

COMPARAISON D'UNE MÉTHODE D'ÉCHANTILLONNAGE MODIFIÉE DE TILLÉ ET DE L'ÉCHANTILLONNAGE DE POISSON

John Slanta¹, Gary Kusch²

RÉSUMÉ

Depuis la fin des années 1950, les échantillons des enquêtes probabilistes sur le secteur de la fabrication réalisées par la Manufacturing and Construction Division (MCD) étaient sélectionnés presque exclusivement par la méthode de Poisson. Selon cette méthode, la probabilité de sélection d'une unité est proportionnelle à une mesure de la taille. L'avantage de l'échantillonnage de Poisson est la grande simplicité du calcul de la variance. Il a toutefois l'inconvénient d'ajouter une composante (habituellement positive) supplémentaire à la variance des estimations, car la taille de l'échantillon est une variable aléatoire. Lors de l'année d'enquête 1998, la MCD a commencé à utiliser la méthode d'échantillonnage modifiée de Tillé pour certaines de ses enquêtes. Elle recourt à cette méthode lorsque les probabilités de sélection sont inégales et que la taille de l'échantillon est fixe. Dans le présent article, nous décrivons brièvement la méthode modifiée et certaines de ses caractéristiques, et comparons, pour diverses enquêtes, les variances obtenues pour la méthode Tillé à celles obtenues pour la méthode de Poisson utilisée antérieurement.

MOTS CLÉS : Échantillonnage de Poisson; échantillonnage de Tillé; estimateur d'Horvitz-Thompson.

1. INTRODUCTION

Avant 1998, la Manufacturing and Construction Division (MCD) du U.S. Bureau of the Census utilisait principalement la méthode de Poisson pour sélectionner, pour ses enquêtes sur le secteur de la fabrication, des échantillons avec probabilités inégales. La méthode d'échantillonnage de Poisson sous-entend que chaque unité d'échantillonnage est sélectionnée dans l'échantillon, ou en est rejetée, indépendamment de toute autre unité, ce qui simplifie le calcul de la variance. Toutefois, l'un des inconvénients de cette méthode est que la taille de l'échantillon varie. Or, dans certaines situations, cette variabilité augmente la variance des estimations.

Les scénarios de tirage d'échantillons de taille fixe par échantillonnage avec probabilités inégales étaient bien connus (Hanif et Brewer, 1983), mais difficiles à mettre en œuvre dans le cas d'opérations d'échantillonnage à grande échelle. En outre, par dessus tout, la variance de l'estimateur non biaisé de Yates-Grundy-Sen (Cochran, 1977) prenait des valeurs négatives. En 1997, la MCD a mis au point une version modifiée d'une méthode d'échantillonnage élaborée par Yves Tillé (Tillé, 1996) qui était facile à mettre en œuvre et permettait de contourner le problème des estimations négatives de la variance (Slanta et Fagan, 1997). La MCD a appliqué cette méthode d'échantillonnage à la Manufacturing Energy Consumption Survey (MECS) de 1998, les Industrial Research and Development (R&D) Surveys de 1998,

¹ John Slanta, U.S. Bureau of the Census, Manufacturing and Construction Division, Salle 2225-4, Washington, D.C. 20233, USA

² Gary Kusch, U.S. Bureau of the Census, Manufacturing and Construction Division, Salle 2225-4, Washington, D.C. 20233, USA

Le présent article présente les résultats de travaux de recherche et d'analyse entrepris par les employés du Census Bureau. Il a fait l'objet d'un examen de portée plus limitée que celui auquel sont soumises les publications officielles du Census Bureau. Le présent rapport est diffusé en vue de tenir les parties intéressées au courant des travaux en cours et de favoriser les discussions à leur sujet.

1999 et 2000, la Survey of Plant Capacity (MQ-C1) de 1999 et la Pollution Abatement Capital Expenditures (PACE) survey de 1999.

Le présent article présente les résultats d'une étude comparative des coefficients de variation (c.v.) calculés pour les estimations obtenues d'après ces enquêtes lorsque la taille de l'échantillon est fixe à ceux qui auraient été obtenus si l'on avait appliqué l'échantillonnage de Poisson. Nous examinons les résultats pour certaines années pour trois des quatre enquêtes (les données finales n'étaient pas disponibles pour la PACES). Ces enquêtes présentent chacune des caractéristiques uniques si bien que les résultats diffèrent de l'une à l'autre.

