

# L'impact du profilage sur la refonte du plan de sondage des Enquêtes Sectorielles Annuelles

Ronan Le Gleut, Thomas Merly-Alpa



Journées de méthodologie statistique de l'Insee 2018 - Paris

14 juin 2018

- 1 Introduction
- 2 Définition de l'exhaustif
- 3 Calcul des allocations
- 4 Résultats
- 5 Conclusion et perspectives

# Sommaire

- 1 Introduction
- 2 Définition de l'exhaustif
- 3 Calcul des allocations
- 4 Résultats
- 5 Conclusion et perspectives

## Deux principales enquêtes entreprises à l'Insee

Deux enquêtes entreprises faisant partie du dispositif Esane permettent de produire des statistiques structurelles :

**ESA** Enquête Sectorielle Annuelle :

**Champ** : activités de commerce, de construction, de services et de transport

# 116 000 unités interrogées en France métropolitaine

**EAP** Enquête Annuelle de Production

**Champ** : secteur industriel

# 35 000 unités interrogées en France métropolitaine

## Deux principales enquêtes entreprises à l'Insee

Deux enquêtes entreprises faisant partie du dispositif Esane permettent de produire des statistiques structurelles :

**ESA** Enquête Sectorielle Annuelle :

**Champ** : activités de commerce, de construction, de services et de transport

# 116 000 unités interrogées en France métropolitaine

**EAP** Enquête Annuelle de Production

**Champ** : secteur industriel

# 35 000 unités interrogées en France métropolitaine

L'un des principaux objectifs de ces deux enquêtes est de repérer les différentes activités exercées par les entreprises, via la ventilation de leur chiffre d'affaires (CA) en branche, et en déduire leur activité principale (APE).

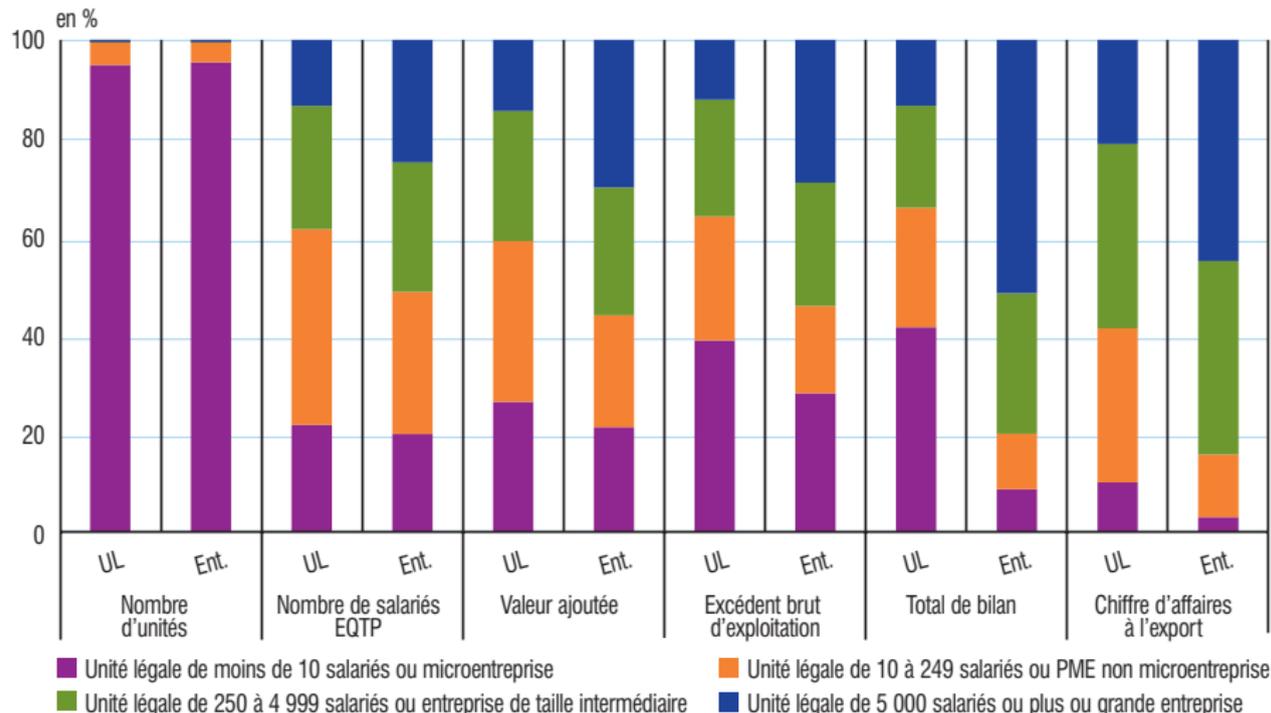
## Contexte

Dans de nombreux pays européens, la statistique d'entreprise est en grand changement.

En France, les enquêtes entreprises portent actuellement sur des établissements ou des **unités légales**.

A partir de maintenant, afin de répondre au règlement européen SBS, les statistiques d'entreprises seront de plus en plus basées sur la notion d'**entreprise**.

