

D

DOCUMENT D'ÉTUDES

LA NOUVELLE MÉTHODE D'ÉCHANTILLONNAGE DE L'ENQUÊTE TRIMESTRIELLE ACEMO DEPUIS 2006

AMÉLIORATION DE L'ALLOCATION DE NEYMAN

Par
*Malik KOUBI**
*Sandrine MATHERN**

N° 146
Février 2009

**Les documents d'études sont des documents de travail ;
à ce titre, ils n'engagent que leurs auteurs
et ne représentent pas la position de la DARES.**

* DARES, au moment de la rédaction de cet article.



RÉSUMÉ

Dans un plan de sondage stratifié, l'allocation de Neyman est une méthode souvent utilisée pour échantillonner les enquêtes. Elle estime avec la plus grande précision possible, la moyenne d'une variable d'intérêt sur l'ensemble de la population étudiée. Si cette méthode est optimale pour l'estimateur au niveau de l'ensemble de la population, elle peut souffrir d'imprécision lorsqu'on veut obtenir des estimations sur des sous-groupes de la population.

On s'intéresse ici à la précision obtenue dans des « regroupements de publication », c'est-à-dire un regroupement de strates d'échantillonnage au niveau duquel on souhaite publier les résultats issus d'une enquête.

Dans ce cadre, le document présente une méthode qui améliore l'allocation de Neyman en assurant une précision « locale » minimale dans les regroupements de publication, sans trop détériorer sa précision « globale » au niveau de l'ensemble de la population.

Le dilemme entre précision globale et précision locale est illustré par une frontière d'efficacité. Cette frontière permet d'évaluer la perte en précision globale lorsque l'on tient à fixer une précision minimale dans des regroupements de publication. Elle constitue de ce fait une aide à la décision précieuse pour l'échantillonnage, car elle synthétise les différents arbitrages devant lesquels le statisticien est placé.

Depuis 2006, cette nouvelle méthode d'échantillonnage est appliquée à l'enquête trimestrielle sur l'activité et les conditions d'emploi de la main-d'œuvre (ACEMO).

SOMMAIRE

INTRODUCTION	7
NOTATIONS	8
GLOSSAIRE	9
HYPOTHÈSE FONDAMENTALE DE LA MÉTHODE PRÉSENTÉE	9
I. PLAN DE SONDAGE STRATIFIÉ ET LIMITES DE L'ALLOCATION DE NEYMAN	11
I.1. Estimateur et précision dans un sondage stratifié	11
I.2. Échantillonnage : l'allocation optimale de Neyman et ses limites	14
I.3. Une solution de compromis entre précision globale et locale.....	16
I.4. Autres méthodes empiriques pour améliorer la précision locale d'un estimateur	18
II. OPTIMISATION DE NEYMAN AVEC CONTRAINTE DE PRÉCISION DANS LES REGROUPEMENTS DE PUBLICATION	21
II.1. Formalisation du problème à résoudre	21
II.2. Schéma global de la résolution du problème en deux étapes	22
II.3. Traduction plus simple des contraintes de précision locale (Étape 1).....	23
II.4. Résolution de l'optimisation sous contrainte (Étape 2).....	26
III. CAS PRATIQUE DE LA MÉTHODE D'OPTIMISATION SOUS CONTRAINTE	31
III.1. Base de sondage et description du programme de résolution.....	31
III.2. Illustration : taille minimale d'échantillon nécessaire pour une précision locale souhaitée	32
III.3. Comparaison de l'allocation de Neyman et de l'allocation sous contrainte de précision locale.....	33
III.4. Concept et illustration de la frontière d'efficacité.....	36
CONCLUSION	41
BIBLIOGRAPHIE	42
ANNEXES	42

INTRODUCTION

En 2005-2006, à l'occasion de la refonte des enquêtes sur l'activité et les conditions d'emploi de la main-d'œuvre (ACEMO)¹, une étude approfondie a été menée sur l'ensemble des points méthodologiques du dispositif. Au cœur de celui-ci, l'enquête trimestrielle ACEMO mesure principalement les évolutions conjoncturelles des salaires de base, de l'emploi salarié et le niveau de la durée hebdomadaire collective du travail. Elle a subi de profondes modifications, notamment concernant la méthode de tirage de l'échantillon d'entreprises à interroger. L'allocation de Neyman est la méthode qui a d'abord été envisagée pour échantillonner cette enquête. Elle permet d'estimer avec la plus grande précision possible la moyenne globale d'une variable d'intérêt sur l'ensemble de la population étudiée. Mais si cette méthode est optimale pour l'estimateur global, elle peut souffrir d'imprécision lorsqu'on souhaite obtenir des estimateurs sur des sous-groupes de cette population.

On s'intéresse ici plus précisément à des « regroupements de publication », définis comme un regroupement de strates d'échantillonnage au niveau duquel on souhaite publier les résultats de l'enquête ACEMO. **La méthode présentée permet de tenir compte du double objectif de précision globale (au niveau de la population) et de précision locale (au niveau des regroupements de publication) et constitue le sujet de cette étude.**

L'objectif de ce document est plus précisément de proposer une solution pour améliorer la précision « locale » de l'allocation de Neyman, c'est-à-dire la plus mauvaise précision obtenue pour les estimateurs correspondant aux regroupements de publication, sans trop détériorer sa précision « globale », correspondant à l'estimateur au niveau de l'ensemble de la population. Nous développons ici une méthode qui formalise et automatise le travail effectué quelquefois « à la main » par le statisticien, lorsqu'il réalloue à la marge cette allocation afin de combler les lacunes de celle-ci dans certains petits regroupements de publication. Empiriquement, on parvient à décrire pour une taille d'échantillon donnée, une frontière d'efficacité entre les objectifs de précision globale et de précision locale. Ce concept est très utile pour évaluer la perte en précision globale lorsque l'on tient à fixer un seuil de précision minimale dans les regroupements de publication.

Le point technique principal de l'étude est l'ajout de contraintes bien choisies à l'optimisation de Neyman traditionnelle, contraintes qui traduisent l'exigence de précision dans les regroupements de publication. On s'attache ensuite à proposer une méthode pratique, la plus efficace possible, qui permette de résoudre ce problème.

Dans la première partie, on rappelle les éléments de la théorie des sondages relatifs à un plan stratifié avec tirage aléatoire simple au sein de chaque strate d'échantillonnage et allocation optimale de Neyman. On aborde les limites de cette allocation de Neyman et différentes solutions qui permettent d'améliorer la précision locale d'un estimateur sans trop dégrader sa précision globale.

La deuxième partie est consacrée à la présentation de notre méthode d'optimisation qui prend en compte la précision locale dans l'optimisation de Neyman. Elle comprend deux étapes. Premièrement, on montre comment les contraintes de précision locale ajoutées à ce programme d'optimisation peuvent être traduites en termes de nombre de questionnaires à tirer dans les strates d'échantillonnage. Deuxièmement, on décrit l'algorithme qui permet de résoudre le problème d'optimisation avec les contraintes reformulées.

La troisième partie se veut un cas pratique de notre méthode d'optimisation appliquée aux données issues des déclarations annuelles de données sociales (DADS) 2004. On compare les résultats de l'allocation sous contrainte de précision locale à ceux de l'allocation de Neyman lors d'un tirage d'échantillon. On illustre notre propos par divers graphiques, dont le graphe du concept de frontière d'efficacité, qui constitue la synthèse des résultats issus de la méthode d'optimisation avec contraintes.

¹ Dispositif d'enquêtes du Ministère du travail, des relations sociales, de la famille et de la solidarité

NOTATIONS

Y : variable d'intérêt que l'on cherche à estimer sur une population

\widehat{Y} : estimateur de la moyenne de Y

N : taille de la population étudiée

n : taille globale de l'échantillon sélectionné

h : indice des strates d'échantillonnage

H : nombre de strates d'échantillonnage

p : indice des regroupements de publication

k : nombre de regroupements de publication

N_h : nombre d'individus dans la strate h

n_h : nombre d'individus sélectionnés dans la strate h

S_h² : variance de la variable y dans la strate h

CV : coefficient de variation d'un estimateur

CV_p : coefficient de variation d'un estimateur au sein d'un regroupement de publication p

GLOSSAIRE

Sondage stratifié : méthode d'échantillonnage d'une enquête, consistant à regrouper les unités de la population finie dont on dispose, selon des critères fortement corrélés à la quantité étudiée, afin d'améliorer la précision obtenue lors de l'estimation globale. On tire ensuite un échantillon de manière aléatoire et indépendante dans chacun des regroupements. Les groupes homogènes créés forment une partition de la population et sont appelés strates.

Strates d'échantillonnage : regroupements de la population créés lors d'un sondage stratifié, desquels sont tirées les unités interrogées. Exemple utilisé dans ce document : secteur d'activité en 36 postes (nes36) x taille d'établissement en 6 modalités.

Regroupement de publication : niveaux d'agrégation de la population auxquels sont diffusés les résultats issus d'une enquête. Dans ce document, on suppose que chaque regroupement de publication est un ensemble de strates d'échantillonnage (on exclut les domaines quelconques). Exemple utilisé dans ce document : secteur d'activité en 16 postes (nes16), taille d'établissement, secteur d'activité en nes16 x taille d'établissement, etc.

Allocation de Neyman : dans un sondage stratifié avec sondage aléatoire simple dans chaque strate, allocation particulière qui permet d'obtenir une précision maximale sur l'estimateur de la moyenne globale pour un budget donné ou un nombre de questionnaires donné.

Précision globale : précision d'un estimateur au niveau de l'ensemble de la population. Elle est mesurée par le coefficient de variation CV_{global} .

Précision locale : précision d'un estimateur au niveau des regroupements de publication. Elle est mesurée par le plus mauvais des coefficients de variation des estimateurs dans les regroupements de publication $CV_{\text{local maximal}}$
 $= \text{Max}_{p \in \text{pub}} CV_p$.

Frontière d'efficacité : ensemble des allocations d'échantillon (n_1, \dots, n_H) à taille n fixée tel qu'on ne peut augmenter la précision locale dans les regroupements de publication, sans détériorer la précision globale au niveau de l'ensemble de la population d'un estimateur (et inversement). La frontière d'efficacité est représentée dans le repère $(CV_{\text{global}}, CV_{\text{local maximal}})$. Par abus de langage, on appellera frontière d'efficacité, l'ensemble des couples $(CV_{\text{global}}, CV_{\text{local maximal}})$ réalisables pour une taille d'échantillon donnée.

Domaine : sous-population d'intérêt particulier pour une enquête.

HYPOTHÈSE FONDAMENTALE DE LA MÉTHODE PRÉSENTÉE

On suppose que chaque regroupement de publication est la réunion d'une ou plusieurs strates d'échantillonnage.
On exclut les domaines quelconques.

I. PLAN DE SONDAGE STRATIFIÉ ET LIMITES DE L'ALLOCATION DE NEYMAN

Cette partie regroupe les résultats de la théorie des sondages qui servent de point de départ à ce document et qui y seront utilisés sans démonstration. Il s'agit de l'allocation de Neyman dans le cadre d'un plan de sondage stratifié par taille d'établissement et secteur d'activité, avec un tirage aléatoire simple dans chaque strate. Elle aborde ensuite les limites de cette allocation et diverses solutions pour y remédier.

I.1. ESTIMATEUR ET PRÉCISION DANS UN SONDAGE STRATIFIÉ

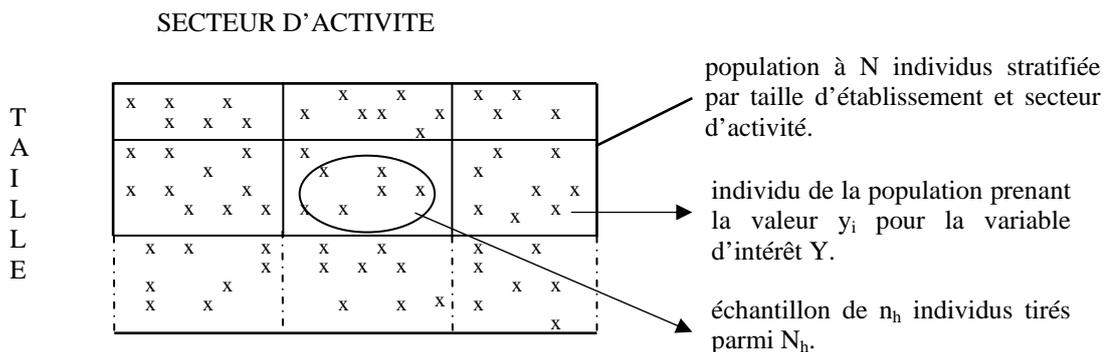
La variable d'intérêt à estimer est la moyenne globale \bar{Y} de la variable Y. Dans la base de sondage, la valeur prise par l'individu i pour cette variable est notée y_i .

Le principe de la stratification est d'utiliser les informations contenues dans la base de sondage afin d'améliorer la précision de l'estimateur. Pour cela, on divise la population de taille N en H groupes de tailles respectives N_1, N_2, \dots, N_H

telles que $N = \sum_{h=1}^H N_h$. Dans chacune des strates, on tire un échantillon de taille n_h par un sondage aléatoire simple tel

que $n = \sum_{h=1}^H n_h$.

Stratification des établissements français selon leur secteur d'activité et leur taille (en nombre de salariés)



Sondage aléatoire simple

La probabilité d'appartenir à un échantillon de taille n, sélectionné dans une population de N individus selon un tirage

aléatoire simple est de $\frac{n}{N}$.

L'estimateur sans biais d'Horvitz-Thompson du total global T de la variable d'intérêt Y est de la forme : $\hat{T} = \frac{\sum_{i \in \text{ech}} y_i}{P_i}$

avec P_i la probabilité d'inclusion dans l'échantillon, soit $\frac{n}{N}$.

L'estimateur de la moyenne de Y est par conséquent :

$$\bar{Y} = \frac{T}{N} \quad \text{donc} \quad \hat{\bar{Y}} = \frac{\hat{T}}{N} = \frac{\sum_{i \in \text{ech}} y_i}{n} = \bar{y}$$

L'estimateur de la moyenne globale de Y selon un sondage aléatoire simple est la moyenne \bar{y} dans l'échantillon sélectionné.

Cet estimateur est sans biais :

$$E(\bar{y}) = E\left(\frac{\hat{T}}{N}\right) = \frac{T}{N} = \bar{Y}$$

La variance de cet estimateur vaut :

$$V(\bar{y}) = \left(1 - \frac{n}{N}\right) * \frac{S^2}{n}$$

avec S^2 la variance de la variable Y dans la population totale.

Sondage stratifié avec un tirage aléatoire simple dans chaque strate

On effectue dans chaque strate, un tirage aléatoire simple sans remise et des tirages indépendants les uns des autres. On a :

$$\bar{Y} = \sum_{h=1}^H \frac{N_h}{N} * \bar{Y}_h$$

La moyenne globale de la variable Y est une moyenne pondérée des moyennes de chaque strate, où le poids représente l'importance de la strate en termes d'effectifs.

L'estimateur de la moyenne de Y est de la forme :

$$\hat{\bar{Y}} = \sum_{h=1}^H \frac{N_h}{N} * \hat{Y}_h = \sum_{h=1}^H \frac{N_h}{N} * \bar{y}_h$$

avec \bar{y}_h , l'estimateur sans biais de \bar{Y}_h , la moyenne de la variable Y dans la strate h, car on tire un échantillon dans chacune des strates selon un sondage aléatoire simple.

Cet estimateur de la moyenne globale de Y est sans biais :

$$E(\hat{\bar{Y}}) = E\left(\sum_{h=1}^H \frac{N_h}{N} * \bar{y}_h\right) = \sum_{h=1}^H \frac{N_h}{N} * \bar{Y}_h = \bar{Y}$$

Sa variance vaut :

$$V(\hat{\bar{Y}}) = V\left(\sum_{h=1}^H \frac{N_h}{N} * \bar{y}_h\right) = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 * \left(1 - \frac{n_h}{N_h}\right) * \frac{S_h^2}{n_h}$$

avec S_h^2 la variance de la variable Y dans la strate h.

Les termes de covariance entre les différentes strates h s'annulent car les tirages sont indépendants les uns des autres.

Interprétation

À la vue de l'expression de la variance de l'estimateur de Y, on constate que pour obtenir une estimation la plus précise possible, il suffit de choisir une stratification qui minimise les dispersions S_h^2 , autrement dit de choisir des strates homogènes en interne selon la variable Y.

Or la dispersion sur la population totale est une grandeur fixe et s'écrit sous la forme :

$$S^2 = \left[\sum_{h=1}^H \frac{(N_h - 1)}{N_h} * \frac{N_h}{N} * S_h'^2 + \sum_{h=1}^H \frac{N_h}{N} (\bar{Y}_h - \bar{Y})^2 \right] * \frac{N}{N-1}$$

avec $S_h'^2$ la variance corrigée de Y dans h.

Soit en considérant les N_h grands pour tout h :

$$S^2 \approx \underbrace{\sum_{h=1}^H \frac{N_h}{N} * S_h'^2}_{\text{variance intra-strates}} + \underbrace{\sum_{h=1}^H \frac{N_h}{N} * (\bar{Y}_h - \bar{Y})^2}_{\text{variance inter-strates}}$$

Cette formule, appelée formule de décomposition de la variance, montre que la dispersion globale S^2 est la somme de la variance intra-strates et inter-strates. S^2 étant une grandeur fixe, minimiser les variances intra-strates $S_h'^2$, revient à maximiser la variance entre les différentes strates.

Une bonne stratification s'obtient donc en formant des strates homogènes en interne (comportement moyen des individus de la strate semblable) et très différentes les unes des autres (comportement moyen au sein de chaque strate, le plus différent possible d'une strate à l'autre).

La prise en compte du taux de réponse à l'enquête

Si l'on tient compte du taux de réponse dans les strates h, la variance d'un estimateur se décline comme suit :

$$V(\hat{Y}) = V\left(\sum_{h=1}^H \frac{N_h}{N} * \bar{y}_h\right) = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 * \left(1 - \frac{t_h n_h}{N_h}\right) * \frac{S_h^2}{t_h n_h}$$

Par la suite pour ne pas alourdir les notations, nous considérons $t_h = 1$. Si l'on souhaite utiliser le taux de réponse dans les calculs, il suffit de remplacer n_h par le nombre effectif de réponses $n_h * t_h^2$.

² Tous ces résultats de la théorie des sondages sont développés dans : **Ardilly P.**, *Les techniques de sondage*, Editions technip, Paris, 2006.