2. MÉTHODE DE COMPARAISON DES VARIANCES

Au départ, nous voulions comparer les variances pour des périodes consécutives pour chaque enquête, la première période coïncidant avec la dernière utilisation du panel sélectionné selon la méthode de Poisson et la deuxième correspondant à la première utilisation du panel sélectionné par la méthode à échantillon de taille fixe. Par exemple, nous avons comparé les c.v. calculés pour la MECS de 1994 à ceux obtenus pour l'enquête de 1998 et avons constaté une réduction considérable de la variance. Ces résultats étaient fort encourageants, mais nous craignons que d'autres facteurs que l'adoption d'un échantillon de taille fixe jouent un rôle dans la diminution des c.v., surtout dans le cas de la MECS ou l'intervalle entre les panels est de quatre ans. Nous voulions comparer des pommes et des pommes, plutôt que des pommes et des poires. Nous avons finalement conclu que, dans le cas de chacune de ces enquêtes, les écarts de période en période dépendent vraisemblablement de l'effet d'autres facteurs que l'utilisation d'un échantillon de taille fixe et que ces effets rendent les comparaisons difficiles.

Comme les variances réelles sont des fonctions mathématiques, nous avons décidé de comparer les estimations de ces fonctions. Les variances réelles dépendent des valeurs des données (en grande partie inconnues), de l'estimateur utilisé, ainsi que des probabilités et probabilités conjointes de sélection des unités d'échantillonnage.

L'estimateur d'Horvitz-Thompson est celui qui est choisi pour la plupart de ces enquêtes (Cochran, 1977). Toutefois, dans le cas de la Survey of Plant Capacity, on utilise un estimateur par quotient dont le numérateur et le dénominateur sont l'un et l'autre des estimateurs d'Horvitz-Thompson. La variance de l'estimateur d'Horvitz-Thompson est donnée par :

$$\sigma^2(\hat{Y}_{HT}) = \sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) y_i^2 + 2 \sum_{i=2}^N \sum_{j=1}^{i-1} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} \quad (1)$$

où y_i représente la valeur étudiée pour la $i^{\text{ème}}$ unité d'échantillonnage;
 B_i représente la probabilité de sélection de la $i^{\text{ème}}$ unité d'échantillonnage;
 B_{ij} représente la probabilité de sélection simultanée des $i^{\text{ème}}$ et $j^{\text{ème}}$ unités d'échantillonnage;
 N représente la taille de la population.

Deux relations doivent être mentionnées lorsque $i \dots j$. La première est que, si l'on utilise l'échantillonnage de Poisson, $B_{ij} = B_i B_j$ et le terme de double somme de l'équation de la variance devient nul. La deuxième est que, si l'on utilise l'échantillonnage de Tillé, $B_{ij} \neq B_i B_j$. Par conséquent, si toutes les données (valeurs étudiées) sont non négatives, alors la variance réelle dans les conditions d'échantillonnage de Tillé sera systématiquement inférieure ou égale à la variance réelle dans les conditions d'échantillonnage de Poisson.

Yates, Grundy et Sen montrent que, pour des échantillons de taille fixe, la formule de la variance susmentionnée peut aussi s'exprimer sous la forme :

$$\begin{aligned}\sigma^2\left(\hat{Y}_{HT, \text{taille fixe d'échantillon}}\right) &= \sum_{i=2}^N \sum_{j=1}^{i-1} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \\ &= \sum_{i=2}^N \sum_{j=1}^{i-1} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i^2}{\pi_i^2} - 2 \frac{y_i}{\pi_i} \frac{y_j}{\pi_j} + \frac{y_j^2}{\pi_j^2} \right)\end{aligned}\quad (2)$$

Notons que le terme croisé qui figure dans (2) est égal au terme de double somme qui figure dans (1). Par conséquent, si la taille de l'échantillon est fixe :

$$\sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) y_i^2 = \sum_{i=2}^N \sum_{j=1}^{i-1} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} \right) \quad (3)$$