# Unité légale VS entreprise

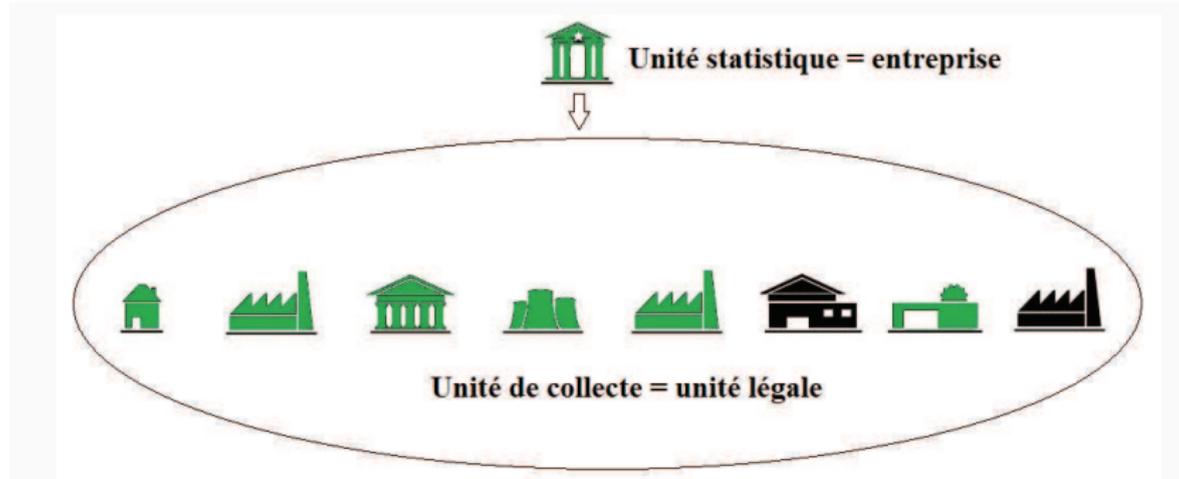


# Répartition des unités par service de gestion d'entreprise / d'unité légale

Service gestion du groupe	Service gestion des UL																		
	Dom	A2	A3	A4	B1	B2	B3	C1	C2	C3	D1	D3	D4	Commerce détail	Commerce gros	Construction IAA	Services entreprises	Services particuliers	Transports
A2	0,36	58,07	3,75	2,27	1,77	1,91	0,79	0,14	0,79	0,40	0,22	0,07	0,29	1,12	5,77	6,17	12,92	2,92	0,25
A3	0,08	2,96	67,96	2,33	1,18	2,13	0,73	0,05	0,13	0,35	0,20	0,18	0,50	0,80	5,04	1,08	10,96	3,06	0,28
A4	0,44	2,07	2,62	53,53	1,90	1,93	0,68	0,51	0,37	0,71	0,98	0,20	1,56	1,90	9,06	3,67	13,34	3,46	0,85
B1	0,36	1,85	1,35	0,96	59,82	2,19	0,51	0,62	0,62	0,28	0,22	0,42	0,48	4,01	8,08	2,81	11,45	3,45	0,51
B2	0,29	1,47	2,36	1,33	2,54	55,45	1,66	0,29	0,22	0,37	0,29	0,04	1,07	1,58	12,52	2,21	12,96	3,13	0,22
B3	0,00	0,67	1,28	0,33	0,57	2,23	64,43	1,00	0,19	0,05	0,24	0,00	0,24	1,00	5,99	0,52	15,83	5,23	0,19
C1	0,12	0,24	0,31	0,47	0,94	0,12	2,24	62,63	0,43	1,29	0,27	0,24	0,08	7,65	8,67	0,63	9,37	3,80	0,51
C2	0,06	0,73	0,36	0,85	0,85	0,12	0,18	0,79	58,32	1,09	0,85	0,24	0,06	4,86	8,93	6,20	11,42	3,10	0,97
C3	0,76	0,22	0,44	0,51	0,70	0,35	0,13	0,82	4,11	50,49	7,90	0,76	0,22	2,53	8,66	4,42	12,10	3,29	1,61
D1	1,50	0,62	0,23	0,69	0,73	0,77	0,35	0,12	0,42	0,65	53,94	0,96	2,88	1,61	8,73	6,81	13,19	4,58	1,23
D3	0,39	0,08	0,00	0,24	1,18	0,24	0,16	1,26	0,16	2,04	1,57	46,70	0,24	3,62	18,87	2,83	13,13	6,53	0,79
D4	0,58	0,73	1,26	2,38	1,17	1,65	0,68	0,15	0,19	0,78	2,38	0,19	55,50	1,17	10,75	3,55	12,60	3,11	1,17
Commerce détail	0,32	0,03	0,07	0,06	0,23	0,06	0,02	0,20	0,05	0,10	0,07	0,02	0,04	79,84	3,41	1,25	7,05	6,67	0,50
Commerce gros	0,44	0,19	0,22	0,32	0,34	0,43	0,23	0,37	0,19	0,45	0,48	0,27	0,30	5,52	75,27	2,31	8,15	3,45	1,06
Construction IAA	0,36	0,58	0,16	0,17	0,25	0,27	0,07	0,04	0,44	0,34	0,77	0,11	0,78	2,20	4,70	70,60	10,08	7,25	0,82
Services entreprises	0,43	0,12	0,08	0,23	0,16	0,17	0,26	0,10	0,10	0,26	0,51	0,10	0,15	1,63	2,70	2,97	82,43	6,91	0,72
Services particuliers	0,36	0,02	0,02	0,04	0,02	0,03	0,12	0,04	0,03	0,03	0,06	0,02	0,04	1,33	1,41	4,47	8,99	82,69	0,31
Transports	0,97	0,03	0,05	0,34	0,26	0,19	0,01	0,03	0,18	0,29	0,20	0,05	0,33	5,61	2,21	2,10	11,53	9,21	66,41

## Unité statistique VS unité interrogée

Étant donné que l'unité statistique (entreprise) est maintenant différente de l'unité interrogée (unité légale), le plan de sondage peut être vu comme un **tirage en grappe**.



# Objectifs

Les principales contraintes sont donc les suivantes :

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .
- . . . pour une diffusion des résultats en entreprises. . .

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .
- . . . pour une diffusion des résultats en entreprises. . .
- . . . tout en pouvant exploiter les résultats en unités légales . . .

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .
- . . . pour une diffusion des résultats en entreprises. . .
- . . . tout en pouvant exploiter les résultats en unités légales . . .
- . . . la diffusion se faisant sur deux domaines de publication . . .