I.2. ÉCHANTILLONNAGE : L'ALLOCATION OPTIMALE DE NEYMAN ET SES LIMITES

Une fois les strates clairement définies, il est nécessaire de déterminer le nombre d'unités à sélectionner au sein de chacune d'elles. Nous nous attacherons ici uniquement à l'échantillonnage optimal, sans détailler les autres solutions possibles, qui donnent des résultats en termes de précision moins bons que ce premier.

L'allocation optimale de Neyman

Dans un sondage stratifié, l'allocation de Neyman est celle qui minimise la variance de l'estimateur de la moyenne de Y, dans le cas d'un sondage aléatoire simple dans chacune des strates, pour un budget donné ou une taille d'échantillon fixée.

Formellement, on résout :

$$\left\{ \begin{array}{l} \text{Min } V(\widehat{Y}) = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 * \left(1 - \frac{n_h}{N_h}\right) * \frac{S_h^2}{n_h} \\ \text{s/c } \sum_{h=1}^H n_h = n \end{array} \right.$$

À l'aide d'un Lagrangien, on trouve les n_h de la forme :

$$n_h = \frac{N_h * S_h}{\sum_{h=1}^H N_h * S_h} * n$$

Le nombre d'unités tirées dans chaque strate h est proportionnel à la grandeur de la strate h en termes d'effectifs et à la dispersion de la variable d'intérêt dans cette strate.

Deux problèmes potentiels liés à l'allocation de Neyman

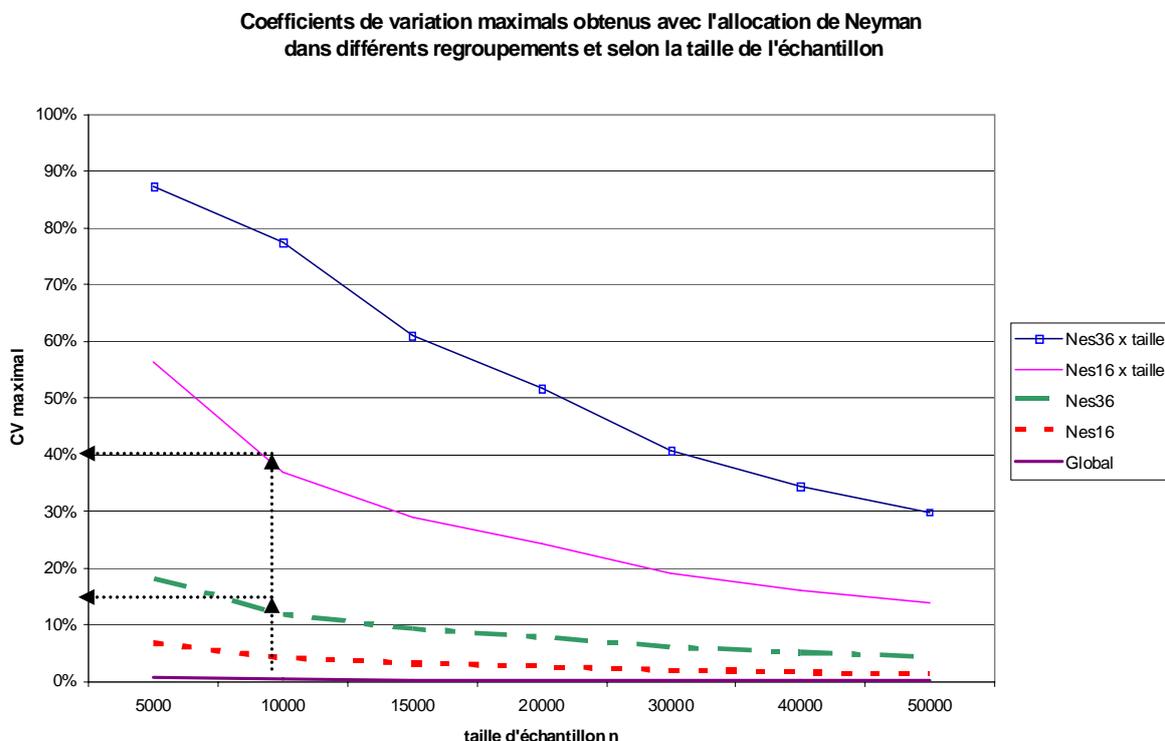
En pratique, on peut trouver un (ou plusieurs) n_h optimal(aux) supérieur(s) au nombre d'individus N_h présents dans la strate h. Ce premier problème peut être résolu en imposant dès le départ de la minimisation $n_h = N_h$ pour les strates concernées.

Le deuxième problème est plus conséquent : l'allocation de Neyman optimise la variance d'un estimateur portant sur *l'ensemble de la population* mais peut introduire des différences notables entre les variances *au niveau des regroupements de publication* : on ne maîtrise pas la précision dans ces regroupements. Or ce point est gênant lorsque, comme dans l'enquête trimestrielle ACEMO, on publie à la fois des indicateurs globaux, tels que les évolutions du salaire mensuel de base (SMB), du salaire horaire de base ouvrier (SHBO) au niveau de l'ensemble de la population, et des indicateurs locaux, tels que ces évolutions de salaires par catégorie socioprofessionnelle, par secteur d'activité et par taille d'établissement.

Le graphique ci-contre illustre cette deuxième limite

Le graphique ci-contre donne le plus mauvais coefficient de variation obtenu avec l'allocation de Neyman dans différents niveaux de publication (publication par secteur d'activité en 16 postes, en 36 postes, par secteur d'activité x taille d'établissement) et selon une taille d'échantillon fixée.

On se base sur les données issues des DADS concernant la masse salariale versée par les établissements français en 2004.



Note de lecture : pour un échantillon de taille n=10 000 établissements, le coefficient de variation maximal de notre estimateur dans les regroupements en nes16 x taille d'établissement est de 36,8 %.

L'allocation de Neyman donne un estimateur très précis au niveau global, le coefficient de variation reste inférieur à 1 % quel que soit l'échantillon de taille supérieure à 5 000. Par contre, si l'on souhaite publier les résultats en nes16, le plus mauvais CV d'un estimateur est de 12 % pour un échantillon de taille n=10 000. Pour des résultats dans un regroupement encore plus fin, comme en nes16 x taille d'établissement et pour une taille d'échantillon de 10 000, le plus mauvais CV atteint 36 %.

De manière générale, même avec une taille d'échantillon « raisonnable » telle que n=30 000³, la précision dans certains regroupements de publication n'est pas toujours satisfaisante à des niveaux fins d'agrégations.

L'objectif de Neyman consistant à minimiser la variance globale d'un estimateur revient à minimiser la somme pondérée par $(\frac{N_h}{N})^2$ des variances dans les différentes strates. Une première solution pour améliorer la précision locale de cette allocation est de pondérer différemment ces strates.

³ La taille d'échantillon n=30 000 établissements est celle utilisée dans le cadre de l'enquête trimestrielle ACEMO, à laquelle nous souhaitons appliquer la méthode permettant d'améliorer l'allocation de Neyman dans les regroupements de publication.

I.3. UNE SOLUTION DE COMPROMIS ENTRE PRÉCISION GLOBALE ET LOCALE

Une première possibilité pour tenir compte de la précision locale est de modifier les pondérations associées aux différentes strates, pour obtenir de meilleurs résultats que ceux de l'allocation de Neyman dans certains regroupements de publication. L'avantage de cette méthode est de préserver le caractère différentiable du problème à résoudre, qui est de la forme suivante :

$$\left\{ \begin{array}{l} \text{Min } \sum_{h=1}^H C_h * \left(1 - \frac{n_h}{N_h}\right) * \frac{S_h^2}{n_h} \\ \text{s/c } \sum_{h=1}^H n_h = n \text{ et } \sum_{h=1}^H C_h = 1 \end{array} \right.$$

avec C_1, \dots, C_H , un ensemble de coefficients explicitant l'importance que l'on souhaite attribuer à chacune des précisions dans les différentes strates.

On obtient alors des allocations n_h de la forme :

$$n_h = \frac{\sqrt{C_h} * S_h}{\sum_{h=1}^H \sqrt{C_h} * S_h} * n$$

L'allocation dans chaque strate dépend de la dispersion dans celle-ci et du poids que l'on souhaite lui attribuer.

Ce principe de pondérer différemment les strates a été retenu par M.D. Bankier

En 1988, M.D. Bankier (Statistiques Canada)⁴ réussit à définir une allocation qui offre un compromis entre une précision d'un estimateur au niveau de la population et une précision identique dans chacune des strates h.

Il minimise :

$$\left\{ \begin{array}{l} \sum_{h=1}^H [(X_h)^\alpha * CV(\hat{T}_h)]^2 \\ \text{s/c } \sum_{h=1}^H n_h = n \end{array} \right. \quad \text{soit} \quad \left\{ \begin{array}{l} \sum_{h=1}^H [(X_h)^\alpha * CV(\bar{y}_h)]^2 \\ \text{s/c } \sum_{h=1}^H n_h = n \end{array} \right.$$

$$\text{car } CV^2(\hat{T}_h) = \frac{V(\hat{T}_h)}{T_h^2} = \frac{V(N_h * \hat{Y}_h)}{(N_h * \bar{Y}_h)^2} = \frac{V(\hat{Y}_h)}{\bar{Y}_h^2} = CV^2(\hat{Y}_h) = CV^2(\bar{y}_h)$$

avec : \hat{T}_h l'estimateur du total de la variable étudiée dans la strate h

X_h une variable auxiliaire connue, qui mesure l'importance de la strate h
(par exemple l'effectif de la strate)

α un paramètre tel que $0 \leq \alpha \leq 1$ quel que soit h

$CV(\bar{y}_h)$ le coefficient de variation de l'estimateur de la moyenne dans h.

$$[CV(\bar{y}_h)]^2 = \frac{V(\bar{y}_h)}{\bar{Y}_h^2} = \frac{1 - n_h/N_h}{n_h} * \left(\frac{S_h}{\bar{Y}_h}\right)^2$$

⁴ **Bankier M.D.**, *Power Allocations : Determining sample Sizes for Sub-national Areas*, The American Statistician, 1988, vol.42 p.174-177.

On obtient une allocation appelée «power allocation », en référence à la puissance α , de la forme :

$$n_h = \frac{\frac{X_h^\alpha}{\bar{Y}_h} * S_h}{\sum_{h=1}^H \frac{X_h^\alpha}{\bar{Y}_h} * S_h} * n$$

Interprétation :

Si $\alpha = 1$ et $X_h = N_h * \bar{Y}_h = T_h$, le total de Y dans la strate h, on retrouve l'allocation de Neyman qui assure la meilleure précision globale.

Si $\alpha = 0$, l'allocation n_h est proportionnelle au vrai coefficient de variation dans h. Il est, en première approximation, peu sensible à h. L'allocation obtenue donne donc à peu près, la même précision dans chacune des strates h.

Si $0 < \alpha < 1$, on se trouve dans une situation de compromis entre la précision globale de Neyman et une même précision au niveau des strates, pour l'estimateur de la moyenne de Y.

I.4. AUTRES MÉTHODES EMPIRIQUES POUR AMÉLIORER LA PRÉCISION LOCALE D'UN ESTIMATEUR

Hormis la méthode de M.D. Bankier, on peut imaginer d'autres solutions pour améliorer la précision locale d'un estimateur. Nous en proposons ici quelques-unes.

Une amélioration « brutale »

On souhaite estimer la masse salariale brute distribuée par les établissements français au cours de l'année 2004 dans le secteur tertiaire marchand (EJ - EP). Pour cela, on sélectionne un échantillon de taille $n=5\ 000$ dans notre population stratifiée par taille d'établissement et par secteur d'activité.

Secteur d'activité	Base de sondage			Allocation de Neyman		Allocation à meilleure précision locale	
	N_h	S_h^2	$N_h * S_h$	n_h	CV	n_h	CV
Commerce : EJ	58 222	$5 * 10^{12}$	$13 * 10^{10}$	1 088	1,3 %	1 088	1,3 %
Transport : EK	16 781	$72 * 10^{12}$	$14 * 10^{10}$	519	1,5 %	519	1,5 %
Banques : EL	8 759	$236 * 10^{12}$	$13 * 10^{10}$	653	1,8 %	653	1,8 %
Immobilier : EM	4 043	$13 * 10^{12}$	$1 * 10^{10}$	132	4,6 %	252	2,5 %
Serv. aux entreprises : EN	40 275	$68 * 10^{12}$	$33 * 10^{10}$	2 120	1,1 %	2 000	1,3 %
Serv. aux particuliers : EP	24 890	$6 * 10^{12}$	$6 * 10^{10}$	488	2,9 %	488	2,9 %
Tertiaire au Global	152 970	$43 * 10^{12}$	$100 * 10^{10}$	5 000	0,68 %	5 000	0,71 %

Lors du calcul de la précision obtenue dans ces strates, l'allocation de Neyman donne une bonne très précision globale (CV=0,68 %), mais certaines strates comme le secteur de l'immobilier (EM) sont beaucoup moins précises (CV=4,6 %).

Une amélioration « brutale » de ce résultat (c'est-à-dire ne respectant pas la frontière d'efficacité), consiste à augmenter arbitrairement le nombre d'établissements sélectionnés dans cette strate au détriment de la strate la plus précise (les services aux entreprises EN avec un CV=1,1 %). On ajoute par exemple 120 établissements à la strate EM que l'on retranche à EN.

La précision dans le secteur de l'immobilier EM s'améliore alors de plus de 2 points alors qu'elle ne se dégrade que de 0,2 point dans les services aux entreprises. Pour autant la précision globale n'en est presque pas affectée : l'écart est de 0,03 point entre la précision de Neyman et celle de notre allocation choisie a posteriori.

Remarque : on constate ici que le minimum de la variance globale est un minimum plat, c'est-à-dire qu'il varie peu lorsque l'allocation reste dans un voisinage proche de celle de Neyman.

Une approche plus précise de cette méthode « brutale »

Cette méthode « brutale » pourrait encore être facilement améliorée, en fixant la taille de l'échantillon dans l'immobilier EM à 252 et en optimisant selon Neyman sur les autres secteurs hors immobilier. On aboutirait ainsi à une meilleure précision globale que celle obtenue avec la précédente méthode, car l'allocation trouvée serait sur la frontière d'efficacité.

Une méthode itérative « douce »

De même, on peut mettre en œuvre une méthode itérative « douce », ayant pour but de pondérer plus fortement les strates où la précision n'est pas satisfaisante. Elle consiste :

- à partir des poids de Neyman

$$\sum_j a_j^{(1)} V(\hat{Y}_j) \text{ avec } a_j^{(1)} \text{ de Neyman de la forme } \left(\frac{N_j}{N}\right)^2$$

- après optimisation, on augmente le poids de la plus mauvaise strate (en termes de précision) par exemple de 5 %

j_0 la plus mauvaise strate :

$$\left\{ \begin{array}{l} a_{j_0}^{(2)} = a_{j_0}^{(1)} * 1,05 \quad (\text{le poids augmente}) \\ a_{j, j \neq j_0}^{(2)} = a_j^{(1)} \quad (\text{les autres poids restent inchangés}) \end{array} \right.$$

- on recalcule l'optimisation et sa précision locale puis on réitère l'étape précédente jusqu'à ce que la précision soit satisfaisante dans tous les regroupements de publication.

$$\sum_j a_j^{(2)} V(\hat{Y}_j)$$

Le principe de cette méthode a été retenu par Joseph B. Kadane en 2005⁵. Il ajoute une unité à sélectionner dans une strate au détriment d'une autre, jusqu'à ce que la variance dans chacune des strates soit acceptable ou jusqu'à ce que la contrainte de coût soit saturée.

Cette méthode itérative « douce » assurerait d'être plus proche de la frontière d'efficacité entre précision globale et précision locale (cf. partie III.4). Cependant, elle est très coûteuse en temps et donc peu opérationnelle en pratique. D'où la nécessité de trouver une approche plus efficace du problème.

⁵ Kadane J.B., *Optimal Dynamic Sample Allocation Among Strata*, Journal of Official Statistics, 2005, vol. 21, N°4, p531-541.

II. OPTIMISATION DE NEYMAN AVEC CONTRAINTE DE PRÉCISION DANS LES REGROUPEMENTS DE PUBLICATION

Cette partie est consacrée à la présentation de notre méthode d'optimisation qui prend en compte la précision locale dans l'optimisation de Neyman. Tout au long de celle-ci, nous distinguons plusieurs degrés d'agrégation de la population étudiée :

- **les strates d'échantillonnage** (par exemple nes36 x taille d'établissement), desquelles sont tirées les unités interrogées
- **les regroupements de publication** (par exemple nes16), auxquels sont diffusés les résultats issus de l'enquête.

Hypothèse : on suppose pour simplifier, que chaque regroupement de publication est la réunion d'une ou plusieurs strates d'échantillonnage (on exclut les domaines quelconques).

II.1. FORMALISATION DU PROBLÈME A RÉSOUDRE

Pour une taille d'échantillon n fixée, on se propose de résoudre le programme suivant :

$\left\{ \begin{array}{l} \text{Min } V(\hat{Y}) \\ \text{s/c } \sum_{h=1}^H n_h = n \\ \text{s/c } n_h \leq N_h \\ \text{s/c } \text{Max}_{p \in \text{pub}} CV_p \leq CV_{\text{seuil}} \end{array} \right.$	$\begin{array}{l} (1) \\ (2) \\ (3) \\ (4) \end{array}$
--	---

L'optimisation (1) et la contrainte (2) correspondent au problème de Neyman. Les contraintes (3) et (4) sont nouvelles par rapport à l'approche classique :

- **les contraintes (3)** permettent de s'assurer que le nombre d'unités tirées dans chaque strate est bien inférieur au nombre d'unités présentes dans celle-ci
- **les contraintes (4)** sont les contraintes de précision locale. Elles expriment que le plus grand des coefficients de variation dans les regroupements de publication doit être inférieur à un seuil de précision que l'on s'est fixé à l'avance. C'est la véritable nouveauté qu'apporte cette étude.

Ce problème n'est plus linéaire, ni différentiable à cause de la fonction "maximum". Mais celui-ci est des plus intéressants, puisqu'il résout entièrement la question du dilemme entre la précision globale et la précision locale d'un estimateur.

Nous nous proposons donc de trouver une solution exacte de façon empirique et d'illustrer cette solution à l'aide du concept de frontière d'efficacité entre deux objectifs, en référence à la notion d'efficacité au sens de Pareto en microéconomie.

II.2. SCHEMA GLOBAL DE LA RESOLUTION DU PROBLEME EN DEUX ETAPES

Voici les différentes étapes que nous allons suivre tout au long du document pour résoudre le programme d'optimisation décrit en partie II.1.

