

QUELLES QUESTIONS POSER DANS UNE ENQUÊTE MÉNAGES POUR CLASSER UNE PROFESSION DANS UNE NOMENCLATURE PCS ?

Thomas DENOYELLE (*), Louis MEURIC (*)

(* Insee, Direction des Statistiques Démographiques et Sociales, Division Emploi et Revenus d'Activité)

Résumé

Dans les enquêtes auprès des ménages, en plus du libellé de profession, complété du grade pour les fonctionnaires, le classement dans la nomenclature des professions et catégories socioprofessionnelles (PCS 2003) requiert théoriquement quatorze variables. A partir de toutes ces variables, on procède à un codage automatique grâce au logiciel Sicore puis à une reprise manuelle, en cas d'échec ou de doute.

Quatre de ces variables sont généralement disponibles dans les enquêtes : sexe, département, occupation d'un emploi et statut dans l'emploi [salarie/non salarie]. Une cinquième (statut d'apprenti) est d'utilisation très marginale pour le codage, mais est souvent une variable d'intérêt. Les neuf autres variables sont : Position professionnelle (ouvrier, ingénieurs...), Nature de l'employeur (public/privé), Activité de l'établissement employeur (selon la nomenclature NAF, à 2 ou 5 caractères), Fonction dans l'emploi (production, installation...), Nombre de salariés employés, Taille de l'entreprise, Orientation principale agricole et Surface agricole utilisée.

La présente étude indique si ces neuf variables sont nécessaires au codage de la profession et pour quelle précision de la nomenclature PCS (voir tableau ci-dessous).

Pour répondre à cette question, on a codé la profession en simulant différents scénarios de disponibilité des neuf variables à partir de l'enquête Emploi. Les classements dans la nomenclature PCS ainsi obtenus ont été comparés avec le classement mobilisant les quatorze variables annexes, selon deux critères statistiques, dont voici les principales conclusions :

Premières conclusions : questions recommandées selon le niveau de précision et le nombre de répondants.

Niveau de précision pour la publication	Nombre total de répondants + effectifs correspondants concernés	Variabes annexes nécessaires au codage en plus du libellé, complété du grade
groupe socioprofessionnel GS - 1 ^{er} caractère		Département, Sexe, Occupation d'un emploi, Statut dans l'emploi, <i>Statut d'apprenti</i> , + Position professionnelle, Nature de l'employeur, NAF2
catégorie socioprofessionnelle CS - deux 1ers caractères	< 3 000 <i>1 900 actifs occupés</i>	les précédentes
	> 3 000 <i>1 900 actifs occupés</i>	+ Fonction et Surface agricole utilisée
	> 18 000 (*) <i>11 650 actifs occupés</i>	+ Nombre de salariés employés
profession PCS -4 caractères		+ NAF5 et Taille de l'entreprise

A noter que la NAF sur 5 caractères n'est pas disponible dans les enquêtes s'appuyant sur le tronç commun des enquêtes ménages.

La Fonction (ainsi qu'une autre question sur la Supervision de collègues) est indispensable au codage de la profession en nomenclature CIP (ISCO), sans considération de taille d'échantillon.

Abstract

In household surveys, in addition to the name of the occupation of an individual (completed by his rank/level if he is a civil servant) the coding process of this individual into the french classification of occupations (PCS 2003) requires fourteen additional variables. Among those variables, four are always collected (gender, French department, status [working person/unemployed/retired], status in the job [salaried/nonsalaried] and one is interesting in itself (apprentice or not) but barely useful for the categorization. Our study aimed at assessing the usefulness of the nine remaining variables (professional status [worker, ingeneer...], nature of the employer [public/private], activity of the employing establishment [following two levels of the "NAF" classification : 2 or 5 digits], function in the job, number of salaried persons employed, size of the establishment, type of the agricultural products, agricultural superficity used) for an automatic categorization with the "Sicore-PCS" tool.

Mots-clés

Professions, nomenclature, codage.

Table des matières

1. INFLUENCE DES VARIABLES ANNEXES SUR LE CODAGE.....	5
1.1. DONNÉES ET MÉTHODES	5
1.2. UTILITÉ DES VARIABLES ANNEXES, SELON LE NIVEAU DE PRÉCISION VOULU (GS, CS OU PCS)	6
1.2.1. <i>La variable STATUT</i>	6
1.2.2. <i>Les variables CPF, PUB, NAF2, NAF5 et GRADE</i>	7
1.2.3. <i>La variable ATYPTRAV</i>	8
1.2.4. <i>Les variables T et OPA</i>	8
1.2.5. <i>Les variables NBS, SAU et FONCTION (FN)</i>	9
1.2.6. <i>Première synthèse des résultats obtenus</i>	12
1.3. INFLUENCE MARGINALE DE L'ABSENCE DE CHAQUE VARIABLE ANNEXE : ANALYSE DÉTAILLÉE.....	13
1.3.1. <i>Position professionnelle (CPF)</i>	14
1.3.2. <i>Statut de l'employeur (PUB)</i>	16
1.3.3. <i>Activité principale de l'établissement employeur (NAF2, NAF5)</i>	19
1.3.4. <i>Orientation principale agricole (OPA)</i>	22
1.3.5. <i>Taille de l'entreprise (T)</i>	24
1.3.6. <i>Fonction principale (FN)</i>	25
1.3.7. <i>Surface agricole utilisée (SAU)</i>	27
1.3.8. <i>Nombre de salariés employés (NBS)</i>	28
2. PRÉCONISATIONS COMPLÉMENTAIRES ET TOUT AUSSI UTILES SUR LES QUESTIONS À POSER, SELON LE TYPE DE PROFESSION ET LE MODE DE COLLECTE.....	30
2.1. NIVEAU DE PRÉCISION ATTEIGNABLE POUR LA DIFFUSION	30
2.1.1. <i>Le recueil ou non de la NAF5 conditionne fortement la diffusion de la PCS à 2 ou 4 positions...sauf pour certaines catégories de professions</i>	30
2.1.2. <i>Le calcul ou non de la CIP/ISCO impacte la précision à arrêter pour le « codage » de la PCS</i>	31
2.1.3. <i>Professions des parents ou d'un tiers absent du ménage</i>	31
2.2. PRISE EN COMPTE DES PRINCIPAUX MODES DE COLLECTE	32
2.2.1. <i>Enquêtes utilisant un questionnaire papier</i>	32
2.2.2. <i>Enquêtes utilisant Capi avec Sicore-PCS embarqué</i>	33
2.2.3. <i>Enquêtes utilisant Cawi (Blaise-web) avec Sicore-PCS embarqué</i>	33
2.3. AUTRE APPLICATION PRATIQUE : PRÉCONISATIONS POUR LA PROFESSION ANTÉRIEURE AU RP	34
ANNEXE 1 : NOMENCLATURE DES PROFESSIONS ET CATÉGORIES SOCIOPROFESSIONNELLES.....	36
ANNEXE 2 : NOTICE GÉNÉRALE SUR LES TROIS CATÉGORIES D'AGRICULTEURS 11, 12 ET 13 (EXTRAIT DU GUIDE DE LA PCS 2003).....	37
ANNEXE 3 : DÉTAIL DU CALCUL D'UNE TAILLE MINIMALE D'ÉCHANTILLON $N_{A,I}$ POUR UNE VARIABLE ANNEXE A ET POUR UN NIVEAU DE PRÉCISION I (1 OU 2 CHIFFRES DE LA PCS)	39
ANNEXE 4 : EXEMPLE DE L'ENQUÊTE EMPLOI 2015 : LISTE DES VARIABLES ANNEXES COLLECTÉES SELON LE TYPE DE PROFESSION	41
ANNEXE 5 : EXEMPLE DE L'ENQUÊTE EMPLOI 2015 : CONSIGNES PARTICULIÈRES DE COLLECTE POUR CERTAINES QUESTIONS NÉCESSAIRES POUR CODER LA PROFESSION PRINCIPALE ACTUELLE.....	42
ANNEXE 6 : LES QUESTIONS SERVANT AU CODAGE DE LA PROFESSION ET COLLECTÉES DANS LE TCM.....	47

Introduction

Dans les enquêtes auprès des ménages, en plus du libellé de profession, complété du corps ou grade pour les fonctionnaires, le classement dans la nomenclature des professions et catégories socioprofessionnelles (PCS 2003) requiert théoriquement quatorze variables. On parle de « variables annexes au libellé ».

Le codage des professions dans les nomenclatures PCS et éventuellement CITP (ISCO en anglais, nomenclature internationale des professions) est toujours fait en aval de la collecte, d'abord par codage automatique grâce au logiciel Sicore, puis par reprise manuelle, en cas d'échec ou de doute sur ce codage.

Les variables annexes sont :

Pour tous les individus :

- STRE : occupation d'un emploi ou non (actifs occupés, retraités, chômeurs)
- STATUT : statut dans l'emploi (salariés, indépendants)
- S : sexe
- DEP : département (du lieu de travail, à défaut : du domicile)
- NAF5 : sous-classe d'activité de l'établissement employeur sur cinq positions
- NAF2 : division d'activité de l'établissement employeur sur deux positions
- T : taille de l'entreprise (petite, moyenne, grande)

Pour les salariés :

- FN : fonction professionnelle (fabrication, installation, vente, secrétariat, etc.)
- PUB : statut de l'employeur (public, privé)
- CPF : position professionnelle ou qualification (ouvriers, ingénieurs, etc.)
- SP : statut d'apprenti

Pour les indépendants (ou non-salariés) :

- NBS : nombre de salariés employés (notamment pour la distinction entre artisans, commerçants et chefs d'entreprises)

Pour les exploitants agricoles :

- OPA : orientation principale agricole
- SAU : surface agricole utilisée

On trouvera en annexe les questions du tronc commun des enquêtes ménages correspondant à ces variables (Annexe 6 : Les questions servant au codage de la profession et collectées dans le TCM).

Certaines de ces variables annexes sont toujours collectées ou devraient toujours l'être, pour leur intérêt en soi : DEP, S, STATUT et STRE¹. La variable SP est d'utilisation marginale pour le codage, mais est souvent une variable d'intérêt.

La présente étude indique dans quelle mesure les neuf autres variables annexes (NAF2, NAF5, FN, T, PUB, CPF, NBS, OPA, SAU) sont utiles au codage de la profession dans la nomenclature des PCS, en fonction :

- du niveau de précision prévu pour la diffusion (1, 2 ou 4 caractères de la PCS)
- du type de profession (profession principale actuelle, profession antérieure, des parents,...)
- et du mode de collecte (questionnaire papier, Capi, Cawi).

Les conclusions générales de l'étude figurent intégralement aux chapitres 1.2.6 et 2.

¹ Nous mentionnons cette variable car elle se trouve dans toute la documentation sur Sicore-PCS. Mais elle n'est plus utilisée pour le codage. En pratique, il faut toujours lui imputer la modalité « 7 : autre » (= actif occupé), même pour coder une profession antérieure.

1. Influence des variables annexes sur le codage

1.1. Données et méthodes

Notre étude mobilise des données collectées dans l'enquête emploi en continu (la seule enquête à collecter l'intégralité des 14 variables annexes), sur des périodes comprises entre le début de l'année 2009 et le deuxième trimestre de 2014, auprès des actifs occupés, en première interrogation ou en réinterrogation².

Les milliers de libellés de professions principales recueillis dans les enquêtes emploi successives ont fait l'objet de simulations de codage selon différents scénarios de disponibilité des 9 variables énumérées dans l'introduction en supposant disponibles les 5 autres. Les répartitions dans la nomenclature des PCS³ résultant de ces simulations ont été comparées avec la répartition obtenue en mobilisant les 14 variables annexes (ce scénario de référence codant 98 % des libellés de l'enquête emploi), selon deux critères :

- le taux de mauvais classements (ou de transferts entre catégories de la PCS)
- et la racine carrée de l'erreur quadratique moyenne des écarts de structure, définie ci-dessous :

$EQM_{A,i}$ est l'erreur quadratique moyenne, au niveau de précision i de la PCS (1 ou 2 chiffres) et pour une variable manquante A :

$$(1) EQM_{A,i} = (1/C_i) * \sum_J (p_{J,A} - p_{J,ref})^2$$

Où $p_{J,A}$ = part du groupe socioprofessionnel ou de la catégorie socioprofessionnelle J dans la population en emploi, calculée en l'absence de A , simulé à partir de l'Enquête emploi

$p_{J,ref}$ = part calculée avec le codage de référence (avec les 14 variables annexes)

C_i = nombre de catégories du niveau de précision i (= 6 pour les groupes sociaux, 31 pour les catégories socioprofessionnelles)

Par la suite, afin de répondre à certaines situations rencontrées en pratique, on a aussi réalisé des scénarios expérimentaux en enlevant plusieurs variables simultanément.

Note sur les deux critères : si l'enquête est utilisée majoritairement pour des statistiques descriptives, on privilégiera le critère EQM, si elle est aussi utilisée pour des analyses croisées ou pour de la modélisation, il faut aussi tenir compte de la part des mauvais classements, voire privilégier ce critère.

² On rappelle que l'enquête emploi en continu donne lieu à six interrogations, durant six trimestres.

³ au niveau des groupes socioprofessionnels [GS], des catégories socioprofessionnelles [CS]³ ou des professions [PCS] — i.e. des codes de la PCS à 1, 2 ou 4 caractères.

1.2. Utilité des variables annexes, selon le niveau de précision voulu (GS, CS ou PCS)

1.2.1. La variable STATUT

Comme précisé en introduction, la variable annexe STATUT (salarié/indépendant) est toujours collectée ou devrait toujours l'être, pour son intérêt en soi. Pour se convaincre néanmoins de l'importance de cette variable annexe, le tableau suivant présente le taux de mauvais classements et le critère $\sqrt{EQM_{A,i}}$. À la différence de la définition en 1.1, $EQM_{A,i}$ est ici l'erreur quadratique moyenne, non pas en l'absence d'une variable précise mais pour chacun des deux scénarios de codage (A) :

- Scénario STAT : codage avec seulement la variable STATUT,
- Scénario SANS : codage sans aucune variable annexe⁴.

Impact sur le classement des actifs occupés pour deux scénarios de codage

	Groupe socioprofessionnel		CS	
	Mauvais classements	\sqrt{EQM}	Mauvais classements	\sqrt{EQM}
Scénario SANS : Sans aucune variable annexe, notamment STATUT	21,2 %	2,4 pt	32,0 %	1,1 pt
Scénario STAT : Avec STATUT, mais sans les autres variables	17,5 %	2,2 pt	28,2 %	1,1 pt

Enquête emploi 2013-2014T2 : 63 995 libellés de professions principales collectés en 1ère interrogation pour les actifs occupés et reconnus par Sicore-batch ou Sicore-interactif lors de la reprise manuelle

Conclusion :

L'introduction de la variable STATUT permet de réduire le taux de mauvais classements de 4 points.

⁴ Il a été utile d'examiner ce scénario car c'est exactement le scénario de codage d'une enquête de la Drees auprès de jeunes élèves.

1.2.2. Les variables CPF, PUB, NAF2, NAF5 et GRADE

Le tableau ci-dessous présente l'influence des variables les plus importantes pour le codage de la profession, ainsi que celle de la NAF5, qui est une variable très coûteuse à recueillir et donc rarement collectée ⁵.