Comme des échantillons de taille fixe ont été utilisés pour les enquêtes en question, nous utiliserons la forme de la variance de Poisson qui figure dans le deuxième membre de l'équation (3). Les estimations des fonctions mathématiques (2) et (3), sont, respectivement :

$$\sum_{i=2}^n \sum_{j=1}^{i-1} (\beta_{ij} - 1) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \quad \text{et} \quad \sum_{i=2}^n \sum_{j=1}^{i-1} (\beta_{ij} - 1) \left(\frac{y_i^2}{\pi_i^2} + \frac{y_j^2}{\pi_j^2} \right) \quad (4)$$

où n est la taille fixe d'échantillon et $\beta_{ij} = \frac{\pi_i \pi_j}{\pi_{ij}}$.

Si $B_{ij} \neq B_i B_j$ pour tout i et j de la population lorsque $i \dots j$, alors les estimations sont non négatives; en outre, si $B_{ij} > 0$ pour chaque i et j de la population lorsque $i \dots j$, alors les estimations sont non biaisées. Notons de nouveau que, si toutes les données (valeurs étudiées) sont non négatives, alors l'estimation de (2) sera systématiquement inférieure ou égale à l'estimation de (3), c'est-à-dire une propriété désirable.

Nous avons apporté deux autres modifications à (4); en premier lieu, nous avons simplifié et réduit le nombre d'opérations mathématiques et en deuxième lieu, nous avons procédé à un ajustement proportionnel (par quotient) de ces fonctions afin de réduire la variabilité excessive de ces nombres lorsque l'on tire plus d'un échantillon.

Le premier ajustement découle d'une propriété de l'échantillonnage de Tillé en vertu de laquelle nombre de valeurs de $\$_{ij}$ sont égales à $\$_{i1}$, pour $j < i$. Par conséquent, le premier ajustement est un remplacement de $\$_{ij}$ par $\$_{i1}$.

Le deuxième est un ajustement proportionnel (par quotient) qui compense dans une certaine mesure l'instabilité de la variance d'échantillon de Yates, Grundy et Sen. Si la taille de l'échantillon est fixe et que les probabilités de sélection sont asymétriques, certaines probabilités conjointes, B_{ij} , risquent d'être très faibles, ce qui pourrait donner lieu à une forte surestimation de la variance d'échantillon. L'inverse est également vrai. Certains échantillons pourraient être sélectionnés là où la variance d'échantillon sera

inférieure à une borne inférieure connue (Biyani, 1980). Par conséquent, nous multiplions la variance d'échantillon par :

$$\frac{n^2}{\sum_{i=1}^n \pi_i + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} \beta_{ij}} \quad (5)$$

Notons que la valeur espérée du dénominateur est égale au numérateur. Pour simplifier les calculs, nous remplaçons β_{ij} par β_{i1} . Cette substitution pourrait ne pas résoudre entièrement le problème présenté par Biyani, mais elle rend la variance d'échantillon plus stable.

Pour la Survey of Plant Capacity, on utilise un estimateur par quotient, si bien que nous avons aussi besoin de termes de covariance. La covariance de \hat{Y} et \hat{X} , où ces estimateurs sont des estimateurs d'Horvitz-Thompson, est donnée par:

$$\sigma^2(\hat{Y}_{HT}) = \sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) y_i x_i + 2 \sum_{i=2}^N \sum_{j=1}^{i-1} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{x_j}{\pi_j} \quad (6)$$

Si la taille d'échantillon est fixe, une estimation de la covariance de \hat{Y} et \hat{X} donnée par :

$$\sum_{i=2}^n \sum_{j=1}^{i-1} (\beta_{ij} - 1) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right) \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right), \quad (7)$$

et l'estimation de la fonction $\sum_{i=1}^N \left(\frac{1}{\pi_i} - 1 \right) y_i x_i$, c.-à-d. la covariance de \hat{Y} et \hat{X} dans les conditions de l'échantillonnage de Poisson, est donnée par :

$$\sum_{i=2}^n \sum_{j=1}^{i-1} (\beta_{ij} - 1) \left(\frac{y_i x_i}{\pi_i^2} + \frac{y_j x_j}{\pi_j^2} \right). \quad (8)$$

Nous appliquons aussi à ces estimations les deux ajustements susmentionnés. Il est nécessaire d'utiliser ces estimations de la covariance plutôt que d'autres, pour que la variance de l'estimation par quotient soit systématiquement non négative.