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .
- . . . pour une diffusion des résultats en entreprises. . .
- . . . tout en pouvant exploiter les résultats en unités légales . . .
- . . . la diffusion se faisant sur deux domaines de publication . . .
- . . . en conservant le même nombre d'unités légales qu'actuellement à enquêter pour chacune des deux enquêtes . . .

# Objectifs

Les principales contraintes sont donc les suivantes :

- Réoptimiser le plan de sondage dans ce nouveau contexte. . .
- . . . pour une diffusion des résultats en entreprises. . .
- . . . tout en pouvant exploiter les résultats en unités légales . . .
- . . . la diffusion se faisant sur deux domaines de publication . . .
- . . . en conservant le même nombre d'unités légales qu'actuellement à enquêter pour chacune des deux enquêtes . . .
- . . . et que ce nombre reste stable d'une année sur l'autre !

# Sommaire

- 1 Introduction
- 2 Définition de l'exhaustif**
- 3 Calcul des allocations
- 4 Résultats
- 5 Conclusion et perspectives

# Définition de l'exhaustif

Les règles appliquées au niveau entreprise sont les suivantes :

- Les anciens seuils d'exhaustivité en termes de CA, d'effectifs et de total de bilan par activité ont été repris...
- ... puis modulés par un taux de couverture du CA à couvrir par activité.

## Définition de l'exhaustif

Les règles appliquées au niveau entreprise sont les suivantes :

- Les anciens seuils d'exhaustivité en termes de CA, d'effectifs et de total de bilan par activité ont été repris...
- ... puis modulés par un taux de couverture du CA à couvrir par activité.

Les entreprises de plus de 200 salariés, de plus de 50 000 K€ de CA et celles composées de plus de 20 unités légales sont également forcées dans l'exhaustif.

## Définition de l'exhaustif

Les règles appliquées au niveau entreprise sont les suivantes :

- Les anciens seuils d'exhaustivité en termes de CA, d'effectifs et de total de bilan par activité ont été repris...
- ... puis modulés par un taux de couverture du CA à couvrir par activité.

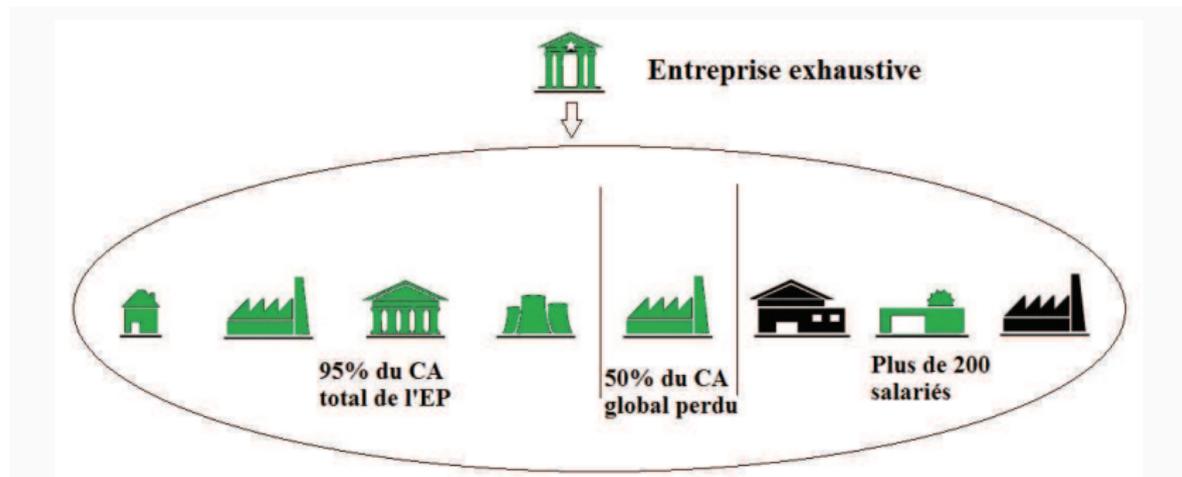
Les entreprises de plus de 200 salariés, de plus de 50 000 K€ de CA et celles composées de plus de 20 unités légales sont également forcées dans l'exhaustif.

Enfin, un cut-off à 95% du CA au sein de chaque entreprise est effectué sauf pour les plus grosses unités légales (voir exemple diapo suivante).

## Exemple d'une entreprise exhaustive

Les unités légales en **vert** font partie de l'exhaustif car :

- elles sont dans le cut-off à 95% du CA au sein de l'entreprise ;
- elles sont cut-offées à 95% mais ont plus de 200 salariés ou font partie des unités légales permettant de récupérer 50% du **total de CA** perdu lors du cut-off.



## Détection des unités atypiques (avant calcul des allocations)

Certaines entreprises peuvent encore avoir un CA atypique par rapport aux autres unités de leur strate de tirage.

⇒ peut poser des problèmes pour le calcul des allocations !

## Détection des unités atypiques (avant calcul des allocations)

Certaines entreprises peuvent encore avoir un CA atypique par rapport aux autres unités de leur strate de tirage.

⇒ peut poser des problèmes pour le calcul des allocations !

Trois méthodes sont exploitées afin d'identifier des unités atypiques dans chaque strate, et de les forcer dans l'exhaustif :

- La contribution d'une unité à la variance ;
- La méthode de [Kokic et Bell \(1994\)](#) ;
- L'algorithme des centres mobiles (k-means).

