Assemblée annuelle de la SSC, Juin 2011
Recueil de la Section des méthodes d'enquête

DÉTERMINATION DE LA TAILLE DE L'ÉCHANTILLON POUR LE PROGRAMME INTÉGRÉ DE LA STATISTIQUE DES ENTREPRISES

N. Gouzi et A. Demnati [1]

RÉSUMÉ

La majorité des enquêtes-entreprises ont des objectifs multiples. Pour atteindre ces objectifs, nous regroupons la population selon chaque objectif. Une entreprise peut appartenir à plusieurs groupes. Nous considérons que chaque groupe est une base de sondage et nous utilisons la théorie des bases multiples. L'une des difficultés éprouvées avec cette approche est la détermination de la taille de l'échantillon. Nous explorons une solution de rechange centrée sur une base unique. Nous évaluons son efficacité en déterminant la taille de l'échantillon qui minimise le coût sujet à des contraintes sur la variance de l'estimation (calibrée) du total de chaque objectif.

Mots-clés: enquêtes-entreprise; bases multiples; base unique.

ABSTRACT

Most business surveys have multiple objectives. In order to reach these objectives, we aggregate the population under each objective. One business can belong to more than one group. We assume that each group has a frame and we use the multiple frames theory. One of the problems met with this approach is the sample size determination. We explore another approach based on a single frame. We evaluate its efficiency with a sample size determination which minimizes the cost under constraints on the variance of calibration total estimation of each objective.

Key words: business survey; multiple frames; single frame.

1. INTRODUCTION

1.1 Description du problème

Statistique Canada a lancé le Programme intégré de la statistique des entreprises (PISE) dont l'objectif principal est l'élaboration d'un système général qui comprend la majorité des enquêtes économiques à Statistique Canada. Ce système vise à uniformiser la méthodologie et l'infrastructure de ces enquêtes. Par ce programme, Statistique Canada veut à la fois remanier l'Enquête unifiée auprès des entreprises (EUE) qui englobe plus de soixante enquêtes économiques et l'élargir pour qu'elle puisse compter environ 120 enquêtes-entreprises, annuelles et infra- annuelles. Afin de répondre à cette exigence, nous sommes à la recherche d'un plan d'échantillonnage assez simple qui permet de faciliter le calcul de l'estimation, le calcul de la variance et la détermination de la taille de l'échantillon.

1.2 Les paramètres d'intérêt

Les enquêtes-entreprises possèdent actuellement une base de sondage F qui contient une liste de toutes les entreprises existantes. Nous supposerons que cette base est complète. Les paramètres d'intérêt de base pour le PISE sont principalement les totaux d'une caractéristique pour plusieurs sous-populations. Une sous-population est définie par la dimension industrielle et la dimension géographique. Supposons que nous avons p industries et q régions géographiques, pour une caractéristique d'intérêt y, nous avons pq paramètres d'intérêt à estimer, le vecteur des paramètres est donc $Y = (Y_1, \dots, Y_I)^T$ où $I = pq$ est le nombre total de sous-populations. Supposons qu'une des variables d'intérêt est y_i qui représente le revenu d'une certaine activité dans une géographie donnée, correspondant à la sous-population i. Le total Y_i d'une sous-population i pour une variable d'intérêt y_i peut s'écrire sous la forme :

[1] Naïma Gouzi, Division des méthodes d'enquêtes auprès des entreprises, Statistique Canada, Canada, K1A 0T6, courriel naima.gouzi@statcan.gc.ca
Abdellatif Demnati, Division des méthodes d'enquêtes auprès des entreprises, Statistique Canada, Canada, K1A 0T6, courriel abdellatif.demnati@statcan.gc.ca

$$Y_i = \sum_k \delta_{ik} y_{ik}\ , i = 1, \dots, I, \qquad (1)$$

où y_{ik} est la valeur de la variable d'intérêt pour l'unité k dans la sous-population i , δ_{ik} est l'indicateur d'appartenance à la sous-population i pour l'unité k, c.-à-d., $\delta_{ik} = 1$ si $k \in$ à la sous-population i, sinon $\delta_{ik} = 0$ et $\sum_k$ représente la sommation sur tous les éléments de la population. Nous supposons que δ_{ik} est connu pour toutes les unités de la base de sondage F.