Impact de l'absence marginale de 5 variables annexes sur le classement des actifs occupés

	Groupe socioprofessionnel		CS		Profession	
	Mauvais classements	√EQM	Mauvais classements	√EQM	Mauvais classements	√EQM
Sans grade	3,42 %	nd	3,92 %	nd	5,51 %	nd
Sans CPF	8,72 %	1,380 pt	12,28 %	0,549 pt	12,51 %	0,066 pt
Sans PUB	1,16 %	0,066 pt	3,11 %	0,249 pt	3,74 %	0,049 pt
Sans NAF5 ni NAF2 ⁶	1,68 %	0,549 pt	6,60 %	0,337 pt	18,75 %	0,193 pt
Sans NAF5 mais avec NAF2	0,68 %	0,099 pt	2,82 %	0,097 pt	10,41 %	0,089 pt

Enquête emploi 2009 et 2010 ⁷ : 81 395 actifs occupés en première interrogation ou ayant changé de profession
Lecture : À la ligne « sans CPF », on a codé le libellé avec toutes les variables annexes sauf la variable annexe CPF.

Conclusions :

Il est important de bien compléter le libellé de profession par celui du grade pour les fonctionnaires.

Les variables annexes PUB, CPF et NAF2 sont indispensables pour bien coder la profession au niveau CS ⁸ et même au niveau groupe socioprofessionnel ⁹. Voir aussi les conclusions dans l'Annexe 3 : Détail du calcul d'une taille minimale d'échantillon $n_{A,i}$ pour une variable annexe A et pour un niveau de précision i (1 ou 2 chiffres de la PCS).

D'après la dernière ligne du tableau, **en l'absence de la variable NAF5** (qui n'est disponible que *via* un appariement avec Sirene) et avec seulement une NAF recueillie au niveau « division » avec Sicore-APE embarqué (que ce soit sous CAPI, CATI ou CAWI), **la qualité de la PCS à 4 positions est bien moindre.**

Dans la suite de ce chapitre, nous supposons donc vouloir coder les professions seulement au niveau groupe socioprofessionnel ou CS (détaillée).

⁵ A la DSDS, cela nécessite en effet une recherche dans le répertoire Sirene à partir des coordonnées de l'établissement, cela n'est réalisé que pour la profession principale au recensement, dans l'enquête emploi et dans l'enquête FQP, ainsi que pour la profession antérieure si elle date de moins de 5 ans pour FQP.

⁶ Si la NAF2 est absente, la NAF5 l'est *a fortiori*.

⁷ Ces chiffres émanent de la note N° 1953/DG75-F230/SLM-RW du 6 octobre 2011 citée en référence. Compte tenu de l'importance des erreurs de classement et de la lourdeur à réaliser de telles simulations, ils n'ont pas été actualisés, contrairement aux simulations suivantes.

⁸ A fortiori au niveau PCS

⁹ De plus, l'absence de la variable CPF augmente le taux de non-codage et donc *le taux de reprise* manuelle de 0,4 %.

1.2.3. La variable ATYPTRAV

La variable ATYPTRAV, que nous n'avons pas encore mentionnée, correspond à une question figurant dans le TCM et dans l'EEC (voir l'Annexe 6 : Les questions servant au codage de la profession et collectées dans le TCM) et décrit le type de travail de l'aide familial : secrétariat, vente, comptabilité / autre. Elle ne fait pas partie des variables annexes nécessaires au codage par Sicore-PCS, mais est utilisée dans l'enquête emploi pour calculer un code PCS sur 4 positions plus précis, mais sans impact sur la CS, ni sur la CITP.

En revanche, s'agissant du questionnaire du TCM qui ne permet pas un codage assez précis de la PCS à 4 positions, on préconise la suppression de la question ATYPTRAV.

1.2.4. Les variables T et OPA

Impact de l'absence marginale des variables T et OPA sur le classement des actifs occupés

	Groupe socioprofessionnel		CS	
	Mauvais classements	√EQM	Mauvais classements	√EQM
Sans OPA <i>Enquête emploi 2013 et 2014T1</i>	0,00 %	0,000 pt	0,00 %	0,001 pt
Sans T, avec NBS <i>Enquête emploi 2009 et 2010</i>	0,00 %	0,005 pt	0,11 %	0,008 pt

Si l'on code la profession à un niveau agrégé (groupe socioprofessionnel ou CS), ou si l'on ne dispose pas d'une NAF sur cinq positions, *la taille de l'entreprise (T) est superflue* (voir partie 1.3.5).

La question sur l'orientation principale agricole (OPA) est peu utile pour la CS, même en l'absence de la NAF5. **D'ailleurs, s'agissant du questionnaire du TCM lui-même, on préconise la suppression de la question OPA.** Il faudrait que le concepteur ait un intérêt d'étude particulier pour les agriculteurs pour qu'il la rétablisse dans son propre questionnaire.

1.2.5. Les variables NBS, SAU et FONCTION (FN)

1.2.5.1. Impact de l'absence marginale de chacune des 3 variables

On se place dans le cas où les variables NAF5, T et OPA sont écartées par le concepteur (situation usuelle dans les enquêtes ménages utilisant le TCM).

Impact de l'absence des variables annexes FN, SAU et NBS (en l'absence de T, NAF5 et OPA) sur le classement des actifs occupés

	Groupe socioprofessionnel		CS	
	Mauvais classements	$\sqrt{\text{EQM}}$	Mauvais classements	$\sqrt{\text{EQM}}$
<i>Impact marginal</i>				
Sans FN	0,69 %	0,137 pt	2,27 %	0,137 pt
Sans SAU	0,00 %	0,000 pt	0,59 %	0,124 pt
Sans NBS	0,00 %	0,001 pt	0,57 %	0,075 pt
<i>Impact groupé</i>				
Sans FN, ni SAU, ni NBS	0,69 %	0,137 pt	3,49 %	0,255 pt

*Hypothèse : les variables T, NAF5 et OPA ne sont pas disponibles
Enquête emploi 2013 et 2014 (1^{er} trimestre) : 46 959 actifs occupés en première interrogation*

Interprétation

Si l'on souhaite publier seulement au niveau groupe socioprofessionnel et si l'on ne doit pas coder en CITP, ces trois variables peuvent être omises dans le questionnaire.

Au niveau CS :

Une lecture rapide de la nomenclature des intitulés de CS peut donner l'impression que la variable NBS est la plus discriminante des 3 variables annexes, et c'est ce qui est indiqué dans la note N° 112 /F420 du 12 mai 2005. En fait, comme l'indiquent les résultats de nos simulations dans le tableau ci-dessus, c'est la variable FN qui est responsable du plus grand nombre de mauvais classements¹⁰.

En outre, la variable FN est indispensable pour coder la CITP (cf. partie 1.3.6).

Du point de vue de l'écart de structure moyen ($\sqrt{\text{EQM}}$), la variable SAU est tout aussi importante que FN¹¹.

Enfin, la variable NBS a une importance moindre que les 2 variables précédentes : au niveau CS, elle permet de distinguer entre les chefs de petites entreprises et les chefs de 10 salariés ou plus.

La question pour le concepteur est alors d'arrêter son exigence de précision du codage. Une telle précision est à mettre en balance avec l'aléa de sondage de l'enquête concernée : plus faible sera l'aléa de sondage et plus élevée pourra alors être l'exigence de précision du codage.

¹⁰ Cela étant, NBS présente probablement un biais de collecte moindre que la question Fonction, question plus complexe à appréhender

¹¹ En l'absence de SAU, Sicore considère fréquemment que les agriculteurs exploitent de petites exploitations (CS 11) et non des exploitations moyennes ou grandes (CS 12 ou 13).

1.2.5.2. Taille minimale d'utilité de chacune des 3 variables

$EQM_{A,i}$, a déjà été défini plus haut : c'est l'erreur quadratique moyenne provoquée par l'omission d'une variable annexe A, et pour l'un des 2 niveaux de précision i de la PCS (1 ou 2 chiffres).

On prend comme exemple ci-dessous le cas où l'on souhaite connaître la précision totale de la variable de profession, sur la population en emploi.

Soit alors V_i la variance moyenne estimée des parts de chaque groupe socioprofessionnel (GS) ou de chaque **CS**, au sein de la population en emploi, et liée à l'aléa de sondage.

$$(0) \quad V_i = 1/TE^2 * V'_i$$

Où TE = taux d'emploi (= 64,1 % en 2013)

V'_i est la variance moyenne estimée des parts des GS ou **CS** des personnes en emploi, ce au sein de la population totale (et non seulement dans la population en emploi, pour la profession principale). Donc :

$$(1) \quad V_i = (1 - n/N) / (n * C_i * TE^2) * \sum_J p_{J,ref} * TE * (1 - p_{J,ref} * TE)$$

(V_i n'est pas liée aux variables annexes)

Où n = nombre total d'individus répondants à l'enquête

N = population totale dans le champ de l'enquête

C_i = toujours le nombre de catégories au niveau de précision i (voir paragraphe 1.1)

= 6 pour les groupes sociaux, 31 pour les catégories socioprofessionnelles

$p_{J,ref}$ = part de la catégorie socioprofessionnelle J dans la population en emploi

Définition :

On calcule alors un nombre minimal de répondants $n_{A,i}$ au-delà duquel il est utile d'introduire la variable annexe A pour le niveau de précision i (1 ou 2 chiffres de la PCS). On peut aussi appeler $n_{A,i}$ la taille minimale d'utilité de la variable A.

$n_{A,i}$ est tel que :

$$(2) \quad \sqrt{EQM_{A,i} + V_i} = 1,1 * \sqrt{V_i}$$

$\sqrt{EQM_{A,i} + V_i}$ correspond à l'erreur quadratique moyenne totale pour l'échantillon considéré

Les justifications théoriques de la formule (2) et le détail du calcul qui s'ensuit sont fournis en annexe 3. On aboutit alors à la formule suivante :

$$(3) \quad n_{A,i} = 0,21 / (C_i * TE^2 * EQM_{A,i}) * \sum_J p_{J,ref} * TE * (1 - p_{J,ref} * TE)$$

Il s'agit toujours d'une formule simplifiée pour une enquête auprès de personnes de 15 à 64 ans vivant en logement ordinaire, en population générale, avec un sondage aléatoire simple (SAS).

Pour la profession antérieure des chômeurs et des anciens actifs par exemple, on remplacera le taux TE par $1 - TE$.

Interprétation

Si le nombre de répondants de l'enquête est inférieur à $n_{A,i}$, (avec donc un écart-type moyen $\sqrt{V_i}$ relativement important des parts des catégories socioprofessionnelles), l'erreur quadratique moyenne totale lorsque manque la variable annexe A ne dépasse pas cet écart-type moyen de plus de 10 % : on peut alors accepter de ne pas inclure la variable annexe A dans le questionnaire.

A l'inverse, au-delà de ce nombre minimal de répondants, on doit inclure la variable annexe A pour réduire l'ampleur de cette erreur quadratique moyenne totale.

Attention :

- Le chiffre de 1,1 ou 10 % (et donc $0,21 = 1,1^2 - 1$) est pris arbitrairement : si on est plus exigeant, avec un seuil de 1,05 par exemple, alors on sera enclin à rajouter une variable annexe pour un nombre minimal de répondants plus faible (dans un rapport de $0,21/0,1025$ dans notre exemple).
- En raison de toutes les précautions évoquées plus haut pour les calculs de variances, les tableaux ci-dessous ne peuvent donner que des nombres minimaux de répondants indicatifs¹² :

**Impact de l'absence des variables FN, NBS, T et SAU
sur le classement des actifs occupés**

	Groupe socioprofessionnel		CS	
	√EQM	Taille minimale d'utilité de la (des) variable (*)	√EQM	Taille minimale d'utilité de la (des) variable (*)
Sans FN	0,137 pt	24 915	0,137 pt	5 458
Sans SAU	0,000 pt	nd	0,124 pt	6 663
Sans FN ni SAU ¹³	0,137 pt	24 915	0,185 pt	3 000
Sans NBS (ni T)	0,001 pt	468 millions	0,075 pt	18 213

*Hypothèse : les variables T, NAF5 et OPA sont par ailleurs non disponibles.
Enquête emploi 2013 et 2014 (1^{er} trimestre) : 46 959 actifs occupés en première interrogation
(*) ou taille minimale d'échantillon, au-delà de laquelle introduire la variable est utile*

Conclusion :

Les deux variables FN et SAU présentent des tailles d'utilité équivalentes : il faut donc les introduire simultanément, dès que le nombre de répondants excède le seuil des 3 000 (ligne « Sans FN ni SAU » du tableau).

Même avec une exigence de précision globale de la CS assez forte, la variable NBS n'est utile que pour des nombres de répondants très importants.

¹² Il appartient au concepteur de corriger les formules ci-dessus d'un ratio tenant compte de différents facteurs. On rappelle que d'après la division sondage, le ratio $\text{ecart_type_estime} / \text{ecart_type_approche}$ (dans le cas d'un SAS non calé) avoisine, respectivement :

- 0,9 pour une variable quantitative déclinée selon 3 tranches d'âges dans l'enquête AES 2012 ;
- 1,3 pour des variables quantitatives dans CVS 2012.

¹³ $\text{EQM}(\text{FN})$ et $\text{EQM}(\text{SAU})$ sont sommables, car leur absence affecte des CS différentes. Donc : $\text{EQM}(\text{FN} ; \text{SAU})_i = \text{EQM}(\text{FN})$ et $\text{EQM}(\text{SAU})$
En toute rigueur, il faudrait aussi tenir compte de $\text{EQM}(\text{NBS})$, mais elle est négligeable

1.2.6. Première synthèse des résultats obtenus

Comme indiqué en introduction, parmi les 14 « variables annexes au libellé » utilisées par Sicore-PCS, 4 variables annexes sont toujours collectées ou devraient toujours l'être, pour leur intérêt en soi : DEP, SEXE, STATUT et STRE. La variable SP est d'utilisation marginale pour le codage, mais est souvent une variable d'intérêt.

Parmi les neuf autres variables annexes (NAF2, NAF5, FN, T, PUB, CPF, NBS, OPA et SAU), le tableau suivant récapitule les variables annexes recommandées en fonction du niveau de précision de la PCS voulu.

Pour le calcul du nombre de répondants, on prend comme exemple le cas où l'on cherche à coder la profession principale pour des actifs occupés, dans une enquête auprès de personnes âgées de 15 à 64 ans.

Niveau de précision pour la publication	Nombre de répondants + (effectifs correspondants concernés)	Variables annexes nécessaires au codage (en plus du libellé complété du grade pour les fonctionnaires)
GS		DEP, SEXE, SP, STRE, STATUT, + CPF, PUB, NAF2
CS	< 3 000 (*) (1 900 actifs occupés)	les précédentes
	> 3 000 (*) (1 900 actifs occupés)	les précédentes + FN (obligatoire pour la CITP /ISCO¹⁴, quelle que soit la taille) et SAU
	> 18 000 (*) (11 650 actifs occupés)	les précédentes + NBS
PCS		les précédentes + NAF5, T

(*) Valeurs calculées pour une enquête auprès de personnes de 15 à 64 ans vivant en logement ordinaire, en population générale (taux d'emploi de 64,1%), avec un sondage aléatoire simple (SAS).

Conclusion

Le chiffre le plus important à prendre en compte dans la colonne (2), ce sont les effectifs correspondants de répondants concernés, entre parenthèses. En première approximation, ces effectifs peuvent être utilisés pour définir des stratégies sur d'autres types de professions : par exemple, la profession antérieure des inactifs et des chômeurs.

Pour une enquête correspondant à moins de 1 900 actifs occupés répondants (ou anciens actifs si on prend la profession antérieure), les variables FN et SAU ne sont même pas indispensables.

Il convient alors, et c'est l'usage, de regrouper l'ensemble des agriculteurs (CS 11 à 13) dans une unique CS égale à 10. Ces résultats reprécisent et corrigent les consignes présentées dans la note N° 112 /F420 du 12 mai 2005.