Étape 1 :

On traduit les contraintes de type (4) de précision locale dans les regroupements de publication en contraintes portant sur les strates d'échantillonnage :

$$CV_p \leq CV_{seuil} \Leftrightarrow \{n_h \geq n_{h \min}(CV_{seuil})\}_{h \in p}$$

Cette équivalence exprime le fait que fixer un seuil de précision dans un regroupement de publication p revient à tirer un nombre minimal de questionnaires dans chaque strate d'échantillonnage qui compose p.

Pour cela :

Étape 1-1 : on démontre l'équivalence ci-dessus.

Étape 1-2 : on explique comment pour une précision CV_{seuil} donnée dans les regroupements de publication, on obtient la valeur des $n_{h \min, h \in p}$ (nombre minimal de questionnaires à sélectionner dans les strates d'échantillonnage) correspondante.

Étape 2 :

Le problème est désormais du type :

$$\left\{ \begin{array}{l} \text{Min}_{n_1, \dots, n_H} V(\hat{Y}) = \sum_h \left(\frac{N_h}{N} \right)^2 \cdot \underbrace{\left(1 - \frac{n_h}{N_h} \right) \cdot \frac{S_h^2}{n_h}}_{\text{var } h} \quad (1) \\ \text{s/c } \sum_{h=1}^H n_h = n \quad (2) \\ \text{s/c } (n_{h \min}(CV_{seuil}) \leq n_h \leq N_h)_{h=1 \dots H} \quad (3+4) \end{array} \right.$$

(4) (3)

où $n_{h \min}(CV_{seuil})$ est supposé connu.

On le résout en montrant les propositions suivantes :

Étape 2-1 : si on connaît à l'avance l'ensemble des contraintes saturantes⁶, on se ramène à un problème classique de Neyman sur les seules strates non saturées.

Étape 2-2 : on peut déterminer l'ordre de saturation des contraintes quand la taille globale de l'échantillon n augmente.

Étape 2-3 : connaissant l'étape 2-2, on peut effectivement déterminer pour chaque taille d'échantillon n, la liste des contraintes saturantes, ce qui en vertu de l'étape 2-1 résout le problème.

⁶ Une contrainte de type (3+4) sature lorsque l'une des deux inégalités est une égalité :

$n_h = n_{h \min}(CV_{seuil})$ ou $n_h = N_h$.

II.3. TRADUCTION PLUS SIMPLE DES CONTRAINTES DE PRÉCISION LOCALE (ÉTAPE 1)

On s'occupe ici à développer l'étape 1 du programme de résolution. On montre qu'une contrainte de précision locale dans *les regroupements de publication* (type (4)) se traduit par un ensemble de contraintes qui portent sur *les strates d'échantillonnage*.

Étape 1-1 : on démontre l'équivalence $CV_p \leq CV_{seuil} \Leftrightarrow \{n_h \geq n_{h \min}(CV_{seuil})\}_{h \in p}$,

c'est-à-dire qu'une précision souhaitée dans un regroupement de publication p, implique de tirer un nombre minimal de questionnaires dans chaque strate d'échantillonnage qui le compose. Pour cela, on montre les faits suivants :

Fait 1 : la précision CV_p obtenue dans un regroupement de publication ne dépend que des $n_{h, h \in p}$.

Fait 2 : l'optimisation globale est équivalente à des optimisations dans chaque regroupement de publication.

Preuve du fait 1 :

Dans un sondage stratifié uniforme dans chaque strate d'échantillonnage, et sous notre hypothèse qu'un regroupement de publication est la réunion d'une ou plusieurs strates d'échantillonnage, on a :

$$CV_p^2 = \frac{\sum_{h \in p} \left(\frac{N_h}{N} \right)^2 * \left(1 - \frac{n_h}{N_h} \right) * \frac{S_h^2}{n_h}}{\bar{Y}_p^2}$$

La précision CV_p ne dépend que de l'ensemble des $n_{h, h \in p}$.

Pour $n_{h, h \in p}$ fixé, il existe une unique valeur de CV_p .

Preuve du fait 2 :

On se place dans l'hypothèse où les regroupements de publication sont chacun un ensemble de strates d'échantillonnage. On montre alors que l'optimisation globale sous contrainte ci-dessous que l'on veut résoudre est à variables séparées selon les regroupements de publication (démonstration en annexe A).

$$\left\{ \begin{array}{ll} \text{Min}_{n_1, \dots, n_H} V(\hat{Y}) & (1) \\ \text{s/c } \sum_h n_h = n & (2) \\ \text{s/c } n_h \leq N_h & (3) \\ \text{s/c } \text{Max}_{p \in \text{pub}} CV_p \leq CV_{seuil} & (4) \end{array} \right. \quad \longrightarrow \quad \begin{array}{l} \text{Problème à variables séparées} \\ \text{selon} \\ \text{les regroupements de publication} \end{array}$$

L'optimisation globale est alors équivalente à des optimisations dans chaque regroupement de publication p. Pour une valeur de CV_{seuil} fixée dans un regroupement de publication, il existe une solution $(n_1, \dots, n_h)_{h \in p}$ unique qui résout le problème d'optimisation dans p.

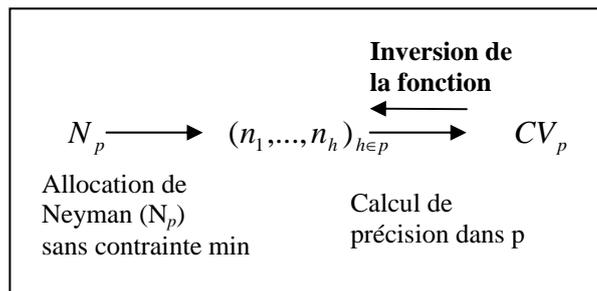
Finalement, la contrainte portant sur un regroupement de publication $CV_{pub} \leq CV_{seuil}$ peut se ramener à une liste de contraintes portant sur les strates d'échantillonnage du type $\{n_h \geq n_{h \min}(CV_{seuil})\}_{h \in p}$. L'étape suivante va permettre de déterminer la valeur des $n_{h \min}(CV_{seuil})$.

Étape 1-2 : à CV_{seuil} fixé dans le regroupement de publication p, on détermine le nombre minimal de questionnaires à tirer $n_{h\min}$ dans les strates d'échantillonnage appartenant à p

On souhaite déterminer les $n_{h\min}$ (CV_{seuil}) correspondants aux contraintes de type (3+4). Pour ce faire, le principe consiste à inverser la fonction qui à un vecteur $(n_1, \dots, n_h)_{h \in p}$ attribue une précision CV_p . On obtient ainsi une table de données liant la précision dans un regroupement de publication et le nombre d'unités à tirer dans chaque strate d'échantillonnage qui le compose.

Principe :

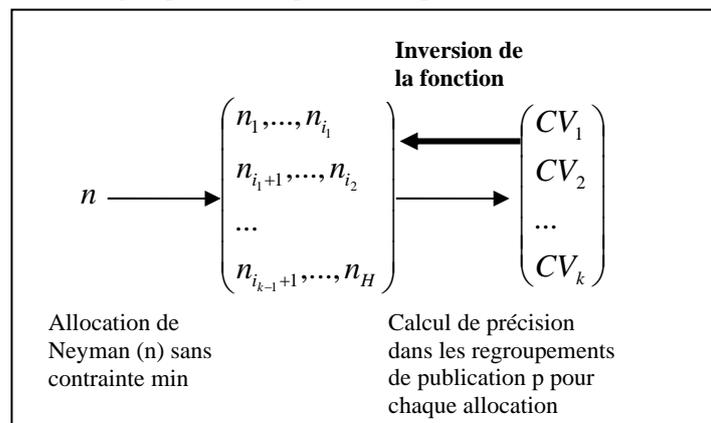
On se place dans un regroupement de publication p et dans le cas d'une allocation de Neyman. On inverse cette fonction au sein de la strate p, grâce au fait 2 établi à l'étape 1.1. (« l'optimisation globale est équivalente à des optimisations dans chaque strate de publication ») qui démontre l'existence une solution unique $(n_1, \dots, n_h)_{h \in p}$ pour un CV_p donné.



- On calcule l'allocation de Neyman pour une liste exhaustive de valeurs de tailles d'échantillon N_p . Pour N_p donné, on obtient ainsi les n_h optimaux, pour toutes les strates d'échantillonnage composant le regroupement de publication p, où l'on s'est placé.
- Pour toutes ces valeurs de N_p , on calcule la précision CV_p dans le regroupement de publication p.
- **Les deux points précédents permettent ainsi d'obtenir une table de données liant la précision dans le regroupement de publication CV_p et le nombre d'unités à tirer dans chaque strate d'échantillonnage qui le compose.** En fixant le *seuil* de précision souhaitée, on obtient ainsi la valeur du vecteur $n_{h\min} = (n_{1\min}, n_{2\min}, \dots, n_{k\min})$, nombre minimal d'unités à interroger dans chaque strate d'échantillonnage de p, pour atteindre la précision CV_{seuil} dans le regroupement de publication.
- On applique cette méthode à tous les regroupements de publication.

En pratique :

On calcule l'allocation de Neyman pour une taille d'échantillon globale n qui varie de 1 000 en 1 000. On calcule ensuite la précision obtenue dans les regroupements de publication pour ces différentes valeurs de n.



En fusionnant ces deux tables sur le regroupement de publication (exemple le secteur d'activité) et la taille d'échantillon n, on obtient une table de données liant la précision dans les regroupements de publication CV_p et le nombre d'unités n_h à tirer dans chaque strate d'échantillonnage qui les composent.

Illustration par un cas pratique :

On se place dans un regroupement de publication (EB) composé de 6 strates d'échantillonnage (nes36 x 6 tailles d'établissement). On souhaite déterminer le nombre minimal de questionnaires à tirer dans chaque strate d'échantillonnage (B0 x taille1) (B0 x taille2) (B0 x taille3) (B0 x taille4) (B0 x taille5) et (B0 x taille6) pour obtenir un CV d'au plus 5 % dans EB. Pour cela, on crée une table de passage entre le CV et le nombre d'unités n_h à sélectionner dans les strates d'échantillonnage.

Table 1 : on calcule l'allocation de Neyman pour différentes tailles d'échantillon en faisant varier n de 1 000 en 1 000 :

Regroupement de publication Nes16	Strates d'échantillonnage h : Nes36 x taillea	N_h	Allocation de Neyman n_h	n
EB	B0x1	4 080	3	1 000
EB	B0x2	2 075	4	1 000
EB	B0x3	854	4	1 000
EB	B0x4	709	7	1 000
EB	B0x5	219	3	1 000
EB	B0x6	78	8	1 000

EB	B0x1	4 080	20	5 000
EB	B0x2	2 075	20	5 000
EB	B0x3	854	22	5 000
EB	B0x4	709	43	5 000
EB	B0x5	219	19	5 000
EB	B0x6	78	47	5 000

EB	B0x1	4 080	46	10 000
EB	B0x2	2 075	47	10 000
EB	B0x3	854	51	10 000
EB	B0x4	709	99	10 000
EB	B0x5	219	44	10 000
EB	B0x6	78	78	10 000

EB	B0x1	4 080	288	50 000
EB	B0x2	2 075	296	50 000
EB	B0x3	854	322	50 000
EB	B0x4	709	622	50 000
EB	B0x5	219	219	50 000
EB	B0x6	78	78	50 000

Table 2 : on calcule ensuite la précision obtenue dans le regroupement de publication (EB) pour ces différentes allocations :

Calcul de précision des n_h dans les regroupements de publication
 Sélection des n_h par $CV_{seuil} = 5\%$

Nes16	n	CV_{EB}
EB	1 000	11,3%
EB	2 000	7,8%
EB	3 000	6,2%
EB	4 000	5,1%
EB	5 000	4,3%
EB	10 000	2,6%
EB	15 000	2,0%
EB	20 000	1,6%
EB	25 000	1,4%
EB	30 000	1,2%
EB	35 000	1,0%
EB	40 000	0,9%
EB	45 000	0,8%
EB	50 000	0,7%

Note de lecture :

Si l'on souhaite un coefficient de variation d'au plus 5 % dans le secteur EB, on choisit dans la table 2 un CV_{EB} de 4,3 %, premier seuil de précision calculé sous la barre des 5 %. Il suffit ensuite de prendre les $n_{h_{min}}$ correspondant à ce CV dans la table 1, soit des $n_{h_{min}}$ égaux à (20, 20, 22, 43, 19, 47) pour les strates d'échantillonnage qui composent EB et pour n=5 000.

Finalement, en appliquant cette méthode à tous les secteurs d'activité, on obtient l'ensemble des $n_{h_{min}}$ nécessaires pour respecter notre seuil de précision de 5 % dans tous les regroupements de publication. Ces $n_{h_{min}}$ vont ensuite être injectés dans notre programme d'optimisation qui ne compte désormais plus que des contraintes portant sur les n_h .

II.4. RÉOLUTION DE L'OPTIMISATION SOUS CONTRAINTE (ÉTAPE 2)

La traduction plus simple des contraintes de type (4), nous amène à la résolution du programme suivant :

$$\left\{ \begin{array}{l} \text{Min}_{n_1, \dots, n_H} V(\hat{Y}) = \sum_h \left(\frac{N_h}{N} \right)^2 \cdot \underbrace{\left(1 - \frac{n_h}{N_h} \right) \cdot \frac{S_h^2}{n_h}}_{\text{var } h} \quad (1) \\ \text{s/c } \sum_{h=1}^H n_h = n \quad (2) \\ \text{s/c } (n_{h \min}(CV_{\text{seuil}}) \leq n_h \leq N_h)_{h=1 \dots H} \quad (3+4) \end{array} \right.$$

(4) (3)

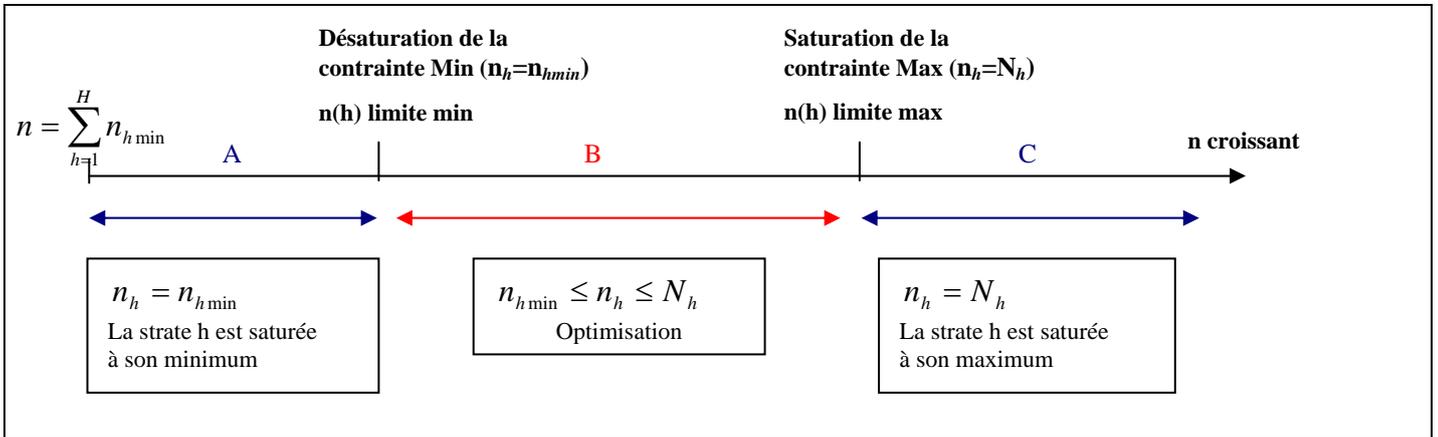
avec $n_{h \min}(CV_{\text{seuil}})$ connu quelle que soit la valeur de h grâce à l'étape 1, la valeur de CV_{seuil} étant fixée une fois pour toute ; N_h connu pour tout h par les caractéristiques de la population.

Afin de résoudre ce problème d'optimisation désormais simplifié, on s'attache dans cette étape 2, à étudier les contraintes de type (3+4). Pour cela, on fait varier la taille d'échantillon n et on visualise la désaturation et saturation⁷ successive de ces contraintes lorsque n augmente.

Le but est de connaître, en fonction de la taille d'échantillon n retenue pour notre échantillonnage final, l'ensemble des strates saturées, pour ensuite appliquer l'optimisation de Neyman aux seules strates non saturées.

Visualisation de la contrainte (3+4) liée à une strate h lorsque n augmente :

$$(n_{h \min}(CV_{\text{seuil}}) \leq n_h \leq N_h)$$



Les contraintes (4) ne peuvent être satisfaites que si $n \geq \sum_{h=1}^H n_{h \min}$.

Trois cas de figure apparaissent successivement pour la strate h . Sur la première plage de valeur de n , notée A, la contrainte Min est saturée : $n_h = n_{h \min}$. Elle désature pour une valeur $n(h)$ limite min. Dans la plage B, la valeur n_h résulte de l'optimisation. Enfin, la contrainte Max sature pour une valeur $n(h)$ limite max et la valeur de n_h est alors fixée par $n_h = N_h$ dans la plage C.

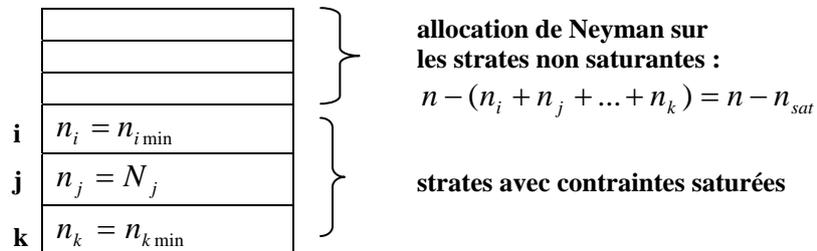
⁷ On rappelle qu'une contrainte de type (3+4) sature lorsque l'une des deux inégalités est une égalité :

$$n_h = n_{h \min}(CV_{\text{seuil}}) \text{ ou } n_h = N_h.$$

Étape 2-1 : si l'on connaît la liste des contraintes saturées...

Si l'on connaît à l'avance la liste des contraintes saturées du système d'équations (3+4), le problème se ramène à une allocation de Neyman classique. En effet, il suffit alors de faire une allocation, sans contrainte de type (3+4), en se restreignant aux strates non saturées (puisque Neyman donne la meilleure allocation possible sans contrainte) et en modifiant la taille de l'échantillon de manière adéquate.