# Sommaire

- 1 Introduction
- 2 Définition de l'exhaustif
- 3 Calcul des allocations**
- 4 Résultats
- 5 Conclusion et perspectives

# Stratification et domaines de diffusion

La stratification est définie par le croisement entre :

- Secteur d'activité de l'entreprise (au niveau APE)
- Effectif de l'entreprise (9 tranches d'effectifs - teff)

Les deux domaines de diffusion sont les suivants :

- APE (activité sur 5 positions)
- Groupe (activité sur 3 positions)  $\times$  teff regroupées

## Allocation de Neyman sous contraintes de coûts

Étant donné que l'unité enquêtée reste l'**unité légale**, la première contrainte du calcul des allocations porte sur le nombre d'unités légales à interroger **pour chaque enquête**, qui ne doit pas dépasser  $N_{UL}$ .

Afin de respecter ces contraintes, nous avons introduit des contraintes de **coûts** dans le calcul de l'allocation de Neyman :

$$\left\{ \begin{array}{l} \min_{n_1, \dots, n_H} \mathbb{V}_p[\hat{t}_{y\pi}] = \sum_{h=1}^H N_h^2 \frac{1 - f_h}{n_h} S_{yh}^2 \\ \text{s.c. } \sum_{h=1}^H C_h n_h = N_{UL} \\ \text{s.c. } n_h \leq N_h \end{array} \right.$$

avec  $C_h = \bar{N}_{UL,h}$  le **coût**, i.e. le nombre moyen d'UL par entreprise dans la strate  $h$ .

# Allocation de Neyman sous contraintes de précisions locales

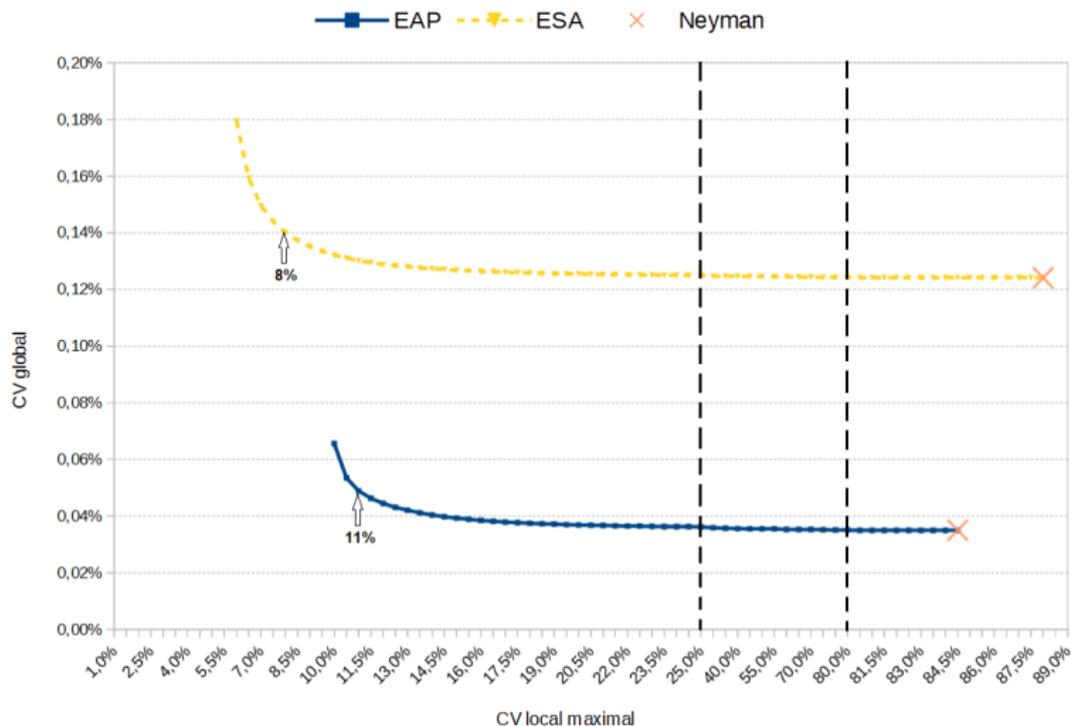
L'objectif est aussi de prendre en considération des contraintes de **précisions locales** sur les deux domaines de diffusion.

Afin de prendre en compte ces contraintes tout en imposant un nombre moyen d'UL à enquêter fixe (contraintes de coûts), nous avons adapté l'algorithme proposé par [Koubi et Mathern \(2009\)](#) :

$$\left\{ \begin{array}{l} \min_{n_1, \dots, n_H} \mathbb{V}_p[\hat{t}_{y\pi}] = \sum_{h=1}^H N_h^2 \frac{1 - f_h}{n_h} S_{yh}^2 \\ \text{s.c. } \sum_{h=1}^H C_h n_h = N_{UL} \\ \text{s.c. } n_h \leq N_h \\ \text{s.c. } \max_{d \in D} CV_d \leq CV_{loc} \end{array} \right.$$

avec  $D$  l'ensemble des domaines de diffusion et  $CV_{loc}$  le coefficient de variation maximum attendu.

## Exemple d'une frontières d'efficacité

Figure – Frontière d'efficacité pour le domaine de diffusion Groupe  $\times$  teff.

## Adaptation de l'algorithme de Koubi et Mathern (2009)

Pour un coût total fixé  $C$ , l'objectif est de déterminer l'ordre de désaturation ou de saturation de chaque strate :

- Pour chaque strate  $h$ , on calcule le nombre minimum d'unités à tirer  $n_h^{min}$  permettant respecter les contraintes précision locales  $CV_{loc}$

## Adaptation de l'algorithme de Koubi et Mathern (2009)

Pour un coût total fixé  $C$ , l'objectif est de déterminer l'ordre de désaturation ou de saturation de chaque strate :

- Pour chaque strate  $h$ , on calcule le nombre minimum d'unités à tirer  $n_h^{min}$  permettant respecter les contraintes précision locales  $CV_{loc}$
- L'allocation de Neyman intégrant les contraintes de coûts sur les strates non saturées  $H'$  (au minimum  $n_h^{min}$  ou au maximum  $N_h$ ) :

$$n_h = (C - C_{sat}) \frac{(N_h S_{yh}) / \sqrt{C_h}}{\sum_{j \in H'} N_j S_{yj} \sqrt{C_j}} \quad \text{ici} \quad (1)$$