¹⁴ En plus de la fonction, une autre question sur la « Supervision de collègues » est indispensable au codage de la profession en nomenclature CITP/ISCO. Voir la note N° 1070/DG75-F230/NR-LM-TD du 7/6/2013 : Nouvelle méthode de codage selon la CITP-08 (ISCO-08) dans les enquêtes auprès des ménages

1.3. Influence marginale de l'absence de chaque variable annexe : analyse détaillée

Chacun des paragraphes suivants est consacré à une variable annexe : on y mesure l'effet marginal de l'absence de chaque variable annexe, pour la (ou les) populations concernées, sur le codage :

- du groupe socioprofessionnel (le premier chiffre de la PCS),
- de la CS (les 2 premiers chiffres),
- et de la profession (les 4 caractères).

On y examine notamment les différences de répartition systématiques par rapport au codage de référence ainsi que les transferts les plus fréquents (et parfois aussi systématiques) en l'absence de la variable annexe.

Pour les variables CPF, PUB, NAF5, NAF2 et T, les chiffres émanent de la note N° 1953/DG75-F230/SLM-RW du 6 octobre 2011 citée en référence. Compte tenu de l'importance des erreurs de classement et de la lourdeur à réaliser de telles simulations, ils n'ont pas été actualisés.

Le fichier Excel ci-dessous détaille l'ensemble des erreurs de codage au niveau PCS en l'absence de chaque variable annexe.



Table.xls

1.3.1. Position professionnelle (CPF)

La position professionnelle est une variable renseignée uniquement pour les salariés. Elle indique en effet la position hiérarchique qu'un salarié occupe dans son travail. Elle correspond notamment aux principaux niveaux des conventions collectives pour les salariés d'entreprise. C'est cette variable qui permet d'ordonner les quatre groupes sociaux composés principalement de salariés. Dans Sicore PCS, la position professionnelle se décompose en neuf modalités :

- 1 manœuvre ou ouvrier spécialisé ;
- 2 ouvrier qualifié, ouvrier hautement qualifié ou technicien d'atelier ;
- 3 agent de maîtrise ;
- 4 directeur général ou adjoint direct au directeur ;
- 5 technicien, dessinateur, VRP ;
- 6 instituteur, assistante sociale, infirmière ; personnel de catégorie B de la fonction publique ;
- 7 ingénieur ou cadre ;
- 8 professeur, personnel de catégorie A de la fonction publique ;
- 9 employé de bureau, de commerce, agent de service, aide soignante, gardienne d'enfants, personnel de catégorie C ou D de la fonction publique.

Dans le tronc commun des enquêtes ménages, la question est aménagée selon le statut du salarié concerné :

Dans votre emploi, êtes-vous classé(e) comme...

Si salarié d'entreprise :

1. Manœuvre ou ouvrier(e) spécialisé(e)
2. Ouvrier(e) qualifié(e) ou hautement qualifié(e), technicien(ne) d'atelier
3. Technicien(ne)
5. Agent de maîtrise, maîtrise administrative ou commerciale, VRP (non cadre)
7. Ingénieur, cadre (à l'exception des directeurs généraux ou de ses adjoints directs)
9. Employé(e) de bureau, employé(e) de commerce, personnel de services
10. Directeur général, adjoint direct

Si salarié de l'État, des collectivités locales, des HLM et des hôpitaux publics :

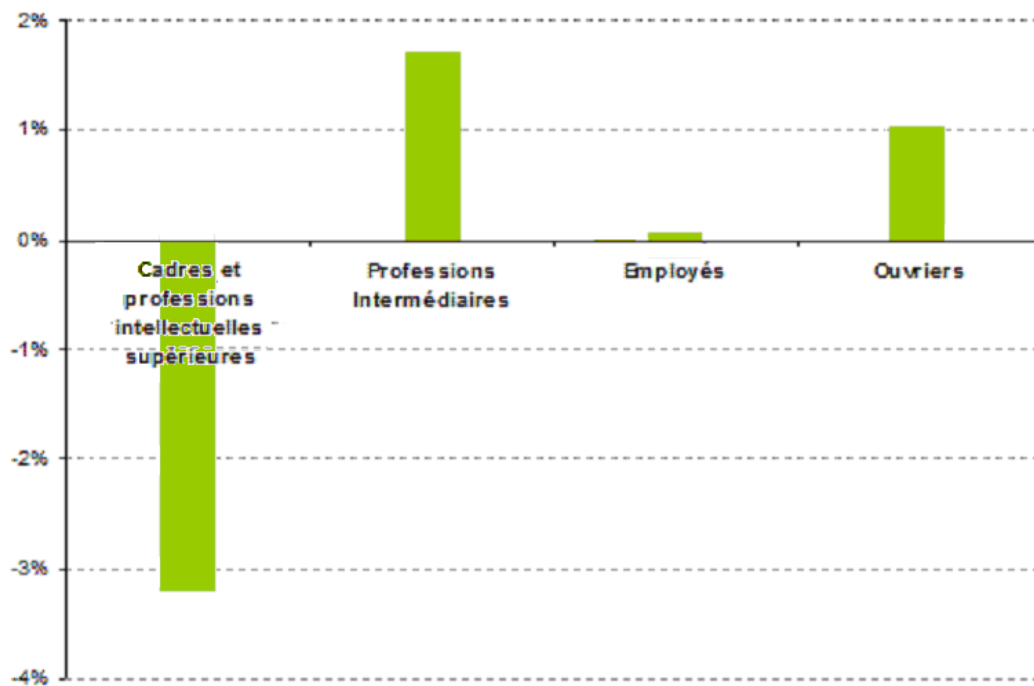
1. Manœuvre ou ouvrier(e) spécialisé(e)
2. Ouvrier(e) qualifié(e) ou hautement qualifié(e)
3. Technicien(ne)
4. Personnel de catégorie B ou assimilé
6. Personnel de catégorie A ou assimilé
8. Personnel de catégorie C ou assimilé

Le groupe socioprofessionnel

La position professionnelle est nécessaire pour déterminer le groupe socioprofessionnel des salariés. Cela signifie que la position professionnelle est une variable dont il ne faut jamais se passer pour étudier les professions et catégories socioprofessionnelles, quel que soit le niveau de détail choisi : 9,8 % des salariés sont mal classés si la position professionnelle est négligée.

De plus, selon les règles de classement dans la PCS, en l'absence d'information sur la position professionnelle, il faut coder au plus faible niveau de qualification possible pour un métier donné : 3,3 % de cadres sont donc classés en professions intermédiaires et 2,8 % de professions intermédiaires sont classées avec les ouvriers ou les employés. Ces effets sont loin d'être compensés par le surclassement de 2 % d'ouvriers et d'employés en professions intermédiaires et de 0,7 % de professions intermédiaires en cadres.

Différence de répartition des salariés dans les groupes sociaux par rapport au codage de référence avec CPF (en points)



Enquêtes emploi 2009 et 2010 : 72 327 salariés interrogés pour la 1^{ère} fois ou ayant changé de profession

Lecture : en codant sans la variable CPF, la part de professions intermédiaires parmi les salariés est relevée artificiellement de 1,7 points : elle passe de x % à x + 1,7 %.

La CS

L'absence de variable de position professionnelle n'étant pas acceptable au niveau groupe, elle l'est donc encore moins au niveau de la CS.

1.3.2. Statut de l'employeur (PUB)

Le statut de l'employeur n'est demandé qu'aux salariés.

Variable SICORE	Question du TCM 2014
PUB	F4 (F) STATUTECL Statut (éclaté) dans l'emploi principal
1 État ; 2 collectivités territoriales, HLM, hôpitaux ; 3 sécurité sociale ; 4 entreprises publiques nationalisées ; 5 privé (y compris particuliers).	<i>Êtes-vous :</i> 1. Salarié de l'État ¹⁵ 2. Salarié d'une collectivité territoriale ¹⁶ 3. Salarié d'une entreprise, d'un artisan, d'une association 4. Salarié d'un ou plusieurs particulier(s) 5. Vous aidez un membre de votre famille dans son travail, sans être rémunéré 6. Chef d'entreprise salarié, PDG, gérant(e) minoritaire, associé(e) 7. Indépendant(e) ou à son compte 9. Salarié des hôpitaux publics ou du secteur public social et médico-social ¹⁷

Le fait de ne pas isoler la sécurité sociale dans la question ci-dessus dégrade très peu le codage de la profession.

Le groupe socioprofessionnel (GS)

Le statut public ou privé de l'établissement employeur est un grand clivage de la PCS qui intervient essentiellement dans la formation des catégories socioprofessionnelles. Mais dans certains cas (1,3 %), cette variable est nécessaire à la formation des groupes sociaux : elle permet surtout de mieux séparer les ouvriers des employés (0,78 %). *Il faut donc utiliser PUB pour déterminer le groupe socioprofessionnel.* Et ce d'autant plus que, pour poser la question sur la position professionnelle (CPF) de façon plus adaptée, des filtres utilisant la variable PUB sont recommandés.

La CS

La variable PUB détermine le bon classement de 3,5 % des salariés au niveau de la CS. En l'absence de statut, les salariés passent des CS de la fonction publique aux catégories équivalentes du secteur privé. Le phénomène inverse se produit aussi, mais dans une moindre mesure. Par conséquent, *l'absence de la variable PUB conduit à sous-estimer le nombre de fonctionnaires :*

¹⁵ Indications supplémentaires données à l'enquêteur dans le TCM :

y compris établissements publics administratifs (EPA) nationaux, établissements publics d'enseignement

¹⁶ y compris OPHLM, établissements publics administratifs (EPA) des collectivités territoriales, yc les assistantes maternelles agréées...

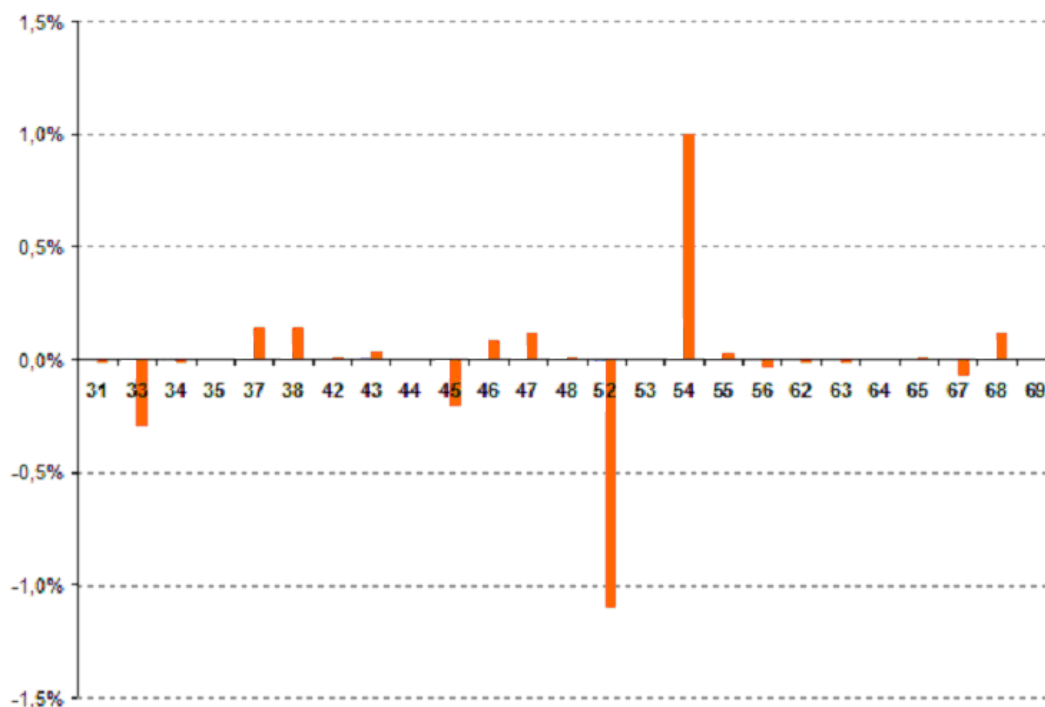
¹⁷ y compris établissements d'hébergement pour personnes âgées

**Transferts les plus fréquents observés en l'absence de PUB
niveau CS – parmi les salariés**

CS chiffrée avec PUB	CS chiffrée sans PUB	
52 Employés civils et agents de service de la fonction publique	54 Employés administratifs d'entreprise	0,969 pt
	63 Ouvriers non qualifiés de type artisanal	0,408 pt
	46 Professions intermédiaires administratives et commerciales des entreprises	0,129 pt
45 Professions intermédiaires administratives de la fonction publique	46 Professions intermédiaires administratives et commerciales des entreprises	0,183 pt
	54 Employés administratifs d'entreprise	0,152 pt
33 Cadres de la fonction publique	38 Ingénieurs et cadres techniques d'entreprise	0,140 pt
	37 Cadres administratifs et commerciaux d'entreprise	0,106 pt
54 Employés administratifs d'entreprise	52 Employés civils et agents de service de la fonction publique	0,104 pt
63 Ouvriers non qualifiés de type artisanal	52 Employés civils et agents de service de la fonction publique	0,297 pt
46 Professions intermédiaires administratives et commerciales des entreprises	45 Professions intermédiaires administratives de la fonction publique	0,257 pt

Enquêtes emploi 2009 et 2010 : 72 327 salariés interrogés pour la 1^{ère} fois ou ayant changé de profession
Lecture : en l'absence de la variable PUB, 0,969 % des salariés, devant être classés dans la CS 52, se retrouvent classés dans la CS 54.

Différence de répartition des salariés dans les CS par rapport au codage de référence avec PUB (en points)



Enquêtes emploi 2009 et 2010 : 72 327 salariés interrogés pour la 1^{ère} fois ou ayant changé de profession

La profession

Au niveau profession, le statut de l'employeur joue pour 4,2 % sur le classement des salariés. En plus, son omission modifie trop la répartition des PCS. *On ne peut pas se passer de PUB pour coder au niveau détaillé.* La variable PUB donne ainsi des informations sur le type de fonction publique : pour les cadres A, il y a confusion entre les collectivités locales et l'État :

Transferts les plus fréquents observés en l'absence de PUB niveau profession – parmi les salariés

Profession chiffrée avec PUB	Profession chiffrée sans PUB	
523a Adjoint administratifs de la fonction publique	542a Secrétaires	0,376 pt
	543d Employés administratifs divers d'entreprises	0,325 pt
	543a Employés des services comptables ou financiers	0,071 pt
684a Nettoyeurs	525c Agents de service de la fonction publique	0,294 pt
525c Agents de service de la fonction publique	684a Nettoyeurs	0,245 pt
	681b Ouvriers non qualifiés du second œuvre du bâtiment	0,158 pt
461d Maîtrise et techniciens des services financiers ou comptables	451e Autre personnel administratif de catégorie B de l'État	0,241 pt
451f Personnel administratif de catégorie B des collectivités locales et des hôpitaux	451e Autre personnel administratif de catégorie B de l'État	0,158 pt
	542a Secrétaires	0,075 pt
333f Personnel administratif de catégorie A des collectivités locales et hôpitaux publics	451e Autre personnel administratif de catégorie A de l'État	0,134 pt

*Enquêtes emploi 2009 et 2010 : 72 327 salariés interrogés pour la 1^{ère} fois ou ayant changé de profession
Lecture : en l'absence de la variable PUB, 0,376 % des salariés, devant être classés dans la PCS 523a, se retrouvent classés dans la PCS 542a.*

1.3.3. Activité principale de l'établissement employeur (NAF2, NAF5)

L'activité principale de l'établissement employeur est une variable essentielle pour coder selon la PCS, pour les salariés comme pour les indépendants. La NAF correspondante peut être chiffrée à 5 ou 2 positions selon l'enquête. On compare ici les codages de la profession :