Le schéma de strates d'échantillonnage suivant illustre ce principe :



Les strates i, j et k sont saturées ; i et k sont saturées à leur minimum n_{\min} et j à son maximum N_j .

On fixe donc n_i, n_j, n_k égales à ces valeurs. Il suffit ensuite de calculer l'allocation de Neyman sur les autres strates non saturantes avec une taille d'échantillon de $n - (n_i + n_j + \dots + n_k) = n - n_{sat}$, avec n_{sat} désignant la taille de l'échantillon composé par l'ensemble des strates saturantes.

Étape 2-2 : l'ordre de saturation des contraintes quand la taille globale de l'échantillon augmente peut être déterminé simplement

On cherche à montrer ici que l'on peut déterminer l'ordre de saturation des contraintes à l'aide de paramètres connus au niveau de la strate $(n_{h \min}, N_h, S_h)$. Plus précisément, on montre que l'ordre de saturation est donné par l'ordre des quantités suivantes :

$$\boxed{Ordre_{h,\max} = \frac{N_h}{N_h * S_h}} \quad \text{et} \quad \boxed{Ordre_{h,\min} = \frac{n_{h \min}}{N_h * S_h}}$$

$Ordre_{h,\max}$ donne l'ordre de saturation des contraintes Max ($n_h = N_h$), et $Ordre_{h,\min}$ donne l'ordre de désaturation des contraintes Min ($n_h = n_{h \min}$).

Lorsque la valeur de n augmente, la première contrainte qui sera saturée est celle pour laquelle $Ordre_{h,\max}$ est la plus petite et la dernière celle pour laquelle $Ordre_{h,\max}$ est la plus grande. Il en va de même pour la désaturation des contraintes donnée par $Ordre_{h,\min}$ (car ces deux quantités sont fonction croissantes de n, cf. équation (*) ci-dessous).

Preuve :

les n_h optimaux pour les strates non saturantes sont de la forme :

$$n_h = \frac{N_h * S_h}{\sum_{h \in H_{\text{nonsat}}} N_h * S_h} * (n - n_{\text{sat}}) \quad \text{avec } h \text{ strate non saturante.}$$

La strate h_0 sature lorsque :

$$n_{h_0} = \frac{N_{h_0} * S_{h_0}}{\sum_{h \in H_{\text{nonsat}}} N_h * S_h} * (n - n_{\text{sat}}) = N_{m_0} \quad \text{avec } N_{m_0} = n_{h_0 \min} \text{ ou } N_{h_0} \text{ selon le cas envisagé}$$

$$\text{soit } \frac{N_{m_0}}{N_{h_0} * S_{h_0}} = \frac{(n - n_{\text{sat}})}{\sum_{h \in H_{\text{nonsat}}} N_h * S_h} \quad (*) \quad \text{et}$$

$$\boxed{n(h_0)_{\text{limite}} = \underbrace{n_{\text{sat}}}_{\text{Indépendant de } h_0} + \underbrace{\frac{N_{m_0}}{N_{h_0} * S_{h_0}} * \sum_{h \in H_{\text{nonsat}}} N_h * S_h}_{\text{Indépendant de } h_0}} \quad (**)$$

On détermine ainsi la taille d'échantillon limite $n(h_0)_{\text{limite}}$, pour laquelle la contrainte de la strate h_0 sature (ou désature selon le cas envisagé). Or, le seul terme qui dépend de h_0 dans cette égalité est la variable $Ordre$. Il suffit donc de connaître sa valeur pour chacune des strates d'échantillonnage, pour en déduire immédiatement dans quel ordre les contraintes saturent lorsque la taille de l'échantillon globale n augmente.

Étape 2-3 : détermination de la taille d'échantillon n pour laquelle chaque contrainte sature

Pour chaque contrainte, on calcule la taille d'échantillon n qui sature cette contrainte. Pour ce faire, il suffit d'utiliser l'équation (**) et l'ordre de saturation donnés par l'étape 2-2.

Exemple illustratif avec 3 strates

On considère 3 strates avec les caractéristiques suivantes :

Strate	n_{hmin}	N_h	S_h
1	2	8	1/4
2	9	30	1/10
3	32	48	1/12

Les contraintes de notre programme d'optimisation de la variance globale sous contrainte de précision locale sont : $(2 \leq n_1 \leq 8)$ et $(9 \leq n_2 \leq 30)$ et $(32 \leq n_3 \leq 48)$.

Calculons les valeurs de la variable *Ordre*, pour définir l'ordre de saturation des 6 contraintes.

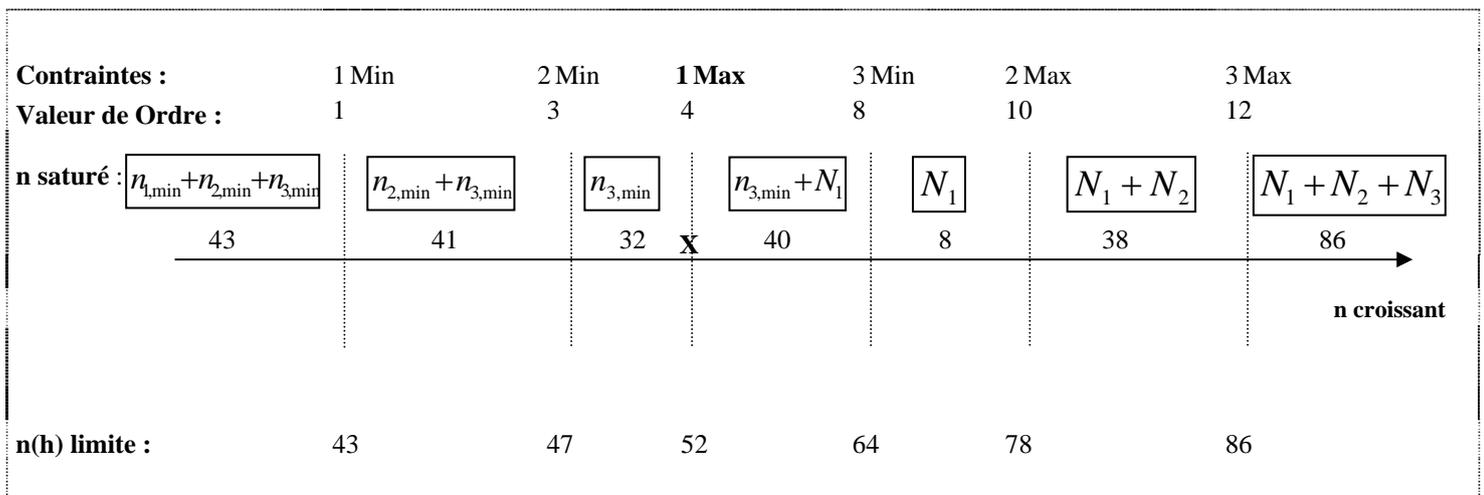
Contrainte	Valeur de la variable ordre	Ordre de saturation
1 Min	1	1
1 Max	4	3
2 Min	3	2
2 Max	10	5
3 Min	8	4
3 Max	12	6

Exemple détaillé pour la contrainte 1 :

$$Ordre_{1min} = n_{1min} / (N_1 * S_1) \\ = 2 / (8 * 1/4) = 1$$

$$Ordre_{1max} = N_1 / (N_1 * S_1) = 1/S_1 \\ = 4$$

Visualisation de l'ordre de saturation des contraintes



En connaissant l'ordre de saturation des contraintes, on peut calculer pas à pas la valeur de n_{sat} , la taille de l'échantillon saturé.

Pour une taille d'échantillon globale n assez grande permettant de respecter les contraintes de précision locale, on a tout d'abord l'ensemble des contraintes saturées à leur minimum ($n_{sat} = n_{1min} + n_{2min} + n_{3min} = 43$). Puis lorsque n augmente, la contrainte 1 Min désature la première, n_{sat} vaut alors $n_{sat} = n_{2min} + n_{3min} = 41$. La contrainte 2 Min désature en deuxième, n_{sat} vaut alors $n_{sat} = n_{3min} = 32$, et ainsi de suite.

Ensuite, connaissant la valeur de la variable *Ordre* et celle de n_{sat} on peut, à l'aide de l'équation (), déterminer la taille d'échantillon $n(h)_{limite}$, pour laquelle les différentes contraintes désaturent etaturent.**

On se place par exemple lors de la saturation de la contrainte 1 Max (par la gauche) : la seule strate saturée à ce moment est la strate 3 : $n_3 = n_{3min} = n_{sat} = 32$. Par conséquent, on peut déterminer la taille d'échantillon $n(1)_{limite\ max}$, pour laquelle la contrainte 1ature à son maximum :

$$n(1)_{limite\ max} = n_{sat} + \frac{N_1}{N_1 * S_1} * (N_1 * S_1 + N_2 * S_2) = 32 + 4 * [(8 * 1/4) + (30 * 1/10)] = 52$$

Remarque : par la droite : $n_{sat} = n_{3min} + N_1 = 40$ et

$$n(1)_{limite\ max} = n_{sat} + \frac{N_1}{N_1 * S_1} * (N_2 * S_2) = 40 + 4 * (30 * 1/10) = 52.$$

Finalement, en fonction de la taille d'échantillon *n* retenue pour notre échantillonnage final, on peut connaître l'ensemble des strates saturées, et appliquer l'optimisation de Neyman aux seules strates non saturées.

Si l'on souhaite tirer un échantillon de taille $n=50$, la strate 3 sera saturée à son minimum et $n_3 = n_{3min} = n_{sat} = 32$. Il suffira de réaliser une allocation de Neyman sur les strates 1 et 2 non saturantes pour cette valeur de n , avec une taille d'échantillon de $50-32=18$.

En conclusion, pour obtenir une allocation de Neyman dont les résultats sont meilleurs localement pour une taille d'échantillon globale n , il suffit de repérer les strates saturées et de faire une allocation de Neyman sur les strates non saturantes, avec une taille d'échantillon n , de laquelle on ôte la taille de l'échantillon saturé. Ne reste qu'à observer les résultats numériques de cette méthode sur un cas concret.

III. CAS PRATIQUE DE LA MÉTHODE D'OPTIMISATION SOUS CONTRAINTE

On se propose ici de comparer les résultats de l'allocation de Neyman à ceux de l'allocation sous contrainte de précision locale lors d'un tirage d'échantillon. Cette partie est également illustrée par différents graphiques, notamment celui de la taille d'échantillon minimale nécessaire pour atteindre la précision CV_{seuil} fixée et celui du concept de frontière d'efficacité, dont le graphe constitue la synthèse des résultats issus de la méthode d'optimisation avec contraintes.

III.1. BASE DE SONDAGE ET DESCRIPTION DU PROGRAMME DE RÉOLUTION

On souhaite estimer la masse salariale versée par les établissements français en 2004 à partir d'un échantillon de 10 000 établissements avec un seuil de précision $CV_{seuil} = 2\%$, 3% ou 5% dans les regroupements de publication.

La base de sondage est une table des déclarations annuelles de données sociales (DADS 2004). Cette table donne pour chaque établissement français (entre autre) son APE et la masse salariale brute distribuée au cours de l'année 2004. La population est stratifiée par secteur d'activité en nes36 et par taille d'établissement en six modalités selon les tranches : (1) 10-19 salariés, (2) 20-49 salariés, (3) 50-99 salariés, (4) 100-249 salariés, (5) 250-499 salariés, (6) 500 salariés et plus. Cette base de sondage est complétée par le calcul des statistiques par strate d'échantillonnage (nes36 x taille d'établissement), telles que le nombre d'unités présentes N_h , la variance S_h^2 et la masse salariale brute par strate. Ces statistiques sont les seuls paramètres utilisés dans le processus d'optimisation.

Extrait de la base de sondage : établissements stratifiés selon leur secteur d'activité et leur taille

Secteur d'activité Nes16 Nes36		Taille	S_h^2	N_h	Masse salariale brute Y_h
EB	B0	1	$3,7 \times 10^{10}$	4 080	1 012 874 000
...					
EH	H0	1	$1,6 \times 10^{10}$	17 109	5 097 404 000
EH	H0	2	$8,5 \times 10^{10}$	9 116	6 611 719 000
EH	H0	3	$29,7 \times 10^{10}$	1 926	3 270 161 000
EH	H0	4	$438,9 \times 10^{10}$	780	3 033 315 000
EH	H0	5	$1564,7 \times 10^{10}$	101	1 036 341 000
EH	H0	6	$21121,4 \times 10^{10}$	35	898 759 000
...					
ER	R2	6	$13889,9 \times 10^{10}$	12	185 627 000

Notre programme de résolution appliqué à cette base de données comporte deux phases.

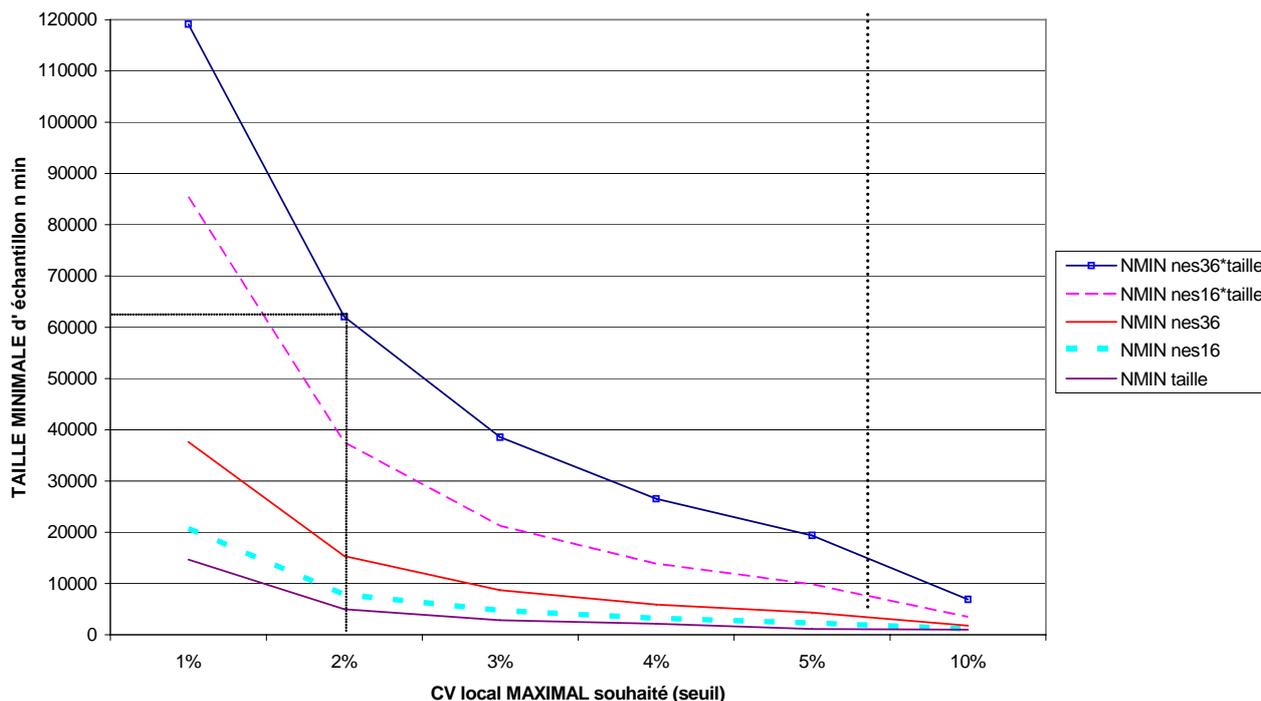
La première consiste en une série d'optimisations classiques de Neyman (intégrant seulement les contraintes Max ($n_h \leq N_h$)), obtenue en faisant varier la taille de l'échantillon n . Elle a pour but de créer la table de correspondance, au niveau de chaque regroupement de publication, entre la précision voulue CV_{seuil} et le nombre minimal d'unités $n_{h \min}(CV_{seuil})$ nécessaires dans chaque strate d'échantillonnage.

Ces relations sont ensuite reprises sous la forme de contraintes dans la deuxième phase concernant l'optimisation sous contrainte de précision locale. On y détermine la valeur des $n_{h \min}(CV_{seuil})$ puis on calcule les tailles d'échantillon limites conduisant à la saturation des contraintes de la forme ($n_{h \min}(CV_{seuil}) \leq n_h \leq N_h$) de chaque strate d'échantillonnage. Enfin, on calcule l'allocation optimale $(n_h)_{h=1, \dots, H}$ sous ces contraintes de précision locale pour une taille d'échantillon n donnée. Pour cela, on prend en compte les strates saturées pour cette taille d'échantillon n en les fixant à leur valeur de saturation et on fait une optimisation de Neyman sur les strates non saturantes (cf. annexes B et C pour la description et le code des macros utilisées lors de la résolution du programme d'optimisation).

III.2. ILLUSTRATION : TAILLE MINIMALE D'ÉCHANTILLON NÉCESSAIRE POUR UNE PRÉCISION LOCALE SOUHAITÉE

Cette partie illustre notre méthode d'optimisation, en donnant pour une précision locale fixée par la valeur CV_{seuil} , la taille d'échantillon $n_{min} = \sum_h n_{h min}(CV_{seuil})$ nécessaire pour l'atteindre dans tous les regroupements de publication.

Taille minimale d'échantillon nécessaire pour obtenir une précision donnée dans différents regroupements



La taille minimale d'échantillon nécessaire pour respecter une précision minimale CV_{seuil} dépend fortement du niveau d'agrégation des regroupements de publication. Pour un coefficient de variation d'au plus 2 % dans chaque taille d'établissement, il suffit d'un échantillon de taille $n=5\,000$, alors qu'il est nécessaire de tirer plus de 62 000 établissements pour obtenir ce même niveau de précision en nes36 x taille d'établissement.

La contrainte de précision locale est très exigeante entre les seuils de 1 % à 2 % de précision, puisque l'échantillon nécessaire au respect de cette contrainte est plus que doublé entre ces deux seuils.

Exemple : pour une publication en Nes16

$n=7\,900$ pour une précision minimale de 2 % dans chaque nes16

$n=20\,700$ pour une précision minimale de 1 % dans chaque nes16.

Remarque : la précision au niveau globale, non représentée ici, est inférieure à 2 % quelle que soit la taille d'échantillon supérieure à 1 000 et inférieure à 1 % pour une taille d'échantillon supérieure à 3 000 unités.

III.3. COMPARAISON DE L'ALLOCATION DE NEYMAN ET DE L'ALLOCATION SOUS CONTRAINTE DE PRÉCISION LOCALE

Nous évaluons l'efficacité de l'allocation contrainte en comparant sa précision avec celle de l'allocation classique de Neyman, pour différents niveaux de publication.