Nous examinons ici deux approches d'échantillonnage, la première se fait à partir des bases de sondage multiples et la seconde se fait à partir d'une base unique. Dans la première approche, nous supposons que nous avons I bases de sondage indépendantes, et que chaque sous-population est représentée par une base de sondage. Les bases peuvent se chevaucher et l'union des bases forme l'ensemble de la population. Un échantillon aléatoire simple est sélectionné de chaque base. Dans la deuxième approche, nous ne considérons qu'une seule base de sondage stratifiée qui couvre l'ensemble de la population. Une entreprise est classifiée dans une seule strate et chaque sous-population représente un domaine d'étude. Un échantillon aléatoire simple est sélectionné de chaque strate. La section 2 donne un bref aperçu de la théorie des bases multiples, la section 3 décrit l'approche à base unique. Finalement, dans la dernière section, nous présentons les résultats d'une étude empirique sur la détermination de la taille de l'échantillon selon l'approche à base unique et nous comparons le tout avec les tailles utilisées actuellement par l'EUE.

2. APPROCHE À BASES MULTIPLES

Une façon naturelle de produire les estimations du total Y_i de plusieurs sous-populations est de construire pour chaque sous-population une base de sondage, voir Kott (1998). Or, une sous-population couvre toutes les entreprises appartenant au même groupe d'industries et à la même zone géographique. En revanche, il est possible qu'une entreprise ayant une structure complexe opère dans plusieurs secteurs d'activités ou dans plusieurs zones géographiques; une entreprise peut donc appartenir à plus d'une base et on se retrouve à créer artificiellement des bases multiples qui se chevauchent. Dans ce qui suit, nous supposons qu'un échantillon aléatoire simple s_i de taille n_i est sélectionné de chaque base de sondage F_i de façon indépendante d'une base à l'autre et nous appliquons la théorie des bases multiples.

La théorie des bases multiples fait état de plusieurs façons de combiner les échantillons s_i qui sont prélevés séparément de chaque base. L'EUE combine présentement ces échantillons de sorte que l'échantillon global, s^*, est l'union de tous ces échantillons, c.-à-d., $s^* = \cup_{i=1}^{I} s_i$. De cette façon, les doubles sont éliminés et la probabilité de sélection d'une entreprise est recalculée en considérant les probabilités de sélection des autres bases d'appartenance comme suit : $\pi_k^* = 1 - \prod_{i=1}^{I}(1 - \pi_{ik})$ où $\pi_{ik} = E(a_{ik})$ est la probabilité d'inclusion de l'élément k dans l'échantillon s_i, E dénote l'espérance par rapport au plan d'échantillonnage et a_{ik} est la variable d'indicateur d'appartenance à l'échantillon s_i. Pour plus de détails sur cette approche voir Bankier (1986).

En se basant sur s^*, l'échantillon global sans double, l'estimateur de Y_i à partir des bases multiples s'écrit sous la forme :

$$\hat{Y}_i^* = \sum_k d_k^* \delta_{ik} y_{ik}\ , \qquad (2)$$

où $\sum_k$ représente la sommation de tous les éléments distincts k de la population, $d_k^* = a_k^*/\pi_k^*$, $a_k^* = 1 - \prod_{i=1}^{I}(1 - a_{ik})$ est l'indicateur d'appartenance à l'échantillon s^* pour l'élément k et $\pi_k^* = E(a_k^*) = 1 - \prod_{i=1}^{I}(1 - \pi_{ik})$ est la probabilité d'inclusion de l'élément k dans l'échantillon s^*. La variance d'échantillonnage du total de la population $\hat{Y}_i^*$ est donnée par la forme générale de Horvitz-Thompson (HT) :

$$Var(\hat{Y}_i^*) = \sum_k \sum_l y_{ik} y_{il} (\pi_{kl}^* - \pi_k^* \pi_l^*)/(\pi_k^* \pi_l^*), \qquad (3)$$

où $\pi_{kk}^* = \pi_k^*$, $\pi_{kl}^* = \pi_k^* + \pi_l^* - 1 + \prod_{i=1}^{I}(1 - \pi_{ik})\left(1 - \pi_{il|k \notin s_i}\right)$ pour $k \neq l$ et $\pi_{il|k \notin s_i} = Pr(l \in s_i |\ k \notin s_i)$.