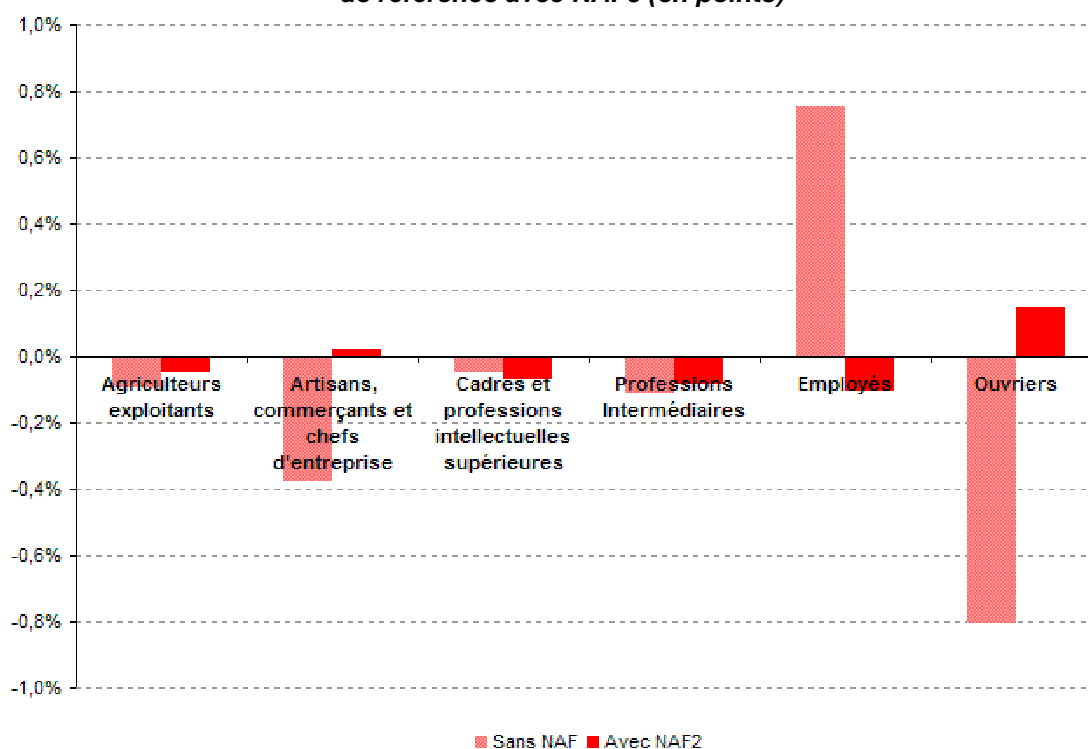
- à partir d'une NAF à 5 positions obtenue par recherche de l'établissement employeur dans le répertoire Sirene ;
- à partir d'une NAF à 2 positions, codée par Sicore-APE en embarqué sur le poste de collecte des enquêteurs ;
- à partir d'une NAF2 codée par Sicore APE en batch à partir d'un libellé d'activité, en aval de la collecte, pour les enquêtes utilisant des questionnaires papier par exemple ;
- sans même le code d'activité NAF2.

Le groupe socioprofessionnel

Sans le code d'activité NAF2, 1,7 % d'actifs sont mal classés. Plus précisément, 0,8 pt d'ouvriers sont mal classés parmi les employés et 0,2 pt d'employés mal classés parmi les ouvriers.

Si en revanche on utilise une NAF2 calculée par Sicore APE au lieu d'une NAF5, les effets sont bien plus faibles : moins de 1 % des individus sont mal classés au niveau groupe socioprofessionnel. *On pourra donc se passer de la NAF5 et n'utiliser qu'une NAF2 pour coder au niveau groupe. On pourra éventuellement se passer totalement de NAF au prix d'une qualité plus faible.*

Différence de répartition des actifs occupés dans les groupes sociaux par rapport au codage de référence avec NAF5 (en points)



Enquêtes emploi 2009 et 2010 : 81 395 actifs occupés interrogés pour la 1^{ère} fois ou ayant changé de profession
La NAF2 est recueillie avec Sicore APE embarqué

La CS

L'absence complète du code NAF (NAF2 ou NAF5) nuit sérieusement à la qualité du codage au niveau CS : 6,6 % des individus sont mal classés et la répartition des catégories sociales s'en trouve largement déformée. En particulier, beaucoup d'ouvriers non qualifiés de type artisanal (0,7 %) se retrouvent classés parmi le personnel des services directs aux particuliers.

Si on ne connaît que la NAF2, les erreurs de classement en CS représentent :

- 2,8 % des individus avec une NAF2 recueillie avec Sicore-APE embarqué
- et 4,1 % avec une NAF2 calculée par Sicore-APE batch en aval de la collecte.

Les confusions les plus courantes se font entre cadres techniques et cadres administratifs ou commerciaux, entre ouvriers de type artisanal et ouvriers de type industriel, entre artisans et commerçants.

Malgré tout, les erreurs de classement sont assez bien réparties entre les différentes CS. Autrement dit, aucune catégorie n'est trop affectée par l'absence d'information détaillée sur l'activité.

Pour ces raisons, et parce que la NAF5 est une information lourde à récupérer, *on pourra déterminer la CS en utilisant uniquement la NAF2*, qu'elle soit recueillie en cours de collecte avec Sicore-APE embarqué ou calculée par Sicore-APE batch en aval¹⁸.

Transferts les plus fréquents observés en l'absence de NAF5 niveau CS – parmi tous les actifs occupés

CS chiffrée avec la NAF5	CS chiffrée sans la NAF5, mais avec la NAF2 recueillie avec Sicore APE embarqué	
38 Ingénieurs et cadres techniques d'entreprise	37 Cadres administratifs et commerciaux d'entreprise	0,305 pt
62 Ouvriers qualifiés de type industriel	63 Ouvriers qualifiés de type artisanal	0,303 pt
68 Ouvriers non qualifiés de type artisanal	67 Ouvriers non qualifiés de type industriel	0,269 pt
63 Ouvriers qualifiés de type artisanal	62 Ouvriers qualifiés de type industriel	0,253 pt
21 Artisans	22 Commerçants et assimilés	0,115 pt

Enquêtes emploi 2009 et 2010 : 81 395 actifs occupés interrogés pour la 1^{ère} fois ou ayant changé de profession
Lecture : en l'absence de la variable NAF5, 0,305 % des salariés, devant être classés dans la CS 38, se retrouvent classés dans la CS 37.

¹⁸ Dans le cas de Sicore APE batch en aval, le taux de codage de la NAF2 s'élève à 76 %, mais le taux de bons codages est de 63 % au niveau sous-classe et de 80 % au niveau division (NAF2).
Malgré tout, la CS est juste dans 96 % des cas et la profession sur quatre positions dans 86 % des cas. Par ailleurs les écarts observés sont répartis à peu près uniformément entre les différentes CS.

La profession

Il est indispensable de connaître la sous-classe d'activité (NAF5) pour déterminer la profession sur quatre positions. En effet, plus de 10 % des individus sont mal classés en l'absence de cette donnée. L'impact de la NAF est plus fort sur les petits indépendants car dans ce cas, l'activité de l'entreprise correspond en général avec la déclaration de profession et prime sur cette dernière. En l'absence de cette information, les transferts se font entre professions très proches car l'activité de l'établissement est un clivage qui intervient principalement au dernier niveau de la nomenclature :

Transferts les plus fréquents observés en l'absence de NAF5 niveau profession – parmi tous les actifs occupés

Profession chiffrée avec la NAF5	Profession chiffrée avec seulement la NAF2 recueillie avec Sicore APE embarqué	
554e Vendeurs en habillement et articles de sport	554a Vendeurs en alimentation	0,36 pt
111a Agriculteurs sur petite exploitation de céréales-grandes cultures	111f Agriculteurs sur petite exploitation sans orientation dominante	0,34 pt
553a Vendeurs non spécialisés	554a Vendeurs en alimentation	0,29 pt
131d Éleveurs d'herbivores, sur grande exploitation	131f Agriculteurs sur grande exploitation sans orientation dominante	0,29 pt
223d Détaillants en habillement et articles de sport, de 0 à 9 salariés	222a Petits et moyens détaillants en alimentation spécialisée, de 0 à 9 salariés	0,21 pt
463c Techniciens commerciaux et technico-commerciaux, représentants en biens de consommation auprès d'entreprises	463b Techniciens commerciaux et technico-commerciaux, représentants en biens d'équipement, en biens intermédiaires, commerce interindustriel	0,20 pt
627d Ouvriers qualifiés de scierie, de la menuiserie industrielle et de l'ameublement	632d Menuisiers qualifiés du bâtiment	0,16 pt
131a Agriculteurs sur grande exploitation de céréales-grandes cultures	131f Agriculteurs sur grande exploitation sans orientation dominante	0,14 pt
633a Électriciens qualifiés de type artisanal (y compris bâtiment)	622b Câbleurs qualifiés, bobiniers qualifiés	0,14 pt
111a Agriculteurs sur petite exploitation de céréales-grandes cultures	111f Agriculteurs sur petite exploitation sans orientation dominante	0,14 pt

Enquêtes emploi 2009 et 2010 : 81 395 individus interrogés pour la 1^{ère} fois ou ayant changé de profession
Lecture : en l'absence de la variable NAF5, 0,36 % des salariés, devant être classés dans la PCS 554e, se retrouvent classés dans la PCS 554a.

1.3.4. Orientation principale agricole (OPA)

Cette question est posée, depuis de nombreuses années, dans le tronc commun des enquêtes auprès des ménages : on demande aux exploitants agricoles l'« orientation principale »¹⁹ de leur exploitation.

En revanche, elle ne figurait plus dans l'enquête emploi ces dernières années. Le motif de la suppression de cette question était la toujours plus faible part des agriculteurs exploitants parmi les actifs occupés (environ 1 % actuellement). Mais elle a été réintroduite en 2013.

Les modalités qu'utilise Sicore-PCS sont identiques à celles de la question posée dans les enquêtes :

- 1 polyculture (culture des terres labourables) ;
- 2 maraîchage ou horticulture ;
- 3 vigne ou arbres fruitiers ;
- 4 élevage d'herbivores (bovins, ovins...) ;
- 5 élevage de granivores (volailles, porcs...) ;
- 6 polyculture-élevage ;
- 7 élevage d'herbivores et de granivores ;
- 8 autre.

La réintroduction de cette variable en 2013 dans l'enquête emploi nous permet de produire des statistiques complètes sur la petite population des agriculteurs exploitants. S'agissant de la profession principale, seules 470 personnes ont été interrogées sur leur orientation principale agricole entre le premier trimestre de 2013 et le premier trimestre de 2014 (inclus), dont 444 au premier rang d'interrogation, auxquels nous allons nous intéresser.

On a établi l'importance de la prise en compte de la variable OPA par Sicore-PCS dans trois scénarios :

- 1° lorsque la NAF (l'activité sur cinq positions) et la NAF2 (l'activité sur deux positions) sont utilisées par Sicore (dans le cas où elles sont disponibles, ce qui est presque toujours le cas de ces 444 agriculteurs exploitants) ;
- 2° lorsque la NAF2 est utilisée (si elle est disponible), mais pas la NAF ;
- 3° lorsque ni la NAF ni la NAF2 ne sont disponibles.

Pourcentage d'erreurs de classement des exploitants agricoles en cas de non-utilisation d'OPA

Scénario	Erreurs au niveau PCS	Erreurs au niveau CS
1° si la NAF5 est utilisée	2,3 %	≈ 0 %
2° si seule la NAF2 est utilisée	57,9 %	7,9 %
3° si la NAF2 n'est même pas utilisée	58,5 %	8,2 %

Enquêtes emploi 2013 et 2014 (1^{er} trimestre) : 444 agriculteurs exploitants interrogés pour la 1^{ère} fois

Le taux d'erreurs au niveau PCS est très élevé dans les scénarios 2 et 3 : un mauvais code est attribué plus d'une fois sur deux. On en déduit qu'en l'absence de la variable NAF à 5 positions (cas des enquêtes ménages sans appariement avec Sirene), la variable OPA a une grande importance pour coder la PCS. Toutefois, on a vu plus haut qu'en l'absence de NAF5, il est impossible de coder correctement la PCS en général.

En revanche, en présence de la NAF5, la variable OPA est peu utile.

Au niveau CS, la variable OPA est moyennement importante en l'absence de NAF5. Vu le petit nombre d'agriculteurs exploitants parmi la population française, cette variable pourra donc être omise des enquêtes ménages.

¹⁹ On trouve parfois aussi « orientation des productions agricoles ».

En conclusion, que la variable NAF5 soit disponible ou non, la variable OPA n'a pas une grande utilité et peut être omise dans les questionnaires des enquêtes ménages.

Voyons néanmoins les conséquences de l'omission de la variable OPA (en l'absence de NAF5).

Dans les 2^e et 3^e scénarios, la CS ne change que pour environ 8 % des exploitants agricoles. Il s'agit le plus souvent d'un passage de la CS 11 vers la CS 12, ou bien d'un passage des CS 11 et 12 vers la CS 13, c'est-à-dire un *agrandissement de la taille de l'exploitation* ! Ces changements de taille d'exploitation peuvent étonner mais en fait, la *taille* servant au codage de la CS n'est pas la surface directement recueillie en collecte : *c'est une notion plus complexe que la simple surface physique mesurée en hectares ou en nombre de têtes de bétail et elle utilise notamment l'orientation agricole comme critère de construction* (voir annexe 2 extraite du guide de la nomenclature PCS 2003).

Au niveau PCS, les changements les plus fréquents sont donnés par le tableau suivant :

**Transferts les plus fréquents observés en l'absence d'OPA
niveau profession – parmi les exploitants agricoles
(scénario n°3 : NAF5 et NAF2 ne sont pas disponibles)**

Profession codée avec OPA	Profession codée sans OPA	
131a Agriculteurs sur grande exploitation de céréales-grandes cultures	131f Agriculteurs sur grande exploitation sans orientation dominante	14,64 pt
131d Éleveurs d'herbivores, sur grande exploitation	131f Agriculteurs sur grande exploitation sans orientation dominante	13,29 pt
111d Éleveurs d'herbivores, sur petite exploitation	111f Agriculteurs sur petite exploitation sans orientation dominante	6,98 pt
111a Agriculteurs sur petite exploitation de céréales-grandes cultures	111f Agriculteurs sur petite exploitation sans orientation dominante	4,05 pt
121d Éleveurs d'herbivores sur moyenne exploitation	121f Agriculteurs sur moyenne exploitation sans orientation dominante	3,15 pt
121a Agriculteurs sur moyenne exploitation de céréales-grandes cultures	121f Agriculteurs sur moyenne exploitation sans orientation dominante	2,48 pt
111c Viticulteurs, arboriculteurs fruitiers, sur petite exploitation	111f Agriculteurs sur petite exploitation sans orientation dominante	1,80 pt

Enquêtes emploi 2013 et 2014 (1^{er} trimestre) : 444 agriculteurs exploitants interrogés pour la 1^{ère} fois
Lecture : en l'absence de la variable OPA, 14,64 % des exploitants agricoles, devant être classés dans la PCS 131a, se retrouvent classés dans la PCS 131f.

NB : Les résultats du scénario 3 sont très proches de ceux du scénario 2, où serait utilisée la NAF2.

Au vu de ce tableau, on peut dire qu'en l'absence de NAF5 l'utilisation de la variable OPA permet systématiquement un codage plus précis de la profession au niveau PCS. D'ailleurs, presque aucun des codes PCS de la colonne de droite n'est présent parmi les 444 codes PCS obtenus lorsque la variable OPA n'est pas utilisée.

1.3.5. Taille de l'entreprise (T)

La taille de l'entreprise est utilisée à la fois pour les salariés et les indépendants.
Pour les indépendants elle sert principalement à pallier l'absence du nombre de salariés.
Pour les salariés, T n'a d'influence que sur quelques professions.

Voici les modalités de T dans Sicore-PCS :

- de 0 à 9 salariés ;
- P de 10 à 49 salariés ;
- M de 50 à 499 salariés ;
- G plus de 500 salariés.

