\right),$$

où $1 - \pi_i$ est le paramètre d'échelle. Cette construction vérifie

$$E(a_i^*) = 1, \quad \text{Var}(a_i^*) = 1 - \pi_i,$$

tout en garantissant des facteurs strictement positifs [Beaumont et Patak, 2012].

Construction binomiale : Dans le cas du sondage de Poisson, une construction consiste à générer indépendamment les facteurs bootstrap selon

$$a_i^* = \text{Bin}\left(\left\lfloor \frac{1}{\pi_i} \right\rfloor, \pi_i\right) + \alpha_i \text{Bin}(1, p_i),$$

avec

$$\alpha_i = (1 - \pi_i) + \left(1 - \left\lfloor \frac{1}{\pi_i} \right\rfloor \pi_i\right),$$

et

$$p_i = \frac{1 - \left\lfloor \frac{1}{\pi_i} \right\rfloor \pi_i}{\alpha_i}.$$

Dans ce cas particulier, cette construction est proche de l'approche par pseudo-population (équivalente quand les poids sont entiers), sans construire explicitement des pseudo-populations [Beaumont et Patak, 2012].

Construction Bernoulli : Dans le cas du sondage de Poisson, les facteurs bootstrap peuvent être générés selon (voir [Antal & Tillé, 2011])

$$a_i = A_i + 2(1 - A_i)B_i,$$

où

$$A_i \stackrel{ind}{\sim} \text{Bernoulli}(\pi_i),$$

et

$$B_i \stackrel{iid}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right).$$

Ces trois constructions présentent l'avantage de produire des facteurs bootstrap positifs. Dans le cas particulier des plans à sélections indépendantes, tels que le sondage de Poisson, les indicatrices de sélection sont indépendantes et les facteurs bootstrap peuvent être générés indépendamment, ce qui simplifie considérablement la mise en œuvre.

3.4.3 Estimation de la variance

À partir des B jeux de poids répliqués, on calcule les estimateurs bootstrap $\hat{\theta}_b^*$. Une forme courante de l'estimateur de variance est

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

Pour les méthodes structurées (BRR, Fay ou jackknife), des facteurs d'ajustement spécifiques doivent être appliqués conformément à la méthode utilisée [Wolter, 2007].

Dans le cas d'estimateurs non linéaires, cette estimation peut être interprétée comme une approximation basée sur une linéarisation implicite [Beaumont et Patak, 2012, Section 5].

3.4.4 Propriétés et avantages

Cette approche présente plusieurs avantages :

- elle ne nécessite pas la construction explicite d'une pseudo-population;
- elle est généralement efficace sur le plan computationnel;
- elle s'adapte naturellement à la diffusion de poids répliqués;
- elle est compatible avec de nombreux estimateurs complexes utilisés en statistique officielle.

3.4.5 Limitations

La méthode comporte toutefois certaines limites :

- la construction des facteurs a_i^* dépend directement du plan d'échantillonnage;
- la mise en oeuvre peut devenir complexe pour certains plans contraints, notamment les plans à taille fixe ou les plans à plusieurs degrés;
- les propriétés exactes sont principalement établies pour les estimateurs linéaires;
- certaines méthodes nécessitent la connaissance des probabilités d'inclusion d'ordre deux;
- certaines formulations très générales du generalized bootstrap peuvent produire des facteurs bootstrap négatifs et donc des poids répliqués négatifs. Les constructions fondées sur les lois Gamma, binomiale ou sur la méthode d'Antal et Tillé permettent au contraire de garantir des poids bootstrap positifs.

3.5 Intervalles de confiance

L'une des principales applications du bootstrap est la construction d'intervalles de confiance (IC) pour un paramètre estimé à partir d'un échantillon. L'idée consiste à exploiter la distribution empirique des répliqués bootstrap pour approcher la distribution de l'estimateur [Davison & Hinkley, 1997, Chapitre 2].

Dans le cadre des enquêtes par échantillonnage, lorsque la méthode bootstrap est adaptée au plan de sondage, les répliqués bootstrap permettent généralement d'approcher certaines caractéristiques de la distribution sous le plan de l'estimateur, en particulier sa variance et, dans certains cas, sa distribution asymptotique [Mashreghi et al., 2016, Section 4]. Cette propriété permet de construire des intervalles de confiance sans recourir à des approximations analytiques, souvent difficiles à établir dans des plans complexes.

On distingue principalement deux approches : l'approximation normale et les intervalles basés sur les quantiles empiriques.

3.5.1 IC par approximation normale

Lorsque l'estimateur satisfait une propriété de normalité asymptotique, un intervalle de confiance à niveau $1 - \alpha$ peut être construit comme :

$$IC_{1-\alpha} = \left[\hat{\theta} - z_{1-\alpha/2} \widehat{SE}, \hat{\theta} + z_{1-\alpha/2} \widehat{SE} \right],$$

où

$$\widehat{SE} = \sqrt{\widehat{\text{Var}}(\hat{\theta})}.$$

Cette approche correspond à l'intervalle de type Wald. Elle est simple à mettre en oeuvre, mais repose sur une approximation normale qui peut être inadéquate lorsque la distribution de l'estimateur est asymétrique ou lorsque l'estimateur est biaisé [Davison & Hinkley, 1997, Chapitre 5].

Elle est généralement adaptée lorsque la distribution des répliqués bootstrap est approximativement symétrique et que la taille d'échantillon est suffisamment grande.

3.5.2 IC par quantiles empiriques (percentile)

Une alternative consiste à utiliser directement les quantiles empiriques des répliqués bootstrap :

$$IC_{1-\alpha} = \left[q_{\alpha/2}(\hat{\theta}_b^*), q_{1-\alpha/2}(\hat{\theta}_b^*) \right].$$

Cette approche, appelée intervalle percentile, ne repose pas sur une hypothèse de normalité et permet de mieux capturer l'asymétrie de la distribution de l'estimateur [Davison & Hinkley, 1997, Chapitre 5, Sections 5.2–5.4].

Dans le contexte des enquêtes, cette méthode doit toutefois être utilisée avec précaution. Lorsque le bootstrap repose sur plusieurs pseudo-populations, la construction d'intervalles percentile pourrait demander une attention particulière, cf. [Mashreghi et al., 2016, Section 4.1.2].