Supposons que nous voulons déterminer la taille d'échantillon optimale n_i tiré de la base de sondage F_i en sachant que le coût total est :

$$C = c_0 + \sum_i c_i n_i. \qquad (4)$$

Nous minimisons ce coût sous les contraintes de i variances:

$$Var(\hat{Y}_i^*) \leq V_i\ ,\ i = 1, \dots, I, \qquad (5)$$

$$2 \leq n_i \leq N_i,\ i = 1, \dots, I,$$

où c_0 est le coût initial de l'enquête, c_i est le coût par unité dans la base F_i et V_i est une tolérance spécifiée. Comme on peut le voir, la détermination des tailles des échantillons revient à minimiser la fonction du coût sujet à des contraintes hautement non linéaires des paramètres à estimer. On constate également que la forme de la variance (3) n'est pas simple et que la probabilité de sélection conjointe (π_{kl}^*) est complexe et difficile à dériver. Ce qui rend difficile la convergence du processus de minimisation. Par conséquent, la détermination de la taille d'échantillon suivant cet estimateur est presque impossible.

3. APPROCHE À BASE UNIQUE

Pour simplifier le mécanisme d'échantillonnage, le Comité consultatif des méthodes statistiques de Statistique Canada a suggéré de contourner l'approche à bases multiples pour choisir plutôt l'approche à base unique stratifiée, voir Statistics Canada (2010). Les entreprises complexes apparaîtront à un seul emplacement, soit à une seule industrie et province et chaque sous-population sera traitée comme un domaine d'étude. À ce moment, l'approche à bases multiples se réduit à l'approche d'une base unique stratifiée où les bases formées par industrie et province ne se chevauchent plus; elles deviennent alors des strates primaires.

Pour faciliter l'illustration de l'emplacement d'une entreprise complexe à une strate primaire, nous supposerons que le nombre de strates H est égal au nombre de domaines I et qu'une entreprise k a I proportions $(p_{1k}, \dots, p_{Ik})$, où p_{ik} est la proportion de l'entreprise k dans le domaine i. L'entreprise k peut être placée à la strate h si sa proportion dans la strate h est la plus grande de toutes les proportions, soit $p_{hk} = \max(p_{1k}, \dots, p_{Ik})$. Pour ce projet, nous avons considéré deux façons de dériver ces proportions. La première, que nous avons nommée par la règle de dominance, est basée sur la contribution de l'entreprise k dans le domaine i par rapport à la contribution totale de l'entreprise k dans tous les domaines : $p_{ik} = \delta_{ik}\, y_{ik} / \sum_i \delta_{ik} y_{ik}$, $i = 1, \dots, I$. La deuxième, que nous avons nommée par le total du domaine, est basée sur la contribution de l'entreprise k dans le domaine i par rapport à la contribution totale de toutes les entreprises appartenant au même domaine: $p_{ik} = \delta_{ik} y_{ik} / \sum_k \delta_{ik} y_{ik}$, $i = 1, \dots, I$.

3.1 Estimateur à base unique stratifiée

Le total de la population (1) à partir d'une base unique stratifiée peut s'écrire sous la forme :

$$Y_i = \sum_h \sum_k j_{hk} \delta_{ik} y_{ik} \;,\; i = 1, \dots, I \text{ et } h = 1, \dots, H \;, \tag{6}$$

où $\sum_h$ représente la sommation sur toutes les strates et j_{hk} est l'indicateur d'appartenance à la strate h pour l'élément k, c.-à-d., $j_{hk} = 1$ si $k \in h$, $j_{hk} = 0$ autrement. On suppose qu'un échantillon aléatoire s_h est choisi de la strate h et que les paramètres d'intérêt sont les totaux dans chaque domaine. Dans ce contexte, l'estimateur de Y_i prend la forme d'un estimateur HT stratifié :

$$\hat{Y}_i = \sum_h \sum_k d_{hk} \delta_{ik} y_{ik} \;,\; i = 1, \dots, I \text{ et } h = 1, \dots, H \;, \tag{7}$$