La taille de l'entreprise peut être recueillie de deux manières :

- soit en récupérant la tranche d'effectif de l'entreprise dans le répertoire Sirene, mais cela nécessite de connaître le SIRET ;
- soit en demandant directement à l'enquêté d'estimer la taille de son entreprise. On peut par exemple compléter la question sur le nombre de salariés de la sorte, notamment pour des questionnaires papier :
 - 1 aucun salarié ;
 - 2 un ou deux salariés ;
 - 3 de trois à neuf salariés ;
 - 4 de 10 à 49 salariés ;
 - 5 de 50 à 499 salariés ;
 - 6 plus de 500 salariés.

Il faut alors poser cette question aux salariés comme aux indépendants, en prenant soin de supprimer la modalité « aucun salarié » dans le cas d'un salarié.

Le groupe socioprofessionnel et la CS

La taille de l'entreprise n'a pas d'impact sur la constitution des groupes socioprofessionnels ni des catégories socioprofessionnelles des salariés.

La profession

Certaines professions de la nomenclature des PCS font explicitement référence au concept de petite, moyenne et grande entreprise. En l'absence de l'information sur la taille de l'entreprise :

- les indépendants sont plutôt considérés comme des chefs de petite entreprise (de 10 à 49 salariés) ;
- et les salariés (les cadres en l'occurrence) plutôt comme des cadres de moyenne entreprise (de 50 à 499 salariés).

La variable de taille est indispensable pour bien classer les chefs d'entreprise et certains cadres des moyennes et grandes entreprises. Mais il s'agit de professions rares (1 % des indépendants et 0,3 % des salariés). *On conseillera donc d'utiliser la variable T afin d'obtenir une PCS très précise même pour ces professions. On pourra se passer de la taille de l'entreprise dans tous les autres cas.* Voici les transferts les plus fréquents engendrés par l'absence de la variable T :

Transferts les plus fréquents observés en l'absence de T niveau profession – parmi les salariés

Profession chiffrée avec T	Profession chiffrée sans T	
374c Cadres commerciaux des grandes entreprises (hors commerce de détail)	374d Cadres commerciaux des petites et moyennes entreprises (hors commerce de détail)	0,087 pt
373b Cadres des autres services administratifs des grandes entreprises	373d Cadres des autres services administratifs des petites et moyennes entreprises	0,072 pt
373a Cadres des services financiers ou comptables des grandes entreprises	373c Cadres des services financiers ou comptables des petites et moyennes entreprises	0,038 pt

Enquêtes emploi 2009 et 2010 : 72 327 salariés interrogés pour la 1^{ère} fois ou ayant changé de profession
Lecture : en l'absence de la variable T, 0,087 % des salariés, devant être classés dans la PCS 374c, se retrouvent classés dans la PCS 374d.

1.3.6. Fonction principale (FN)

La fonction principale *ne doit être collectée que pour les salariés.*

Variable SICORE	Question du TCM 2014
FN	F15 (SO) FONCTION Fonction principale
	Dans votre emploi, quelle est votre fonction principale ?
1 production, fabrication, chantiers ;	1. Production, chantier, exploitation
2 installation, réparation, maintenance ;	2. Installation, réparation, maintenance
3 nettoyage, gardiennage, entretien ménager ;	3. Gardiennage, nettoyage, entretien ménager
4 manutention, magasinage, logistique ;	4. Manutention, magasinage, logistique
5 secrétariat, saisie, accueil ;	5. Secrétariat, saisie, accueil
6 gestion, comptabilité ;	6. Gestion, comptabilité
7 commerce, vente, technico-commercial ;	7. Commercial, technico-commercial
8 études, recherche et développement, méthodes ;	8. Études, recherche et développement, méthodes
0 autres.	9. Enseignement
	10. Soins des personnes
	11. Autre fonction

On examine les conséquences de l'absence de FN sur le classement des salariés, sous l'hypothèse que les variables T, NAF5 et OPA sont non disponibles (cas du

Le groupe socioprofessionnel

La fonction n'est pas nécessaire pour déterminer le groupe socioprofessionnel. Son absence n'entraîne qu'une erreur de 0,76 % dans le classement des salariés ; principalement des ouvriers classés comme employés et inversement.

La CS

Au niveau CS en revanche, 2,51 % de salariés sont mal classés en l'absence de la fonction principale. La plupart des transferts s'effectuent de la manière suivante :

Transferts les plus fréquents observés en l'absence de FN niveau CS – parmi les salariés

CS chiffrée avec FN, mais sans T, NAF5 et OPA	CS chiffrée sans FN, ni T, NAF5 ni OPA	
68 Ouvriers non qualifiés de type artisanal	67 Ouvriers non qualifiés de type industriel	0,345 pt
37 Cadres administratifs et commerciaux d'entreprise	38 Ingénieurs et cadres techniques d'entreprise	0,240 pt
38 Ingénieurs et cadres techniques d'entreprise	37 Cadres administratifs et commerciaux d'entreprise	0,163 pt
46 Professions intermédiaires administratives et commerciales des entreprises	48 Contremaîtres, agents de maîtrise	0,161 pt
68 Ouvriers non qualifiés de type artisanal	52 Employés civils et agents de service de la fonction publique	0,156 pt
62 Ouvriers qualifiés de type industriel	63 Ouvriers qualifiés de type artisanal	0,151 pt
68 Ouvriers non qualifiés de type artisanal	54 Employés administratifs d'entreprise	0,139 pt

Enquêtes emploi 2013 à 2014 (1^{er} trimestre) : 41 658 salariés interrogés pour la 1^{re} fois
Lecture : en l'absence de la variable FN, 0,345 % des salariés, devant être classés dans la CS 68, se retrouvent classés dans la CS 67.

La profession

La fonction influe à hauteur de 6,69 % sur le codage des salariés au niveau profession et son omission déforme nettement la répartition des professions.

Pour déterminer une PCS de bonne qualité il est donc utile de connaître la fonction pour les salariés. D'ailleurs, lorsque l'enquête est réalisée avec SICORE-PCS embarqué (comme dans la nouvelle enquête Emploi), il est possible de ne poser la question sur la fonction que dans un nombre précis de cas.

Transferts les plus fréquents observés en l'absence de FN niveau profession – parmi les salariés

Profession chiffrée avec FN, mais sans T, NAF5 et OPA	Profession chiffrée sans FN, ni T, NAF5 ni OPA	
388a Ingénieurs et cadres d'étude, recherche et développement en informatique	388c Chefs de projets informatiques, responsables informatiques	0,955 pt
684a Nettoyeurs	676e Ouvriers non qualifiés divers de type artisanal	0,314 pt
372a Cadres chargés d'études économiques, financières, commerciales	384b Ingénieurs et cadres de fabrication en mécanique et travail des métaux	0,240 pt
478c Techniciens d'installation, de maintenance, support et services aux utilisateurs en informatique	478a Techniciens d'étude et de développement en informatique	0,182 pt
545a Employés administratifs des services techniques de la banque	545b Employés des services commerciaux de la banque	0,168 pt
681b Ouvriers non qualifiés du second œuvre du bâtiment	525c Agents de service de la fonction publique (sauf écoles, hôpitaux)	0,156 pt
477c Techniciens d'installation et de maintenance des équipements non industriels	474c Techniciens de fabrication et de contrôle-qualité en construction mécanique et travail des métaux	0,146 pt

Enquêtes emploi 2013 à 2014 (1^{er} trimestre) : 41 658 salariés interrogés pour la 1^{re} fois
Lecture : en l'absence de la variable FN, 0,955 % des salariés, devant être classés dans la PCS 388a, se retrouvent classés dans la PCS 388c.

Si on recherche une qualité légèrement moindre, on pourra néanmoins se permettre de négliger la fonction professionnelle pour le calcul de la PCS. Certaines professions se vident alors au profit d'une autre, très proche. Souvent les trois premiers chiffres de la PCS sont identiques.

* *
*

Importance de la fonction pour le codage dans la CITP

L'outil Sicore-CITP utilise la fonction comme variable annexe. En l'absence de cette variable, un nombre non négligeable de salariés se trouvent mal classés²⁰ :

- 2,17 % sont mal classés au niveau « grand groupe » (code CITP à une position) ;
- 4,27 % sont mal classés au niveau « sous-grand groupe » (deux positions) ;
- 5,59 % sont mal classés au niveau « groupe de base » (quatre positions)²¹.

Ceci s'explique par la structure de cette nomenclature, qui prend davantage en compte la dimension « métier », donc la fonction, que ne le fait la nomenclature des PCS, plus axée sur le statut et la position professionnelle.

Ce phénomène est encore plus marqué lorsque les variables T, NAF5 et OPA manquent en plus de la fonction (cas du TCM).

Si les résultats de l'enquête doivent être codés dans la CITP, la FONCTION est indispensable.

²⁰ La variable annexe PCS utilisée par Sicore-CITP est elle-même issue d'un codage avec Sicore-PCS n'utilisant pas la fonction.

²¹ Chiffres obtenus pour les salariés au premier rang d'interrogation en 2013.

1.3.7. Surface agricole utilisée (SAU)

La surface agricole utilisée (SAU) ne concerne que les agriculteurs. Dans Sicore PCS, cette variable peut prendre une centaine de valeurs différentes, s'échelonnant entre moins de 20 ares et plus de 400 hectares.

On fait ici l'hypothèse que les variables T, NAF5 et OPA sont non disponibles.

Le groupe socioprofessionnel

La surface agricole utilisée n'influe pas au niveau du groupe socioprofessionnel.

La CS

Comme la variable NBS, la surface agricole intervient au deuxième niveau de la nomenclature : elle permet de hiérarchiser les exploitants agricoles entre eux.

Elle assure le classement des agriculteurs suivant la taille de l'exploitation : petite, moyenne ou grande (relativement à l'orientation principale agricole ; cf. Annexe 2). En l'absence de cette information, Sicore considère très largement que les agriculteurs exploitent de petites exploitations. En l'absence de SAU, la CS est mal codée pour 61 % des agriculteurs exploitants (soit pour 0,6 % des actifs occupés).

Transferts les plus fréquents observés en l'absence de SAU niveau CS – parmi les agriculteurs exploitants

Profession chiffrée avec SAU, mais sans T, NAF5 et OPA	Profession chiffrée sans SAU, ni T, NAF5 ni OPA	
13 Agriculteurs sur grande exploitation	11 Agriculteurs sur petite exploitation	43,04 pt
12 Agriculteurs sur moyenne exploitation	11 Agriculteurs sur petite exploitation	14,87 pt
	13 Agriculteurs sur grande exploitation	1,57 pt
11 Agriculteurs sur petite exploitation	13 Agriculteurs sur grande exploitation	1,57 pt

*Enquête emploi 2013 et 2014 (1^{er} trimestre) : 444 agriculteurs exploitants interrogés pour la 1^{ère} fois
Lecture : en l'absence de la variable SAU, 43,04 % des exploitants agricoles, devant être classés dans la CS 13, se retrouvent classés dans la CS 11.*

La profession

Puisque la variable SAU spécifie la taille de l'exploitation et puisque la taille est discriminante au niveau CS et uniquement à ce niveau-là, l'influence de la variable SAU au niveau fin de la nomenclature est identique à son influence au niveau CS, d'où 61 % de mal classés en son absence. Les transferts les plus fréquents sont le passage de 131f à 111f (66 % des transferts) ou de 121f à 111f (17 % des transferts) ; toutes ces rubriques se rapportant à des exploitations sans orientation dominante.

1.3.8. Nombre de salariés employés (NBS)

Le nombre de salariés employés (NBS) ne concerne que les indépendants. Dans Sicore PCS, cette variable se décline en quatre modalités :

- 1 aucun salarié ;
- 2 un ou deux salariés ;
- 3 de trois à neuf salariés ;
- 4 dix salariés et plus.

En l'absence de cette variable, Sicore utilise la taille de l'entreprise (T) (cf. partie 1.3.5) .

On fait ici l'hypothèse que les variables T, NAF5 et OPA sont non disponibles

Le groupe socioprofessionnel

Ni le nombre de salariés employés, ni la surface agricole n'influent au niveau du groupe socioprofessionnel. Seuls 0,02 % des indépendants sont mal classés en leur absence.

La CS

Comme la variable SAU, le nombre de salariés intervient au deuxième niveau de la nomenclature pour hiérarchiser les indépendants. Cette variable permet de séparer les chefs d'entreprise (qui emploient plus de dix salariés) des artisans et des commerçants. En l'absence de l'orientation principale agricole²², elle permet aussi de distinguer les agriculteurs sur petite, moyenne ou grande exploitation.

L'absence de la variable NBS (en plus de T, NAF5 et OPA) conduit au mauvais classement de 4,92 % des indépendants. *Le nombre de salariés est donc nécessaire pour déterminer la CS des indépendants.* En l'absence d'information, *le classement par défaut se fait surtout vers le bas de la hiérarchie* : les chefs d'entreprise deviennent artisans ou commerçants et les agriculteurs exploitent des petites surfaces.

Transferts les plus fréquents observés en l'absence de NBS niveau CS – parmi les indépendants

CS chiffrée avec NBS et T, mais sans NAF5 et OPA	CS chiffrée sans NBS ni T, NAF5 ni OPA	
13 Agriculteurs sur grande exploitation	11 Agriculteurs sur petite exploitation	2,26 pt
23 Chefs d'entreprise de 10 salariés ou plus	22 Commerçants et assimilés	1,09 pt
	21 Artisans	0,90 pt
21 Artisans	23 Chefs d'entreprise de 10 salariés ou plus	0,34 pt
22 Commerçants et assimilés	23 Chefs d'entreprise de 10 salariés ou plus	0,30 pt

Enquêtes emploi 2013 à 2014 (1^{er} trimestre) : 5 301 indépendants interrogés pour la 1^{er} fois
Lecture : en l'absence de la variable NBS, 2,26 % des indépendants, devant être classés dans la CS 13, se retrouvent classés dans la CS 11.

²² La notion de taille de l'exploitation est plus élaborée que la simple surface physique mesurée en hectares. Elle doit permettre de comparer entre elles des exploitations de nature différente. C'est la marge brute standard qui est traditionnellement utilisée. Dans la PCS, on l'obtient en croisant l'orientation principale agricole (OPA) et la surface agricole utilisée (SAU). Si une de ces deux variables manque, on approxime la marge brute standard par le nombre de salariés.

La profession

Au niveau fin de la nomenclature, 6,38 % des indépendants sont mal classés lorsque la variable NBS manque en plus de T, NAF5 et OPA. Voici les plus fréquents changements :

Transferts les plus fréquents observés en l'absence de NBS niveau profession – parmi les indépendants

Profession chiffrée avec NBS, mais sans T, NAF5 et OPA	Profession chiffrée sans NBS, ni T, NAF5 ni OPA	
131f Agriculteurs sur grande exploitation sans orientation dominante	111f Agriculteurs sur petite exploitation sans orientation dominante	1,49 pt
224d Exploitants de café, restaurant, hôtel, de 3 à 9 salariés	224a Exploitants de petit restaurant, café-restaurant, de 0 à 2 salariés	0,77 pt
131c Viticulteurs, arboriculteurs fruitiers, sur grande exploitation	111c Viticulteurs, arboriculteurs fruitiers, sur petite exploitation	0,55 pt
223a Détaillants en ameublement, décor, équipement du foyer, de 0 à 9 salariés	227d Autres indépendants divers prestataires de services, de 0 à 9 salariés	0,36 pt
233c Chefs d'entreprise commerciale, de 10 à 49 salariés	227d Autres indépendants divers prestataires de services, de 0 à 9 salariés	0,32 pt
233a Chefs d'entreprise du bâtiment et des travaux publics, de 10 à 49 salariés	214f Autres artisans de fabrication (y compris horlogers, matériel de précision)	0,25 pt
233c Chefs d'entreprise commerciale, de 10 à 49 salariés	222a Petits et moyens détaillants en alimentation spécialisée, de 0 à 9 salariés	0,21 pt
233d Chefs d'entreprise de services, de 10 à 49 salariés	227d Autres indépendants divers prestataires de services, de 0 à 9 salariés	0,21 pt

Enquêtes emploi 2013 à 2014 (1^{er} trimestre) : 5 301 indépendants interrogés pour la 1^{re} fois

Lecture : en l'absence de la variable NBS, 1,49 % des indépendants, devant être classés dans la PCS 131f, se retrouvent classés dans la PCS 111f.

