

Enseignements sur les estimations d'emploi de l'appariement entre l'enquête Emploi et les bases administratives en 2012

Claude PICART



Institut national de la statistique et des études économiques

F1906

**Enseignements sur les estimations d'emploi de
l'appariement entre l'enquête Emploi et les bases
administratives en 2012**

Claude PICART*

Septembre 2019

Direction des Statistiques Démographiques et Sociales - Timbre F001
88, avenue Verdier - CS 70058 - 92541 MONTRouGE CEDEX - France
Tél. : 33 (1) 87 69 62 82 - E-mail : dg75-f001@insee.fr - Site Web Insee : <http://www.insee.fr>

Ces documents de travail ne reflètent pas la position de l'Insee et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only their author's views.

* Insee-DSDS - Direction des Statistiques Démographiques et Sociales – Département de l'Emploi et des Revenus d'Activité

Enseignements sur les estimations d'emploi de l'appariement entre l'enquête Emploi et les bases administratives en 2012

Synthèse.....	2
Introduction.....	5
Présentation de la démarche.....	8
I - Description de l'appariement.....	11
A - Appariement avec les DADS.....	11
1 - Fichiers en entrée.....	11
2 - Choix des identifiants.....	11
3 - Quel critère de qualité des appariements ?.....	14
B - Description de l'appariement avec les bases non-salariés.....	17
C - Fusion des appariements.....	20
D - Restriction de l'appariement DADS aux présents dans le fichier poste.....	21
1 - Examen interne aux DADS des raisons de l'élimination des postes.....	22
2 - Parmi ces individus dont tous les postes DADS ont été supprimés, certains sont salariés selon l'EEC.....	23
II – Analyse du nombre de personnes en emploi au moins une fois dans l'année.....	25
A - Population de référence dans les sources administratives : partage entre principalement salariés et principalement non-salariés.....	25
1 - Identification des sources de biais.....	25
2 - Le partage entre principalement salariés et principalement non-salariés.....	26
B - Population de référence dans l'EEC : correction du biais d'attrition par calage du panel EEC...	27
C - Le calage de l'échantillon apparié sur les bases administratives.....	29
D - Résultats – les personnes « Jamais en emploi dans l'EEC ».....	29
1 - Pour les salariés et principalement salariés.....	29
2 - Sur l'ensemble du champ.....	30
E - Résultats - PEA2.....	31
1 - Démarche.....	31
2 - Résultats par groupe.....	32
3 - Par âge : une sous-pondération des jeunes en emploi dans l'EEC.....	36
F - Ajout des trous de « collecte » des bases administratives.....	37
G - Synthèse sur les PEA.....	40
III – Passage à l'estimation de l'emploi moyen annuel.....	42
A - Démarche, concept d'emploi et préliminaires.....	42
1 - Un traitement spécifique pour les salariés purs.....	42
2 - Une différence de concept : les absences de longue durée.....	43
3 - Emploi DADS : moyenne annuelle vs moyenne des semaines de référence.....	44
B - La question des dates et des fusions de période dans les DADS.....	45
1 - Le décalage de paie.....	45
2 - Durée, nombre d'heures et densité.....	45
C - Traitement des salariés purs.....	46
1 - Tout emploi déclaré dans l'EEC peut être supposé comme étant de l'emploi.....	47
2 - Absences EEC : des facteurs explicatifs significatifs mais trop ciblés.....	48
3 - Les Jemp sont-ils de l'emploi ?.....	51
4 – Affectation des absences EEC.....	54
D - Traitement des non-salariés.....	58
IV – Enseignements de l'appariement.....	59
A - Retour sur la démarche.....	59
B – Enseignements sur l'enquête Emploi.....	61
C – Enseignements sur les sources administratives.....	62
D – Synthèse : écart entre EEC et estimations annuelles d'emploi.....	64
Annexes.....	66
Annexe 1 – Compléments (détails) au document.....	66
Annexe 2 – Glossaire des sigles.....	80

Synthèse

Le niveau d'emploi des Estimations annuelles d'emploi (EAE), estimé à partir des sources administratives, est de 26,5 millions en 2012, en France métropolitaine ; il est supérieur de 0,7 million à celui de l'enquête Emploi en continu (EEC) : 25,8 millions. Ces deux mesures ne portent toutefois pas sur le même concept (emploi *annuel moyen au sens du BIT* pour l'EEC, emploi *rémunéré au 31/12* pour les EAE) ni sur le même champ (*au lieu de résidence* et hors communautés pour l'EEC, *au lieu de travail* pour les EAE).