où $d_{hk} = j_{hk}\, a_{hk} / \pi_{hk}$ est le poids d'échantillonnage attaché à l'élément k associé à l'échantillon s_h, a_{hk} est l'indicateur d'appartenance à l'échantillon s_h pour l'élément k, c.-à-d., $a_{hk} = 1$ si $\in s_h$, $a_{hk} = 0$ autrement et π_{hk} est la probabilité d'inclusion de l'élément k dans l'échantillon s_h. La variance d'échantillonnage du total de la population $\hat{Y}_i$ est donnée par $Var(\hat{Y}_i) = \sum_h \boldsymbol{V}_h(y_i)$ où $\boldsymbol{V}_h(u)$ dénote la variance de l'estimateur $\hat{U}_h = \sum_k d_{hk} u_k$ du total U_h. Sous un plan aléatoire simple (PAS) ou sous un plan de Bernoulli (PB), cette variance peut être exprimée par :

$$\boldsymbol{V}_h(u) = \alpha_h + \beta_h / n_h, \tag{8}$$

où pour un PAS, $\alpha_h = -N_h S_h^2(u)$, $\beta_h = N_h^2 S_h^2(u)$, $S_h^2(u) = \sum_k j_{hk}(u_k - \bar{U}_h)^2 / (N_h - 1)$ et $\bar{U}_h = \sum_k j_{hk} u_k / N_h$. Pour un PB $\alpha_h = -\sum_k j_{hk} u_k^2$ et $\beta_h = N_h \sum_k j_{hk} u_k^2$.

Nous pouvons aussi considérer l'estimateur par le quotient qui est l'estimateur de HT ajusté par le total du domaine

$$\hat{Y}_{iR} = N_i(\hat{Y}_i / \hat{N}_i) \text{ où } \hat{N}_i = \sum_h \sum_k d_{hk} \delta_{ik} \text{ et } N_i = \sum_k \delta_{ik}. \tag{9}$$

3.2 Calage

Nous pouvons améliorer l'efficacité de l'estimateur HT donné par (7) en calant par les tailles connues de chaque domaine. L'estimateur (9) peut être vu comme un estimateur calé par le total du domaine avec le facteur de correction donné par $g_k = (N_i / \hat{N}_i)$. Pour faire le calage à toutes les tailles des domaines en même temps, nous pouvons utiliser l'estimateur par la régression généralisée ERG qui a la forme suivante : $\tilde{Y}_i = \sum_k d_k g_k \delta_{ik} y_{ik}$, où le facteur g_k est donné par $g_k = 1 + (X - \hat{X})^T \left(\sum_k d_k x_k x_k^T\right)^{-1} x_k$ où $X = (N_1, \dots, N_I)^T$ est un vecteur de dimension I qui correspond aux totaux connus des variables de calage $x = (x_1, \dots, x_I)^T$ et le vecteur $x_k = (\delta_{1k}, \dots, \delta_{Ik})^T$ contient les indicateurs des domaines pour chaque unité k. L'estimateur satisfait la contrainte de calage $\tilde{X} = \sum_k d_k g_k x_k = X$. Pour ce qui est de Bernoulli, on peut faire un calage par les tailles des strates afin de rendre l'estimation proche de l'aléatoire simple.

3.3 Détermination de la taille de l'échantillon

Supposons que nous considérons I caractéristiques d'intérêt $y_1, \ldots, y_I$. Sous un échantillonnage aléatoire simple stratifié et sous un échantillonnage de Bernoulli stratifié, nous pouvons exprimer la variance d'échantillonnage de la $i^{\text{ème}}$ variable d'intérêt y_i comme suit : $Var(\hat{Y}_i) = \sum_h (\alpha_h + \beta_h / n_h)$. Cette variance a une forme simple puisqu'elle s'écrit sous une forme séparable, ce qui facilite le processus de minimisation. Nous voulons déterminer $n = (n_1, \ldots, n_H)^T$, le vecteur de tailles d'échantillon optimales en sachant que le coût donné par

$$C = c_0 + \sum_h c_h n_h \tag{10}$$

est minimisé sous les contraintes de i variances:

$$Var(\hat{Y}_i) \leq V_i\ ,\ i = 1, \ldots, I, \tag{11}$$
$$2 \leq n_h \leq N_h,\ h = 1, \ldots, H,$$