Taille d'échantillon n=10 000 et publication par secteur d'activité (Nes16)

Nes16	CV Neyman	CV Allocation contrainte au seuil de 2%
Global	0,44%	0,46%
EB	2,6%	2,0%
EC	1,9%	1,9%
ED	1,4%	1,5%
EE	1,5%	1,7%
EF	1,3%	1,4%
EG	2,6%	2,0%
EH	1,8%	1,9%
EJ	1,2%	1,4%
EK	1,4%	1,5%
EL	1,6%	1,8%
EM	4,2%	2,0%
EN	1,0%	1,1%
EP	2,7%	2,0%
EQ	1,6%	1,8%
ER	2,2%	2,0%

Pour un échantillon de taille n=10 000, l'allocation de Neyman donne de bons résultats tant au niveau global qu'au niveau des regroupements de publication en Nes16. L'allocation contrainte permet néanmoins de diviser par deux le coefficient de variation de certains petits secteurs, comme celui de l'immobilier (EM). L'augmentation du CV global qui en résulte est faible (+0,02 point).

L'efficacité de l'allocation contrainte augmente lorsque la taille de l'échantillon est faible, car l'allocation de Neyman donne alors de moins bons résultats au niveau local.

Taille d'échantillon n=10 000 et publication par secteur d'activité (Nes36)

Nes36	CV Neyman	CV Allocation contrainte au seuil de 3%	n _{strate_pub} Neyman	n _{strate_pub} Allocation contrainte au seuil de 3%	DELTA
GLOBAL	0,44%	0,49%	10 000	10 000	/
B0	2,6%	3,0%	365	298	-67
C1	6,5%	2,9%	49	187	138
C2	3,6%	2,9%	136	184	47
C3	2,7%	2,9%	163	152	-12
C4	4,1%	3,0%	87	150	63
D0	1,4%	1,7%	147	129	-18
E1	2,0%	2,5%	88	77	-11
E2	2,3%	3,0%	260	173	-87
E3	3,0%	2,8%	238	259	21
F1	3,8%	3,0%	88	132	44
F2	6,0%	3,0%	46	161	115
F3	4,2%	3,0%	95	176	81
F4	2,3%	2,8%	260	200	-61
F5	2,5%	2,8%	257	210	-47
F6	3,0%	3,0%	154	154	0
G1	4,2%	2,9%	22	28	6
G2	3,0%	3,0%	177	177	0
H0	1,8%	2,3%	486	323	-163
J1	3,2%	2,9%	144	177	34
J2	1,8%	2,3%	685	469	-216
J3	1,9%	2,4%	455	303	-152
K0	1,4%	1,7%	584	446	-137
L0	1,6%	2,0%	742	552	-190
M0	4,2%	2,9%	152	278	126
N1	1,0%	1,3%	322	293	-29
N2	1,5%	1,9%	1 321	927	-394
N3	3,1%	2,9%	706	789	83
N4	2,3%	2,9%	87	71	-16
P1	3,1%	2,9%	219	244	26
P2	4,8%	3,0%	335	715	381
P3	11,9%	3,0%	21	254	232
Q1	4,3%	3,0%	159	311	152
Q2	1,8%	2,2%	577	384	-193
R1	2,1%	2,9%	199	132	-67
R2	5,3%	3,0%	174	484	311

À un niveau de publication encore plus fin, l'allocation contrainte apporte un vrai gain en terme de précision notamment pour les petits secteurs tels que l'habillement et cuir C1, l'industrie textile F2, les services personnels et domestiques P3 et les activités associatives R2. Avec une allocation classique de Neyman, leur coefficient de variation peut aller jusqu'à près de 12 % (secteur P3) ; celui-ci peut être réduit à 3 %, grâce à une meilleure redistribution de l'allocation. On tire ainsi 254 établissements dans P3 contre seulement 21 avec l'allocation classique de Neyman et trois fois plus d'établissements dans les secteurs C1, F2 et R2. Ce gain d'établissements dans ces secteurs d'activité se fait au détriment des secteurs les plus précis L0, J2, N2, Q2, qui voient leur nombre d'établissements interrogés diminuer, mais tout en leur préservant une bonne précision au niveau local.

La précision globale reste quant à elle, peu affectée par cette redistribution de l'allocation de Neyman : le coefficient de variation est respectivement de 0,44 % pour l'allocation selon Neyman et de 0,49 % pour l'allocation contrainte par un seuil de précision local de 3 %.

Taille d'échantillon n=10 000 et publication par secteur d'activité x taille d'établissement (Nes16 x taille)

Nes16 x taille	CV Neyman	CV Allocation contrainte au seuil de 5 %	n _{strate pub} Neyman	n _{strate pub} Allocation contrainte au seuil de 5 %	DELTA
GAIN DE PRÉCISION en CV_p > 5 points					
Global	0,44%	0,59%	10 000	10 000	-0,15pts
EB x 1	11,5%	4,9%	46	234	6,6pts
ED x 1	36,8%	5,0%	7	179	31,8pts
ED x 2	20,1%	5,0%	8	95	15,1pts
ED x 3	16,8%	4,9%	4	32	11,9pts
ED x 4	10,7%	4,9%	17	57	5,8pts
EG x 1	15,1%	5,0%	16	134	10,1pts
EM x 1	12,2%	5,0%	42	234	7,2pts
EM x 3	12,2%	5,0%	18	83	7,2pts
EM x 5	13,2%	4,7%	10	31	8,5pts
ER x 1	10,0%	5,0%	68	265	5pts
PERTE DE PRÉCISION en CV_p > 2 points					
EB x 6	0,0%	3,5%	78	71	-3,5pts
EC x 6	1,6%	5,0%	102	59	-3,4pts
EE x 6	0,4%	4,4%	171	101	-4pts
EF x 6	0,5%	4,2%	212	146	-3,7pts
EG x 6	1,8%	4,8%	42	29	-3pts
EM x 6	0,0%	3,7%	22	21	-3,7pts
EN x 4	2,7%	5,0%	381	125	-2,3pts
EN x 5	2,8%	5,0%	249	107	-2,2pts
EN x 6	0,3%	3,0%	483	298	-2,7pts
ER x 6	1,4%	4,7%	101	66	-3,3pts
VARIATIONS DE N_{strate pub} > 100					
EB x 1	11,5%	4,9%	46	234	188
ED x 1	36,8%	5,0%	7	179	172
EG x 1	15,1%	5,0%	16	134	118
EL x 1	7,9%	4,9%	105	260	155
EL x 2	7,7%	4,9%	161	358	197
EM x 1	12,2%	5,0%	42	234	191
ER x 1	10,0%	5,0%	68	265	197
EF x 4	2,9%	4,6%	186	80	-106
EJ x 1	2,9%	4,4%	292	126	-166
EJ x 2	2,4%	4,3%	351	115	-236
EJ x 3	3,0%	4,6%	196	85	-112
EJ x 4	2,9%	4,6%	231	100	-131
EN x 2	3,2%	4,5%	527	285	-242
EN x 3	3,4%	4,8%	274	148	-126
EN x 4	2,7%	5,0%	381	125	-256
EN x 5	2,8%	5,0%	249	107	-142
EN x 6	0,3%	3,0%	483	298	-185

Lorsque l'on passe à une publication en nes16 x taille d'établissement, les gains de précision avec l'allocation contrainte sont toujours plus accentués (1^{ère} partie du tableau) mais la perte de précision globale est également plus importante. Il y a un vrai arbitrage entre précision locale et globale.

Tous les niveaux de publication peuvent être ramenés à un seuil maximal de variation de 5 %. En contrepartie, certains secteurs d'activités voient leur précision diminuer (2^{ème} partie du tableau) jusqu'à un maximum de 4 points par rapport à l'allocation de Neyman.

Enfin, la 3^{ème} partie du tableau met en évidence les plus grandes redistributions de questionnaires entre les strates d'échantillonnage et donc dans les regroupements de publication associés.

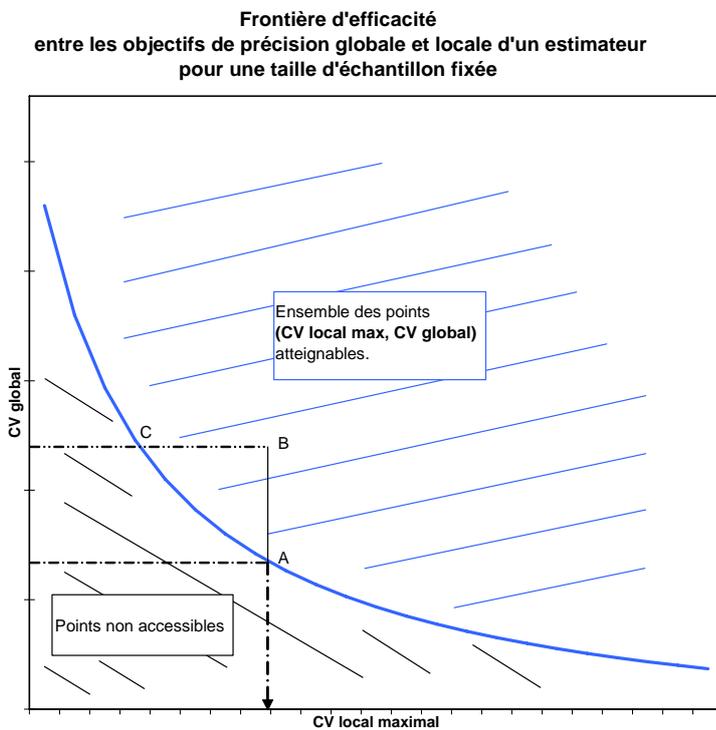
III.4. CONCEPT ET ILLUSTRATION DE LA FRONTIÈRE D'EFFICACITÉ

Le dilemme entre précision globale et locale d'un estimateur que rencontre le statisticien est illustré ici par le graphique de la frontière d'efficacité entre deux objectifs, en référence à la notion d'efficacité au sens de Pareto en microéconomie. Cette représentation constitue la synthèse des résultats issus de notre programme d'optimisation sous contrainte d'un seuil de précision locale à respecter.

Frontière d'efficacité

On appelle frontière d'efficacité, l'ensemble des allocations d'échantillon (n_1, \dots, n_H) à taille n fixée tel qu'on ne peut augmenter la précision locale dans les regroupements de publication, sans détériorer la précision globale au niveau de l'ensemble de la population d'un estimateur (et inversement). Cette frontière d'efficacité est représentée dans le repère $(CV_{global}, CV_{local\ maximal})$.

Schéma :



Par abus de langage, nous parlerons de la frontière d'efficacité comme l'ensemble des couples $(CV_{global}, CV_{local\ maximal})$ réalisables pour une taille d'échantillon fixée, de la courbe ci-dessus. Ce sont les points optimaux.

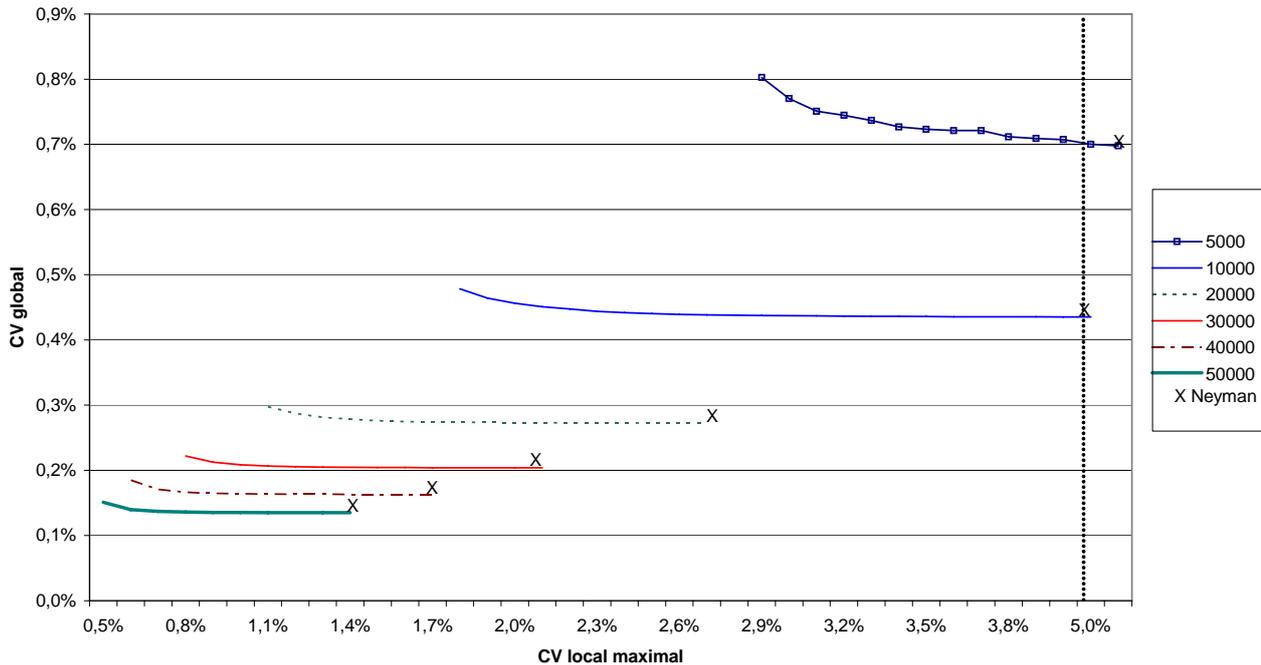
Ceux au-dessus de la courbe, comme le point B, sont possibles mais des situations meilleures peuvent être atteintes : pour le même niveau de précision globale que B, la précision locale maximale peut être réduite à celle de C ; pour le même CV local maximal que B, on peut atteindre une meilleure précision globale (point A).

Cette courbe constitue une aide à la décision précieuse pour l'échantillonnage car elle synthétise les différents arbitrages devant lesquels le statisticien est placé.

En pratique, la courbe est obtenue en faisant varier CV_{seuil} et en enregistrant pour chaque valeur de cette variable, le résultat de l'optimisation CV_{global} .

Regroupements de publication : par secteur d'activité (Nes16)

Frontière d'efficacité et allocation de Neyman pour différentes tailles d'échantillon et une publication en nes16



Ce graphique montre que les précisions globales ainsi que les précisions locales d'un estimateur sont fonctions croissantes de la taille de l'échantillon tiré (translation des frontières d'efficacité en bas à gauche avec n croissant). La précision globale d'un estimateur est très satisfaisante puisqu'elle reste sous le seuil de 1 % quelle que soit la taille d'échantillon de 5 000 à 50 000 établissements ; la précision locale par contre est plus dispersée, puisque le $CV_{local\ maximal}$ varie de plus de 5 % pour un échantillon de taille n=5 000 à 0,5 % pour n=50 000.

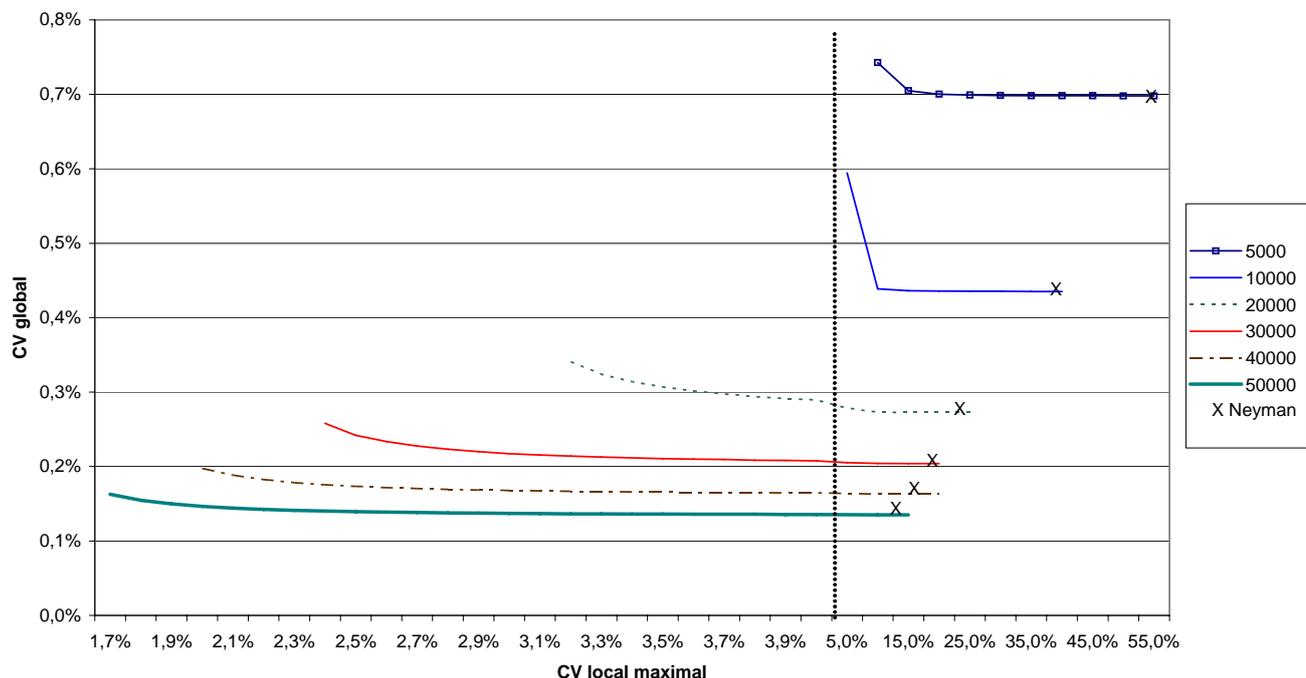
La précision locale obtenue avec l'allocation de Neyman (repérée par x) peut être sensiblement améliorée au prix d'une dégradation minimale de la précision globale de notre estimateur : la précision globale est en effet décroissante avec la précision locale, mais se stabilise rapidement quel que soit le $CV_{local\ maximal}$ (courbes plates). Le minimum de la variance globale est un minimum plat.

Exemple : pour un échantillon de 20 000 établissements, le CV local maximal peut être diminué de 2,7 % (Neyman) à 1,1 % (allocation sous contrainte), en impactant la précision globale de seulement 0,03 point (précision globale de 0,27 % à 0,30 %).

Mieux vaut alors s'assurer la meilleure précision locale, à savoir choisir l'échantillon tel que le pire des CV locaux soit de 1,1 %, au détriment de l'allocation de Neyman à partir de laquelle il n'y a plus de contrainte de précision locale. Des précisions locales inférieures au seuil de 1,1 % ne sont pas atteignables, car la contrainte $n \geq \sum_h n_{h\min}$ n'est alors plus vérifiée.