## Adaptation de l'algorithme de Koubi et Mathern (2009)

Pour un coût total fixé  $C$ , l'objectif est de déterminer l'ordre de désaturation ou de saturation de chaque strate :

- Pour chaque strate  $h$ , on calcule le nombre minimum d'unités à tirer  $n_h^{min}$  permettant respecter les contraintes précision locales  $CV_{loc}$
- L'allocation de Neyman intégrant les contraintes de coûts sur les strates non saturées  $H'$  (au minimum  $n_h^{min}$  ou au maximum  $N_h$ )

$$n_h = (C - C_{sat}) \frac{(N_h S_{yh}) / \sqrt{C_h}}{\sum_{j \in H'} N_j S_{yj} \sqrt{C_j}} \quad \text{ici} \quad (1)$$

- La formule (1) donne les valeurs limites du coût  $C$  pour chaque strate (au minimum si  $a_h = n_h^{min}$  ou au maximum si  $a_h = N_h$ ) :

$$C_{lim}(h) = C_{sat} + \frac{a_h \sqrt{C_h}}{N_h S_{yh}} \sum_{j \in H'} N_j S_{yj} \sqrt{C_j} \quad (2)$$

## Adaptation de l'algorithme de Koubi et Mathern (2009)

Pour un coût total fixé  $C$ , l'objectif est de déterminer l'ordre de désaturation ou de saturation de chaque strate :

- Pour chaque strate  $h$ , on calcule le nombre minimum d'unités à tirer  $n_h^{min}$  permettant respecter les contraintes précision locales  $CV_{loc}$
- L'allocation de Neyman intégrant les contraintes de coûts sur les strates non saturées  $H'$  (au minimum  $n_h^{min}$  ou au maximum  $N_h$ )

$$n_h = (C - C_{sat}) \frac{(N_h S_{yh}) / \sqrt{C_h}}{\sum_{j \in H'} N_j S_{yj} \sqrt{C_j}} \quad \text{ici} \quad (1)$$

- La formule (1) donne les valeurs limites du coût  $C$  pour chaque strate (au minimum si  $a_h = n_h^{min}$  ou au maximum si  $a_h = N_h$ ) :

$$C_{lim}(h) = C_{sat} + \frac{a_h \sqrt{C_h}}{N_h S_{yh}} \sum_{j \in H'} N_j S_{yj} \sqrt{C_j} \quad (2)$$

- Le terme dépendant de  $h$  en **bleu** dans la formule (2) donne l'ordre de désaturation de  $n_h^{min}$  ou de saturation à  $N_h$  de chaque strate

## Exemple

Strate	$n_h^{min}$	$N_h$	$S_{yh}$	$C_h$
1	2	8	0.25	2
2	10	30	0.1	1.7
3	32	48	0.08	1.5

L'ordre de désaturation ou de saturation de chaque strate est donné par  $ord_h^{min} = n_h^{min} \sqrt{C_h} / (N_h S_{yh})$  ou  $ord_h^{max} = N_h \sqrt{C_h} / (N_h S_{yh})$  :

Contrainte	$ord_1^{min}$	$ord_1^{max}$	$ord_2^{min}$	$ord_2^{max}$	$ord_3^{min}$	$ord_3^{max}$
Valeur	1.41	5.66	4.35	13.04	10.21	15.31
Ordre	1	3	2	5	4	6

## Exemple

Contraintes :

1, min

2, min

1, max

3, min

2, max

3, max

 $C_{sat}$  :

$$C_1 n_1^{min} + C_2 n_2^{min} + C_3 n_3^{min}$$

69

$$C_2 n_2^{min} + C_3 n_3^{min}$$

65

$$C_3 n_3^{min}$$

48

$$C_1 N_1 + C_3 n_3^{min}$$

64

$$C_1 N_1$$

16

$$C_1 N_1 + C_2 N_2$$

67

$$C_1 N_1 + C_2 N_2 + C_3 N_3$$

139

 $C$  $C_{lim}(h)$  :

69

77.3

86.1

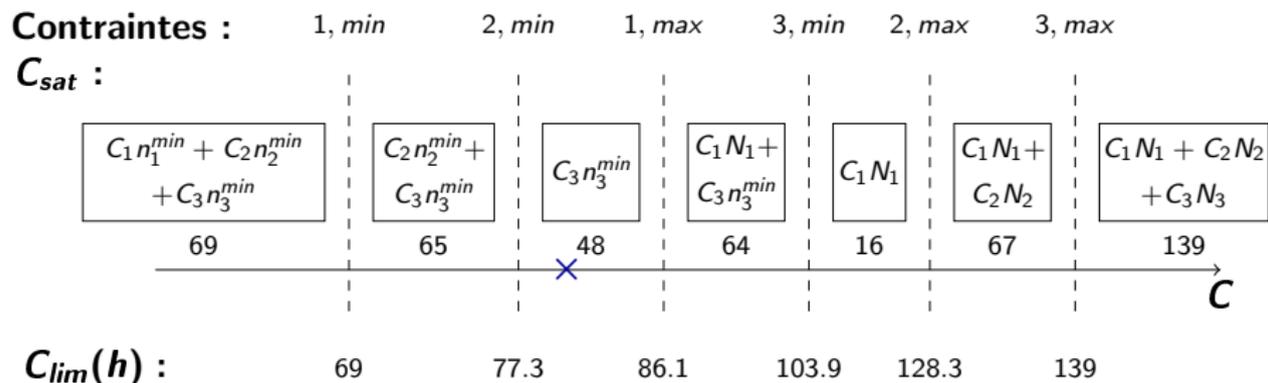
103.9

128.3

139

Pour un coût total fixé  $C$ , on sait à partir de chaque  $C_{lim}(h)$  quelles sont les strates saturées au minimum ou au maximum.

