

---

## COMMENT REDRESSER UN ÉCHANTILLON D'UNITÉS LÉGALES TIRÉES VIA LEURS ENTREPRISES ?

L'ADAPTATION DE LA CORRECTION DE LA NON-RÉPONSE, DU TRAITEMENT DES VALEURS  
INFLUENTES ET DU CALAGE DE L'ENQUÊTE SECTORIELLE ANNUELLE ET DE L'ENQUÊTE ANNUELLE DE  
PRODUCTION À LEUR NOUVEAU PLAN DE SONDAGE

Arnaud FIZZALA

Insee, Direction de la méthodologie et de la coordination statistique et internationale,  
Département des méthodes statistiques, Division Sondages.

[arnaud.fizzala@insee.fr](mailto:arnaud.fizzala@insee.fr)

**Mots-clés** : non-réponse, calage, winsorisation, entreprises, profilage

---

### Résumé

Depuis le millésime 2016, le tirage des échantillons de l'enquête sectorielle annuelle (ESA) et de l'enquête annuelle de production (EAP), qui font partie du système d'élaboration des statistiques annuelles d'entreprises (ESANE) [1] est réalisé au niveau des entreprises profilées. Lorsqu'une entreprise profilée est tirée, toutes<sup>1</sup> les unités légales relevant du champ de l'enquête (en tant qu'unité légale) qui lui sont rattachées sont sélectionnées dans l'échantillon d'unités légales correspondant. On envoie alors un questionnaire aux unités légales de cet échantillon, et les réponses des entreprises profilées sont ensuite « reconstituées » à partir des retours de questionnaires des unités légales.

Les premiers résultats concernant l'année 2016 ont été diffusés début 2018, au niveau unités légales seulement, le règlement européen sur les statistiques structurelles annuelles n'exigeant l'élaboration de statistiques au niveau des entreprises qu'à partir de l'année 2017. A partir du millésime 2017, les résultats d'ESANE seront élaborés au niveau des entreprises afin de répondre à ce règlement européen... Mais aussi au niveau des unités légales car le règlement européen encadrant les comptes nationaux n'est lui pas passé au concept d'entreprises profilées.

Cette perspective de devoir continuer, pour au moins quelques années, à produire des résultats au niveau des unités légales malgré le passage du plan de sondage à un niveau entreprises a conduit à adapter différents aspects du processus d'élaboration des résultats d'ESANE.

L'article présentera l'adaptation au nouveau plan de sondage des traitements post-collecte (traitement de la non-réponse, traitement des valeurs influentes, calage) d'ESANE en se limitant<sup>2</sup> à l'élaboration des résultats au niveau unités légales.

C'est notamment le traitement des valeurs influentes qui a du être adapté car la méthode utilisée jusqu'ici - Winsorisation avec des seuils constitués selon la méthode de Kokic et Bell [2] - a été conçue pour s'appliquer dans le strict cadre d'échantillons aléatoires simples stratifiés<sup>3</sup>. En effet, d'un « point de vue unités légales », le tirage aléatoire simple stratifié d'entreprises profilées

---

<sup>1</sup>En pratique, toutes les unités légales ne sont pas forcément interrogées et on procède alors par imputation pour les celles contribuant le moins au chiffre d'affaires de l'entreprise profilée.

<sup>2</sup> L'adaptation des traitements post-collecte pour l'élaboration des résultats au niveau des entreprises est encore en cours d'étude et fait l'objet d'une autre proposition d'article aux JMS 2018.

<sup>3</sup> Cadre qui est en effet rencontré de façon systématique ou presque dans les enquêtes alimentant la statistique d'entreprises française.

correspondant au nouveau plan de sondage s'apparente à un tirage d'unités légales en grappes, les grappes correspondant aux entreprises profilées.

Plusieurs stratégies ont été envisagées afin d'adapter le traitement des valeurs influentes au nouveau plan de sondage en entreprises :

- Appliquer la Winsorisation avec les seuils de Kokic et Bell « en faisant comme si » l'échantillon d'unités légales était tiré selon un plan de sondage aléatoire simple stratifié d'unités légales ;
- Appliquer des méthodes alternatives se basant sur le biais conditionnel [3] et s'adaptant facilement au nouveau plan de sondage.

Une comparaison de ces stratégies a pu être réalisée en amont de la production des premiers résultats 2016 en s'appuyant sur des variables fiscales disponibles pour l'ensemble des unités légales<sup>4</sup>, et sur des simulations de tirage d'échantillons d'entreprises selon le nouveau plan de sondage de l'ESA et de l'EAP. Cette comparaison, dont la présentation détaillée constituera l'essentiel de l'article, a mené à retenir la Winsorisation avec les seuils de Kokic et Bell « en faisant comme si » l'échantillon d'unités légales était tiré selon un plan de sondage aléatoire simple stratifié d'unités légales. Et c'est effectivement la méthode qui a été appliquée pour l'élaboration des premiers résultats 2016.

Les autres traitements post-collecte nécessitent moins d'ajustements pour s'adapter au nouveau plan de sondage et seront donc évoqués de façon plus succincte dans l'article. Quelques tests de convergence du calage ont tout de même été réalisés et laissent penser que le calage devrait pouvoir continuer à être pratiqué en utilisant les mêmes variables et les mêmes marges que « sous » l'ancien plan de sondage<sup>5</sup>. Le traitement de la non-réponse ne pouvait pas être étudié en amont de la réception des premières données et n'a été modifié qu'à la marge pour l'élaboration des premiers résultats : l'indicatrice d'appartenance à une entreprise profilée a été ajoutée comme variable mobilisable pour constituer les groupes de réponses homogènes.

## Bibliographie

[1] P. Brion, 2011 *“Esane, le dispositif rénové de production des statistiques structurelles d'entreprises”* Courrier des statistiques n°130.

[2] T. Deroyon, 2015 *“Traitement des valeurs atypiques d'une enquête par winsorization - application aux enquêtes sectorielles annuelles”*. Acte des Journées de Méthodologie Statistique 2015.

[3] C. Favre Martinoz, D. Haziza, J-F. Beaumont, 2015 *“A method of determining the winsorization threshold, with an application to domain estimation”* Survey Methodology, vol. 41, n°1 (June): 57-77;

---

<sup>4</sup> Il n'était donc pas nécessaire d'attendre les résultats de l'enquête pour commencer à étudier les méthodes.

<sup>5</sup> En pratique, le calage s'est effectivement déroulé sans difficulté particulière pour l'élaboration des premiers résultats au niveau unités légales portant sur 2016.