Un appariement pour expertiser les facteurs d'écart entre emploi annuel de l'enquête Emploi et les bases administratives

Pour expertiser les facteurs d'écart entre ces deux sources, un appariement entre l'EEC et les bases statistiques (Données Annuelles de Déclarations Sociales – DADS – et « non-salariés ») alimentées par les données administratives a été réalisé pour les années 2011, 2012 et 2013. Les estimations proposées dans ce document de travail portent sur 2012, tout en mobilisant marginalement des informations sur 2011 et 2013.

Le concept commun le plus pertinent, pour cette comparaison, est celui de l'**emploi annuel moyen au sens du BIT** ; en effet ramener les données de l'EEC à un concept d'emploi en fin d'année aurait considérablement réduit l'échantillon de cette dernière ; en outre l'EEC est calée et pondérée pour être utilisée en moyenne trimestrielle. Certaines périodes d'emploi rémunérées ne correspondent toutefois pas à de l'emploi au sens du BIT : 0,2 million d'emplois des DADS correspondent à des congés de longue durée dans l'enquête Emploi et sont donc exclus du champ (ligne (3) du tableau de synthèse suivant).

Le champ de comparaison est celui de l'intersection des deux sources :

- les frontaliers, résidents français travaillant à l'étranger, d'environ 0,4 million, sont ainsi exclus de l'EEC car non présents dans les DADS ;
- le cas symétrique, résidents étrangers travaillant en France, est négligeable dans les DADS.

Pour être sur un champ commun, il faudrait exclure les actifs vivant en communauté (hors champ de l'EEC) dans les DADS mais on ne sait pas les distinguer dans cette source. Le recensement de la population permet d'estimer leur niveau à 0,2 million. Au total ces deux différences de champ contribuent pour -0,2 million à l'écart d'ensemble (ligne (2))

L'appariement entre les deux sources s'est fait sur les variables de prénom, sexe, date et département de naissance, ainsi que la commune de résidence et le numéro de voie de l'adresse. Il porte sur des individus : en 2012, sur le champ précisé ci-dessus, il y a 32 millions de personnes en emploi au moins une fois dans l'année dans les bases issues des sources administratives. Environ 88 000 répondants à l'EEC déclarent être en emploi à au moins une des quatre interrogations de l'année : 81 % d'entre eux sont retrouvés dans au moins une source administrative.

La stratégie d'estimation des écarts d'emploi se fait en deux étapes :

- la première consiste à **comparer dans les deux sources le nombre de personnes en emploi au moins une fois dans l'année** (« PEA ») ; cette comparaison permet de détecter des défauts de couverture dans les deux sources, et, au niveau des individus, d'affecter un nouveau poids.
- la seconde étape permet d'**estimer la durée moyenne dans l'année (DMA) de chaque individu apparié** en confrontant, à chacune des 4 semaines de référence, l'emploi dans l'enquête et l'emploi de la source administrative. Elle permet de détecter les absences à tort dans l'EEC, les présences à tort dans les DADS, et sous un certain nombre d'hypothèses, de proposer une méthode arbitrant l'« emploi annuel moyen » de chaque personne en emploi.

La combinaison de ces deux étapes permet de proposer un emploi annuel moyen (EMA) repondéré à partir de chacune des deux sources. Ainsi, en supposant représentatifs les individus appariés, les chiffrages proposés par la suite sont établis directement à partir de la base d'appariement, après repondération des individus et arbitrage, chaque trimestre, entre l'emploi EEC et les sources administratives.

Un écart entre EEC et estimations annuelles qui résulte pour partie d'une sous-déclaration dans l'EEC

Les différences de champ (communautés, transfrontaliers) et de concept d'emploi (congrés maladie) se compensent globalement (tableau).