## Exemple



Pour un coût total fixé  $C$ , on sait à partir de chaque  $C_{lim}(h)$  quelles sont les strates saturées au minimum ou au maximum.

Par exemple, pour sélectionner un échantillon avec  $C = 80$  :

- la strate 3 est saturée au minimum et  $C_3 n_3^{min} = C_{sat} = 48$
- une optimisation classique de Neyman est appliquée sur les strates non saturées 1 et 2 avec  $80 - 48 = 32$  unités restantes à tirer [ici](#)

# Sommaire

- 1 Introduction
- 2 Définition de l'exhaustif
- 3 Calcul des allocations
- 4 Résultats**
- 5 Conclusion et perspectives

## Estimateur du nombre d'unités légales à enquêter

L'allocation de Neyman intègre un **coût moyen par strate**, ce qui conduit en **moyenne** au nombre d'unités légales voulu  $N_{UL}$ .

## Estimateur du nombre d'unités légales à enquêter

L'allocation de Neyman intègre un **coût moyen par strate**, ce qui conduit en **moyenne** au nombre d'unités légales voulu  $N_{UL}$ .

Mais le nombre d'unités légales qui seront réellement enquêtées est aléatoire et dépend de l'échantillon sélectionné :

$$\hat{N}_{UL} = \sum_{h=1}^H \sum_{k \in S_h} N_{UL,k} ,$$

avec  $N_{UL,k}$  le nombre d'unités légales de l'entreprise  $k$ .

## Estimateur du nombre d'unités légales à enquêter

L'allocation de Neyman intègre un **coût moyen par strate**, ce qui conduit en **moyenne** au nombre d'unités légales voulu  $N_{UL}$ .

Mais le nombre d'unités légales qui seront réellement enquêtées est aléatoire et dépend de l'échantillon sélectionné :

$$\hat{N}_{UL} = \sum_{h=1}^H \sum_{k \in S_h} N_{UL,k} ,$$

avec  $N_{UL,k}$  le nombre d'unités légales de l'entreprise  $k$ .

Cette quantité peut se réécrire comme l'estimateur d'Horvitz-Thompson de la variable  $z_k = \pi_k N_{UL,k}$  :

$$\hat{N}_{UL} = \sum_{h=1}^H \sum_{k \in S_h} N_{UL,k} = \sum_{h=1}^H \sum_{k \in S_h} \frac{z_k}{\pi_k} = \hat{t}_{z\pi}$$

## Variance de cet estimateur et variabilité des résultats

En conséquence, la variance de cet estimateur est :

$$\mathbb{V}_p \left[ \hat{N}_{UL} \right] = \sum_{h=1}^H n_h (1 - f_h) S_{N_{UL},h}^2$$

avec  $S_{N_{UL},h}^2 = \frac{1}{N_h - 1} \sum_{k \in U_h} (N_{UL,k} - \bar{N}_{UL,h})^2$  la variance empirique de  $N_{UL,k}$  dans la strate  $h$ .

## Variance de cet estimateur et variabilité des résultats

En conséquence, la variance de cet estimateur est :

$$\mathbb{V}_p \left[ \hat{N}_{UL} \right] = \sum_{h=1}^H n_h (1 - f_h) S_{N_{UL},h}^2$$

avec  $S_{N_{UL},h}^2 = \frac{1}{N_h - 1} \sum_{k \in U_h} (N_{UL,k} - \bar{N}_{UL,h})^2$  la variance empirique de  $N_{UL,k}$  dans la strate  $h$ .

	<i>EAP</i>	<i>ESA</i>	<i>Total</i>
$n_{ent}$	27 000	82 700	109 700
$\mathbb{E}_p \left[ \hat{N}_{UL} \right] = N_{UL}$	35 000	116 000	151 000
$IC_{95\%}(N_{UL})$	[34 970 ; 35 030]	[115 840 ; 116 160]	[150 830 ; 151 170]

**Table** – Résultats relatifs au nombre d'entreprises à tirer ( $n_{ent}$ ) et au nombre d'unités légales à enquêter ( $N_{UL}$ ).

## Précisions pour une diffusion au niveau entreprise

Niveaux	Domaines de diffusion					
	APE (1)			Groupe × teff (2)		
	$n_{ent,1}$	$n_{ent,2}$	$n_{ent,mix}$	$n_{ent,1}$	$n_{ent,2}$	$n_{ent,mix}$
100% Max	5%	<b>74.4%</b>	<b>23.1%</b>	<b>89.3%</b>	11%	<b>43.1%</b>
90%	5%	<b>9%</b>	<b>6.3%</b>	<b>20.8%</b>	11%	<b>12.5%</b>
75% Q3	5%	4.9%	4.4%	9.2%	8%	8.9%
50% Médiane	2%	2%	2%	4.2%	4.6%	4.2%
25% Q1	0.9%	0.8%	0.8%	0.1%	0.2%	0.2%
10%	0.2%	0.1%	0.2%	0%	0%	0%
0% Min	0%	0%	0%	0%	0%	0%

**Table** – Distribution des CV locaux des estimations de CA **au niveau entreprise** selon l'allocation et le domaine considérés (sans la strate exhaustive des plus de 200 salariés pour le domaine Groupe × teff).