2. Préconisations complémentaires et tout aussi utiles sur les questions à poser, selon le type de profession et le mode de collecte

2.1. Niveau de précision atteignable pour la diffusion

Avant même de poser la question du niveau de précision atteignable/souhaitable pour la diffusion d'une variable de profession (1, 2 ou 4 caractères), il convient de s'interroger sur l'utilité de collecter ce type de profession. À titre d'exemple, les professions collectées dans le module F du tronc commun des enquêtes ménages sont les suivantes : la profession principale, la profession secondaire, la profession antérieure et la profession du conjoint décédé. Seule la profession principale est obligatoire dans ce module, les autres sont optionnelles.

Une fois le niveau de précision fixé pour chacune des professions collectées, le concepteur peut arrêter la liste des questions nécessaires au codage, d'après les conclusions de la partie 1.2. Cependant, on module ces conclusions par quelques compléments ci-dessous.

2.1.1. Le recueil ou non de la NAF5 conditionne fortement la diffusion de la PCS à 2 ou 4 positions...sauf pour certaines catégories de professions

Rappel du 1.2.2 : la variable NAF5 principale de l'établissement employeur ne peut être obtenue qu'au moyen d'une recherche dans le répertoire Sirene à partir de ses coordonnées. En général, faute de NAF5, le concepteur ne peut publier au mieux que la CS à 2 positions.

Cependant, pour coder certaines catégories de professions sur 4 positions, telles que les enseignants de l'élémentaire, du secondaire et du supérieur, ou les professions de policiers et militaires, ou encore certaines PCS agrégées au niveau 3 positions, la NAF5 n'est pas utile.

Donc, même si le concepteur ne dispose pas de la NAF5, s'il a besoin pour *certaines catégories de professions* d'un codage plus précis que la CS (en général pour la profession actuelle), il peut aussi choisir de coder la profession sur 4 positions et de la diffuser, *moyennant un avertissement*. À cet effet, le fichier Excel inclus au début de la partie 1.3. détaille notamment l'ensemble des erreurs de codage au niveau PCS en l'absence de NAF5.

2.1.2. Le calcul ou non de la CITP/ISCO impacte la précision à arrêter pour le « codage » de la PCS

Quel que soit le niveau de précision choisi pour la diffusion, on rappelle que :

- l'outil Sicore-PCS (qui code 80 % à 95 % des libellés de profession) attribue systématiquement un code PCS sur 4 positions²³ ;
- le niveau de précision choisi pour la diffusion doit bien sûr être celui demandé pour la reprise manuelle de la profession par le pôle PCS²⁴ qui traite des libellés non reconnus par SICORE (5 % à 20 %) ²⁵ ;
- l'outil informatique Sicore-CITP (ou Sicore-ISCO, qui code dans la nomenclature internationale des professions) attribue systématiquement un code CITP/ISCO sur 4 positions²⁶ ; pour cela, il a impérativement besoin en entrée d'un code PCS sur 4 positions, faute de quoi il lui est impossible de coder.

Donc même si le concepteur ne dispose pas de la NAF5 mais qu'il doit coder la CITP, il doit néanmoins **coder** systématiquement la profession sur 4 positions (quitte à ne **diffuser** que la CS) : pour cela, il doit spécifier aux gestionnaires de reprise manuelle de coder sur 4 positions les 5-20 % de libellés à reprendre.

2.1.3. Professions des parents ou d'un tiers absent du ménage

Si l'on prend l'exemple de la profession des parents à l'adolescence collectée dans le TCM et l'enquête emploi, l'imprécision constatée dans les réponses à l'enquête emploi, notamment au sujet de la position professionnelle, qui est une information très discriminante, oblige à se contenter de publier au niveau du groupe socioprofessionnel (GS, le premier caractère).

Selon leur éloignement dans le temps et dans l'espace, la conclusion peut être similaire pour d'autres professions observées, pour la profession d'un tiers absent du ménage par exemple.

²³ Mais avec une qualité variant selon les variables annexes disponibles.

²⁴ Situé à la Direction Régionale de Franche-Comté

²⁵ Le niveau de précision demandé pour la reprise manuelle fait partie des paramètres à spécifier pour le poste de reprise spécifique à l'enquête et qu'utilisera le le pôle PCS.

²⁶ Mais avec une qualité variant selon les variables annexes disponibles et aussi selon la précision avec laquelle la PCS a été codée.

2.2. Prise en compte des principaux modes de collecte

Selon le mode de collecte, le coût d'une question peut varier ; le concepteur peut donc légitimement se demander s'il doit appliquer les conclusions de l'ensemble de la partie 1 à la lettre quel que soit le mode de collecte.

2.2.1. Enquêtes utilisant un questionnaire papier

Les contraintes du questionnaire papier sont :

- l'espace restreint (4 pages, 6 pages...) limitant le nombre de questions et d'explications
- la simplicité requise pour les filtres des questions : si les questions NBS, SAU peuvent bénéficier de filtres efficaces (les mêmes que sous CAPI, d'ailleurs), en revanche, la question sur la FONCTION ne peut être filtrée efficacement qu'au moyen de Sicore embarqué.

Dans le cadre d'une enquête utilisant un questionnaire papier et présentant des contraintes de place, on doit privilégier les questions ayant la plus grande influence sur le codage de la PCS.

S'il n'y a pas de contrainte de place, les conclusions de la partie 1.2 s'appliquent à la lettre, sauf pour la question sur la FONCTION, dont le coût et la complexité en auto-administré peuvent se révéler prohibitifs. Si le concepteur décide de ne pas poser la question sur la FONCTION, l'intérêt d'introduire la question SAU seule doit alors s'apprécier au regard d'un seuil plus élevé, de 4 260 répondants concernés (pour la profession principale par exemple, il s'agit des actifs occupés).

Éventuellement, il est possible de résumer la variable NBS en deux modalités (plus ou moins de 10 salariés) et la variable FONCTION en quatre modalités.

Incidemment, on rappelle (cf. partie 1.3.3) que dans le cas d'une enquête en dépôt-retrait sans appariement avec le répertoire Sirene, il est possible de collecter un libellé d'activité et de le coder en NAF2 avec Sicore-APE en aval de la collecte, tout cela pour un niveau de qualité suffisant de la CS (à 2 chiffres).

Pour mémoire, le questionnaire papier du recensement de la population, qui contient les questions principales sur la profession, peut être consulté à l'adresse :

http://www.insee.fr/fr/methodes/sources/pdf/questionnaire_rp_bulletin_individuel.pdf

2.2.2. Enquêtes utilisant Capi avec Sicore-PCS embarqué

L'utilisation de Sicore-PCS embarqué sur le poste de collecte des enquêteurs a deux avantages :

- lorsque le libellé de profession saisi n'est pas reconnu dans les bases de Sicore, une reformulation est demandée
- certaines questions qui ne sont pas utilisées pour coder la PCS d'un libellé (d'un individu) donné ne sont pas posées (sauf s'il s'agit de variables d'intérêt). C'est le cas, notamment, de la question sur la fonction (variable FN).

Outre Sicore-PCS embarqué, Sicore-APE embarqué code l'activité de l'établissement employeur au niveau « division » de la NAF (2 positions) et, puisqu'on peut corriger le libellé saisi, donne lieu à un codage de l'activité de meilleure qualité que Sicore-APE lancé en batch après la collecte.

Avec ces outils de collecte (Capi et Sicore-PCS embarqué) qui libèrent complètement des deux contraintes du questionnaire papier et pour une enquête n'utilisant pas le TCM, les conclusions de la partie 1.2 s'appliquent à la lettre. Si les 2 questions SAU et FONCTION sont posées, elles doivent être filtrées selon la méthode du TCM.

2.2.3. Enquêtes utilisant Cawi (Blaise-web) avec Sicore-PCS embarqué

Les enquêtes par internet étant auto-administrées, il y a un risque pour que l'enquêté comprenne mal certaines questions, comme dans le cas d'un questionnaire papier. Pour y remédier, l'utilisation de Sicore installé sur Cawi doit être envisagée dans les toutes prochaines années, mais ce n'est pas encore le cas aujourd'hui²⁷. Des tests sont en cours, dans le cadre de l'enquête emploi, pour valider la qualité des libellés de profession, d'activité et de diplômes recueillis avec Sicore installé sur Cawi. On prévoit que la saisie d'un libellé de profession inconnu de la base Sicore-PCS, ou comportant une faute d'orthographe ou trop flou, déclenche l'affichage de messages tels que « veuillez reformuler votre réponse ». Les recherches internationales en la matière laissent cependant entrevoir que la qualité du recueil est toujours meilleure avec un enquêteur formé à cet effet.

Si les tests de Sicore installé sur Cawi s'avèrent concluants, l'utilisation de Sicore-PCS embarqué permettra, pour chaque individu, de ne poser que les questions nécessaires au codage de sa profession, plutôt que de les poser toutes. Les questions effectivement posées seraient donc les mêmes que sur le poste Capi.

²⁷ Cela permettrait d'ailleurs de faire du multi-mode Internet-face à face avec des questionnaires quasi identiques.

2.3. Autre application pratique : préconisations pour la profession antérieure au RP

La profession antérieure au recensement de la population est codée pour tous les individus sans emploi. Pour ce faire, on utilise uniquement le libellé de profession déclaré et le statut (ce qui correspond au scénario STAT étudié dans la partie 1.2.1). Au RP, 90 % des libellés de professions sont reconnus par Sicore-batch, ou Sicore-interactif lors de la reprise manuelle.

Le tableau de la partie 1.2.1 montre un biais de codage important de la profession antérieure au RP, de 17,5 % de mauvais codages au niveau des groupes socioprofessionnels et 28,2 % au niveau des catégories socioprofessionnelles.

Par ailleurs, on étudie ci-dessous l'équilibre entre aléa de sondage et biais du codage pour deux sources concurrentes sur la profession antérieure, le RP (et plus précisément la dernière EAR) et l'enquête emploi.

On définit V''_i et $EQMT_{A,i}$:

V''_i s'apparente au V_i du chapitre 1, mais est calculé sur la population des chômeurs ou inactifs.

V''_i est la variance moyenne estimée des parts de chaque groupe socioprofessionnel (GS) ou de chaque CS, au sein de la population sans emploi mais en ayant déjà eu un, pour l'un des 2 niveaux de précision i de la PCS, 1 ou 2 chiffres.

Si aucune erreur de mesure ne vient par exemple s'ajouter à l'aléa de sondage, V''_i représente aussi l'erreur quadratique moyenne en présence de toutes les variables annexes.

$$(1) \quad V''_i = (1 - n / N) / (n * C_i * TNE^2) * \sum_J p_{J,ref} * TNE * (1 - p_{J,ref} * TNE)$$

Où cette fois, TNE = taux de non emploi des 15 ans et plus vivant en logement ordinaire (= 49,1% en 2013, source Enquêtes emploi)

Il s'agit d'une formule simplifiée pour une enquête auprès de personnes de 15 à 64 ans vivant en logement ordinaire, en population générale, avec un sondage aléatoire simple (SAS).²⁸

$EQMT_{A,i}$ est alors **l'erreur quadratique moyenne totale dans le scénario de codage A :**

$$(2) \quad EQMT_{A,i} = EQM_{A,i} + V''_i$$

où $EQM_{A,i}$ est la grandeur définie en première partie, soit l'erreur quadratique moyenne pour le scénario de codage (A).

On peut alors faire les comparaisons dans le tableau ci-dessous :

²⁸ Il appartient au concepteur de corriger les formules ci-dessus d'un ratio tenant compte de différents facteurs (plan de sondage stratifié, éventuel effet grappe, calage). En général, les calculs de variance de la division sondage présentent deux indicateurs très utiles : `ecart_type_estime` (celui de l'enquête après calage) et `ecart_type_approche` (dans le cas d'un SAS non calé). Pour les enquêtes AES 2012 et CVS 2012, le ratio des deux écart-types avoisine, respectivement :

- 0,9 pour une variable quantitative déclinée selon 3 tranches d'âges dans AES 2012, ces tranches d'âges ayant été calées sur des données démographiques ; cf document de travail sur la précision d'enquêtes tirées dans Octopusse, du 21/08/2014
- et 1,3 pour des variables quantitatives dans CVS 2012, le calage, réalisé au niveau ménage, apportant un gain faible pour les statistiques individuelles ; cf note interne du 18/01/2013.

**Erreurs quadratiques moyennes totales sur la profession antérieure pour deux sources :
l'enquête emploi et une EAR.**

Source et scénario de codage	Erreurs $\sqrt{EQM}_{A,i}$ (*)		Nombre de répondants de 15 ans et plus dans la source correspondante	Erreurs totales ($\sqrt{EQMT}_{A,i}$)	
	Groupe socioprofessionnel	CS		Groupe socioprofessionnel	CS
Scénario STAT (de type RP, **)	2,2 pt	1,1 pt	7 285 137	2,2 pt	1,1 pt
Scénario sans NAF5 ni T (type EEC, ***)	0,1 pt	0,1 pt	63 850 ²⁹	0,2 pt	0,1 pt

(*) résultats de simulations sur l' *Enquête emploi 2013-2014T2* :

(**) = méthode de codage du RP, chiffres issus du paragraphe 1.2.1

(***) = méthode de codage de l'Enquête emploi, chiffres comparables au scénario de codage sans NAF5, paragraphe 1.2.2

Interprétation :

Compte tenu de la taille d'une enquête annuelle de recensement, l'aléa de sondage, au niveau national (mesuré par la grandeur V''), est très faible, si bien que l'EQMT est quasi égal à l'EQM.

Pour l'enquête emploi en revanche, l'EQM est du même ordre de grandeur que la variance V''_i car le codage utilise un plus grand nombre de variables annexes (il ne manque que la NAF sur 5 positions et la taille de l'entreprise).

Mais surtout, les écarts entre les erreurs quadratiques moyennes totales (calculés sous hypothèse de SAS non calés) sont très nettement en faveur de l'enquête emploi. Ils sont si importants que la prise en compte des méthodes d'échantillonnage et de calage dans l'enquête emploi, qu'il faudrait en toute rigueur introduire dans l'analyse (voir la note de bas de page précédente sur les ratio $\text{ecart_type_estime} / \text{ecart_type_approche}$ et calculés par la division sondages), peuvent difficilement conduire à modifier sensiblement ce constat. Il conviendrait même de regarder si l'enquête emploi n'est pas une source plus précise que le RP dans certaines régions.

²⁹ Sur une année donnée, 164 300 individus différents de 15 ans et plus répondent à l'enquête à un voire plusieurs des quatre trimestres, l'échantillon étant renouvelé par 1/6^e chaque trimestre. Si l'on s'intéresse à des séries annuelles, on doit tenir compte de ce qu'une partie de ces individus répond aussi l'année suivante (et même l'année d'après pour ceux recrutés au 4^e trimestre). Une année d'enquête emploi présente quatre sous-échantillons entrants, soit 63 850 répondants de 15 ans et plus en 2013 : c'est le nombre annuel de répondants sans doublon « sur le long terme ». Pour une variable aussi stable que la profession antérieure et en première approximation, la variance est directement liée à cet effectif.