3. RÉSULTAT DES ENQUÊTES

3.1 Enquête sur la R-D

L'Enquête sur la R-D (Industrial Research and Development Survey), qui est une enquête auprès des entreprises réalisée annuellement, est conçue pour recueillir, entre autres, des renseignements sur les dépenses au titre de la R-D, selon la catégorie de R-D, la branche d'activité et l'État. L'une des caractéristiques importantes de cette enquête est qu'une faible proportion seulement des entreprises qui y participent entreprennent effectivement des activités de R-D, si bien que la plupart ne déclarent aucune dépense. Notre étude porte sur l'un des principaux tableaux publiés, A-2, qui présente les estimations de l'emploi sur le marché intérieur, des ventes nettes sur le marché intérieures, du nombre de scientifiques et d'ingénieurs, de la R-D financée par l'administration fédérale, de la R-D financée par l'entreprise et de la R-D totale. Ces six postes sont ventilés selon la branche d'activité et selon la catégorie d'emploi. Les contraintes d'échantillonnage sont univariées et fondées sur la R-D totale selon la branche d'activité. Notre

étude porte sur l'enquête réalisée pour 1999 et ne couvre pas les regroupements de branches d'activité sauf le total général.

La première partie de l'étude a trait exclusivement à la R-D totale selon la branche d'activité détaillées. Au départ, nous avons contrôlé 48 cellules du tableau pour l'échantillonnage et nous avons exclu de l'étude toute cellule pour laquelle l'estimation ou la variance estimée pour la méthode d'échantillonnage de Poisson était nulle. Nous avons ainsi retenu 46 cellules pour l'étude. Ensuite, nous avons calculé les estimations de la variance pour la méthode d'échantillonnage de Tillé et pour celle de Poisson. Puis, nous avons calculé le « taux de réduction » défini comme étant :

$$100 \left(\frac{\hat{\sigma}_{Poisson} - \hat{\sigma}_{Tille}}{\hat{\sigma}_{Poisson}} \right) \quad (9)$$

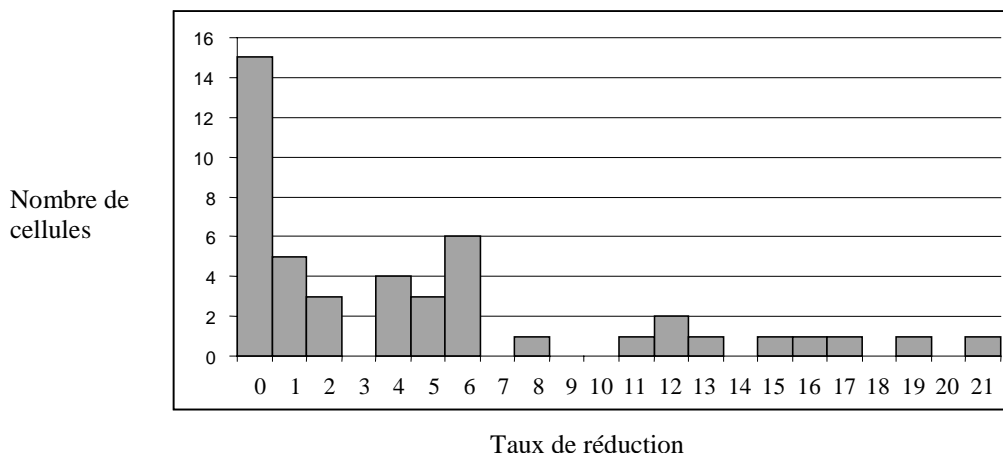
Enfin, nous avons produit des histogrammes où le « taux de réduction » est représenté sur l'axe horizontal et le nombre de cellules du tableau associées au « taux de réduction », sur l'axe vertical. Dans les graphiques qui suivent, tout « taux de réduction » dont l'écart par rapport à une valeur donnée est inférieur ou égal à une demi largeur de bande sur l'axe horizontal est regroupé avec cette valeur. Les statistiques descriptives du graphique 1 sont énumérées ci-dessous.