Cette approche est particulièrement pertinente lorsque la distribution bootstrap présente une asymétrie marquée. Elle peut toutefois devenir instable lorsque la taille d'échantillon est faible ou lorsque le nombre de répliquats bootstrap est insuffisant. Un nombre suffisamment élevé de répliquats est nécessaire pour stabiliser l'estimation des quantiles. En pratique, les intervalles percentiles reposent souvent sur un nombre de répliquations supérieur à celui requis pour la seule estimation de la variance [Davison & Hinkley, 1997, Chapitre 5].

Par ailleurs, en présence de poids fortement variables ou dans des domaines de petite taille, la distribution des répliquats peut être instable, ce qui peut affecter la qualité des intervalles, en particulier pour les méthodes basées sur les quantiles.

Il convient toutefois de noter que les intervalles percentiles supposent que les répliquats bootstrap fournissent une approximation raisonnable de l'ensemble de la distribution de l'estimateur. Cette hypothèse est généralement plus naturelle dans les approches fondées sur le rééchantillonnage direct ou sur les pseudo-populations. En revanche, certaines méthodes de répliquat par poids sont principalement construites pour estimer la variance. Dans ce cas, l'utilisation directe des quantiles empiriques des répliquats doit être considérée avec prudence et nécessite une validation spécifique.

3.6 Différents aspects concernant la mise en œuvre

Dans les enquêtes par échantillonnage, les poids finaux $w_i^{(cal)}$ résultent de plusieurs étapes successives : pondération de sondage, correction de la non-réponse et calage sur des totaux connus de la population.

La question méthodologique centrale est de déterminer à quel niveau de pondération le bootstrap doit intervenir et quelles étapes doivent être reproduites dans chaque répliquat afin que la variance estimée corresponde bien à celle de l'estimateur effectivement publié.

3.6.1 Principe général

Le principe général du bootstrap en sondage consiste à reproduire, autant que possible, l'ensemble des étapes ayant conduit à l'estimateur final. Autrement dit, le bootstrap doit être appliqué à l'estimateur complet (incluant les étapes de correction et d'ajustement), selon une approche de type « plug-in ».

Dans le cas contraire, la variance estimée correspondrait à celle d'un estimateur intermédiaire (non-calé ou partiellement ajusté), et non à celle de l'estimateur diffusé.

Comme le souligne Davison & Sardy [2007, p. 8], les méthodes de rééchantillonnage peuvent être appliquées directement à un estimateur complexe (incluant imputation et calage), à condition de reproduire l'ensemble des étapes dans chaque répliquat.

3.6.2 Prise en compte du calage

Le calage constitue un élément central des estimateurs utilisés en statistique officielle. Lorsque l'estimateur final repose sur des poids calés, il est en général nécessaire de reproduire le calage dans chaque répliquat bootstrap.

Dans le cadre des méthodes par pseudo-population, [Chauvet \[2007, p. 177\]](#) montre que le calage peut être pris en compte en redressant chaque réplikat sur la pseudo-population correspondante, suivant une approche similaire à celle proposée par [Canty & Davison \[1999\]](#). Cette procédure permet d'intégrer correctement l'effet du calage dans l'estimation de la variance.

Plus généralement, [Chauvet \[2007, p. 178\]](#) montre que la réduction de variance induite par le calage est bien prise en compte dans le cadre du bootstrap, à condition que celui-ci soit reproduit à chaque étape selon une approche de type plug-in.

Ne pas reproduire le calage conduit à estimer la variance d'un estimateur non-calé. Si le calage a un effet important sur les poids alors des écarts significatifs par rapport à la variance de l'estimateur effectivement publié peuvent apparaître.

3.6.3 Prise en compte de la non-réponse

Les enquêtes par échantillonnage sont généralement affectées par des phénomènes de non-réponse, partielle ou totale, qui modifient la structure de l'échantillon initial [\[Chauvet, 2007, p. 42\]](#). Cette non-réponse est habituellement corrigée par pondération pour la non-réponse totale ou par imputation pour la non-réponse partielle.

Une approche consiste à utiliser directement les poids corrigés de la non-réponse pour la génération des réplikats bootstrap (construction des pseudo-populations, tirage des échantillons bootstrap, génération de poids bootstrap, etc.). Une approche plus complète, mais plus coûteuse, consiste à reproduire explicitement le mécanisme de non-réponse dans chaque réplikat. L'estimation de variance obtenue dépend alors du mécanisme de non-réponse retenu et de la manière dont celui-ci est modélisé à partir de l'échantillon observé.

3.6.4 Prise en compte de l'imputation

Lorsque des méthodes d'imputation sont utilisées, l'estimation de la variance devrait aussi intégrer cette étape.

Une alternative plus simple consiste à considérer l'estimateur imputé comme un estimateur standard et à appliquer directement le bootstrap. Cette approche peut toutefois sous-estimer la variance lorsque l'incertitude liée à l'imputation n'est pas correctement prise en compte. Plus généralement, l'intégration correcte de l'imputation dans une procédure bootstrap est souvent délicate, car le modèle d'imputation est lui-même estimé à partir des données observées. La variabilité associée à cette étape n'est pas nécessairement capturée lorsque le mécanisme d'imputation n'est pas reproduit dans les réplikats [\[Davison & Sardy, 2007, p. 8\]](#).

3.6.5 Diagnostics des réplikats bootstrap

Au-delà de l'estimation de la variance, l'analyse des réplikats bootstrap constitue un outil de diagnostic utile. L'examen de leur distribution permet notamment d'identifier des réplikats fortement influents ou des comportements atypiques liés à certaines unités de grande taille. De telles situations sont fréquentes dans les enquêtes auprès des entreprises ou dans les domaines de petite taille et méritent une analyse spécifique afin de distinguer les réalisations plausibles du plan d'échantillonnage d'éventuels problèmes de mise en œuvre de la procédure bootstrap.