Annexe 1 : Nomenclature des professions et catégories socioprofessionnelles

La nomenclature des professions et catégories socioprofessionnelles (PCS) est composée de 486 postes pour les actifs occupés. Chaque poste est identifié par un code sur quatre positions. Les postes sont agrégés sous forme de catégories socioprofessionnelles (CS) à deux chiffres (en 31 postes pour les actifs occupés).

Niveau agrégé (8 postes dont 6 pour les actifs)		Niveau détaillé (42 postes dont 31 pour les actifs occupés)	
1	Agriculteurs exploitants	11	Agriculteurs sur petite exploitation
		12	Agriculteurs sur moyenne exploitation
		13	Agriculteurs sur grande exploitation
2	Artisans, commerçants et chefs d'entreprise	21	Artisans
		22	Commerçants et assimilés
		23	Chefs d'entreprise de 10 salariés ou plus
3	Cadres et professions intellectuelles supérieures	31	Professions libérales
		33	Cadres de la fonction publique
		34	Professeurs, professions scientifiques
		35	Professions de l'information, des arts et des spectacles
		37	Cadres administratifs et commerciaux d'entreprise
		38	Ingénieurs et cadres techniques d'entreprise
4	Professions Intermédiaires	42	Professeurs des écoles, instituteurs et assimilés
		43	Professions intermédiaires de la santé et du travail social
		44	Clergé, religieux
		45	Professions intermédiaires administratives de la fonction publique
		46	Professions intermédiaires administratives et commerciales des entreprises
		47	Techniciens
		48	Contremaîtres, agents de maîtrise
5	Employés	52	Employés civils et agents de service de la fonction publique
		53	Policiers et militaires
		54	Employés administratifs d'entreprise
		55	Employés de commerce
		56	Personnels des services directs aux particuliers
		62	Ouvriers qualifiés de type industriel
6	Ouvriers	63	Ouvriers qualifiés de type artisanal
		64	Chauffeurs
		65	Ouvriers qualifiés de la manutention, du magasinage et du transport
		67	Ouvriers non qualifiés de type industriel
		68	Ouvriers non qualifiés de type artisanal
		69	Ouvriers agricoles

Annexe 2 : Notice générale sur les trois catégories d'agriculteurs 11, 12 et 13 (extrait du guide de la PCS 2003)

Ce groupe socioprofessionnel est composé exclusivement d'individus qui exercent à titre professionnel une activité agricole, soit en qualité de chef d'exploitation, soit en qualité d'associé d'exploitation ou d'aide familial non salarié. Par conséquent, cette activité ne comprend que des indépendants.

Les agriculteurs sont classés selon la taille et l'orientation de leur exploitation. L'orientation représente le type d'activité dans laquelle l'exploitation s'est spécialisée : culture de céréales – grandes cultures, maraîchage – horticulture, élevage de bovins, viticulture... La taille est une notion plus complexe que la simple surface physique mesurée en hectares ou en nombre de têtes de bétail. Elle doit permettre de comparer entre elles des exploitations de nature très différente. L'indicateur habituellement utilisé pour cet exercice est la marge brute standard, notion quasiment équivalente à une valeur ajoutée comptable. Ainsi, elle représente, pour une exploitation donnée, une sorte de valeur ajoutée normalisée, qui peut être décomposée par produit. Elle dépend de la taille physique de l'exploitation et de sa localisation géographique. Le tableau ci-après s'inspire largement de ces notions pour classer les agriculteurs. Cela explique, par exemple, pourquoi une exploitation de vignes et fruits de 7 hectares est considérée comme moyenne en zone 2 et 3 et grande en zone 1.

Les aides familiaux non salariés (voir annexe 1) sont classés comme la personne qu'ils aident, c'est-à-dire en exploitants agricoles. Les aides familiaux d'agriculteurs qui se déclarent salariés doivent être classés dans la catégorie 69 « Ouvriers agricoles ».

Le groupe socioprofessionnel des agriculteurs comporte trois catégories d'agriculteurs :

- sur petite exploitation ;
- sur moyenne exploitation ;
- sur grande exploitation.

La séparation des catégories est effectuée selon le critère de taille défini précédemment, la marge brute standard, qui tient compte de la surface agricole utilisée ou SAU de l'exploitation et de la zone géographique.

Délimitation des zones géographiques :

Maraîchage :

zone 1 : Alsace, Champagne-Ardenne, Centre, Haute-Normandie, Basse-Normandie, Bretagne, Poitou-Charentes, Aquitaine, Midi-Pyrénées, Auvergne, Languedoc-Roussillon, Corse

zone 2 : Nord-Pas-de-Calais, Picardie, Île-de-France, Lorraine, Franche-Comté, Bourgogne, Rhône-Alpes, Provence-Alpes-Côte d'Azur, Pays-de-Loire, Limousin.

Vignes et fruits :

zone 1 : Île-de-France, Champagne-Ardenne, Picardie, Basse-Normandie, Bourgogne, Nord- Pas-de-Calais, Alsace, Bretagne, Limousin.

zone 2 : Haute-Normandie, Centre, Lorraine, Franche-Comté, Pays-de-Loire, Poitou-Charentes, Aquitaine, Rhône-Alpes, Provence-Alpes-Côte d'Azur.

zone 3 : Midi-Pyrénées, Auvergne, Languedoc-Roussillon, Corse.

Herbivores :

zone 1 : Champagne-Ardenne, Lorraine, Franche-Comté, Bourgogne, Centre, Limousin, Auvergne, Rhône-Alpes, Midi-Pyrénées, Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Corse.

zone 2 : Basse-Normandie, Île-de-France, Alsace, Pays-de-Loire, Poitou-Charentes, Aquitaine.

zone 3 : Nord-Pas-de-Calais, Picardie, Haute-Normandie, Bretagne.

CHIFFREMENT DES AGRICULTEURS EXPLOITANTS			
ORIENTATION DE L'EXPLOITATION ET ZONE (EVENTUELLEMENT)	SUPERFICIE LIMITE ET NUMÉRO DE RUBRIQUE (superficie en hectares SAU)		
	PETITE EXPLOITATION (moins de 20 ha équivalent-blé)	MOYENNE EXPLOITATION (de 20 à 40 ha équivalent-blé)	GRANDE EXPLOITATION (plus de 40 ha équivalent-blé)
GRANDES CULTURES	moins de 20 ha 111a	de 20 à 40 ha 121a	40 ha et plus 131a
MARAICHAGE - ZONE 1	moins de 0,5 ha 111b	de 0,5 à 1 ha 121b	1 ha et plus 131b
MARAICHAGE - ZONE 2	-	moins de 0,5 ha 121b	0,5 ha et plus 131b
VIGNES ET FRUITS - ZONE 1	moins de 1 ha 111c	de 1 à 2 ha 121c	2 ha et plus 131c
VIGNES ET FRUITS - ZONE 2	moins de 5 ha 111c	de 5 à 10 ha 121c	10 ha et plus 131c
VIGNES ET FRUITS - ZONE 3	moins de 7 ha 111c	de 7 à 14 ha 121c	14 ha et plus 131c
HERBIVORES - ZONE 1	moins de 30 ha 111d	de 30 à 60 ha 121d	60 ha et plus 131d
HERBIVORES - ZONE 2	moins de 25 ha 111d	de 25 à 50 ha 121d	50 ha et plus 131d
HERBIVORES - ZONE 3	moins de 20 ha 111d	de 20 à 40 ha 121d	40 ha et plus 131d
GRANIVORES	moins de 5 ha 111e	de 5 à 10 ha 121e	10 ha et plus 131e
POLYCULTURE - ELEVAGE	moins de 20 ha 111f	de 20 à 40 ha 121f	110 ha et plus 131f

Annexe 3 : Détail du calcul d'une taille minimale d'échantillon $n_{A,i}$ pour une variable annexe A et pour un niveau de précision i (1 ou 2 chiffres de la PCS)

3.1. Justification du calcul de $n_{A,i}$

On définit ci-dessous $EQM_{A,i}$, V_i et $EQMT_{A,i}$:

1) $EQM_{A,i}$ est l'erreur quadratique moyenne provoquée par l'omission d'une variable annexe A, et pour l'un des 2 niveaux de précision i de la PCS (1 ou 2 chiffres).

$$(1) \quad EQM_{A,i} = (1/C_i) * \sum_J (p_{J,A} - p_{J,ref})^2$$

Où $p_{J,A}$ = part du groupe socioprofessionnel ou de la catégorie socioprofessionnelle J dans la population en emploi, calculée selon le scénario de codage A, simulé à partir de l'Enquête emploi
 $p_{J,ref}$ = part calculée avec le codage de référence (avec les 14 variables annexes)
 C_i = nombre de catégories au niveau de précision i (= 6 pour les groupes sociaux, 31 pour les catégories socioprofessionnelles)

2) On prend comme exemple ci-dessous le cas où l'on souhaite connaître la précision totale de la variable de profession, sur la population en emploi.
 Soit alors V_i la variance moyenne estimée des parts de chaque groupe socioprofessionnel (GS) ou de chaque **CS**, au sein de la population en emploi, et liée à l'aléa de sondage.

$$(2) \quad V_i = 1/TE^2 * V'_i$$

Où TE = taux d'emploi (= 64,1 % en 2013)
 V'_i est la variance moyenne estimée des parts des GS ou **CS** des personnes en emploi, ce au sein de la population totale (et non seulement dans la population en emploi, pour la profession principale). Donc :

$$(2bis) \quad V_i = (1 - n / N) / (n * C_i * TE^2) * \sum_J p_{J,ref} * TE * (1 - p_{J,ref} * TE)$$

(V_i n'est pas liée aux variables annexes)

Où n = nombre total d'individus répondants à l'enquête
 N = population totale dans le champ de l'enquête

$p_{J,ref}$ = part de la catégorie socioprofessionnelle J dans la population en emploi

Il s'agit d'une formule simplifiée pour une enquête auprès de personnes de 15 à 64 ans vivant en logement ordinaire, en population générale, avec un sondage aléatoire simple et sans tenir compte d'un éventuel effet grappe ni du calage a posteriori de l'enquête.

Pour la profession antérieure des chômeurs et des anciens actifs par exemple, on remplacera le taux TE par $1 - TE$.

Il appartient au concepteur de corriger les formules ci-dessus d'un ratio tenant compte de ces différents facteurs. En général, les calculs de variance de la division sondage présentent deux indicateurs très utiles : *ecart_type_estime* (celui de l'enquête après calage) et *ecart_type_approche* (dans le cas d'un SAS non calé). Pour les enquêtes AES2012 et CVS2012, le ratio des deux écart-types avoisine, respectivement :

- 0,9 pour une variable quantitative déclinée selon 3 tranches d'âges dans AES2012, ces tranches d'âges ayant été calées sur des données démographiques ; cf. document de travail sur la précision d'enquêtes tirées dans Octopusse, du 21/08/2014
- et 1,3 pour des variables quantitatives dans CVS2012, le calage, réalisé au niveau ménage, apportant un gain faible pour les statistiques individuelles ; cf. note interne du 18/01/2013.

V_i est supposée représenter l'erreur quadratique moyenne en présence de toutes les variables annexes (aucune erreur de mesure ne viendrait par exemple s'ajouter à l'aléa de sondage).

3) EQMT_{A,i} est l'erreur quadratique moyenne totale lorsque manque une variable annexe A :

$$(3) \quad EQMT_{A,i} = EQM_{A,i} + V_i$$

4) On définit alors une taille critique (ou minimale) d'échantillon $n_{A,i}$ pour chaque variable annexe A et pour chacun des 2 niveaux de précision i (1 ou 2 chiffres de la PCS). $n_{A,i}$ est tel que :

$$(4) \quad \sqrt{EQMT_{A,i}} = 1,1 * \sqrt{V_i}$$

Tant que la taille n de l'échantillon de l'enquête est inférieure à $n_{A,i}$, et donc que l'écart-type moyen des parts de chaque catégorie socioprofessionnelle ($\sqrt{V_i}$) est relativement important, l'erreur quadratique moyenne totale lorsque manque la variable annexe A ne dépasse pas l'écart-type moyen de plus de 20 % : on peut alors accepter de ne pas inclure la variable annexe A dans le questionnaire. À l'inverse, au-delà de cette taille d'échantillon, on doit inclure la variable annexe A.

Pour une enquête ménages classique, on peut assimiler à 1 le facteur d'exhaustivité $(1 - n / N)$, ce qui simplifie les formules.

En conséquence de (4) :

$$(5) \quad n_{A,i} = 0,21 / (C_i * TE^2 * EQM_{A,i}) * \sum_J p_{J,ref} * TE * (1 - p_{J,ref} * TE)$$

Le chiffre de 1,1 ou 10% (et donc $0,21 = 1,1^2 - 1$) est pris arbitrairement : si on est plus exigeant, avec un seuil de 1,05 par exemple, alors on sera enclin à rajouter une variable annexe pour une taille d'échantillon plus faible (dans un rapport de 0,21/0,1025 dans notre exemple).

3.2. Application aux variables CPF, PUB et NAF2

Revenons maintenant aux variables CPF, PUB et NAF2 étudiées dans la partie 1.2.2. Sous les hypothèses précédentes, on aboutit aux valeurs suivantes pour ces variables :

Impact de l'absence marginale des variables CPF, PUB et NAF2 sur le classement des actifs occupés

	Groupe socioprofessionnel		CS	
	\sqrt{EQM}	Taille minimale d'utilité de la variable	\sqrt{EQM}	Taille minimale d'utilité de la variable
Sans CPF	1,380 pt	246	0,549 pt	340
Sans PUB	0,066 pt	107 352	0,249 pt	1 652
Sans NAF5 ni NAF2 ³⁰	0,549 pt	1 552	0,337 pt	902

Enquête emploi 2009 et 2010 : 81 395 actifs occupés en première interrogation ou ayant changé de profession

Les tailles d'échantillon dans la deuxième colonne confirment la conclusion de la partie 1.2.2.

³⁰ Si la NAF2 est absente, la NAF5 l'est *a fortiori*.

Annexe 4 : Exemple de l'enquête emploi 2015 : liste des variables annexes collectées selon le type de profession

Variables annexes toujours demandées dans l'Enquête emploi, pour tous les Types de profession

STRE	Occupation : travaille / travaillait
STATUT	Statut (indépendant / salarié)
PUB	Nature de l'employeur (public / privé)
CPF	Classification (= position professionnelle)
S	Sexe
NBS	Nombre de salariés

Autres variables annexes demandées diversement selon le type de profession

		Type de profession					
		principale	à l'entrée dans l'entreprise	un an auparavant	antérieure	secondaires	des parents
T	Taille	x					
NAF	Naf	x					
NAF2	Naf 2	x	x	x	x	x	
FN	Fonction	x		x	x		
SP	Statut d'apprenti	x	x	x			
OPA	Orientation agricole	x	+/- (*)		x		
SAU	Surface agricole	x	+/-		x		
ENCADR	Supervision	x		x	x		
TYPETENS	Type d'établissement scolaire	x	x	x	x	x	
GRADEDEC	Grade à part de la profession	x					
DEP	Département						

(*) uniquement pour les agriculteurs n'ayant pas changé de situation

Le type d'établissement scolaire sert uniquement si les enseignants n'ont pas indiqué leur grade en clair.

Seules les professions principale, un an auparavant et antérieure sont codées en CITP/ISCO.

Annexe 5 : Exemple de l'enquête emploi 2015 : consignes particulières de collecte pour certaines questions nécessaires pour coder la profession principale actuelle

Libellé de profession

BA10 Quelle est votre profession (principale) ?