Statistiques descriptives du graphique 1 :

Nombre total de cellules	46	
Nombre de cellules à réduction nulle	6	(Nombre de cellules pour lesquelles le « taux de réduction » est exactement égal à zéro.)
Nombre de cellules à réduction positive	40	
Taux de réduction le plus faible	0	
Taux de réduction le plus élevé	21	
Taux de réduction médian	3	

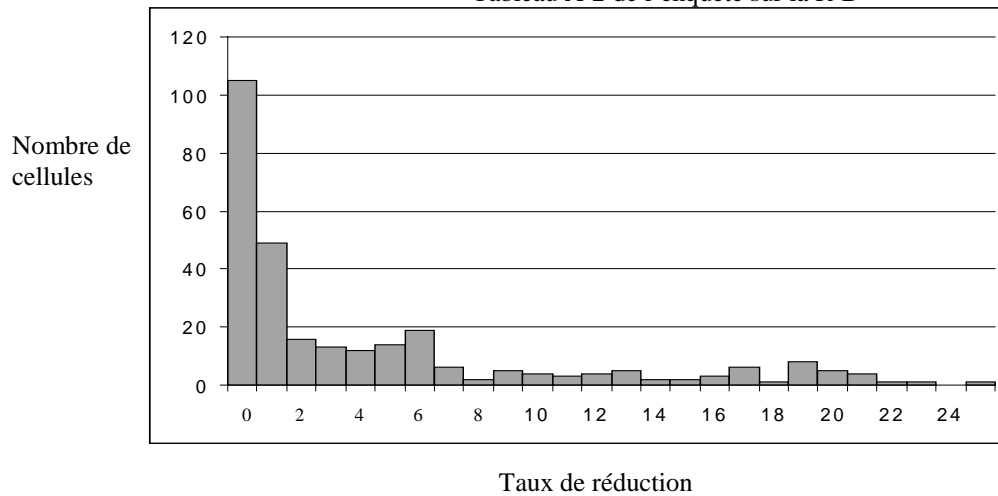
L'amélioration est faible dans la plupart des cas, mais elle est importante pour quelques cellules, particulièrement celle pour laquelle le « taux de réduction » est de 21 %. Soulignons que, dans aucun cas, la méthode d'échantillonnage de Tillé ne produit une augmentation de la variance comparativement à la méthode d'échantillon de Poisson.

Graphique 1
R-D totale selon la branche d'activité



Le graphique 2 résume les résultats pour toutes les cellules du tableau publié A-2, excluant les regroupements de branches d'activité sauf le total général.

Graphique 2
Tableau A-2 de l'enquête sur la R-D



Statistiques descriptives du graphique 2 :

Nombre total de cellules	291	Taux de réduction le plus faible	0,0
Nombre de cellules à réduction nulle	50	Taux de réduction le plus élevé	24,7
Nombre de cellules à réduction positive	241	Taux de réduction médian	1,3

Les résultats sont comparables, mais le nombre de valeurs nulles est beaucoup plus élevé. Une valeur nulle, qui signifie que l'estimation de la variance pour la méthode de Tillé est égale à l'estimation de la variance pour la méthode de Poisson, ne peut être observée que si une seule unité d'échantillonnage à sélection sans certitude a une valeur non nulle. Comme une faible proportion seulement d'entreprises non sélectionnées avec certitude déclarent des dépenses de R-D non nulles, cet événement n'est pas inhabituel.