3.6.6 Recommandations opérationnelles à l'OFS

Dans le contexte des enquêtes de l'OFS listées dans la section [4](#), la plupart des pratiques suivantes ont été retenues :

1. utiliser les poids corrigés pour la non-réponse pour la construction des pseudo-populations, tirage des échantillons bootstrap, génération de poids bootstrap ;
2. reproduire le calage dans chaque réplicat lorsque l'estimateur final repose sur des poids calés pour prendre en compte le gain dû au calage, notamment si le calage a un effet important sur les poids ;
3. dans le cas des méthodes par pseudo-population, effectuer le calage sur les totaux de la pseudo-population correspondante. Dans certains contextes, il peut également être pertinent d'évaluer l'intérêt d'un calage sur les totaux de la vraie population plutôt que sur ceux de la pseudo-population, notamment lorsque la construction de cette dernière introduit une variabilité non négligeable ;
4. documenter précisément les étapes reproduites (non-réponse, imputation, calage) ainsi que la version des poids utilisée ;
5. contrôler que la moyenne des estimations bootstrap reste suffisamment proche de l'estimation observée, un écart important peut constituer un signal d'alerte concernant le nombre de réplifications ou la validité pratique de la procédure bootstrap ;
6. contrôler la présence de réplicats bootstrap fortement influents et documenter leur impact sur l'estimation de la variance ;
7. dans le cas des estimateurs SAE reposant sur une prédiction pour les unités non échantillonnées, il convient de définir explicitement la population de référence utilisée pour les prédictions dans les réplicats bootstrap. L'expérience acquise dans plusieurs projets de l'OFS a montré que ce choix méthodologique peut avoir un impact non négligeable sur les résultats obtenus.

3.6.7 Reproductibilité

Pour garantir la reproductibilité des résultats, il est recommandé de dissocier l'étape de génération des réplicats de l'étape d'estimation, et de fixer une graine (seed) aléatoire.

Selon la méthode retenue, cela implique de conserver :

- soit les indices des unités sélectionnées dans chaque réplicat ;
- soit les facteurs multiplicatifs utilisés pour générer les poids bootstrap ;
- soit les réalisations simulées dans le cas d'un bootstrap paramétrique.

Cette séparation permet de reproduire exactement les estimations obtenues, indépendamment de l'environnement informatique, et évite de devoir recalculer l'ensemble des réplicats pour chaque nouvelle estimation.

4 Application du bootstrap à l'OFS

Afin de faciliter la lecture de cette section et de situer clairement les différences méthodologiques entre les projets menés à l'OFS, le tableau 1 présente une synthèse des approches bootstrap mises en œuvre, en distinguant le type de méthode, la nature des estimateurs ainsi que les principales difficultés rencontrées.

TABLE 1 Comparaison synthétique des projets bootstrap de l'OFS

Projet	Méthode bootstrap	Estimateur	Données	Difficultés principales
SAE-SE	Paramétrique de type PDB	EBLUP, modèle linéaire mixte	RS 2012–2016, AVS	Très petits domaines; forte hétérogénéité des tailles d'échantillon; sensibilité au modèle.
SAE-BESTA	Pseudo-populations non paramétriques	Estimateurs SAE fondés sur un modèle linéaire mixte	BESTA, registre des entreprises	Plan en grappes; probabilités d'inclusion inégales; hétérogénéité des établissements; domaines cantonaux et NOGA.
STATENT	Pseudo-populations non paramétriques	Estimateur complexe multi-sources des EPT	AVS, registres, BESTA, ERST et autres enquêtes	Processus de production complexe; grandes entreprises influentes; calage et harmonisation multi-sources.
PIAAC	Réplication des poids, méthode de Fay	Moyennes, scores et autres indicateurs	Enquête PIAAC	Standards OCDE; poids répliqués; variables auxiliaires sensibles non transmises.

Les sous-sections suivantes présentent plus en détail chacun de ces projets, en mettant l'accent sur les choix méthodologiques retenus et les principaux enseignements tirés de leur mise en œuvre.

4.1 Projet SAE-SE

Le Relevé structurel ¹ (RS) fournit chaque année des informations détaillées sur la population résidente permanente âgée de 15 ans ou plus. Avec un échantillon d'environ 200 000 personnes, il permet des estimations directes fiables pour les grandes communes. En revanche, pour la majorité des communes, la taille des échantillons est souvent trop faible pour se baser sur l'estimateur direct.

Dans ce contexte, l'OFS a collaboré avec la professeure Isabel Molina afin d'évaluer l'utilisation de méthodes d'estimation sur petits domaines (SAE) pour la production du taux d'activité au niveau communal.

Les travaux reposent sur les données du RS appariées aux données administratives AVS de l'année précédente, et s'appuient sur :

- un modèle linéaire à effets mixtes utilisant des variables auxiliaires démographiques et structurelles;
- un bootstrap paramétrique de type *parametric design bootstrap* (PDB), permettant d'estimer la précision des prédictions SAE.

Le modèle SAE utilisé suit la structure classique de l'EBLUP, dans laquelle l'estimation pour un domaine d combine un estimateur direct fondé sur les données observées du domaine et une composante synthétique issue du modèle et des variables auxiliaires. La combinaison relative entre ces deux composantes dépend de la taille d'échantillon n_d du domaine : pour les petits domaines, la composante modélisée domine, tandis que l'estimateur direct prend davantage de poids lorsque n_d augmente.

La mesure de précision retenue dans ce projet est l'erreur quadratique moyenne sous le plan de sondage (design-MSE), qui combine les effets de la variance sous le plan d'échantillonnage et des erreurs induites par la modélisation [Dupraz et al., 2018, p. 5]. L'estimation de cette quantité est réalisée à l'aide d'un bootstrap paramétrique de type PDB, dont les performances se sont révélées satisfaisantes principalement pour des domaines dont la taille d'échantillon est de l'ordre de $n_d \geq 100$ (voir section 3.3).

1. <https://www.bfs.admin.ch/bfs/fr/home/bases-statistiques/recensement-population/quatre-elements-cles/releve-structurel.html>

Afin d'améliorer la qualité des estimations de la MSE, un pooling temporel des données du RS est mis en œuvre, consistant à regrouper plusieurs années d'échantillons en un seul jeu de données. Cette approche permet d'augmenter la taille effective d'échantillon au niveau communal.