## Précisions pour une diffusion au niveau unité légale

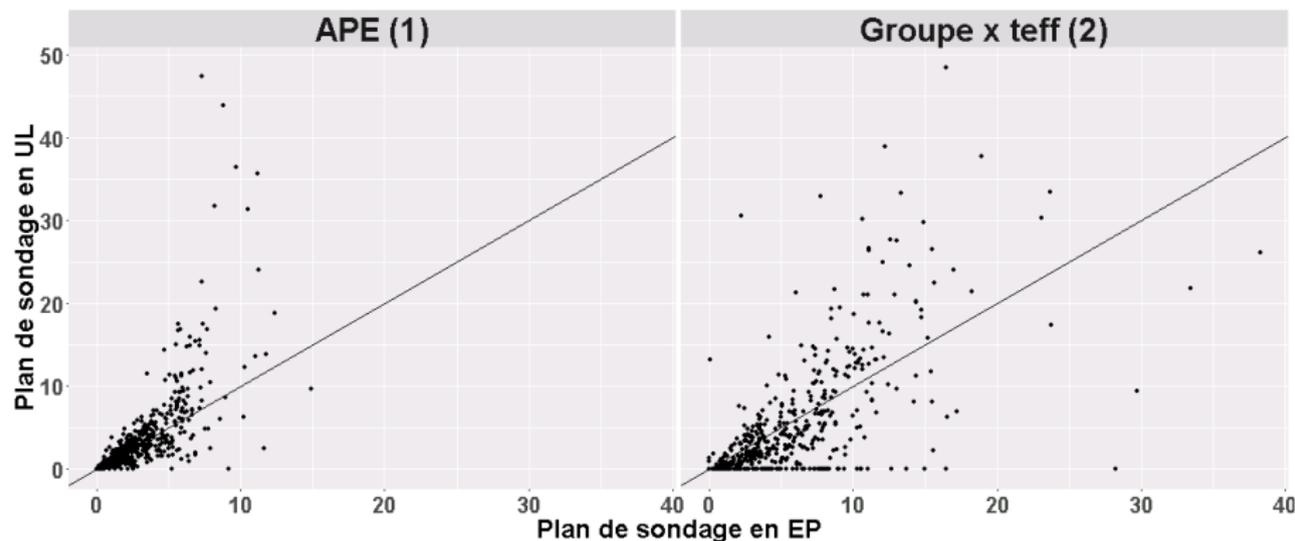


Figure – CV locaux des estimations de CA **au niveau unité légale** selon le plan de sondage (UL/EP) et le domaine considérés. les critères d'exhaustivité (unités de + de 100 salariés, etc.); [ici](#)

# Sommaire

- 1 Introduction
- 2 Définition de l'exhaustif
- 3 Calcul des allocations
- 4 Résultats
- 5 Conclusion et perspectives**

# Conclusion

La prise en compte du concept économique d'entreprise dans les statistiques structurelles constitue un changement de paradigme majeur. . .

# Conclusion

La prise en compte du concept économique d'entreprise dans les statistiques structurelles constitue un changement de paradigme majeur. . .

. . . qui implique une refonte du plan de sondage des Enquêtes Sectorielles Annuelles. . .

- unités de collecte différentes des unités statistiques ;
- optimisation du plan de sondage pour une diffusion au niveau entreprise sous des contraintes pesant sur les unités de collecte ;
- un échantillon permettant une exploitation au niveau unité légale avec une qualité satisfaisante.

# Conclusion

La prise en compte du concept économique d'entreprise dans les statistiques structurelles constitue un changement de paradigme majeur. . .

. . . qui implique une refonte du plan de sondage des Enquêtes Sectorielles Annuelles. . .

- unités de collecte différentes des unités statistiques ;
- optimisation du plan de sondage pour une diffusion au niveau entreprise sous des contraintes pesant sur les unités de collecte ;
- un échantillon permettant une exploitation au niveau unité légale avec une qualité satisfaisante.

. . . et qui permet une meilleure vision du tissu productif français.

## Perspectives

Quelques changements ont depuis eu lieu dans le plan de sondage sur :

- les critères d'exhaustivité (unités de + de 100 salariés, etc.) ; [ici](#)
- l'ajout d'un troisième domaine de diffusion (groupe d'activité) dans le calcul des allocations.

## Perspectives

Quelques changements ont depuis eu lieu dans le plan de sondage sur :

- les critères d'exhaustivité (unités de + de 100 salariés, etc.) ; [ici](#)
- l'ajout d'un troisième domaine de diffusion (groupe d'activité) dans le calcul des allocations.

D'autres analyses plus approfondies peuvent être menées sur :

- l'amélioration de la stratification (Dalenius et Hodges Jr, 1959 ; Gunning et Horgan, 2004 ; Lavallée et Hidiroglou, 1988) ;
- la recherche de facteurs optimaux ( $\alpha, 1 - \alpha$ ) dans le calcul de l'allocation mixte (Merly-Alpa et Rebecq, 2016) ;
- l'utilisation d'autres algorithmes de répartition multivariée de l'échantillon (Bethel, 1989 ; Falorsi et Righi, 2015) ;
- la mise en oeuvre d'un plan de sondage rotatif (Demoly et al., 2014).

Vielen Dank für Ihre Aufmerksamkeit !



Bethel, J. W. (1989).

Sample allocation in multivariate surveys.

*Survey Methodology*, 15(1) :47–57.



Chanteloup, G. (2018).

Consolider les réponses des unités légales pour une statistique d'entreprise plus cohérente.

*Journées de Méthodologie Statistique, Paris.*



Dalenius, T. and Hodges Jr, J. L. (1959).

Minimum variance stratification.

*Journal of the American Statistical Association*, 54(285) :88–101.



Demoly, E., Fizzala, A., and Gros, E. (2014).

Méthodes et pratiques des enquêtes entreprises à l'insee.

*Journal de la Société Française de Statistique*, 155(4) :134–159.



Deville, J.-C. and Lavallée, P. (2006).

Indirect sampling : The foundations of the generalized weight share method.