Regroupements de publication : par secteur d'activité (Nes16) x taille d'établissement

Frontière d'efficacité et allocation de Neyman pour différentes tailles d'échantillon et une publication en nes16 x taille d'établissement



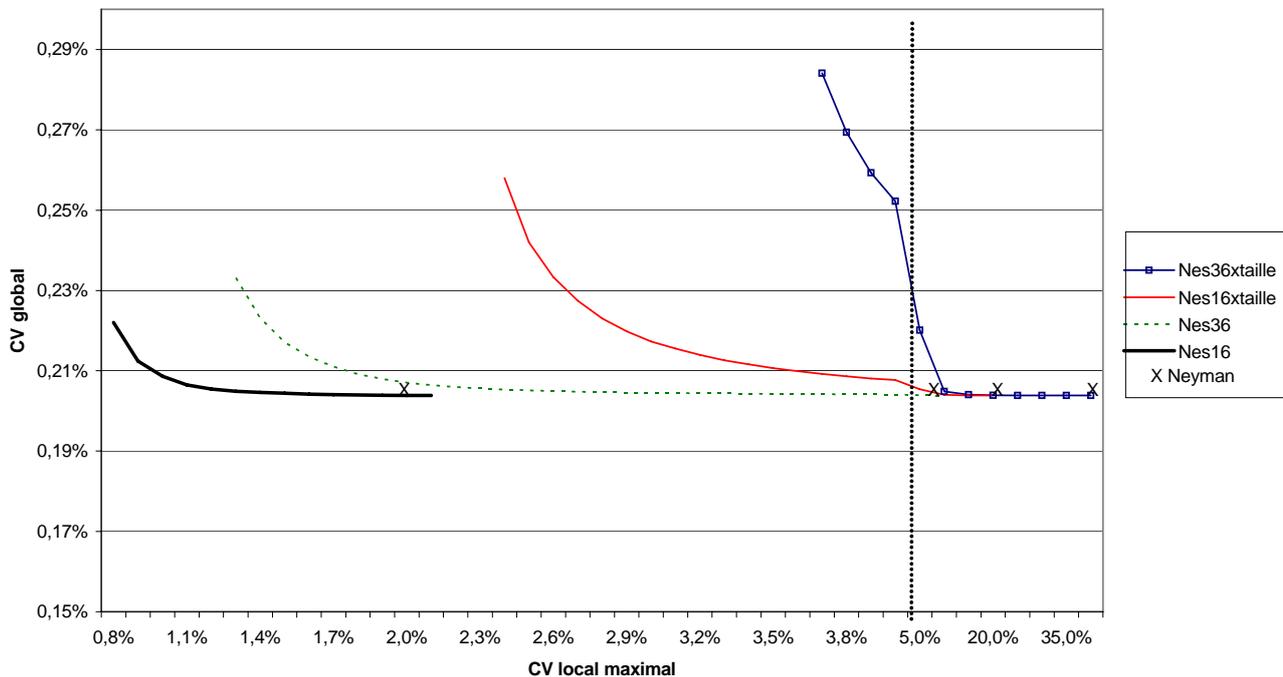
Lorsque l'on publie des résultats dans des niveaux de nomenclature plus fins (nes16 x taille d'établissement), les contraintes de précisions locales deviennent plus exigeantes : les résultats locaux de l'allocation de Neyman peuvent être améliorés mais avec un impact un peu plus fort sur la précision globale.

Exemple : pour $n=20\ 000$, le $CV_{\text{local maximal}}$ peut être ramené de 24,4 % (allocation de Neyman) à 3,2 % (allocation sous contrainte) avec une perte de précision globale de 0,07 point (passage du CV_{global} de 0,27 % à 0,34 %).

Néanmoins, les gains de précisions de l'allocation sous contrainte sont conséquents et ceci en partie à cause de quelques mauvaises précisions locales données par l'allocation de Neyman, dans les petits regroupements de publication.

Résultats pour n=30 000 et différents niveaux de publication

Comparaison des frontières d'efficacité pour une taille d'échantillon n=30 000 et différents niveaux de publication



La méthode d'optimisation sous contrainte de précision locale apporte un fort gain de précision locale par rapport à l'allocation de Neyman dans le cas de regroupements de publication fins. Mais cette amélioration de la précision locale dans les strates fines demande également une détérioration plus conséquente de la précision globale.

Exemple : pour n=30 000, la précision en nes16 x taille d'établissement peut être ramenée de 19,1 % à 2,4 % grâce à l'allocation sous contrainte. La précision globale s'en voit dégrader de 0,06 point (CV_{global} respectivement de 0,20 % avec l'allocation de Neyman et de 0,26 % avec l'allocation sous contrainte). La plus mauvaise précision en nes36 de 6,1 % en CV peut également être améliorée jusqu'au seuil de 2,4 % avec une perte de précision globale de moins de 0,01 point.

La précision locale augmente avec le niveau d'agrégation des strates d'échantillonnage : nos estimateurs sont plus précis en nes16 qu'en nes36, et ceux en nes16 x taille d'établissement sont plus précis qu'en nes36 x taille d'établissement.

Cette courbe de frontière d'efficacité est très utile pour l'échantillonnage d'une enquête car elle permet d'évaluer la perte en précision globale d'un estimateur lorsque l'on fixe un seuil de précision minimale dans les regroupements de publication.

CONCLUSION

En 2006, une nouvelle méthode d'échantillonnage a été appliquée à l'enquête trimestrielle ACEMO.

Cette nouvelle méthode permet d'améliorer l'allocation classique de Neyman qui optimise l'estimateur « global » d'une variable au niveau de l'ensemble de la population, par la prise en compte de la précision « locale » des estimations dans les regroupements de publication.

Tout au long du document, on distingue ainsi les strates d'échantillonnage et les regroupements de publication. Les strates d'échantillonnage sont constituées en amont du tirage afin de répartir les individus dans des groupes homogènes du point de vue de la variable d'intérêt. Les regroupements de publication sont les niveaux d'agrégation auxquels on diffuse les indicateurs issus d'une enquête (par secteur d'activité, par taille d'entreprise). On se place dans le cadre simplifié où chaque regroupement de publication est la réunion d'une ou de plusieurs strates d'échantillonnage.

On ajoute au programme d'optimisation de Neyman une contrainte de précision locale, afin que la plus mauvaise précision obtenue pour les estimateurs dans les regroupements de publication soit inférieure à un seuil que l'on s'est fixé. Cette contrainte de précision dans les regroupements de publication se laisse ensuite traduire en un nombre minimal de questionnaires à tirer dans les strates d'échantillonnage. Le problème d'optimisation avec ces contraintes reformulées peut alors être résolu grâce à une méthode algorithmique.

Cette optimisation est illustrée par le concept de frontière d'efficacité, c'est-à-dire l'ensemble des couples (précision globale, précision locale minimale) réalisables pour une taille d'échantillon donnée. Celle-ci constitue une aide à la décision précieuse pour l'échantillonnage, car elle synthétise les différents arbitrages devant lesquels le statisticien est placé.

BIBLIOGRAPHIE

Ardilly P., *Les techniques de sondage*, Éditions technip, Paris, 2006.

Bankier M.D., *Power Allocations : Determining sample Sizes for Sub-national Areas*, *The American Statistician*, 1988, vol.42 p.174-177.

Cochran W., *Sampling Techniques* (3^{ème} édition), John Wiley, New York, 1977.

Desabie J., *Théorie et pratique des sondages*, Dunod, Paris, 1966.

Grosbras J-M., *Méthodes statistiques des sondages*, Economica, Paris, 1987.

Kadane J.B., *Optimal Dynamic Sample Allocation Among Strata*, *Journal of Official Statistics*, 2005, vol. 21, N°4, p531-541.

Tillé Y., *Théorie des sondages : Échantillonnage et estimation en populations finies : cours et exercices*, Dunod, Paris, 2001.

Divers documents issus du groupe de travail « méthodologie de tirage d'échantillon » lors de la refonte des enquêtes ACEMO.

ANNEXES

Annexe A : Démonstration de la proposition : « l'optimisation globale est équivalente à des optimisations dans chaque regroupement de publication ».

p 43

Annexe B : Description des macros sas utiles à la résolution du programme d'optimisation.

p 44

Annexe C : Macros sas de résolution du programme d'optimisation.

p 46

Annexe A : démonstration de la proposition : l'optimisation globale est équivalente à des optimisations dans chaque regroupement de publication.

On se place dans l'hypothèse où chaque regroupement de publication est la réunion d'une ou plusieurs strates d'échantillonnage. Le problème que l'on résout est le suivant :

$$(I) \begin{cases} \text{Min}_{n_1, \dots, n_H} V(\hat{Y}) & (1) \\ \text{s/c } \sum_{h=1}^H n_h & (2) \\ \text{s/c } n_h \leq N_h & (3) \\ \text{s/c } \text{Max}_{p \in \text{pub}} CV_p \leq CV_{\text{seuil}} & (4) \end{cases}$$

Notons par ailleurs

$r_1 = (n_1, \dots, n_{i_1})$, $r_2 = (n_{i_1+1}, \dots, n_{i_2})$, ..., $r_k = (n_{i_{k-1}+1}, \dots, n_{i_k})$, le regroupement des strates d'échantillonnage en k regroupements de publication.

$s_1 = (n_1 + \dots + n_{i_1})$, ..., $s_k = (n_{i_{k-1}+1} + \dots + n_{i_k})$, le nombre d'unités dans chaque regroupement de publication.

Soit (n_1^0, \dots, n_H^0) la solution du problème (I), (s_1^0, \dots, s_k^0) les sommes partielles correspondantes.

Alors la solution de (I) est également solution du problème suivant, auquel on a ajouté des contraintes vérifiées par la solution :

$$\begin{cases} \text{Min } \sum_h \left(\frac{N_h}{N} \right)^2 * \left(1 - \frac{n_h}{N_h} \right) * \frac{S_h^2}{n_h} = \varphi(r_1, \dots, r_k) \\ \text{s/c } \left(\sum_{h \in p} n_h = s_p^0 \right)_{p=1, \dots, k} \\ \text{s/c } (n_h \leq N_h |_{h \in p})_{p=1, \dots, k} \\ \text{s/c } (CV_p(n_h |_{h \in p}) \leq CV_{\text{seuil}})_{p=1, \dots, k} \end{cases}$$

Or ce problème est à variables séparables. En effet posons :

$$\varphi_1(n_1, \dots, n_{i_1}) = \sum_{h=1}^{i_1} \left(\frac{N_h}{N} \right)^2 * \left(1 - \frac{n_h}{N_h} \right) * \frac{S_h^2}{n_h}$$

$$\varphi_k(n_{i_{k-1}+1}, \dots, n_{i_k}) = \sum_{h=i_{k-1}+1}^{i_k} \left(\frac{N_h}{N} \right)^2 * \left(1 - \frac{n_h}{N_h} \right) * \frac{S_h^2}{n_h}$$

On voit que $\varphi(r_1, \dots, r_k) = \varphi_1(r_1) + \dots + \varphi_k(r_k)$.

Par conséquent, le problème est équivalent à la résolution des k problèmes suivants ($p=1, \dots, k$) :

$$\text{Problème } p \begin{cases} \text{Min } \sum_{h=i_{p-1}+1}^{i_p} \left(\frac{N_h}{N} \right)^2 * \left(1 - \frac{n_h}{N_h} \right) * \frac{S_h^2}{n_h} \\ \text{s/c } \left(\sum_{h=i_{p-1}+1}^{i_p} n_h = s_p^0 \right) \\ \text{s/c } (n_h \leq N_h)_{h=i_{p-1}+1, \dots, i_p} \\ \text{s/c } (CV_p(n_{i_{p-1}+1}, \dots, n_{i_p}) \leq CV_{\text{seuil}}) \end{cases}$$

Il traduit le problème de Neyman dans le regroupement de publication p.

Annexe B : Description des macros sas utiles à la résolution du programme d'optimisation.

Optimisation classique

Une première série d'optimisations classiques de Neyman, obtenue en faisant varier la taille de l'échantillon n , a pour but de créer la table de correspondance, au niveau de chaque regroupement de publication, entre la précision voulue CV_{seuil} et le nombre minimal d'unités $n_{h\min}(CV_{seuil})$ nécessaires dans chaque strate d'échantillonnage. Ces relations seront reprises sous la forme de contraintes dans la partie optimisation avec contraintes. Les macros utilisées sont les suivantes :

➤ **La macro PRECISION (regroupement de publication, table de sortie)**

Elle calcule la précision obtenue dans chaque regroupement de publication pour une allocation $(n_h)_{h=1,\dots,H}$ donnée.

En entrée : les regroupements de publication et la table des allocations AL

En sortie : une table RESPRH avec la précision par regroupement de publication.

➤ **La macro ALLOC (n)**

Elle calcule l'allocation optimale $(n_h)_{h=1,\dots,H}$ pour une taille d'échantillon n donnée.

En entrée : la taille de l'échantillon n

En sortie : la table AL avec la valeur des $(n_h)_{h=1,\dots,H}$ pour une taille d'échantillon n .

➤ **La boucle BOUCLENC (ndeb, nbiter, pas)**

Cette boucle non contrainte, contient la macro ALLOC, et calcule l'allocation optimale pour différentes tailles d'échantillon n , avec une première taille d'échantillon ($ndeb$), un nombre d'itérations ($nbiter$) et un intervalle de progression (pas).

En entrée : la valeur de $ndeb$, $nbiter$, pas

En sortie : la table RESAL donnant pour chaque n , l'allocation optimale. Cette table est en fait un empilement des tables des allocations AL (en sortie de la macro ALLOC).

On peut ensuite calculer la précision dans les regroupements de publications donnée par chacune des allocations optimales obtenues.

Remarque : c'est cette boucle qui constitue la partie du programme la plus longue à s'exécuter, puisqu'on calcule un grand nombre d'allocation optimale selon l'incrément de la taille d'échantillon n . Par exemple, le calcul de l'allocation optimale pour des tailles d'échantillon variant de 1 000 à 276 000 avec un pas de 1 000 prend 3 minutes environ.

Optimisation sous contraintes de précision locale

Cette partie constitue le cœur de notre apport à l'optimisation classique de Neyman. On s'appuie ici sur les résultats issus de l'étape 1 pour introduire les contraintes traduisant la précision locale souhaitée dans les regroupements de publication. Pour se faire on utilise ici :

➤ **La macro CONT_MIN (CV_seuil)**

Elle donne l'effectif minimal $n_{h\min}(CV_{seuil})$ par strate d'échantillonnage nécessaire pour atteindre le niveau de précision CV_{seuil} dans chaque regroupement de publication.

En entrée : le CV_{seuil} souhaité, la table des allocations optimales pour différents n (RESAL) et la table des précisions obtenues dans les regroupements de publication pour ces allocations (RESPRH)

En sortie : la table NHMIN donnant le nombre minimal d'unités à tirer par strates h pour respecter la contrainte de précision locale.

➤ **La macro LIMSAT**

Elle calcule les tailles d'échantillon limites conduisant à la saturation des contraintes $(n_{h\min}(CV_{seuil}) \leq n_h \leq N_h)$ de chaque strate d'échantillonnage (cf. schéma de la partie II.4).

En entrée : la base de sondage (strates d'échantillonnages et N_h), complétée par la variable $n_{h\min}(CV_{seuil})$ issue de la contrainte de précision locale de la table NHMIN

En sortie : la table LIMSAT avec les valeurs de n(h) limite min et n(h) limite max pour chaque strate d'échantillonnage.

➤ **La macro ALLOC (n)**

Elle calcule l'allocation optimale $(n_h)_{h=1,\dots,H}$ pour une taille d'échantillon n donnée. Si $n_{h\min}(CV_{seuil})=0$, on obtient l'allocation classique de Neyman ; sinon, on obtient l'allocation sous contrainte de précision dans les regroupements de publication.

En entrée : la taille de l'échantillon n, la table LIMSAT (calcul de l'allocation en fonction des strates saturantes ou non saturantes)

En sortie : la table AL avec la valeur des $(n_h)_{h=1,\dots,H}$ pour une taille d'échantillon n fixée.

➤ **La boucle BOUCLEC (ndeb, nbiter, pas, nbCV, pasCV)**

Cette boucle contrainte, contient les macros CONT_MIN, LIMSAT et ALLOC et calcule l'allocation sous contrainte de précision locale pour différentes tailles d'échantillon et différents CV_{seuil} . La taille d'échantillon est pilotée par la première taille d'échantillon (*ndeb*), un nombre d'itérations (*nbiter*) et un intervalle de progression (*pas*) ; le CV_{seuil} donnant la précision locale est pilotée par un nombre d'itérations de celui-ci (*nbCV*) et un intervalle de progression (*pas*).

En entrée : la valeur de *ndeb*, *nbiter*, *pas*, *nbCV*, *pasCV*, un CV_{seuil} , les tables des allocations (RESAL) et celle des précisions associées (RESPRH) issues de la première boucle

En sortie : la table RESALC donnant pour chaque n, l'allocation optimale sous contrainte de précision locale.