Les analyses empiriques montrent que le pooling constitue un levier essentiel pour l'extension du champ de publication. Avec un pooling de trois ans, environ 60 % des communes atteignent les critères de publication, couvrant environ 94 % de la population cible [Meskaldji et al., 2021, p. 2]. Ce pourcentage augmente à environ 76 % des communes et 98 % de la population avec un pooling de cinq ans [Meskaldji et al., 2021, p. 12].

Sous l'hypothèse que l'erreur est principalement due à la variance d'échantillonnage et que les informations apportées par les différentes années sont indépendantes, le pooling permet une réduction théorique de la RMSE pouvant atteindre $1 - \sqrt{1/T}$ pour un pooling sur T années [Meskaldji et al., 2021, p. 11]. En pratique, ce gain peut être plus faible lorsque d'autres sources d'erreur, notamment liées à la modélisation, demeurent relativement constantes au cours du temps. Toutefois, les analyses montrent que l'amélioration effective de la précision reste modérée et que l'estimation de la RMSE par la méthode PDB peut présenter une certaine instabilité [Meskaldji et al., 2021, p. 12].

Ainsi, l'expérience du projet SAE-SE montre que le pooling constitue un élément déterminant pour rendre l'approche SAE opérationnelle au niveau communal. Il améliore la stabilité des estimations et permet d'augmenter de manière conséquente le nombre de communes publiables.

4.2 Projet SAE-BESTA

La Statistique de l'emploi² (BESTA) est une enquête trimestrielle dont les origines remontent à 1925 et qui couvre les secteurs secondaire et tertiaire. Elle repose sur un échantillon d'environ 18 000 entreprises, représentant environ 65 000 établissements. La BESTA fournit des estimations du nombre d'emplois et d'équivalents plein temps, ventilées par sexe et par branche d'activité (Nomenclature générale de l'activité : NOGA).

L'échantillon est tiré selon un plan de sondage en grappes, les établissements étant regroupés au sein des entreprises. Le tirage des entreprises repose sur un plan de Poisson à probabilités d'inclusion inégales, défini à partir du cadre de sondage de l'OFS basé sur le Registre des Entreprises et des Établissements (REE). Les informations sur les unités échantillonnées ainsi que celles du cadre de calage sont actualisées trimestriellement.

Dans plusieurs domaines, notamment pour certains cantons et regroupements NOGA2 (NOGA à deux chiffres), les tailles d'échantillon sont trop faibles pour assurer une précision satisfaisante des estimations directes. C'est dans ce contexte qu'a été lancé le projet pilote d'innovation SAE-BESTA, dont l'objectif principal était d'évaluer le potentiel des méthodes d'estimation sur petits domaines pour produire des estimations cantonales de l'emploi total, ventilées par sexe. Un objectif complémentaire consistait à étendre l'analyse aux branches économiques NOGA2 [Massiani et al., 2024].

L'approche SAE utilisée repose sur un modèle linéaire à effets mixtes, dans lequel l'emploi relevé dans l'enquête est modélisé en fonction de variables auxiliaires connues pour toutes les unités de la population. Les variables auxiliaires utilisées incluent la grande région, la section d'activité économique, ainsi que l'emploi issu du cadre de sondage. Cette variable peut différer de l'emploi relevé lors de l'enquête en raison de décalages temporels ou d'effets de mesure [Massiani et al., 2024, Saucy et al., 2024].

Le modèle comporte des effets aléatoires pour les cantons et pour les groupes NOGA2. Il est ajusté sur une série temporelle de 20 trimestres couvrant la période du deuxième trimestre 2015 au quatrième trimestre 2019.

2. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/industrie-services/enquetes/besta.html>

La variance des estimations SAE est évaluée au moyen d'un bootstrap non paramétrique basé sur la méthode des pseudo-populations (voir section 3.2). Cette approche permet de reproduire la variabilité sous le plan de sondage en tenant compte de la structure du tirage en grappes, des probabilités d'inclusion et des principales étapes de pondération.

Les résultats montrent que l'approche SAE permet un gain substantiel en termes de réduction de variance, en particulier pour les petits cantons (cantons à faible taille d'échantillon), pour lesquels les estimations directes présentent souvent des coefficients de variation élevés [Massiani et al., 2024, Saucy et al., 2024]. Pour de nombreux cantons, les coefficients de variation des estimations SAE sont nettement plus faibles que ceux des estimations directes.

En revanche, pour les cantons disposant de tailles d'échantillon importantes, pour lesquels l'estimation directe est déjà précise, les gains sont plus limités. Pour les regroupements NOGA2, les résultats sont plus contrastés : certains présentent une amélioration marquée, tandis que d'autres montrent des réductions de variance plus modestes. Ces résultats restent globalement encourageants, mais ne permettent pas, dans l'état actuel du projet, une diffusion systématique de résultats détaillés pour l'ensemble de ces domaines.

Les résultats obtenus ont été jugés suffisamment robustes pour permettre la diffusion, sous forme de statistiques expérimentales, d'estimations SAE cantonales du total de l'emploi et de l'emploi ventilé par sexe³.

L'utilisation d'un bootstrap par pseudo-population était particulièrement adaptée dans ce contexte, car elle permettait de reproduire la complexité du plan d'échantillonnage de la BESTA sans nécessiter le développement de formules analytiques de variance difficiles à établir.

4.3 Projet STATENT

La STATENT⁴ se base principalement sur les données des registres de l'assurance vieillesse et survivants (AVS), ainsi que sur les informations contenues dans le registre des entreprises et des établissements (REE) de l'OFS. Ces données sont complétées par différentes enquêtes auprès des entreprises.

L'estimation des équivalents plein temps (EPT), décrite dans Nedyalkova & Assoulin [2024], repose sur la combinaison de plusieurs sources de données (registres et enquêtes) et comprend plusieurs étapes successives d'harmonisation, d'estimation et d'ajustement destinées à assurer la cohérence globale des résultats.

En raison de la complexité de cet estimateur, qui ne peut pas être exprimé sous une forme analytique simple, l'estimation de la variance par des approches classiques, telles que la linéarisation, est difficile à mettre en œuvre en pratique. L'OFS a donc retenu une approche bootstrap basée sur la méthode des pseudo-populations.

