



Février 2016

Fiche méthodologique

Cadre de sondage STATEM

Statistique de l'emploi (STATEM)

Fiche méthodologique

Cadre de sondage 2015

Principales caractéristiques

Univers de référence :	516'000* entreprises englobant 592'000 établissements (unités locales du REE) 4'800'000 emplois ou 3'804'000 EPT
Taille de l'échantillon :	18'000 entreprises (3,5%) englobant 65'000 établissements (11,0%) 2'171'000 emplois (45%) ou 1'785'000 EPT (47%)

* Tous les chiffres sont arrondis au millier

Révision 2015

La révision de la Statistique de l'emploi s'inscrit dans le cadre du Programme général de statistiques des entreprises de l'OFS, avec pour buts, entre autres, la réduction de la charge auprès des entreprises et l'utilisation des données administratives. Dans ce contexte, le cadre de sondage et l'échantillon de la STATEM ont été adaptés au 2^{ème} trimestre 2015. L'univers de référence a été adapté à la nouvelle Statistique structurelle des entreprises (STATENT) qui est elle-même fondée sur la source AVS (Assurance vieillesse et survivants) et qui remplace les Recensements fédéraux des entreprises (RFE). Le cadre de sondage s'est donc considérablement agrandi. Il comprend désormais les entreprises dans lesquelles on travaille moins de 20h par semaine, ainsi que les emplois de moins de 6h hebdomadaires qui étaient absents des RFE. Pour cette raison, l'ensemble des séries publiées ont été révisées.

Définition de l'emploi

La définition de l'emploi a été adaptée à la notion d'emploi selon l'AVS dès 2011. Selon cette définition, il est demandé aux entreprises d'annoncer tous leurs emplois selon quatre taux d'occupation, y compris les emplois de moins de 6h par semaine (taux d'occupation inférieur à 15%), ainsi que le total des emplois convertis en emplois équivalents plein temps (EPT). Les séries révisées de l'emploi total se composent de la somme de ces quatre catégories d'emplois.

Plan d'échantillonnage 2015

Dans la nomenclature NOGA2008, le plan de sondage a été établi selon le standard OFS50 correspondant aux divisions économiques (NOGA2 digits) ou à des regroupements de divisions économiques. La STATEM demeure une statistique dont les résultats sont calculés et diffusés au niveau des établissements (ou unités locales). Le tirage de l'échantillon suit un plan à deux degrés. On procède tout d'abord à un tirage d'entreprises (premier degré), puis, lorsqu'une entreprise est tirée, tous ses établissements sont inclus dans l'échantillon (second degré). Ceci permet d'intégrer la STATEM dans le Système de coordination de tirage des enquêtes auprès des entreprises récemment mis en œuvre à l'OFS. Afin de réduire la taille de l'échantillon, les critères du plan de sondage ont été revus et simplifiés. Ses caractéristiques seront décrites dans un prochain rapport.

Compléments d'échantillons régionaux

Depuis la révision 2000, les cantons et les grandes villes ont la possibilité de financer des compléments d'échantillon. Ces cas sont intégrés dans le plan d'échantillonnage.

Modes d'enquête

L'enquête est conduite au niveau des entreprises, selon 3 canaux (questionnaire papier, enquête en ligne et livraison de données sous forme électronique pour les grandes entreprises multi-établissements). Plus de 80% des données sont récoltées de manière électronique. Afin de limiter la charge auprès des entreprises, les diverses enquêtes de l'OFS portant sur l'emploi sont coordonnées entre elles.

Entreprises et établissements du profiling : sont inclus dans cette catégorie les établissements de grandes entreprises qui livrent leurs données sous forme électronique (fichiers) de manière exhaustive pour l'ensemble de leurs établissements.

Entreprises multi-établissements en dehors du profiling: on demande aux entreprises qui comportent plus d'un établissement de livrer leurs données de manière agrégée pour l'ensemble de tous leurs établissements. Les emplois sont ensuite répartis dans les établissements qu'ils représentent. L'échantillon reste toutefois un échantillon d'établissements et seules les données des établissements initialement inclus dans l'échantillon sont utilisées pour l'estimation des résultats.

Traitement de la non-réponse : la non-réponse est traitée de trois manières :

- a) par imputation
- b) par correction pour la non-réponse
- c) par calage

On entend par **imputation**, l'attribution de données à certaines entreprises qui n'ont pas fourni de données pour le trimestre en cours. On procède à des imputations pour les entreprises du profiling, ainsi que pour celles qui sont identifiées comme très grandes dans leur strate de tirage et pour lesquelles on attend un taux de réponse de 100%. On reprend dans ces cas les données du cadre de sondage ou du trimestre précédent corrigées de l'évolution moyenne de la branche à laquelle ils appartiennent.