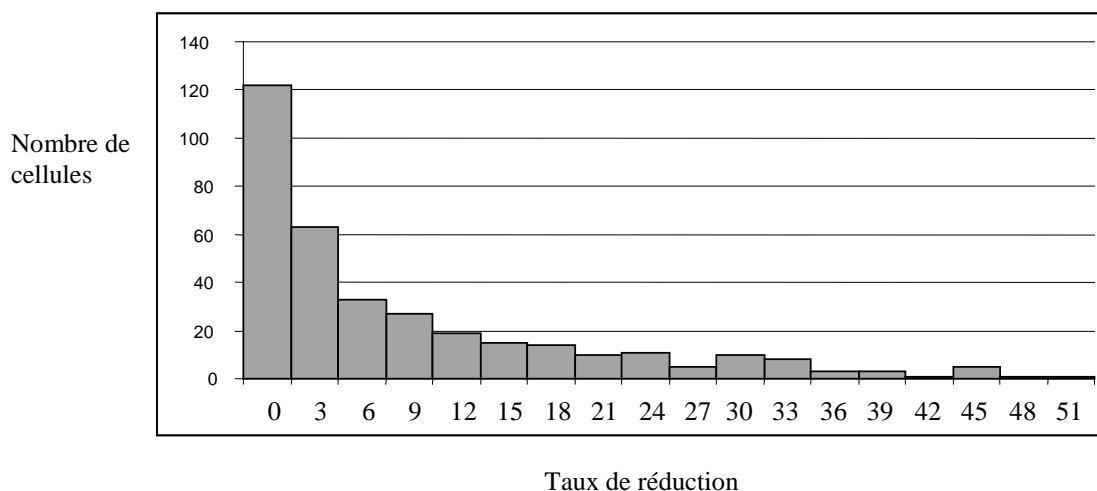
3.2 MECS

La MECS est une enquête auprès des établissements conçue pour recueillir des renseignements sur l'énergie consommée lors de l'utilisation de diverses catégories de combustible. Ces dernières années, l'enquête a été réalisée à intervalle de quatre ans, le cycle le plus récent ayant eu lieu en 1998. Notre étude porte sur le tableau N3.1 publié pour cette année de référence. Il peut être consulté sur le site Web

<http://www.eia.doe.gov/emeu/mecs/mecs98/datatables/contents.html>.

Ce tableau donne la consommation d'énergie selon la catégorie de combustible et selon la région. Nous avons examiné uniquement le niveau total pour les États-Unis et n'avons pas tenu compte des données ventilées selon la région. Les catégories de combustible dont la consommation nette est donnée dans le tableau sont l'électricité, le mazout lourd, le mazout léger, le gaz naturel, le gaz naturel liquéfié et les liquides du gaz naturel, le charbon, le coke et le poussier, les autres combustibles et l'ensemble des combustibles. Les statistiques descriptives du graphique 3 sont énumérées plus bas et la largeur de chaque barre de l'histogramme est égale à 3.

Graphique 3
Tableau N3.1 de la MECS



Statistiques descriptives du graphique 3 :

Nombre total de cellules	351	Taux de réduction le plus faible	! 0,05
Nombre de cellules à réduction négative	3	Taux de réduction le plus élevé	51,4
Nombre de cellules à réduction nulle	33	Taux de réduction médian	3.3
Nombre de cellules à réduction positive	315		

Nous avons examiné les trois cellules pour lesquelles la variance a augmenté lors de l'utilisation de la méthode d'échantillonnage modifiée de Tillé. Nous avons constaté qu'un mélange de microdonnées positives et négatives avait servi au calcul des estimations. Les analystes chargés de l'enquête ont confirmé qu'il pourrait exister des situations légitimes où les données sont négatives. Par conséquent, il n'est plus possible de garantir que la sélection d'un échantillon de taille fixe réduira la variance. Néanmoins, même sans cette garantie, les résultats semblent meilleurs dans le cas de la MECS que dans celui de l'enquête sur la R-D, situation qui tient au fait que, pour la seconde, la valeur d'une forte proportion de microdonnées était nulle, ce qui a rendu les deux estimations de la variance plus comparables.