où c_0 est le coût initial pour l'enquête, c_h est le coût par unité dans la strate h et V_i est une tolérance spécifiée. Supposons que le coefficient de variation de $\hat{Y}_i$ ne doit pas dépasser 5 %, c.-à-d., $Var(\hat{Y}_i)^{1/2} / E(\hat{Y}_i) \leq 0{,}05$, alors $V_i = (Y_i \times 0{,}05)^2$. Puisque nous avons une population à plusieurs variables où chaque variable correspond a un domaine d'estimation, la taille d'échantillon doit tenir compte simultanément des variances de tous les domaines.

4. ÉTUDE EMPIRIQUE

L'objectif principal de cette étude est de comparer le concept d'une base unique par rapport au concept des bases multiples par l'entremise du processus de détermination des tailles d'échantillon. Pour mener cette comparaison, nous avons utilisé les données de l'EUE. L'EUE comprend 60 sous-enquêtes. Chaque sous-enquête englobe un ou plusieurs secteurs d'activités et provinces. Chaque secteur d'activité et province est stratifié comme suit : une strate à tirage obligatoire qui contient des unités prédéterminées à l'avance par leur structure, une strate à tirage complet qui contient des unités ayant un grand revenu, une strate à tirage partiel qui contient les unités ayant un revenu moyen et une deuxième strate à tirage partiel qui contient des unités ayant un revenu petit. L'intérêt principal est de produire le total des estimations annuelles des activités économiques des entreprises qui pratiquent leurs activités dans un groupe d'industries et une province en particulier. Le coefficient de variation visé dans cette étude est de 5 %.

4.1 L'approche actuelle de l'EUE

L'EUE est une enquête qui utilise actuellement la théorie des bases multiples où chaque type d'industrie et province constitue une base de sondage. Malheureusement, la répartition de la taille de l'échantillon selon ce plan d'échantillonnage à bases multiples est presque impossible étant donnée sa complexité. Alors, l'EUE se sert de plusieurs enquêtes indépendantes pour déterminer uniquement les tailles de l'échantillon requis. Cette approche suppose qu'aucune relation n'existe entre les sous-populations, c'est-à-dire qu'une unité dans une sous-population n'a aucun lien avec les unités des autres sous-populations et que chaque sous-population fait l'objet d'une enquête indépendante. Supposons maintenant que nous avons I enquêtes indépendantes. Dans ce contexte, les données de chaque sous-population i sont recueillies par l'entremise de l'enquête i. Donc, nous avons plusieurs enquêtes fondées sur plusieurs bases indépendantes stratifiées selon les strates définies plus haut.

Nous supposons qu'un échantillon aléatoire simple stratifié est choisi de la base de sondage correspondante à l'enquête i. Le paramètre d'intérêt de l'enquête i est le total de la sous-population i de la variable d'intérêt y_i. Sous un plan stratifié, l'estimateur HT de Y_i pour l'enquête i est donné par :

$$\hat{Y}_i^E = \sum_h \sum_k d_{ihk} \delta_{ik} y_{ik}\ ,\ i = 1, \ldots, I, \tag{12}$$

où $\delta_{ik} = 1$ pour tous les éléments de l'enquête i, $d_{ihk} = j_{ihk}\, a_{ihk} / \pi_{ihk}$ est le poids d'échantillonnage de l'enquête i attaché à l'élément k associé à l'échantillon s_{ih} , a_{ihk} est l'indicateur d'appartenance à l'échantillon s_{ih} pour l'élément k, c.-à-d., $a_{ihk} = 1$ si $k \in s_{ih}$, $a_{ihk} = 0$ autrement, π_{ihk} est la probabilité d'inclusion de l'élément k dans l'échantillon s_{ih}, j_{ihk} est l'indicateur d'appartenance à la strate h de l'enquête i pour l'élément k, c.-à-d., $j_{ihk} = 1$ si $k \in h$, $j_{ihk} = 0$ autrement.