La profession principale pour les personnes qui exercent plusieurs professions est celle qu'elles déclarent comme telle. Si l'enquêté ne sait pas se déclarer, considérer la profession qui prend le plus de temps ou en cas d'égalité, la plus rémunératrice.

On entend par profession le métier exercé par la personne enquêtée à son poste de travail. Ce n'est pas l'activité de l'établissement, ni la fonction du service dans lequel il exerce, ni le grade.

À la lecture de l'intitulé de profession, il doit être possible de se représenter le plus précisément possible son activité sur son poste de travail c'est à dire les tâches réalisées. Il vaut mieux être trop précis que pas assez. Attention : l'ensemble des informations qui font l'objet des questions suivantes (statut, classification professionnelle ...) ne doivent pas figurer dans le libellé de la profession. Par exemple : « fonctionnaire », « cadre », « ouvrier » ou « chez Marty ».

La déclaration spontanée de l'enquêté doit être privilégiée. Cependant il faut être attentif à des confusions possibles :

- La profession peut être différente de la qualification ou du « métier » de l'enquêté :

Par exemple, un individu qui a une qualification de peintre et qui est actuellement employé de conditionnement peut spontanément indiquer qu'il est peintre et non sa profession principale d'employé de conditionnement.

- Une personne ayant fait des études d'infirmière va peut-être se déclarer « infirmière » alors qu'elle fait provisoirement autre chose.

Il faut donc indiquer la profession exercée durant la semaine de référence et non la qualification habituelle de l'enquêté.

Certains libellés sont trop vagues, par exemple informaticien.

Dans tous les cas, l'enquêteur inscrit le libellé de profession avec l'accord de l'enquêté.

STATUT _ Statut vis-à-vis de l'emploi

BA13 Vous exercez cette profession :

1. A votre compte
2. Salarié chef d'entreprise, gérant mandataire, PDG
3. Salarié (autre que chef d'entreprise)
4. Vous travaillez pour un membre de votre famille sans être salarié

Dans le cas d'un stage, choisir la modalité 3 : « salarié ».

Les apprentis sous contrat, les travailleurs à domicile, les stagiaires rémunérés sont considérés comme des salariés,

Les prêtres, religieux, rabbins, pasteurs seront conventionnellement classés comme salariés.

SP _ Emploi précaire

BB8 Quel est le type de votre contrat de travail ?

0. Pas de contrat de travail
1. Contrat à durée indéterminée (CDI)
2. Contrat à durée déterminée (CDD) autre que saisonnier
3. Contrat saisonnier
4. Contrat d'intérim ou de travail temporaire
5. Contrat d'apprentissage

NAF / NAF2 _ Activité de l'entreprise (pour les chefs d'entreprise) ou de l'établissement employeur

BA32a Quelle est l'activité économique principale de l'entreprise , de l'établissement qui vous emploie ?

L'activité économique à libeller est celle du lieu ou des lieux d'exercice de la profession principale, généralement celle de l'établissement. Par exemple, pour un peintre à l'entretien des bâtiments dans une usine de textile, on ne répondra pas « peinture », qui est son occupation professionnelle, ni même « entretien des bâtiments » qui est l'activité de son service, mais « industrie textile » parce que cette activité est celle de l'usine tout entière.

Attention : l'activité de l'établissement dans lequel travaille l'enquêté n'est pas toujours celle de l'entreprise. Par exemple : les salariés d'une société de crédit appartenant à une entreprise automobile ne répondront pas « construction automobile » mais « distribution de crédit ».

Lorsque l'établissement a plusieurs activités, indiquez uniquement la principale.

Le libellé de l'activité doit être le plus précis possible, sans mot inutile. La zone a une longueur de 50 caractères maximum. Efforcez-vous de bien distinguer les caractéristiques de la production et de la vente (commerce de détail de gros, industriel ou artisanal).

Par exemple, le terme de « Commerce » est trop vague, vous préciserez « Commerce de gros en fruits et légumes ». Pour boulangerie ou charcuterie, précisez industrielle ou artisanale.

PUB _ Nature de l'établissement employeur

BA30 Quelle est la nature de votre employeur (principal) ?

ou : **Quelle est la nature de vos employeurs ?**

1. Entreprise privée ou association
2. Entreprise publique (EDF, La Poste, SNCF, etc.)
3. État
4. Collectivités territoriales
5. Hôpitaux publics
6. Sécurité sociale
7. Particulier

Dans la mesure du possible, l'enquêteur doit renseigner seul cette question à partir des informations recueillies sur le nom de l'employeur s'il est sûr de ne pas se tromper.

La catégorie employeurs particuliers ne dépend pas du type d'activité exercée par l'employeur lui-même, mais de sa qualité de particulier qui emploie la personne. Par exemple, dans le cas d'une femme de ménage employée par un avocat, l'employeur avocat sera classé « particulier » si la salariée travaille au domicile de l'avocat, et « entreprise privée » si elle travaille au cabinet de l'avocat. Les grandes entreprises possédées ou contrôlées par l'État, ou les établissements publics sont à classer dans la modalité « Entreprise publique (EDF, La Poste, SNCF, CEA, RATP, ADP , BDF, DCNS , France Télévision et ONF etc.) », les associations dans la modalité « Entreprise privée ou association ».

Indications supplémentaires données à l'enquêteur dans le TCM :

1. y compris établissements publics administratifs (EPA) nationaux, les établissements publics d'enseignement
4. y compris OPHLM, des établissements publics administratifs (EPA) des collectivités territoriales, les assistantes maternelles agréées...
5. y compris établissements d'hébergement pour personnes âgées

CPF _ Position professionnelle

Pour les salariés du privé :

BA49 Dans cet emploi de « ... profession », êtes-vous classés comme...

1. Manœuvre ou ouvrier spécialisé
2. Ouvrier qualifié ou hautement qualifié, technicien d'atelier
3. Technicien

4. Agent de maîtrise, maîtrise administrative ou commerciale, VRP (non cadre)
5. Ingénieur, cadre (à l'exception des directeurs ou de ses adjoints directs)
6. Directeur général, adjoint direct
7. Employé de bureau, employé de commerce, personnel de services
8. Autre

Attention à l'appellation « assimilés cadres ». Ce ne sont pas des cadres mais généralement des techniciens, des agents de maîtrise, des VRP, plus rarement des employés. Même si ces salariés cotisent à la caisse de retraite des cadres, ils ne doivent pas être classés en catégorie « cadres ».

Les apprentis sont classés conventionnellement en « ouvrier non qualifié » ou, dans les métiers non ouvriers en « employé ».

Les artistes sont classés comme « ingénieurs, cadre ».

La quasi-totalité des salariés peut être classée dans la nomenclature. La catégorie « Autres » pourra être utilisée dans les cas particuliers des intermittents du spectacle, des sportifs, des religieux, etc.

Les artistes sont classés comme « ingénieurs, cadre ».

Parfois l'enquêteur est face à un enquêté qui sait donner un intitulé plus ou moins précis de son statut ou de son métier dans l'administration, mais ne connaît pas sa propre catégorie parce qu'il y a eu des changements statutaires au cours de l'histoire (passage des instituteurs aux professeurs des écoles, par exemple). Dans ce cas, la règle à suivre est de choisir la catégorie la plus basse de celles entre lesquelles l'enquêté hésite, car c'est comme ça qu'est conçue la codification de la PCS dans Sicore : quand on ne connaît pas la position professionnelle, on code selon la position la plus basse possible pour l'intitulé de métier dont on dispose.

Pour les salariés du public :

BA50 Dans cet emploi, êtes-vous classé comme :

1. Manœuvre ou ouvrier spécialisé
2. Ouvrier qualifié ou hautement qualifié
3. Technicien
4. Personnel de catégorie C
5. Personnel de catégorie B
6. Personnel de catégorie A
7. Autre

Si l'enquêté ne connaît pas sa propre catégorie parce qu'il y a eu des changements statutaires au cours de l'histoire (passage des instituteurs aux professeurs des écoles, par exemple), la règle à suivre par l'enquêteur est de choisir le grade le plus bas de ceux entre lesquels l'enquêté hésite.

Les artistes sont classés comme « ingénieurs, cadre ».

Les élèves ADS (adjoint de sécurité) et les élèves officiers ne sont pas titulaires. Il faut malgré tout les classer dans le grade correspondant à un titulaire.

FONCTION

BA36 Dans cet emploi, quelle est votre fonction principale ?

1. Production, chantier, exploitation
2. Installation, réparation, maintenance
3. Gardiennage, nettoyage, entretien ménager
4. Manutention, magasinage, logistique
5. Secrétariat, saisie, accueil
6. Gestion, comptabilité
7. Commerce (y compris restauration, hôtellerie, tourisme), technico-commercial
8. Études, recherche et développement, méthodes
9. Enseignement, soin aux personnes
10. Autre fonction

Grade

BB3 Vous êtes agent de la fonction publique d'État, territoriale ou hospitalière, veuillez préciser, s'il y a lieu, votre grade, votre corps, votre spécialité.

Il s'agit de prendre la déclaration de l'enquêté. Si ce dernier ne sait pas répondre, il y a la possibilité d'indiquer « NSP » (ne sait pas).

Ne jamais mettre d'autres informations que celles concernant la catégorie (pas de niveau d'indices, d'échelon, etc.).

Ne jamais mettre les lettres A B C seules : inscrire plutôt : « catégorie B » , par exemple.

Les catégories ne peuvent pas être des chiffres.

Contractuel, stagiaire, vacataire, titulaire ne sont pas des grades.

SAU _ Surface agricole utilisée

BB26 Quelle est en la superficie agricole utile de l'exploitation ?

Comptez la superficie agricole utilisée c'est-à-dire celle qui est cultivée, exploitée. Ôtez de la superficie totale les friches, les étangs, les bois, les landes.

Pour les GAEC (Groupe Agricole d'Exploitation en Commun), indiquez leur superficie totale.

Pour les exploitations hors sol, conventionnellement vous inscrirez « 0 ».

Dans l'enquête emploi, les variables NBS et T sont obtenues par consolidation :

- de la taille de l'entreprise issue des données Sirene, pour la profession principale
- et, en cas d'échec d'appariement Sirene ou pour les autres professions, des deux questions ci-dessous :

Nombre de salariés du site

Question utilisée pour les salariés de la fonction publique

BB30 Combien y a-t-il environ de salariés sur le site qui vous emploie ?

NBSALA

1. 1 salarié
2. 2 salariés
3. 3 salariés
4. 4 salariés
5. 5 salariés
6. 6 salariés
7. 7 salariés
8. 8 salariés
9. 9 salariés
10. 10 à 49 salariés
11. 50 à 499 salariés
12. 500 salariés ou plus

Les apprentis sont exclus de ce décompte. La réponse attendue correspond à l'effectif moyen sur la période récente.

Taille de l'entreprise

BB31 Combien de personne employez-vous ?

ou : **Combien de salariés travaillent dans l'entreprise qui vous emploie ?**

0. Aucun salarié
1. 1 salarié
2. 2 salariés
3. 3 salariés
4. 4 salariés
5. 5 salariés
6. 6 salariés
7. 7 salariés
8. 8 salariés
9. 9 salariés
10. 10 à 49 salariés
11. 50 à 499 salariés
12. 500 salariés ou plus
13. Sans objet (entreprise avec un seul établissement)

Annexe 6 : Les questions servant au codage de la profession et collectées dans le TCM

Dans le tronc commun des enquêtes ménages, on demande le libellé de profession principale en clair³¹ et on pose plusieurs questions afin de reconstituer la plupart des variables annexes utiles à Sicore-PCS. Le tableau suivant présente ces questions dans la version 2014 du TCM.

Questions du module F du TCM 2014		Variable annexe déduite et utilisée par SICORE
N°	NOM DE LA VARIABLE et Intitulé de la question	
F4 (F)	STATUTECL Statut dans l'emploi principal	PUB et STATUT
F7 (O)	PROFESSION Profession principale de l'enquêté ou de la personne aidée <i>en clair (40 caractères au maximum)</i>	libellé de profession, complété du grade pour les fonctionnaires
F9 (SO)	TYPEEMPLOI Type d'emploi	SP
F12(F)	SUPERVISION Supervision d'autres salariés	SUPV (pour Sicore-CITP)
F13 (F)	CLASSIF Classification dans l'emploi	CPF
F15 (SO)	FONCTION Fonction principale	FN
F16 (F)	SALARIES Nombre de salariés pour les « indépendants »	NBS
F17 (SO), F18 (F), F19 (SO)	Quelle est l'activité de l'établissement qui vous emploie ou que vous dirigez ? Avec Sicore-APE en embarqué	NAF2
F20 (O)	SUPH Superficie de l'exploitation agricole	SAU
F21 (O)	SUPA Superficie en ares	SAU
F22 (F)	OPA Orientation des productions agricoles	OPA
F23 (F)	AFTYPTRAV Type de travail <u>de l'aide familial</u> : secrétariat, vente, comptabilité / autre	pas utilisée par Sicore-PCS ³²

D'autres variables sont également nécessaires pour coder la PCS sur 4 positions. Elles sont collectées par exemple pour la profession principale dans l'enquête emploi, non mais pas dans le TCM. Il s'agit de :

Activité de l'établissement, par appariement Siam	NAF5
Taille de l'entreprise (pour les salariés)	T

Remarque sur la variable STATUTECL :

Cette variable permet, d'une part, de déterminer les valeurs des variables annexes PUB et STATUT et, d'autre part, d'aménager la question sur la position professionnelle (CPF) selon le caractère *public* ou *privé* de l'employeur. C'est la plus importante de toutes les variables annexes.

³¹ Il importe d'ailleurs de rappeler aux enquêteurs de bien recueillir le *grade* en plus du métier dans le libellé de profession (par exemple : *professeur agrégé*) Il est même conseillé d'inclure une question spécifique pour collecter ce grade, comme c'est le cas pour la profession principale dans l'Enquête emploi depuis 2013. Mais cela suppose alors de mettre en place des codages automatiques et manuels plus complexes, disponibles uniquement dans le cadre du nouveau poste de reprise générique des variables d'emploi-activité, et non dans le Poste de Reprise Harmonisé (PRH) : dans le poste générique, on essaie d'abord de coder avec Sicore le libellé de profession auquel on accole le grade et, en cas d'échec du codage, on essaie de coder le libellé sans le grade.

³² On peut cependant en tenir compte dans des traitements postérieurs à Sicore-PCS pour coder au mieux la PCS à 4 positions, à l'intérieur des CS concernées (artisans, commerçants, professions libérales)

F4 (F)	STATUTECL Statut dans l'emploi principal (pour tous les actifs occupés)
--------	---

Êtes-vous :

- 1. Salarié(e) de l'État ?
- 2. Salarié(e) d'une collectivité territoriale ?
- 3. Salarié(e) d'une entreprise, d'un artisan, d'une association ?
- 4. Salarié(e) d'un ou plusieurs particulier(s) ?
- 5. Aide un membre de votre famille dans son travail, sans être rémunéré
- 6. Chef d'entreprise salarié, PDG, gérant minoritaire, associé
- 7. Indépendant ou à son compte
- 9. Salarié(e) des hôpitaux publics ou du secteur public social et médico-social ?

Bibliographie

Note N° 1953/DG75-F230/SLM-RW du 6 octobre 2011 : Influence des variables annexes dans le codage de la profession selon la nomenclature des PCS.

Note N° 1070/DG75-F230/NR-LM-TD du 7 juin 2013 : Nouvelle méthode de codage selon la CITP-08 (ISCO-08) dans les enquêtes auprès des ménages : guide à destination des concepteurs et des CPOS.

Note N° 553/DG75-F230/SM : Utilité de la question sur le « grade en clair déclaré » dans les enquêtes-ménages du 25 février 2008.

Note N° 112 /F420 du 12 mai 2005 : Codification de la profession avec Sicore – Les variables nécessaires et leurs modalités.