3.3 Enquête sur la capacité de production des usines

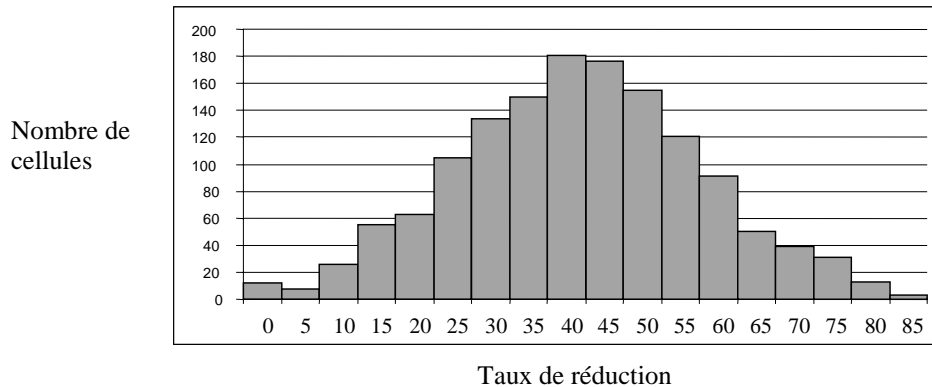
L'enquête sur la capacité de production des usines (Survey of Plant Capacity) est également une enquête auprès des établissements conçue pour recueillir des renseignements sur le niveau de production. Nous nous sommes concentrés sur l'enquête réalisée pour l'année de référence 1999. Les trois variables que nous avons examinées sont les suivantes :

1. valeur marchande de la production réelle au quatrième trimestre (que nous désignons par Réelle);
2. valeur marchande estimée qu'aurait eu la production si l'usine avait tourné à pleine capacité (que nous désignons par Pleine);
3. valeur estimée qu'aurait eu la production si l'usine avait tourné dans des conditions d'urgence nationale (que nous désignons par Urgence).

Les taux d'utilisation sont calculés d'après ces variables. Le premier est le taux d'utilisation de la pleine capacité (Réelle sur Pleine) et le deuxième, le taux d'utilisation de la capacité d'urgence (Réelle sur Urgence). Les statistiques publiées d'après l'enquête n'incluent que les taux d'utilisation; les estimations des niveaux d'utilisation ne sont pas publiées, mais sont parfois offertes sous forme de totalisations spéciales. Notre analyse porte à la fois sur les estimations des niveaux et des taux d'utilisation, mais uniquement pour les branches d'activité du niveau à six chiffres du SCIAN (classe nationale).

Le graphique 4 donne le « taux de réduction » pour les estimations du niveau selon le code à six chiffres du SCIAN.

Graphique 4
Capacité de production - Réelle, Pleine, Urgence - Niveau d'utilisation

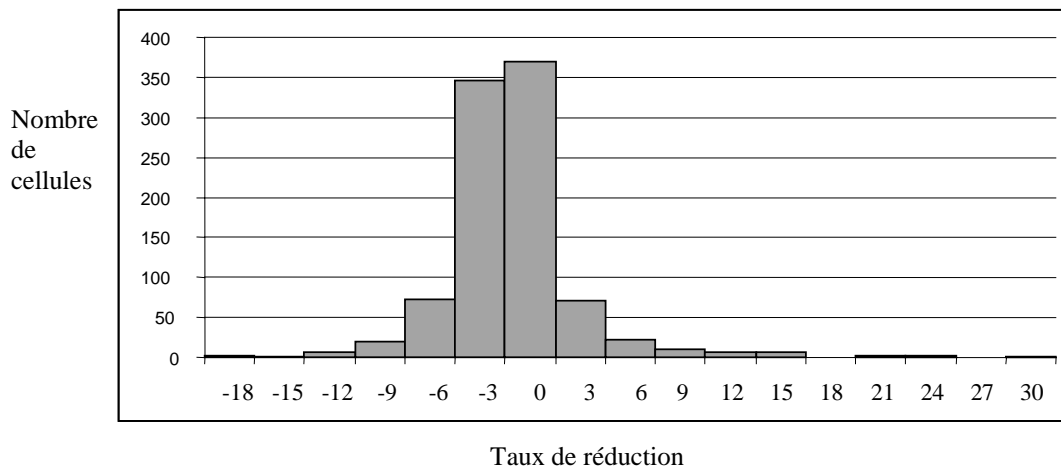


Statistiques descriptives du graphique 4 :

Nombre total de cellules	1 413	Taux de réduction le plus faible	0,0
Nombre de cellules à réduction nulle	9	Taux de réduction le plus élevé	85,1
Nombre de cellules à réduction positive	1 404	Taux de réduction médian	42,0

Le graphique 5 donne le « taux de réduction » pour les taux d'utilisation.