La variance d'échantillonnage du total de la population de l'enquête i est exprimée en fonction de $Var(\hat{Y}_i^E) = \sum_h \boldsymbol{V}_{ih}(y_i)$ où $\boldsymbol{V}_{ih}(u)$ dénote la variance de l'estimateur $\hat{U}_{ih} = \sum_k d_{ihk} u_k$ du total U_{ih}. Cette variance peut être écrite pour un plan aléatoire simple sous la forme suivante: $\boldsymbol{V}_{ih}(u) = \alpha_{ih} + \beta_{ih}/n_{ih}$ où $\alpha_{ih} = -N_{ih} S_{ih}^2(u)$, $\beta_{ih} = N_{ih}^2 S_{ih}^2(u)$, $S_{ih}^2(u) = \sum_k j_{ihk}(u_k - \bar{U}_{ih})^2/(N_{ih} - 1)$ et $\bar{U}_{ih} = \sum_k j_{ihk} u_k / N_{ih}$.

Supposons que n_{ih} est la taille de l'échantillon de la strate h dans l'enquête i. L'objectif est de déterminer n_{ih} optimal qui minimise le coût de l'enquête i sous la contrainte que la variance de $\hat{Y}_i^E$ ne doit pas dépasser un seuil de tolérance spécifié V_i. La fonction du coût prend la forme suivant: $C_i = c_{i0} + \sum_h c_{ih} n_{ih}$ où c_{i0} est le coût initial pour l'enquête i et c_{ih} est le coût par unité associé à la strate h et à l'enquête i. Supposons que $c_{i0} = 0$ et $c_{i1} = c_{i2} = \cdots = c_{iH} = 1$, c.-à-d., le coût est le même dans chaque strate et égale à 1. Supposons aussi que le coefficient de variation de $\hat{Y}_i^E$ ne doit pas dépasser 5 %, c.-à-d., $Var(\hat{Y}_i^E)^{1/2}/E(\hat{Y}_i^E) \leq 0{,}05$, donc $V_i = (Y_i \times 0{,}05)^2$. Alors la répartition optimale qui minimise le coût et pour une précision désirée, $Var(\hat{Y}_i^E) = V_i$, est donnée de façon explicite par :

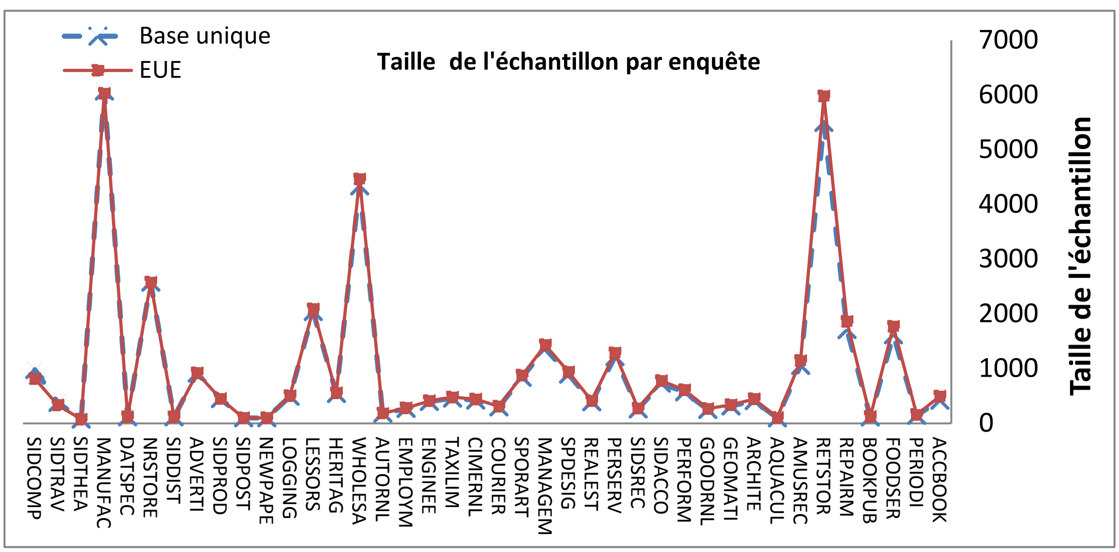
$$n_{ih} = \beta_{ih}^{1/2} \sum_h \beta_{ih}^{1/2} / (V_i - \sum_h \alpha_{ih}). \tag{13}$$