Dans ce cadre, des pseudo-populations sont construites en reproduisant les différentes sources de données et leurs mécanismes de pondération propres à chaque composante. En particulier, les répliqués utilisent des poids corrigés pour la non-réponse dans BESTA, tandis que les entreprises issues des enquêtes considérées comme exhaustives, telles que ERST, sont associées à des poids égaux à 1.

Cette approche permet d'obtenir une estimation empirique de la variance des estimations d'EPT, en tenant compte de l'ensemble du processus de production statistique.

L'estimation de la variance est réalisée à plusieurs niveaux d'agrégation (national, domaines d'activité économique OFS50 et régions NUTS II). Les résultats montrent que la variabilité diffère sensiblement selon les secteurs économiques et les niveaux d'agrégation. Cette hétérogénéité reflète à la fois les différences de qualité des sources utilisées et l'impact des différentes étapes d'estimation et d'ajustement intervenant dans la production des EPT.

3. <https://www.bfs.admin.ch/bfs/fr/home/expstat/projets/jobstat.html>

4. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/industrie-services/enquetes/statent.html>

Cette hétérogénéité confirme la nécessité d'une approche empirique de l'estimation de la variance, adaptée à la complexité du processus de production de la STATENT, dans lequel les modèles d'estimation des EPT et les ajustements successifs influencent directement la variabilité finale des résultats. De plus, la présence d'entreprises de très grande taille peut également conduire à des répliqués fortement influents, ce qui justifie un examen attentif de la distribution des estimations bootstrap.

4.4 Projet PIAAC

Le programme PIAAC ⁵ (Programme for the International Assessment of Adult Competencies) est une étude internationale conduite par l'Organisation de coopération et de développement économiques (OCDE), visant à évaluer les compétences de base des adultes.

L'enquête PIAAC fournit une vue d'ensemble des niveaux de compétences en littéracie, numéracie, et en résolution adaptative de problèmes parmi la population résidente permanente âgée de 16 à 65 ans, tout en permettant des comparaisons internationales. En Suisse, l'enquête est réalisée auprès d'un échantillon brut d'environ 22 000 personnes, avec un taux de réponse d'environ 30 %. Compte tenu des exigences de comparabilité internationale, une approche spécifique a été retenue pour l'estimation de la variance.

Les méthodes utilisées suivent les recommandations de l'OCDE [OECD, 2022a] et sont décrites dans le rapport de méthodes de Bianchi et al. [2025]. L'estimation de la variance repose sur des méthodes de réplification par poids, et plus précisément sur une variante de la méthode BRR (Balanced Repeated Replication) connue sous le nom de méthode de Fay (voir section 3.4).

Dans ce cadre, un ensemble de poids répliqués est fourni aux utilisateurs. Les utilisateurs peuvent ainsi estimer la variance de n'importe quel paramètre en recalculant l'estimateur avec chaque jeu de poids, sans avoir accès aux variables auxiliaires utilisées dans la construction des poids finaux.

Les poids répliqués sont construits selon la méthode de Fay, qui repose sur une perturbation contrôlée des poids de sondage à l'aide d'un facteur d'atténuation. La formule de variance associée s'écrit :

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{(1-K)^2} \cdot \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta})^2,$$

où $\hat{\theta}_b^*$ est l'estimation obtenue avec le b -ième jeu de poids répliqués, B le nombre total de réplifications, et K un facteur d'atténuation fixé selon les recommandations de l'OCDE ($K = 0.3$ dans PIAAC) [OECD, 2022a]. Les utilisateurs n'ont donc pas besoin de reproduire les différentes étapes de pondération ou de calage; celles-ci sont déjà intégrées dans les poids répliqués fournis.

La pondération PIAAC inclut notamment une correction pour l'éligibilité inconnue parmi les non-répondants, des ajustements de non-réponse ainsi qu'un calage sur des totaux issus du Relevé structurel et d'autres sources administratives.

Cette approche présente plusieurs avantages :

- elle fournit aux utilisateurs un mécanisme simple et standardisé pour l'estimation de la variance;
- elle assure la cohérence avec les standards internationaux définis par l'OCDE;
- elle évite aux utilisateurs d'avoir à reproduire les étapes complexes de pondération et de calage;
- elle permet de limiter la diffusion des variables auxiliaires utilisées dans la pondération.

5. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/education-science/enquetes/piaac.html>

5 Conclusions

Ce rapport présente une synthèse des méthodes de bootstrap appliquées au contexte des statistiques officielles de l'OFS, avec un accent particulier sur l'estimation de la variance d'estimateurs complexes. Après un rappel des fondements théoriques et des principales adaptations du bootstrap aux enquêtes par échantillonnage, plusieurs applications concrètes ont été présentées, illustrant la diversité des approches mises en œuvre : méthodes non-paramétriques par pseudo-population, méthodes par génération de poids, ainsi que bootstrap paramétrique dans un cadre de modèle d'estimation sur petits domaines (SAE).

Sur la base des projets SAE-SE, SAE-BESTA, STATENT et PIAAC, plusieurs constats peuvent être mis en avant :

- Le bootstrap constitue une solution opérationnelle pour l'estimation de la variance, en particulier lorsque les formules analytiques sont difficiles à établir.
- Les projets analysés illustrent la diversité des méthodes de bootstrap pouvant être utilisées selon le contexte d'application, qu'il s'agisse d'estimation sur petits domaines, d'estimateurs complexes issus de plusieurs sources de données ou de contraintes liées à la diffusion des résultats.
- Les méthodes par pseudo-populations se révèlent particulièrement utiles lorsque l'estimateur repose sur un processus de production complexe ou sur des prédictions pour les unités non échantillonnées, comme dans certains projets SAE. Les approches par réplique de poids offrent quant à elles une mise en œuvre souvent plus simple et constituent une solution naturelle lorsqu'il est souhaitable de diffuser directement des poids répliqués aux utilisateurs.
- Le bootstrap paramétrique de type PDB permet d'intégrer la composante liée à la modélisation dans l'évaluation de la précision, mais reste sensible aux hypothèses du modèle.