Annexe C : Macros sas de résolution du programme d'optimisation.

```

/*****
/* macro calculant les tailles d'échantillon-limites conduisant à la
/* saturation des différentes strates
*****/
%macro LIMSAT ;

proc sort data=BASE out=BASE2;
by variable &strtech; run;

proc means data=BASE2 sum noprint nway;
var auxh Nhmin;
class variable;
output out=auxtot sum=auxtot Nhmintot; run;

data auxtot;
set auxtot(keep=variable auxtot Nhmintot);run;

data tabmax;
set BASE2(keep=variable &strtech auxh Nhmax th);
by variable;
ord=Nhmax/auxh; /* ordre de saturation des strates */
table="max";run;

data tabmin;
set BASE2(keep=variable &strtech auxh Nhmin th Yh Xh);
by variable;
ord=Nhmin/auxh; /* ordre de saturation des strates */
table="min";run;

data x6;
set tabmax (rename=(Nhmax=Nh)) tabmin(rename=(Nhmin=Nh));
by variable;run;

data x6;
merge x6 auxtot;
by variable;run;

proc sort data=x6;by variable ord;run;

data x7;
set x6;
by variable ord ;
idstrate=_N_;
Nh_1=lag(Nh);auxh_1=lag(auxh);table_1=lag(table);
if first.variable then do;Nh_1=0;auxh_1=0;table_1="min";
end;run;

data STRATES ;
set x7;
by variable idstrate;
retain SNh_1 Sauxh_1 ;
if first.variable then do; SNh_1=Nhmintot; Sauxh_1=auxtot;end;
if table_1="max" then signe=1;
if table_1="min" then signe=-1;
SNh_1=SNh_1+signe*Nh_1; /*effectif des strates saturantes */
Sauxh_1=Sauxh_1+signe*auxh_1;
nlim=SNh_1+Nh*(auxtot-Sauxh_1)/auxh;
/* taille limite d'échantillon induisant la saturation de la strate */
keep variable table signe Nh auxh SNh_1 Sauxh_1
&strtech nlim auxtot Yh Xh idstrate;run;
data tabmax tabmin;
```

```

set STRATES (keep=variable &strtech nlim table aux tot Yh Xh) ;
if table="min" then output tabmin;
else output tabmax;run;

proc sort data=tabmax (keep=variable &strtech nlim aux tot);
    by variable &strtech;run;
proc sort data=tabmin (keep=variable &strtech nlim);
    by variable &strtech;run;

data LIMSAT;
merge
tabmax (rename=(nlim=nlimmax))
    tabmin(rename=(nlim=nlimmin))
    BASE2;
    by variable &strtech;run;
%mend LIMSAT;

/*****
/*macro calculant l'allocation pour une taille d'échantillon donnée */
/*****
%macro ALLOC (n=);
* on extrait des informations de LIMSAT, on en déduit les strates saturantes et non saturantes;
data p1; set STRATES;
by variable idstrate;if nlim>&n;run;

/* Totaux pour les strates saturantes */
data P2; set p1 ;by variable idstrate;
if first.variable;
keep variable Sauxh_1 SNh_1 ;run;

data p3; merge LIMSAT P2; by variable;run;

data NSAT SAT; set p3;by variable;
if nlimmin<&n<nlimmax then output NSAT;
else output SAT;run;

/*calcul de l'allocation pour les strates non saturantes*/
data x1; set NSAT;
nech=&n-SNh_1; /* on modifie la taille de l'échantillon en conséquence */
auxtoth=auxtot- Sauxh_1;
petitnh=nech*auxh/auxtoth;
if petitnh>Nh then petitnh=Nh;
table="NSAT ";run;

/*calcul de l'allocation pour les strates saturantes */
data x2; set sat;
if &n<=nlimmin then petitnh=Nhmin;
else if &n>=nlimmax then petitnh=Nhmax;
table="SAT ";run;

/*synthèse de l'allocation */
data AL ; set x1 x2 ;
txsond=petitnh/Nh;
varh=(1-petitnh*th/Nh)*sh2/(petitnh*th);
keep variable table &strtech petitnh Nh Nhmin Nhmax txsond
th Yh Xh sh2 nes16 nes36 nech auxh auxtoth;run;

proc sort; by variable &strtech;run;
%mend ALLOC;

```

```

/*****
/* macro calculant la précision obtenue par regroupement de publication */
/*****
%macro PRÉCISION (stratpub=,tabsor=);
data x11; set AL ;
varh=(1-petitnh*th/Nh)*sh2/(petitnh*th);
aux1=Nh*Nh*varh;
aux2=Yh;
aux3=Xh;
aux4=petitnh;
aux5=Nh;run;

proc means data=x11 sum noprint nway;
var aux1 aux2 aux3 aux4 aux5;
class variable &stratpub;
output out=res2 sum=;run;

data &TABSOR; set res2;
var=aux1/(aux3*aux3);
cv=sqrt(aux1)/aux2;
petitpub=aux4;
Xpub=aux3;
Ypub=aux2;
Npub=aux5;
keep variable &stratpub cv var petitpub Xpub Ypub Npub;run;
%mend PRÉCISION;

/*****
/***** OPTIMISATION SANS CONTRAINTTE *****/
/*****

%macro BOUCLENC ( ndeb=,nbiter=,pas=);
data RESAL; set vide;run;
%do i=0 %to %eval(&nbiter-1);
%let ntot=%eval(&i*&pas+&ndeb);

%alloc (n=&ntot);
data al; set al;n=&ntot;run;
data RESAL; set RESAL al;run;
%end;
%mend BOUCLENC;

/*****
/***** OPTIMISATION SOUS CONTRAINTTE DE PRÉCISION LOCALE *****/
/*****

/* effectif minimum par regroupement de publication pour différents niveaux de précision. */
%macro CONT_MIN (cvmax=);
data x2;
set RESPRH;
contrainte=(cv<&cvmax);run;

proc sort data=x2;
by variable &strtpub. contrainte descending cv;run;

data x3a x3b;
set x2;
by variable &strtpub. contrainte descending cv;
if contrainte and first.contrainte then output x3a;
/* le plus grand petitnh tel que la contrainte est vérifiée */
else if (not contrainte) and last.contrainte then output x3b;

```

```

/* au cas où la contrainte n'est jamais vérifiée*/
keep variable &strtpub. n contrainte ;run;

data x3;
merge x3b x3a;
by variable &strtpub.;run;

proc freq; tables contrainte;run;

proc sort data=das.resal out=x4;
by variable &strtpub. n;run;

data NHMIN;
merge x3 (in=in1) x4;
by variable &strtpub. n;
if in1;
NHMIN=petitnh;
cvmax=&cvmax;
keep variable &strtech. NHMIN nh cvmax;run;

proc sort;
by variable &strtech.;run;
%mend;

/* boucle sur la taille d'échantillon et la précision locale cvmax en sortie RESALC */
%macro bouclec ( ndeb=,nbiter=,pas=,nbCV=,pasCV=);
data RESALC; set vide;run;
%do j=1 %to %eval(&nbCV);
    %let cvmax=&j*&pasCV;
    %cont_min (cvmax=&cvmax);

    proc means data=nhmin sum noprint nway;
    var nhmin;
    class variable;
    output out=nhmintot sum=nhmintot;run;

    data BASE; /* table BASE avec toutes les stats descriptives
                et les nhmin calculés par strates d échantillonnage */
    merge STAT NHMIN (keep=NHMIN variable &strtech.);
    by variable &strtech.;
    Nhmax=Nh;run;
    %limsat ;
    %do i=0 %to %eval(&nbiter-1); /*boucle sur la taille d'échantillon*/
        %let ntot=%eval(&i*&pas+&ndeb);
        %alloc (n=&ntot.);
        data al2; merge al nhmintot (keep=variable nhmintot);
        by variable;n=&ntot.;cvmax=&cvmax.;
        select=(n>=nhmintot);run;
        data RESALC; set RESALC al2;run;
    %end;
%end;
%mend;

```

- N° 1 *La négociation salariale de branche entre 1985 et 1993*, par Olivier BARRAT (DARES), septembre 1994.
- N° 2 *Créations et suppressions d'emplois en France. Une étude sur la période 1984-1992*, par S. LAGARDE (INSEE), E. MAURIN (DARES), C. TORELLI (INSEE), octobre 1994.
- N° 3 *L'impact du coût sur la substitution capital-travail*, par Ferhat MIHOUBI (DARES), novembre 1994.
- N° 4 *Éducation, expérience et salaire. Tendances et évolutions de long terme*, par D. GOUX (INSEE) et Eric MAURIN (DARES), novembre 1994.
- N° 5 *Origine sociale et destinée scolaire. L'inégalité des chances devant l'enseignement à travers les enquêtes FQP 1970, 1977, 1985 et 1993*, par D. GOUX (INSEE) et Eric MAURIN (DARES), décembre 1994.
- N° 6 *Perception et vécu des professions en relation avec la clientèle*, par Sabine GUYOT et Valérie PEZET (Institut pour l'amélioration des conditions de travail), déc. 1994.
- N° 7 *Collectifs, conflits et coopération dans l'entreprise*, par Thomas COUTROT (DARES), février 1995.
- N° 8 *Comparaison entre les établissements des PME des grandes entreprises à partir de l'enquête RÉPONSE*, par Anna MALAN (DARES) et Patrick ZOUARY (ISMA), septembre 1996.
- N° 9 *Le passage à une assiette valeur ajoutée pour les cotisations sociales : une approche sur données d'entreprises*, par Gilbert CETTE et Élisabeth KREMP (Banque de France), novembre 1996.
- N° 10 *Les rythmes de travail*, par Michel CÉZARD et Lydie VINK (DARES), décembre 1996.
- N° 11 *Le programme d'entretien auprès des 900 000 chômeurs de longue durée - Bilan d'évaluation*, par Marie RUAULT et René-Paul ARLANDIS (DARES), mars 1997.
- N° 12 *Créations et suppressions d'emplois et flux de main-d'oeuvre dans les établissements de 50 salariés et plus*, par Marianne CHAMBAIN et Ferhat MIHOUBI (DARES), avril 1997.
- N° 13 *Quel est l'impact du commerce extérieur sur la productivité et l'emploi ? Une analyse comparée des cas de la France, de l'Allemagne et des États-Unis*, par Olivier CORTES et Sébastien JEAN (CEPII), mai 1997.
- N° 14 *Bilan statistique de la formation professionnelle en 1995-1996* - DARES, mai 1997.
- N° 15 *Les bas salaires en France 1983-1997*, par Pierre CONCIALDI (IRES) et Sophie PONTHEUX (DARES), octobre 1997.
- N° 16 *Les jeunes en difficulté à travers le réseau des missions locales et des PAIO entre 1994 et 1996 - Résultats du panel TERSUD de 1997*, DARES et DIJ, janvier 1998.
- N° 17 *L'impact macro-économique d'une politique de RTT : l'approche par les modèles macro-économiques*, DARES (Mission analyse économique), SEMEF-BDF, OFCE, janvier 1998.
- N° 18 *L'opinion des Français face au chômage dans les années 80-90*, par Jacques CAPDEVIELLE et Arlette FAUGERES (CEVIPOF), janv. 1998.
- N° 19 *Intéressement et salaires : Complémentarité ou substitution ?* par Sylvie MABILE, DARES, mars 1998.
- N° 20 *L'impact économique de l'immigration sur les pays et régions d'accueil : modèles et méthodes d'analyse*, par Hubert JAYET, Université des sciences et technologies de Lille I, avril 1998.
- N° 21 *Analyse structurelle des processus de création et de suppression d'emplois*, par Frédéric KARAMÉ et Ferhat MIHOUBI, DARES, juin 1998.
- N° 22 *Quelles place pour les femmes dans les dispositifs de la politique de l'emploi entre 1992 et 1996 ?*, par Franck PIOT, DARES, août 1998.
- N° 23 *Deux années d'application du dispositif d'incitation à la réduction collective du temps de travail*, par Lionel DOISNEAU, DARES, sept. 1998.
- N° 24 *Le programme « Nouveaux services-Emplois jeunes », d'octobre 1997 à octobre 1998*, par Françoise BOUYGARD, Marie-Christine COMBES, Didier GÉLOT, Carole KISSOUN, DARES, novembre 1998.
- N° 25 *Une croissance plus riche en emplois depuis le début de la décennie ? Une analyse en comparaison internationale*, par Sandrine DUCHÊNE et Alain JACQUOT, DARES et INSEE, mars 1999.
- N° 26 *Stratégies concurrentielles et comportements d'emploi dans les PME - Un état de la littérature*, par Philippe TROUVÉ, avril 1999.
- N° 27 *Effets sur les trajectoires des chômeurs d'un passage dans deux dispositifs de politique d'emploi (CES-SIFE), Rapport final pour la convention du 15/06/98 (n° 98020) passée entre le Gréquam et la Dares*, Christelle BARAILLER, mai 1999.
- N° 28 *Les inégalités salariales entre hommes et femmes dans les années 90*, par Dominique MEURS et Sophie PONTHEUX, ERMES- Paris II et DARES, juin 1999.
- N° 29 *Les allocataires du RMI et l'emploi*, par Dominique ARNOUT (Rapport de stage), juin 1999.
- N° 30 *Les stratégies des entreprises face à la réduction du temps de travail*, par Anne-Lise AUCOUTURIER, Thomas COUTROT (DARES) et Étienne DEBAUCHE (Université Paris X-Nanterre), septembre 1999.
- N° 31 *Le mandatement dans le cadre de la loi du 13 juin 1998*, par Christian DUFOUR, Adelheid HEGE, Catherine VINCENT et Mouna VIPREY (IRES), octobre 1999.
- N° 32 *L'effort financier des collectivités locales dans la lutte contre le chômage et pour l'aide à l'emploi*, par Jacques ABEN, Paul ALLIES, Mohammad-Saïd DARVICHE, Mohammed DJOULDEM, Muriel FROEHLICH, Luis DE LA TORRE, octobre 1999.
- N° 33 *La dynamique asymétrique des flux de création et de suppression d'emplois : une analyse pour la France et les États-Unis*, par Frédéric KARAMÉ (DARES), nov. 1999.
- N° 34 *Évaluation d'une mesure de politique pour l'emploi : la convention de conversion*, par Marc WEIBEL (rapport de stage), janvier 2000.
- N° 35 *Premières évaluations quantitatives des réductions collectives du temps de travail*, par Murielle FIOLE, Vladimir PASSERON et Muriel ROGER, janvier 2000.
- N° 36 *La durée annuelle et l'aménagement du temps de travail en 1994*, par Annie DELORT et Valérie LE CORRE, février 2000.
- N° 37 *Analyse des premiers accords conventionnés de passage à 35 heures - Étude monographique de 12 accords*, par Pierre BOISARD et Jérôme PELISSE, février 2000.
- N° 38 *Syndrome, miracle, modèle polder et autres spécificités néerlandaises : quels enseignements pour l'emploi en France ?*, par Sébastien JEAN (CEPII), août 2000.
- N° 39 *La mise en œuvre de la formation dans les contrats de qualification - Rapport final*, par Marie-Christine COMBES (GPI-MIS), octobre 2000.
- N° 40 *L'impact du développement des services sur les formes du travail et de l'emploi - Rapport final pour la Dares* -, par Christian du TERTRE et Pascal UGHETTO (IRIS-Université Paris-IX-Dauphine), novembre 2000.
- N° 41 *Le suivi du plan social par l'employeur au service de l'amélioration du processus décisionnel : l'apport de trois études de cas*, par Christophe CORNOLT, Yves MOULIN et Géraldine SCHMIDT (Université Nancy II), février 2001.
- N° 42 *L'impact des marchés financiers sur la gestion des ressources humaines : une enquête exportatrice auprès des grandes entreprises françaises*, par Sabine MONTAGNE et Catherine SAUVIAT (IRES), mars 2001.
- N° 43 *L'impact du traitement des activités occasionnelles sur les dynamiques d'emploi et de chômage (Convention d'étude Dares-Ires)*, par Hervé HUYGHUES DESPOINTE, Florence LEFRESNE et Carole TUCHSZIRER, mars 2001.
- N° 44 *L'adaptation des marchés du travail à l'évolution des systèmes de retraite*, par Antoine BOMMIER, Thierry MAGNAC et Muriel ROGER, avril 2001.
- N° 45 *Étude de la démographie des organismes de formation continue*, par Isabelle BAUDEQUIN, Annie CHANUT, Alexandre MELIVA (DARES et CEREQ), juin 2001.
- N° 46 *L'évolution des sorties d'emploi vers la retraite et la préretraite. Une approche par métiers*, par Agnès TOPIOL (DARES), juillet 2001.
- N° 47 *Prospective des métiers à l'horizon 2010 : une approche par familles d'activité professionnelles*, par Agnès TOPIOL (DARES), juin 2001.
- N° 48 *L'évolution des sorties d'emploi vers la retraite et la préretraite*, juillet 2001.
- N° 49 *L'information statistique sur la participation des entreprises à la formation continue : état des lieux et évolutions possibles*, août 2001.
- N° 50 *Base de données des comptes sociaux des entreprises commerciales (fichiers DIANE). Panel DIANE/UNEDIC, période 1991-1999*, par Anne SAINT-MARTIN (DARES), janvier 2002.
- N° 51 *Dynamique des métiers et usage de l'informatique : une approche descriptive*, par Thomas COUTROT (DARES) et Jennifer SIROTEAU, février 2002.
- N° 52 *Licenciements et marchés financiers : les illégitimités de la convention financière*, par Tristan BOYER (FORUM), avril 2002.
- N° 53 *Mécanisme du plan de licenciement : déconstruction d'argumentaires économiques de projets de licenciements*, par Tristan BOYER (FORUM), avril 2002.
- N° 54 *À la recherche du temps gagné : des salariés face aux 35 heures*, par Jérôme PELISSE (CEE), mai 2002.
- N° 55 *La réduction du temps de travail en Lorraine : enjeux, négociations et pratiques des entreprises*, par Lionel JACQUOT (LASTES) et Nora SETTI (GREE), avril 2002.
- N° 56 *Principaux résultats de l'enquête RTT et modes de vie*, par Marc-Antoine ESTRADE et Dominique MEDA (DARES), mai 2002.
- N° 57 *Enquête passages : projets, attitudes, stratégies et accords liés à la généralisation des 35 heures - Guide méthodologique et analyse préliminaires*, par Mathieu BUNEL, juillet 2002.
- N° 58 *Cohésion sociale, emploi et compétitivité : éléments pour un débat*, par Rachel BEAUJOLIN-BELLET, Marc-Antoine ESTRADE, Jean-Yves KERBOUC'H, Tristan KLEIN, Frédéric LERAI, Dominique MEDA, Anne SAINT-MARTIN, Frédéric TRIMOUILLE (DARES), août 2002.
- N° 59 *La politique de l'emploi au prisme des territoires*, par Thierry BERTHET, Philippe CUNTIGH (CERVL-CEREQ) et Christophe GUITTON (DARES), septembre 2002.
- N° 60 *Comparaison internationales de durée et de productivité*, par Odile CHAGNY et Mireille BRUYERE (Observatoire Français des Conjonctures Économiques), sept. 2002.
- N° 61 *L'effet des 35 heures sur la durée du travail des salariés à temps partiel*, par Aline OLIVEIRA (ENSAE) et Valérie ULRICH (DARES), sept. 2002.
- N° 62 *Les effets du dispositif d'intéressement sur l'insertion au marché du travail des bénéficiaires de l'allocation chômage*, par Nadia ALIBAY et Arnaud LEFRANC (Université de Cergy-Pontoise), octobre 2002.
- N° 63 *Normes d'emploi et marché du travail dans les métiers liés aux technologies de l'information*, par Yannick FONDEUR et Catherine SAUVIAT (DARES), nov. 2002.
- N° 64 *Enquête « RÉPONSE » 1998 - Questionnaire « Représentants du personnel » - De la participation au conflit*, par Daniel FURJOT (DARES), déc. 2002.
- N° 65 *Développement et dialogue social - Les TPE face aux 35 heures*, par Pascal CHARPENTIER (CNAM) et Benoît LEPLEY (GIP-MIS), janvier 2003.
- N° 66 *La mobilité professionnelle et salariale des salariés âgés analysée à travers les DADS*, par Frédéric LAINÉ, mars 2003.
- N° 67 *Un indicateur régional d'évolution mensuelle d'emploi dans les établissements de 50 salariés ou plus*, par Magda TOMASINI, avril 2003.
- N° 68 *La réorganisation du travail et son impact sur les performances des entreprises industrielles : une analyse sur données françaises 1995-1999*, par Véronique JANOD et Anne Saint-Martin, avril 2003.
- N° 69 *Discrimination et emploi : revue de la littérature*, par Hélène GARNER-MOYER, mai 2003.
- N° 70 *Impact du traitement des activités occasionnelles sur les dynamiques d'emploi et de chômage - 2ème partie Espagne - Italie*, par Florence LEFRESNE (IRES) et Carole TUCHSZIRER (IRES), mai 2003.
- N° 71 *Souplesse et sécurité de l'emploi : Orientations d'études et de recherches à moyen terme*, coordination par Carole Yerochewski, juin 2003.
- N° 72 *Séries de données sur les mouvements de main-d'oeuvre 1996-2001*, par Lucile Richet-Mastain, juillet 2003.
- N° 73 *35 heures et mise en oeuvre des dispositifs de modulation/annualisation dans les enquêtes REPONSE et PASSAGES*, par Matthieu Bunel, août 2003
- N° 74 *Le licenciement pour motif personnel : une catégorie juridique aux contours flous et difficiles à cerner par les statistiques administratives*, par Maria-Teresa Pignoni et Patrick Zouary (Si2S), octobre 2003
- N° 75 *Plan national d'action pour l'emploi 2003. Annexe statistique. Indicateurs de suivi et d'évaluation*, coordination Christine Charpail et Norbert Holcblat, octobre 2003.
- N° 76 *Les estimations mensuelles d'emploi salarié dans le secteur concurrentiel*, par Raphaël Cancé, octobre 2003.