4.2 L'approche à base unique

Pour le cas d'une base unique, nous supposons qu'un échantillon aléatoire simple est choisi indépendamment dans chaque strate selon le plan de Bernoulli. Le coût initial est nul et le coût dans chaque strate est le même et égale à 1. Notre objectif est de déterminer la taille d'échantillon optimale n_h, $où$ $h = 1, \dots, H$, qui minimise le coût total sous la contrainte selon laquelle la variance de l'estimateur de chaque domaine calibré au niveau de la strate soit plus petite qu'un certain seuil de tolérance, V_i, spécifié à l'avance. Le nombre de combinaisons possibles de groupes d'industrie par province, représente le nombre de contraintes à respecter pour le processus de minimisation. Tandis que le nombre des combinaisons possibles de groupe d'industries, de province et de sous-strates représentent le nombre de paramètres à estimer ou le nombre de tailles à déterminer. La plus grande enquête a 200 domaines à estimer et 500 tailles à déterminer, ces nombres permettent de voir la grandeur du problème de minimisation. Malgré ces chiffres, le processus de minimisation a convergé pour toutes les enquêtes de même que, la solution optimale est obtenue pour chaque enquête.

4.3 Résultat de l'étude empirique

Graphique 1 : Taille de l'échantillon par enquête selon l'EUE et l'approche à base unique



Le graphique 1 présente les tailles de l'échantillon obtenues avec une base unique comparativement aux tailles obtenues selon l'EUE. Dans ce graphique, chaque sous-population est répartie selon les quatre sous-strates, dont la strate à tirage obligatoire qui contient majoritairement les entreprises complexes. Ces entreprises sont placées dans une seule strate par la règle de dominance. Comme, on peut le voir sur le graphique, les deux courbes coïncident. Ces résultats nous montrent qu'il n'y a pratiquement pas de différence entre les deux approches pour ce qui est de la taille d'échantillon. De plus, nous avons refait les mêmes étapes de comparaison sauf que les entreprises complexes ont été placées dans une seule strate selon le total du domaine. Là encore, nous avons obtenu les mêmes résultats.

Nous avons également comparé les deux approches en supposant que la strate à tirage obligatoire est éliminée. Alors, les entreprises de cette strate ont été réparties dans le reste des strates selon leurs revenus. Par la suite, nous avons refait le même processus de détermination des tailles d'échantillon. Nous avons remarqué que, pour les enquêtes ayant plusieurs entreprises complexes, l'approche à base unique est plus performante lorsque ces unités sont placées dans un seul emplacement selon la contribution du total du domaine que lorsque ces entreprises sont placées dans un seul emplacement selon la règle de la dominance.

5. CONCLUSION

De ces résultats, nous pouvons conclure que si nous avons une strate à tirage obligatoire qui contient majoritairement des entreprises complexes, le plan établi sur une base unique est similaire au plan utilisé par l'EUE en ce qui concerne l'objectif des tailles d'échantillon. Cela est valable pour les deux méthodes utilisées pour placer une entreprise à un seul emplacement et former une base unique. Par contre, lorsque la strate à tirage obligatoire est éliminée, la façon dont l'entreprise est placée dans une strate joue un rôle important pour le plan à base unique. Toutefois, nous continuons à explorer d'autres façons de placer une entreprise complexe à un seul emplacement.

En résumé, le plan fondé sur une base unique semble atteindre notre objectif visé pour ce qui est de la taille d'échantillon. En effet, ce plan se démarque par ses propriétés très simples. Il facilite de beaucoup le calcul de la variance et par le fait même le processus de détermination de l'échantillon. Combiné à la méthode d'échantillonnage de Bernoulli et des nombres aléatoires permanents, ce plan comporte beaucoup d'avantages dans le cadre du Programme intégré de la statistique des entreprises, notamment dans la coordination et la rotation des échantillons.

RÉFÉRENCES

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association,* 81, 1074-109.

Kott, P.S., Amrhein J.F. et Hicks, S.D. (1998). Échantillonnage et estimation à partir de bases de sondage listes multiples. *Techniques d'enquête*, 24, 3-10.

Statistics Canada (2010). Discussion at the Statistics Canada's Advisory Committee on Statistical Methods, April.