Certaines limitations doivent néanmoins être soulignées. Les méthodes bootstrap peuvent entraîner un coût computationnel élevé, en particulier lorsque des étapes telles que l'imputation ou le calage doivent être reproduites dans chaque réplique. Par ailleurs, la qualité des estimations de variance obtenues dépend de la capacité du dispositif bootstrap à reproduire correctement la variabilité induite par le plan de sondage et, le cas échéant, par les mécanismes de modélisation. Les différents projets analysés ont également mis en évidence plusieurs choix méthodologiques qui ne sont pas toujours explicitement traités dans la littérature. Ceux-ci concernent notamment la construction des pseudo-populations, le choix des poids utilisés dans les répliques, la prise en compte du calage et de la non-réponse, le traitement des répliques fortement influents ainsi que la définition de la population de référence dans le cas des estimateurs SAE. Ces éléments montrent que la mise en œuvre pratique du bootstrap ne se limite pas à l'application d'un cadre théorique général, mais nécessite souvent des arbitrages méthodologiques ainsi qu'une validation empirique adaptés au contexte de production statistique considéré.

Ce rapport constitue une première synthèse structurée des pratiques de bootstrap à l'OFS. Au-delà de la présentation des méthodes existantes, il vise également à capitaliser l'expérience acquise dans plusieurs projets méthodologiques de l'OFS et à fournir un cadre cohérent pour les futurs développements en matière d'estimation de la variance, notamment dans les situations où une approche par bootstrap est jugée plus adaptée ou plus facile à mettre en œuvre qu'un développement analytique explicite.

Annexes

A Notations

Les notations suivantes sont utilisées dans l'ensemble du rapport.

Symbole	Définition
<i>Population et échantillon</i>	
U	Population finie de taille N
s	Échantillon observé de taille n
$U^*, U^{*(c)}$	Pseudo-population, éventuellement indexée par c
$s^{*(b)}$	Échantillon bootstrap (réplicat b)
$s^{*(b,c)}$	Échantillon bootstrap b tiré dans la pseudo-population c
<i>Plan de sondage et pondération</i>	
π_i	Probabilité d'inclusion de l'unité i
π_{ij}	Probabilité d'inclusion conjointe des unités i et j
$\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$	Covariance des indicatrices d'inclusion sous le plan
w_i	Poids associé à l'unité i (selon le contexte : poids de sondage, corrigé ou calé)
$w_i^{(0)} = 1/\pi_i$	Poids de sondage initial
$w_i^{(nr)}$	Poids après correction de la non-réponse
$w_i^{(cal)}$	Poids final après calage
a_i^*	Facteur multiplicatif utilisé pour générer les poids bootstrap
$w_i^* = a_i^* w_i$	Poids bootstrap
$m_i^{(c)}$	Nombre de duplications de l'unité i dans la pseudo-population c
<i>Estimateurs et réplcats</i>	
$\hat{T}_{HT}$	Estimateur de Horvitz–Thompson d'un total
$\hat{\theta}$	Estimateur du paramètre θ
$\hat{\theta}_b^*$	Estimateur bootstrap (réplicat b)
$\hat{\theta}_{b,c}^*$	Estimateur obtenu sur le réplcat b de la pseudo-population c
$\bar{\theta}^*$	Valeur de référence associée à une population bootstrap
$\bar{\theta}^*$	Moyenne des estimateurs bootstrap
<i>Mesures de précision</i>	
$\widehat{\text{Var}}(\hat{\theta})$	Estimation de la variance
$\widehat{\text{MSE}}(\hat{\theta})$	Estimation de l'erreur quadratique moyenne
$\widehat{\text{SE}}(\hat{\theta})$	Écart-type estimé
$\text{CV}(\hat{\theta})$	Coefficient de variation basé sur la variance
<i>Paramètres du bootstrap</i>	
B	Nombre de réplcats bootstrap
C	Nombre de pseudo-populations (bootstrap non-paramétrique par pseudo-populations)
<i>Modèles SAE</i>	
Y_{di}	Variable d'intérêt pour l'unité i du domaine d
$\mathbf{x}_{di}$	Vecteur de covariables associées à l'unité i du domaine d
d	Indice de domaine

β	Paramètres fixes
$\hat{\beta}$	Estimateur des paramètres fixes
u_d	Effet aléatoire de domaine, $u_d \sim \mathcal{N}(0, \sigma_u^2)$
e_{di}	Erreur individuelle, $e_{di} \sim \mathcal{N}(0, \sigma_e^2)$
u_d^*, e_{di}^*	Effets simulés dans le bootstrap
Y_{di}^*	Valeur simulée
σ_u^2	Variance inter-domaines
σ_e^2	Variance intra-domaines
$\hat{\theta}_{d,\text{FIX}}$	Estimateur direct/sous modèle à effet fixe utilisé comme référence dans la méthode PDB
$\hat{\gamma}_d$	Coefficient de combinaison de l'EBLUP pour le domaine d

Bibliographie

- Antal, E. & Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, 106(494), 534–543.
- Bianchi, B., Leuenberger, M., Massiani, A. & Panchard, C. (2025). *Enquête PIAAC 2022–2023 : Plan de sondage, pondération, estimation de variance et analyse du biais dû à la non-réponse*. Rapport de méthodes, OFS, Neuchâtel.
- Beaumont, J.-F., Haziza, D., & Bocci, C. (2008). On the use of the bootstrap with imputed data. *Survey Methodology*, 34(2), 215–226.
- Beaumont, J.-F. & Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review*, 80, 127–148.
- Canty, A. J. & Davison, A. C. (1999). Resampling-based variance estimation for labour force surveys. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 48(3), 379–391.
- Chauvet, G. (2007). *Méthodes de bootstrap en population finie*. Thèse de doctorat, ENSAI.
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- Davison, A. C. & Sardy, S. (2007). Resampling variance estimation in surveys with missing data. *Journal of Official Statistics*, 23, 371–386.
- Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376–382.
- Dippo, C. S., Fay, R. E., & Morganstein, D. (1984). *Computing Variances from Complex Samples Using Replication Techniques*. In : *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Dupraz, M., Massiani, A. & Kilchmann, D. (2018). *Estimation du taux d'activité au niveau communal dans le cadre du Relevé structurel : Étude de la combinaison des méthodes Small Area Estimation et du pooling*. Notice méthodologique, OFS, Neuchâtel.
- Efron, B. (1979). Bootstrap methods : Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
- Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, 91(434), 490–498.
- Ghosh, M. & Rao, J. N. K. (1994). Small area estimation : An appraisal. *Statistical Science*, 9(1), 55–93.
- Gross, S. (1980). Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.
- Lohr, S. L. (2010). *Sampling : Design and Analysis*. 2nd edition. Brooks/Cole, Boston.