- N° 77 *Les déterminants du jugement des salariés sur la RTT*, par Gilbert CETTE (CEDERS), Nicolas DROMEL (GREQAM) et Dominique Méda (DARES), novembre 2003.
- N° 78 *Trajectoires passées par un emploi à bas salaire. Une étude à partir du panel européen des ménages*, par Bertrand LHOMMEAU (DARES), novembre 2003.
- N° 79 *Evaluation des statistiques administratives sur les conflits du travail*, par Delphine BROCHARD (MATISSE-CNRS), novembre 2003.
- N° 80 *Les disparités de rémunération entre hommes et femmes : la situation de quatre branches professionnelles*, par Fathi FAKHFAKH (Université Paris II - ERMES), Séverine LEMIERE (Université de Littoral - MATISSE), Marie-Pierre MERLATEAU (Université Paris II - ERMES) et Dominique MEURS (Université Paris II - ERMES), janvier 2004.
- N° 81 *Arbitrage entre flexibilité interne et flexibilité externe : une analyse empirique*, par Matthieu BUNEL (IREGE - Université de Savoie), mai 2004.
- N° 82 *Dossier Age et emploi : synthèse des principales données sur l'emploi des seniors*, coordination Frédéric LERAIS et Pierre MARIONI, mai 2004.
- N° 83 *La contribution des femmes à la performance* : une revue de la littérature, par Sophie LANDRIEUX-KARTOCHIAN (Université Paris I - Panthéon Sorbonne, CERGORS), octobre 2004.
- N° 84 *En 2002, l'insertion des jeunes dans l'emploi se fait plus ou moins lentement selon les pays européens*, par François BRUNET, octobre 2004.
- N° 85 *Etude de qualité sur le questionnement relatif au temps de travail dans les enquêtes Acemo*, par l'ENSAE Junior Etudes, octobre 2004.
- N° 86 *Les processus de mise en oeuvre de l'offre de formation Unédic dans le cadre du PARE* (plan d'aide au retour à l'emploi), par Florence LEFRESNE et Carole TUCHSZI RER (IRES), avec la collaboration statistique de Hervé Huyghues Despointes, octobre 2004.
- N° 87 *Quels effets de la négociation salariale d'entreprise sur l'évolution des salaires entre 1999 et 2001 ?*, par Abdenor BRAHAMI et Catherine DANIEL, novembre 2004.
- N° 88 *Plan national d'action pour l'emploi 2004. Annexe statistique. Indicateurs de suivi et d'évaluation*, coordination Christine Charpail, novembre 2004.
- N° 89 *Les expositions aux risques professionnels par secteur d'activités - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, S. HAMON-CHOLET, D. WALTISPERGER (Dares) et E. YILMAZ (stagiaire du DESS «Techniques statistiques et informatiques» Université Panthéon Assas Paris 2), décembre 2004.
- N° 90 *Les pouvoirs du temps. La transformation des régulations dans les organisations du travail après la RTT*, par Michel PEPIN, en collaboration avec Bernard DOERFLINGER, Yves JORAND, Myriam MAUFROY (ESSOR Consultants), janvier 2005.
- N° 91 *Mixité professionnelle et performance des entreprises, le levier de l'égalité*, par Catherine ACHIN, Dominique MEDA, Marie WIERINK, janvier 2005.
- N° 92 *La place du travail dans l'identité*, par Hélène GARNER, Dominique MEDA (Dares), et Claudia SENIK (Delta, Paris IV), janvier 2005.
- N° 93 *Audit de l'enquête sur les mouvements de main-d'oeuvre (EMMO)*, par Heidi WECHTLER, janvier 2005.
- N° 94 *Modalités de passage à 35 heures des TPE*, par Victor DE OLIVEIRA, février 2005.
- N° 95 *Evaluation des politiques d'emploi : la deuxième génération des panels des bénéficiaires*, par Christine CHARPAIL, Tristan KLEI, Serge ZILBERMAN, février 2005.
- N° 96 *Contribution Delalande : quels dispositifs similaires ou alternatifs en Europe du Nord*, par Violaine DELTEIL et Dominique REDOR (GIPMIS), février 2005.
- N° 97 *L'impact des conditions de travail sur la santé : une expérience méthodologique*, par Thomas COUTROT (Dares) et Loup Wolff (Centre d'étude de l'emploi), février 2005.
- N° 97bis *L'impact des conditions de travail sur la santé : une expérience méthodologique. Annexes*, par Thomas COUTROT (Dares) et Loup WOLFF (Centre d'étude de l'emploi), février 2005.
- N° 98 *La mixité professionnelle : les conditions d'un développement durable*, par Michèle FORTE, Myriam NISS, Marie-Claude REBEUH, Emmanuel TRIBY (BETA, Cereq, Université Louis Pasteur de Strasbourg), février 2005.
- N° 99 *Bilan d'activité 2003 des missions locales et des PAIO*, par Camille BONAÏTI (Dares) et Amaria SEKOURI (DGEFP), avril 2005.
- N° 100 *RTT et organisation du travail : l'incidence des lois Aubry II*, par P. CHARPENTIER (GRIOT-LISE, CNAM-CNRS), H. HUYGHUES DESPOINTES, M. LALLÈMENT (GRIOT-LISE, CNAM-CNRS), F. LEFRESNE (IRES et GRIOT-LISE, CNAM-CNRS), J. LOOS-BARON (BETA/LATTS-CNRS, CNAM-CNRS), N. TURPIN-HYARD (GRIOT-LISE, CNAM-CNRS), mai 2005.
- N° 101 *Éléments de bilan sur les travaux évaluant l'efficacité des allègements de cotisations sociales employeurs*, par Véronique REMY, juillet 2005.
- N° 102 *Les réticences à entrer dans le cadre légal des 35 heures*, par Y. JORAND et J.-M. GELIN (Selarj ESSOR), D. TONNEAU et F. FORT (CGS), B. DOERFLINGER, M. PEPIN et M. MAUFROY (Essor Consultants), juillet 2005.
- N° 103 *Allègements généraux de cotisations sociales et emploi peu qualifié : de l'impact sectoriel à l'effet macro-économique*, par Stéphanie JAMET (Dares lors de la réalisation de l'étude), août 2005.
- N° 104 *La négociation de branche sur la formation professionnelle : les apports de la négociation de branche suite à la réforme de la formation professionnelle tout au long de la vie*, par Caroline RIVIER et Carine SEILER, sous la direction de Jean-Marie LUTTRINGER (Circé), septembre 2005.
- N° 105 *Après un contrat aidé : les conditions de vie s'améliorent*, par Emmanuel BERGER et Tristan KLEIN, septembre 2005.
- N° 106 *Difficultés d'emploi, santé et insertion sociale*, par François BRUN, Colette LEYMARIE, Emma MBIA, Patrick NIVOLLE (Centre d'études de l'emploi), collaboration extérieure : Marie MARIN, octobre 2005.
- N° 107 *La sécurisation des trajectoires professionnelles*, par Dominique MEDA et Bertrand MINAULT, octobre 2005.
- N° 108 *Le licenciement des salariés protégés. Processus et enjeux*, par Mario CORREIA (Institut du travail d'Aix-en-Provence, LEST) et Nicole MAGGI-GERMAIN (Institut des sciences sociales du travail, Université Paris I, Panthéon-Sorbonne, DCS), février 2006.
- N° 109 *Les expositions aux risques professionnels par secteur d'activité (nomenclature 2003 niveau 31) - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, S. HAMON-CHOLET, D. WALTISPERGER (Dares), mars 2006.
- N° 110 *Les relations professionnelles dans les pays d'Europe centrale et orientale au tournant de l'entrée dans l'Union européenne. Survey de littérature*, par M. WIERINK, mars 2006.
- N° 111 *Renégocier la RTT. Les enseignements de 16 démarches d'entreprise*, par M. PEPIN, B. DOERFLINGER, Y. JORAND, P. NICOLAS (Essor Consultants) et D. TONNEAU (Ecole des Mines de Paris), avril 2006.
- N° 112 *La mesure d'un effet global du projet d'action personnalisé*, par Etienne DEBAUCHE et Stéphane JUGNOT, avril 2006.
- N° 113 *La politique spécifique de l'emploi et de la formation professionnelle : un profit à moyen terme pour les participants ? Les exemples du CIE, du CES et du SIFE*, par Karl EVEN et Tristan KLEIN, avril 2006.
- N° 114 *Stratégie européenne pour l'emploi. Évaluation des politiques de l'emploi et du marché du travail en France (2000-2004)*, coordination Christine CHARPAIL et Frédéric LERAIS, avril 2006.
- N° 115 *Les expositions aux risques professionnels - Les ambiances et contraintes physiques - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, L. VINCK, D. WALTISPERGER (Dares), juillet 2006.
- N° 116 *Pourquoi les moins qualifiés se forment-ils moins ?*, par Camille BONAÏTI, Aurore FLEURET, Patrick POMMIER, Philippe ZAMORA, juillet 2006.
- N° 117 *Le CDD : un tremplin vers le CDI dans deux tiers des cas... mais pas pour tous*, par Bérangère JUNOD, juillet 2006.
- N° 118 *Les expositions aux risques professionnels - Les produits chimiques - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, L. VINCK, D. WALTISPERGER (Dares), juillet 2006.
- N° 119 *Anticipation et accompagnement des restructurations d'entreprises : dispositifs, pratiques, évaluation*, par R. BEAUJOLIN-BELLET (coordination), Ch. CORNOLTI, J.-Y. KERBOUC'H, A. KUHN, Y. MOULIN (Reims Management School), et la collaboration de J.-M. BERGERE, F. BRUGGEMAN, B. GAZIER, D. PAUCARD, C.-E. TRIOMPHE, octobre 2006.
- N° 120 *Les expositions aux risques professionnels - Les contraintes organisationnelles et relationnelles - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, L. VINCK, D. WALTISPERGER (Dares), octobre 2006.
- N° 121 *Les expositions aux risques professionnels par famille professionnelle - Résultats SUMER 2003*, par Dr R. ARNAUDO, Dr I. MAGAUD-CAMUS, Dr N. SANDRET (DRT - Inspection médicale du travail et de la main-d'oeuvre), M.-C. FLOURY, N. GUIGNON, L. VINCK, D. WALTISPERGER (Dares), décembre 2006.
- N° 122 *Intérim : comparaison de sources*, par Basma SAADAoui, en collaboration avec Nicolas de RICCARDIS, mars 2007.
- N° 123 *Allègements de cotisations sociales et coûts sectoriels. Une approche par les DADS*, par Bertrand LHOMMEAU et Véronique REMY, avril 2007.
- N° 124 *Séries de données régionales sur les mouvements de main-d'oeuvre entre 1996 et 2005*, par Bruno LUTINIER, mai 2007.
- N° 125 *Colloque "Age et emploi". Emploi et travail des seniors : des connaissances à l'action. Synthèse des principales données sur l'emploi des seniors*, coordination Pierre MARIONI, juin 2007.
- N° 126 *Accès à l'emploi et qualité de l'insertion professionnelle des travailleurs handicapés en milieu ordinaire de travail*, par Claire FANJEAU (Université Paris I et Centre d'études de l'emploi), juin 2007.
- N° 127 *Le poids du temps partiel dans les trajectoires professionnelles des femmes*, par Sophie RIVAUD (stagiaires à la Dares) et Valérie ULRICH, juillet 2007.
- N° 128 *Analyse de l'évolution des statistiques de demandeurs d'emploi inscrits à l'ANPE de la mi-2005 à la fin 2006*, par Etienne DEBAUCHE, Thomas DEROYON, Fanny MIKOL et Hélène VALDELIEVRE, août 2007.
- N° 129 *Les déterminants de l'emploi non-salarié en France depuis 1970*, par Grégoire LURTON (EnsaE) et Fabien TOUTLEMONDE (Dares), septembre 2007.
- N° 130 *Revue de littérature : organisations patronales en France et en Europe* par Marion RABIER (ENS/EHESS - Dares), décembre 2007.
- N° 131 *The social multiplier and labour market, participation of mothers*, par Eric MAURIN (PSE) et Julie MOSCHION (CES-Université Paris I, Dares), décembre 2007.
- N° 132 *L'influence causale du nombre d'enfants et de leur âge de première scolarisation sur l'activité des mères : une revue de la littérature*, par Julie MOSCHION (CES-Université Paris I, Dares), décembre 2007.
- N° 133 *Conséquences des fusions-acquisitions sur la gestion de la main-d'oeuvre : une analyse empirique sur les données françaises pour la vague de la fin des années 1990*, par Matthieu BUNEL (CEE, Université de technologie de Belfort-Montbéliard), Richard DUHAUTOIS (CEE, CREST, Université de Marne-la-Vallée), Lucie GONZALEZ (Dares-MAE), janvier 2008.
- N° 134 *Les politiques d'allègements ont-elles un effet sur la mobilité salariale des travailleurs à bas salaires ?*, par Bertrand LHOMMEAU et Véronique REMY, janvier 2008.
- N° 135 *Le recours au chômage partiel entre 1995 et 2005*, par Oana CALAVREZO (LEO et CEE), Richard DUHAUTOIS (CEE, CREST, Université de Marne-la-Vallée) et Emmanuelle WALKOWIAK (LEO et CEE), février 2008.
- N° 136 *Enquête auprès des chômeurs créateurs ou repreneurs d'entreprise ayant bénéficié de l'Accre en 2004*, par Catherine DANIEL, février 2008.
- N° 137 *Les élections aux comités d'entreprise de 1989 à 2004 : une étude de l'évolution des implantations et des audiences syndicales*, par Olivier JACOD avec la collaboration de Rim BEN DHAOU (EnsaE), avril 2008.

- N° 138 **Les disparités spatiales de sortie du chômage : vingt-deux analyses régionales**, par Emmanuel DUGUET, Yannick L'HORTY (Université d'Evry-Val d'Essonne, EPEE, Centre d'études de l'emploi et TEPP), André WISSLER (Centre d'études de l'emploi et TEPP), Florent SARI (Université de Paris-Est, OEP, Centre d'études de l'emploi et TEPP), Jonathan BOUGARD et Luc GOUPIL (Centre d'études de l'emploi), mai 2008.
- N° 139 **Mesurer les grèves dans les entreprises : des données administratives aux données d'enquêtes**, par Alexandre CARLIER, août 2008.
- N° 140 **Evaluation de la loi du 4 mai 2004 sur la négociation d'accords dérogatoires dans les entreprises**, par O. MERRIAUX (Sciences-Po Recherche, IEP Grenoble), J-Y KERBOURC'H (Université de Haute-Alsace) et C. SEILER (Cabinet Circé Consultants), août 2008.
- N° 141 **Les modèles de projections d'emploi par métier à moyen terme. Panorama des expériences menées dans différents pays**, par Laure OMALEK, octobre 2008.
- N° 142 **Retour sur l'évolution du nombre de demandeurs d'emploi inscrits à l'ANPE en 2005 et 2006 : une estimation révisée de l'impact des modifications de la gestion et du suivi des demandeurs d'emploi sur le nombre d'inscrits à l'ANPE, en tenant compte de l'indemnisation**, par E. DEBAUCHE (Insee), Nicolas DERUYON et Fanny MIKOL (Dares), décembre 2008.
- N° 143 **Enquête sur trois secteurs : La Poste, sous-traitance pétrochimique et restauration rapide. Le syndicalisme face aux différentes formes de la flexibilité**, par P. BOUFFARTIGUE et J.-R. PENDARIES (LEST-CNRS, Université de la Méditerranée-Université de Provence), F. PEROUMAL (Université René Descartes-Paris V), E. PERRIN (consultante associée au LEST), avec la participation de J. BOUTREILLER, B. FRIBOURG et T. SAMZUN (LEST-CNRS, Université de la Méditerranée-Université de Provence), S. CONTREPOIS (GTM-CNRS), R. JEAN et E. ORBAN (ASPT-CNRS), décembre 2008.
- N° 144 **Flexibilité et action collective. Salariés précaires et représentation syndicale**, par C. DUFOUR, A. HEGE, J.-M. PERNOT (IRES), S. BEROU (Université Lyon2-Triangle-CNRS), J.-M. DENIS (Université de Marne-la-Vallée - CEE-CNRS), décembre 2008.
- N° 145 **Les retournements de l'improbable. Les conditions de la mobilisation collective des intermittents du spectacle et des salariés de grandes librairies et de centres d'appel**, par A. COLLOVALD (Université de Nantes), L. MATHIEU (CRPS), décembre 2008.
- N° 146 **La nouvelle méthode d'échantillonnage de l'enquête trimestrielle ACEMO depuis 2006. Amélioration de l'allocation de Neyman**, par Malik KOUBI et Sandrine MATHERN, février 2009.