*Survey Methodology*, 32(2) :165–176.



Falorsi, P. D. and Righi, P. (2015).

Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys.

*Survey Methodology*, 41(1) :215–236.



Fizzala, A. (2018a).

Comment redresser un échantillon d'unités légales tirées via leurs entreprises ?

*Journées de Méthodologie Statistique, Paris.*



Fizzala, A. (2018b).

La gestion par partage des poids des changements de contour des entreprises dans l'enquête sectorielle annuelle.

*Journées de Méthodologie Statistique, Paris.*



Gros, E. and Le Gleut, R. (2018).

Sample coordination and response burden for business surveys : methodology and practice of the procedure implemented at insee.

*Cambridge Scholars Publishing, accepted.*



Gunning, P. and Horgan, J. M. (2004).

A new algorithm for the construction of stratum boundaries in skewed populations.

*Survey Methodology*, 30(2) :159–166.



Kokic, P. N. and Bell, P. A. (1994).

Optimal winsorizing cutoffs for a stratified finite population estimator.  
*Journal of Official Statistics*, 10(4) :419–435.

 Koubi, M. and Mathern, S. (2009).  
Résolution d'une des limites de l'allocation de Neyman.  
*Journées de Méthodologie Statistique, Paris*.

 Lavallée, P. and Hidiroglou, M. A. (1988).  
On the stratification of skewed populations.  
*Survey methodology*, 14(1) :33–43.

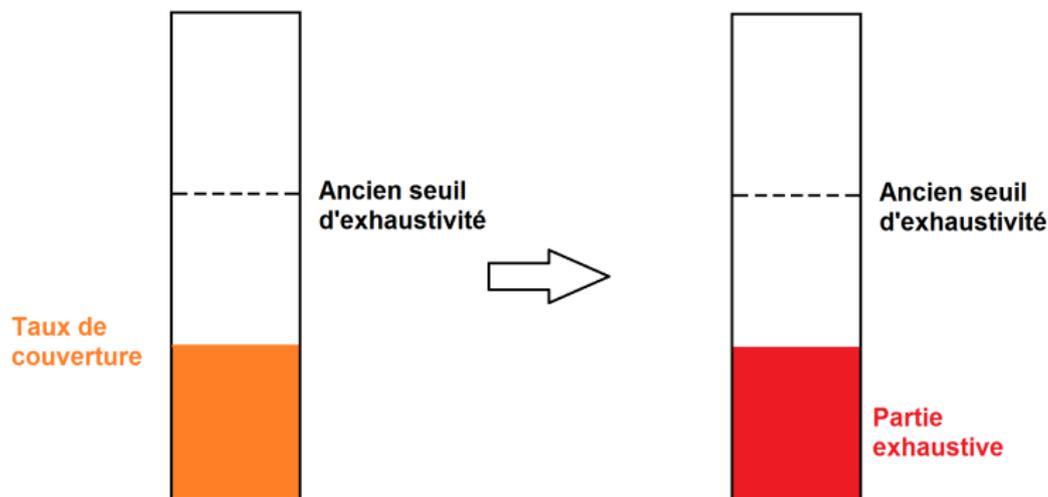
 Merly-Alpa, T. and Rebecq, A. (2016).  
Optimisation d'une allocation mixte.  
*9ème colloque francophone sur les Sondages, Gatineau*.

 Neyman, J. (1934).  
On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection.  
*Journal of the Royal Statistical Society*, 97(4) :558–625.

 Regulation EEC (1993).  
Council Regulation (EEC) 696 / 93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community.



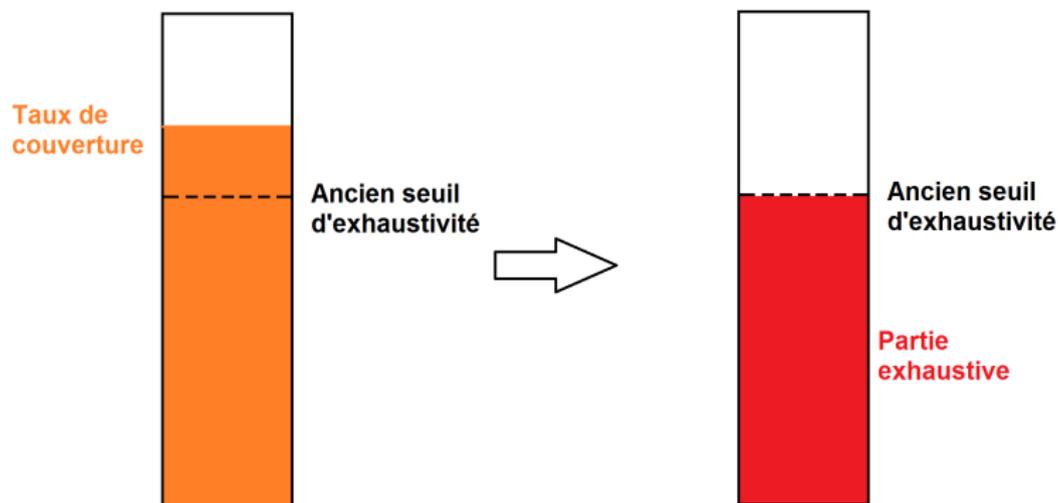
## Définition de l'exhaustif : Cas 1



Le taux de couverture du CA de l'exhaustif en entreprise avec les critères historiques ne dépasse pas l'ancien seuil d'exhaustivité du CA en UL.

⇒ **On ne change rien.**

## Définition de l'exhaustif ESA : Cas 2



Le taux de couverture du CA de l'exhaustif en entreprise avec les critères historiques dépasse l'ancien seuil d'exhaustivité du CA en UL.

⇒ On effectue un cut-off à l'ancien seuil d'exhaustivité du CA parmi les entreprises qui vérifient les critères historiques.