Graphique 5
Capacité de production - Taux d'utilisation de la pleine capacité et de la capacité d'urgence



Statistiques descriptives du graphique 5 :

Nombre total de cellules	942	Taux de réduction le plus faible	! 18,0
Nombre de cellules à réduction négative	725	Taux de réduction le plus élevé	29,7
Nombre de cellules à réduction nulle	18	Taux de réduction médian	! 1,4
Nombre de cellules à réduction positive	199		

Pour les estimations du niveau d'utilisation de la capacité, le « taux de réduction » des coefficients de variation est très bon, la réduction médiane étant de 42 % sans augmentation de la variance. Par contre,

pour les taux d'utilisation de la capacité, le « taux de réduction » des coefficients de variation est médiocre. Dans le cas d'une cellule, le coefficient de variation a augmenté de 18 % et, dans l'ensemble, le coefficient de variation de 77 % de cellules a augmenté. Heureusement, comme le montre le graphique 5, l'augmentation de la variance est faible dans la plupart des cas. Cette augmentation de la variance d'une forte proportion de cellules est due au fait que les coefficients de corrélation entre les estimations de niveau sont élevés dans le cas de l'échantillonnage de Poisson. Puisque, pour cet échantillonnage, la taille de l'échantillon est variable, les estimations de niveau ont tendance à être faibles lorsque la taille de l'échantillon est faible et importantes lorsque la taille de l'échantillon est importante, situation qui produit une forte corrélation. Or, dans le cas d'une estimation par quotient, la variance est d'autant plus faible que la corrélation est forte.

4. CONCLUSION

Dans le cas de l'enquête sur la R-D, nous constatons une amélioration systématique, même si elle est faible dans la plupart des cas. Pour la MECS, les résultats sont similaires, mais les améliorations sont généralement plus importantes que dans le cas de l'enquête sur la R-D. L'étude de la MECS montre néanmoins qu'un mélange de données négatives et positives peut donner lieu à une augmentation de la variance dans le cas de la méthode d'échantillonnage modifiée de Tillé. Toutefois, dans le cas de la MECS, l'augmentation n'est pas significative. L'étude de la Survey of Plant Capacity révèle une très forte réduction de la variance des estimations de niveau qui servent au calcul des taux d'utilisation. Malheureusement, la variance d'environ les trois quarts des taux d'utilisation augmente lorsque l'on passe à la méthode d'échantillonnage modifiée de Tillé. Cette situation tient au fait que les taux d'utilisation sont calculés au moyen d'un estimateur par quotient au lieu d'un estimateur d'Horvitz-Thompson. Dans le cas de l'échantillonnage de Poisson, la forte corrélation des estimations de niveau contribue à la réduction de la variance de l'estimateur par quotient.

Si l'on utilise un estimateur d'Horvitz-Thompson et que les données ne sont pas négatives, il semble que le passage à l'échantillonnage de Tillé ne présente aucun inconvénient et que les avantages puissent être importants si toutes les données, ou presque toutes, sont non nulles pour plusieurs unités d'échantillonnage dont le tirage n'a pas lieu avec certitude. Si les données ne sont négatives que pour quelques unités d'échantillonnage, il semble que la méthode avec sélection d'un échantillon de taille fixe soit avantageuse aussi. Si l'on utilise un autre estimateur que celui d'Horvitz-Thompson, l'adoption d'une méthode prévoyant la sélection d'un échantillon de taille fixe pourrait ne présenter aucun avantage.

BIBLIOGRAPHIE

Biyani, Shriram H. (1980), "On Inadmissibility of the Yates-Grundy Variance Estimator in Unequal Probability Sampling", *Journal of the American Statistical Association*, Volume 75, Number 371, pp. 709-712.

Cochran, William G. (1977), *Sampling Techniques*, New York: John Wiley & Sons.

Hanif, M., et Brewer, K. R. W. (1983), *Sampling with Unequal Probabilities*, New York: Springer-Verlag.

Slanta, John G., et Fagan, James T. (1997), "A Modified Approach to Sample Selection and Variance Estimation with Probability Proportional to Size and Fixed Sample Size", rapport non publié, MCD Working Paper Number: Census/MCD/WP-97/02.

Tillé, Yves (1996), "An Elimination Procedure for Unequal Probability Sampling Without Replacement", *Biometrika*, **83**, 1, pp. 238-241.