Une procédure de **correction pour la non-réponse** est appliquée aux entreprises des autres strates. Elle consiste à adapter les poids de tirage des entreprises répondantes pour compenser pour celles qui n'ont pas répondu au trimestre en cours.

La procédure de **calage** consiste à corriger les poids résultants de l'étape précédente de manière à caler les estimations sur les valeurs du cadre de sondage.

Traitement des valeurs aberrantes : les valeurs extrêmes et peu plausibles sont identifiées et traitées par une procédure de robustification (adaptation des poids). Les cas identifiés sont examinés individuellement.

Précision des résultats et erreurs d'estimation : diverses causes d'erreur peuvent affecter la qualité des résultats publiés et influencer l'adéquation de ces derniers au « réel ». Les plus importantes sont :

- a) **erreurs de mesure** : ces dernières proviennent du fait que les données recueillies durant le trimestre sous revue ne correspondent pas dans tous les cas à la situation réelle des entreprises. Nous ne disposons pas d'estimations pour ce type d'erreur. De telles erreurs, lorsqu'elles sont détectées, peuvent être corrigées à posteriori, par la publication de résultats corrigés et lors des révisions des séries.
- b) **erreurs sur les caractéristiques du cadre de sondage**: de telles erreurs portent sur les caractéristiques des entreprises et des établissements (nomenclature, emplois, localisation), ainsi que sur l'adéquation entre le cadre de sondage et l'état réel du tissu économique. Nous ne disposons pas d'estimations pour ce type d'erreur. Des corrections sont effectuées lors des révisions rétrospectives des séries.
- c) **erreur d'échantillonnage** : dépend de l'incertitude liée à la procédure d'estimation par échantillonnage. Elle est estimée par la variance et présentée sous forme de coefficients de variation (cf. encadré ci-dessous). Ces coefficients sont affectés par les taux de réponse et d'imputation. Une diminution du taux de réponse a pour effet d'augmenter les coefficients de variation. Cet effet est pris en compte dans les estimations de la variance. En revanche, aucune correction n'est introduite pour les valeurs imputées. La variance est par conséquent sous-estimée dans ces cas.

Interprétation de l'erreur d'échantillonnage :

Les estimations de précision sont publiées sous forme de coefficients de variation et reflètent l'erreur d'échantillonnage et, dans une certaine mesure, celle résultant de la non-réponse.

Intervalles de confiance et coefficients de variation

L'incertitude d'un estimateur (par exemple de la variable emploi total) peut s'exprimer en termes d'intervalles de confiance ou de coefficient de variation (CV).

Un **intervalle de confiance à 95%** indique que si l'enquête était répétée un grand nombre de fois indépendamment et dans les mêmes conditions, 95% en moyenne des intervalles obtenus contiendraient effectivement le paramètre à estimer.

Le **coefficient de variation** (défini comme le rapport entre l'écart type et la valeur de la variable estimée) est une estimation relative de la précision de l'estimation qui ne dépend pas de la valeur absolue de la variable estimée. Ainsi, dans le cas de la STATEM, le CV permet de comparer l'incertitude des valeurs estimées entre branches économiques ou régions de taille très différentes. On obtient l'intervalle de confiance à 95% en termes de pourcents de la valeur estimée en multipliant le coefficient de variation par 1,96. Ainsi, un coefficient de variation de 0,5% signifie que l'estimateur sera situé dans un intervalle d'environ +/- 0,98 % de la valeur estimée.

Exemple : si la variable de l'emploi total est estimée à 4'900'000 et que son CV vaut 0.5%, l'intervalle de confiance à 95% est approximativement de:

$$4'900'000 \pm (48'000), \text{ soit } [4'852'000, 4'948'000]$$

Publications méthodologiques :

- Renaud, A. (2008) Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Rapport de méthodes, Office fédéral de la statistique, Neuchâtel, 2008, Numéro de commande: 338-0052
- Renaud, A., Panchard C., Potterat J. (2008) Statistique de l'emploi. Révision 2007 : méthodes d'estimation, Rapport de méthodes, Office fédéral de la statistique, Neuchâtel, 2008, Numéro de commande : 338-0055

Abréviations :

- AVS : Assurance vieillesse et survivants
CV : coefficient de variation
EPT : emplois en équivalence plein-temps
OFS : office fédéral de la statistique
REE : Registre des Entreprises et des Etablissements
RFE : Recensement fédéral des entreprises (jusqu'en 2008)
STATEM : Statistique de l'emploi
STATENT : Statistique structurelle des entreprises (dès 2011)