- Mach, L., Saïdi, A. & Pettapiece, R. (2007). Study of the properties of the Rao–Wu bootstrap variance estimator for survey data. *Proceedings of the Survey Methods Section, Statistical Society of Canada*.
- Mashreghi, Z., Haziza, D., & Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10, 1–52.
- Massiani, A., Meskaldji, D.-E. & Kilchmann, D. (2024). *SAE-BESTA - Estimation sur petits domaines dans cadre de la Statistique de l'emploi*. Rapport de l'OFS, Neuchâtel.
- Meskaldji, D.-E., Massiani, A. & Kilchmann, D. (2021). *Analyse de l'impact du pooling : Estimation du taux d'activité au niveau communal dans le cadre du relevé structurel*. Notice méthodologique, OFS, Neuchâtel.
- Molina, I. & Strzalkowska-Kominiak, E. (2020). Estimation of proportions in small areas : application to the labour force using the Swiss census structural survey. *Journal of the Royal Statistical Society : Series A*, 183(1), 281–310.
- Molina, I. & Strzalkowska-Kominiak, E. (2020). Estimation of proportions in small areas : application to the labour force using the Swiss census structural survey. *Journal of the Royal Statistical Society : Series A*, Supplementary Material.
- Nedyalkova, D. & Assoulin, D. (2024). *STATENT – Estimation des EPT : Méthodologie, modèles et estimation de la variance*. Rapport de méthodes, OFS, Neuchâtel.
- OECD (2022a). *PIAAC Cycle 2 - Technical Standards and Guidelines*. OECD Publishing.
- OECD (2022b). *PIAAC Cycle 2 Sampling Plan, Part II : Weighting and Variance Estimation Plans*. Technical report.
- Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*. 2nd edition. Wiley.
- Rao, J. N. K. & Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401), 231–241.
- Saucy, F., Columbano, A., Kilchmann, D., Ferster, M., Massiani, A., Meskaldji, D.-E. & Schmassmann, S. (2024). *SAE-BESTA : approfondissement des estimations au niveau des cantons et analyse de la pertinence de l'extension aux branches économiques*. Rapport de l'OFS, Neuchâtel.
- Rao, J. N. K., Wu, C. F. J., & Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2), 209–217.
- Shao, J. & Sitter, R. R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 91(435), 1278–1288.
- Shao, J. (2003). Impact of the bootstrap on sample surveys. *Statistical Science*, 18(2), 191–198.
- Shao, J. & Tu, D. (2012). *The Jackknife and Bootstrap*. Springer.
- Sitter, R. R. (2001). Resampling methods in complex surveys : An overview in honour of J.N.K. Rao's retirement. *Proceedings of the Survey Methods Section, SSC Annual Meeting*, 9–14.
- Strzalkowska, E. & Molina, I. (2018). *Small Area Estimation in the Structural Survey*. Rapport de l'OFS, Neuchâtel.
- Wolter, K. M. (2007). *Introduction to Variance Estimation*. 2nd edition. Springer.

Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodenberichte der Sektion Statistische Methoden des BFS
Methodology reports published by the FSO's Statistical Methods Section

- Meskaldji, D.-E., Mariétan, R. (2026). Applications du bootstrap à l'OFS. Estimation de variance à l'aide de méthodes de rééchantillonnage. Numéro OFS : 338-0095
- Chevalier, C. (2025). Statistique sur les télécommunications. Plan de sondage. Numéro OFS : 338-0094
- Potterat, J. (2025). Erhebung leichte Nutzfahrzeuge (LWE) 2023. Stichprobenplan, Gewichtung und Schätzverfahren. Numéro OFS : 338-0093
- Chevalier, C. (2025). Statistique de la valeur ajoutée 2021. Révision de la pondération. Numéro OFS : 338-0092
- Bianchi, B., Leuenberger, M., Massiani, A., Panchard, C. (2025). Enquête PIAAC 2022–2023 : Plan de sondage, pondération, estimation de variance et analyse du biais dû à la non-réponse. Numéro OFS : 338-0091
- Meskaldji, D.-E., Panchard, C., Qualité, L., Saliba, J. (2024). Enquêtes thématiques 2018-2023. Plans de sondage, tirages, pondération et estimation de variance. Numéro OFS : 338-0090
- Chevalier, C. (2024). Statistique de la valeur ajoutée 2021. Révision du plan de sondage. Numéro OFS : 338-0089
- Nedyalkova, D., Assoulin, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalents. FSO number : 338-0088
- Qualité, L. (2023). Relevé structurel de la population suisse 2018. Révision de la pondération et sélections répétées. Numéro OFS : 338-0087
- Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0086
- Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS : 338-0085
- Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS : 338-0084
- Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS : 338-0083
- Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS : 338-0082
- Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS : 338-0081
- Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer : 338-0080
- Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015 : Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0079
- Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS : 338-0078
- Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics. FSO number : 338-0077
- Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. BFS-Nummer : 338-0076
- Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. BFS-Nummer : 338-0073

- Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009 : Statistische Datenaufbereitung und Hochrechnung. BFS-Nummer : 338-0072
- Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro OFS : 338-0071
- Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009 : Stichprobenrahmen und Stichprobenplan. BFS-Nummer : 338-0070
- Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0069
- Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0068
- Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO number : 338-0067-05
- Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0066
- Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0065
- Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. BFS-Nummer : 338-0064
- Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0063
- Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0062
- Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro OFS : 338-0061
- Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO number : 338-0060
- Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro OFS : 338-0059
- Graf, E. (2009). Weightings of the Swiss Household Panel : SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. FSO number : 338-0058
- Graf, E. (2009). Pondérations du Panel Suisse de Ménages : PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro OFS : 338-0057-05
- Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro OFS : 338-0056
- Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro OFS : 338-0055
- Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro OFS : 338-0054
- Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium". Numéro OFS : 338-0053
- Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro OFS : 338-0052
- Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro OFS : 338-0051
- Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer : 338-0050

- Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro OFS : 338-0049
- Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. BFS-Nummer : 338-0048
- Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. BFS-Nummer : 338-0047
- Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer : 338-0046
- Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro OFS : 338-0045
- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro OFS : 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. BFS-Nummer : 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0041
- Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro OFS : 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS : 338-0039
- Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample : Analysis and precision assessment of wage components. FSO number : 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer : 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer : 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer : 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer : 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS : 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS : 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer : 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer : 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number : 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethode der Lohnindizes und deren Varianzschätzer. BFS-Nummer : 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS : 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS : 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro OFS : 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS : 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. BFS-Nummer : 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer : 338-0020.

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro OFS : 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. BFS-Nummer : 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro OFS : 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. BFS-Nummer : 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro OFS : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. BFS-Nummer : 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro OFS : 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro OFS : 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer : 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro OFS : 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro OFS : 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. BFS-Nummer : 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro OFS : 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro OFS : 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro OFS : 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. BFS-Nummer : 338-0004

Programme des publications de l'OFS

En tant que service statistique central de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public. Il utilise à cet effet plusieurs moyens et canaux et diffuse ses informations statistiques par domaines thématiques.

Les domaines statistiques

- 00 Bases statistiques et généralités
- 01 Population
- 02 Espace et environnement
- 03 Travail et rémunération
- 04 Économie nationale
- 05 Prix
- 06 Industrie et services
- 07 Agriculture et sylviculture
- 08 Énergie
- 09 Construction et logement
- 10 Tourisme
- 11 Mobilité et transports
- 12 Monnaie, banques, assurances
- 13 Sécurité sociale
- 14 Santé
- 15 Éducation et science
- 16 Culture, médias, société de l'information, sport
- 17 Politique
- 18 Administration et finances publiques
- 19 Criminalité et droit pénal
- 20 Situation économique et sociale de la population
- 21 Développement durable, disparités régionales et internationales

Sélection de publications

Statistique: 175 ans au service de l'État fédéral moderne



Cette brochure a été éditée à l'occasion de la 10 000^e publication de l'Office fédéral de la statistique. Reprenant dix publications et événements appartenant au passé, elle retrace l'histoire de l'OFS, de ses publications et de la statistique en Suisse. La brochure compte 52 pages et est disponible en quatre langues (français, allemand, italien et anglais).

www.statistique.ch – sources d'informations des plus utiles

Le portail Statistique suisse est un outil moderne et attrayant qui vous donne accès aux informations statistiques actuelles. Nous attirons ci-après votre attention sur les offres les plus prisées.

Catalogue général



Presque tous les documents publiés par l'OFS depuis 1860 sont disponibles gratuitement sous forme électronique sur le portail Web Statistique suisse (www.statistique.ch). Les publications sur papier peuvent être commandées par téléphone au +41 58 463 60 60 ou par e-mail à l'adresse suivante order@bfs.admin.ch.

www.statistique.ch → Statistiques → Catalogue

Vous souhaitez être parmi les premiers informés?



Abonnez-vous à un Newsmail et vous recevrez par e-mail des informations sur les résultats les plus récents et les activités actuelles concernant le thème de votre choix.

www.news-stat.admin.ch

Aperçu des données publiées



Pour donner une meilleure vue d'ensemble de tous les jeux de données publiés en continu par l'OFS, nous les avons regroupés sur cette page. Elle contient également un aperçu des données de la statistique fédérale déjà accessibles sur la plateforme opendata.swiss ainsi que des données de la National Summary Data Page (NSDP).

www.data.bfs.admin.ch

Tableau de bord des votations Suisse



Le tableau de bord de l'OFS utilise les données publiques ouvertes (OGD) pour présenter rapidement les résultats des votations sous une forme visuelle et conviviale. Toutes les données sont librement accessibles et disponibles en cinq langues.

votations.admin.ch/fr/overview

Informations complémentaires

Service de renseignement

+41 58 463 60 11, info@bfs.admin.ch



www.statistique.ch



Facebook



Youtube



X



LinkedIn

Ce rapport de méthodes présente les principales approches de rééchantillonnage mises en œuvre à l'Office fédéral de la statistique (OFS) pour l'estimation de la variance dans le cadre d'enquêtes par échantillonnage. L'accent est mis sur l'utilisation du bootstrap pour estimer la variance des estimateurs lorsque les développements analytiques sont difficiles à obtenir. Après un rappel des principes du bootstrap et de ses variantes (paramétrique ou non-paramétrique), les spécificités de son application dans un contexte d'enquêtes basées sur un échantillon aléatoire sont décrites, notamment en tenant compte des poids d'échantillonnage, de la non-réponse et du calage sur des totaux connus de la population. Des applications concrètes sont ensuite présentées dans le cadre de différentes statistiques de l'OFS. Pour chacune de ces statistiques, les choix méthodologiques effectués, les résultats obtenus, ainsi que les défis rencontrés sont illustrés. L'ensemble du rapport vise à documenter les pratiques en matière d'estimation de variance à l'aide de méthodes de rééchantillonnage dans les statistiques de l'OFS, en tenant compte des contraintes opérationnelles et des exigences de qualité tout en capitalisant l'expérience acquise dans plusieurs projets méthodologiques de l'OFS.

En ligne

www.statistique.ch

Imprimés

www.statistique.ch
Office fédéral de la statistique
CH-2010 Neuchâtel
order@bfs.admin.ch
tél. +41 58 463 60 60

Numéro OFS

338-0095

ISBN

978-3-303-00787-7

Les informations publiées ici contribuent à mesurer
la réalisation des objectifs de développement durable (ODD).



Système d'indicateurs MONET 2030

www.statistique.ch → Statistiques → Développement durable
→ Système d'indicateurs MONET 2030