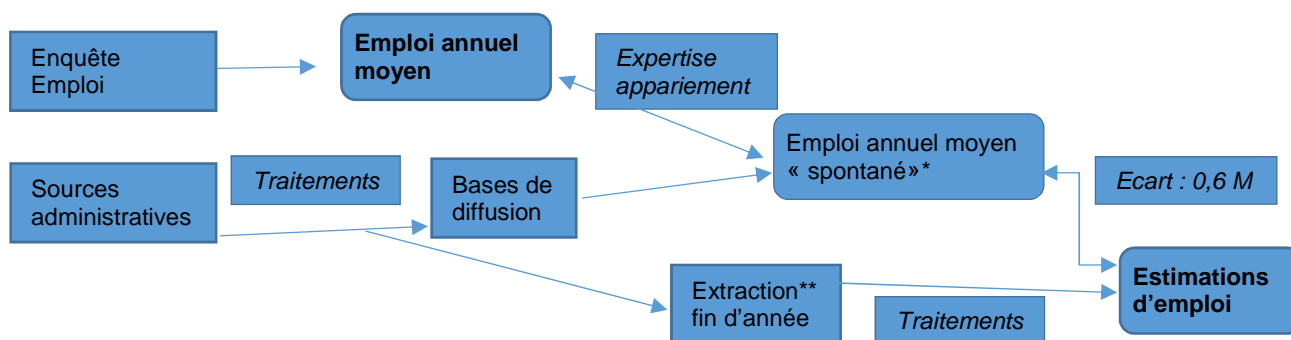
Tableau de synthèse (en milliers d'emplois)

En millions		2012	
Enquête Emploi	(a)	25 754	Moyenne sur l'année, France métropolitaine
Estimation emploi	(b)	26 453	Estimation fin d'année, France métropolitaine
Écart apparent à expliquer	(1) = (b) - (a)	699	
Différences de champ	(2)	-193	
Communautés		200	Source : Recensement de la population
Transfrontaliers		-393	Source : enquête Emploi
Différences de concept d'emploi	(3)	173	
Emploi rémunérés non BIT		173	Congés longue maladie notamment, estimé avec l'appariement
Traitements des sources administratives initiales	(4)	-600	Passage aux Estimations d'emploi + effet (fin vs moyenne annuelle)
Écart à expliquer (à champ et concepts communs)	(5) = (1)-(2)-(3)-(4)	1 319	
Défauts de couverture et trous de collecte des bases administratives initiales	(6)	-658	Source : appariement
Sous-déclaration (totale ou partielle) dans l'EEC	(7)	1 077	Source : appariement
Sous-pondération dans l'EEC	(8)	894	Source : appariement
Autres	(9) = (5) -(6)-(7)-(8)	6	Par solde

Champ : France métropolitaine

Notes : (a) il s'agit de l'estimation lors de la publication, chiffre révisé depuis compte tenu notamment de l'actualisation des pondérations calées sur le recensement.

Le travail présenté dans ce document compare l'emploi de l'enquête Emploi à celui des des bases administratives de diffusion.



* Cet emploi annuel moyen « spontané » est calculé pour les besoins de l'étude. Ce n'est pas une variable de diffusion.

** Pour les salariés, l'extraction se fait à partir d'une version des Dads « grand format » antérieure à celle diffusée

Pour rendre compte de l'écart entre l'enquête Emploi et les estimations d'emploi, il faut intégrer l'effet des traitements propres aux estimations d'emplois (correction de trous de collectes, de doublons, passage de données en moyenne annuelle à des données en fin d'année). Non expertisés dans le cadre de ce travail, ces traitements contribuent pour -0,6 million (ligne (4)), autrement dit le niveau des Estimations d'emploi en fin d'année est inférieur de 0,6 million à celui qui serait obtenu en calculant directement un emploi annuel moyen à partir des bases de diffusion.

C'est à partir de ces bases de diffusion que les appariements avec l'EEC ont été effectués. Les autres facteurs d'écart entre les deux sources peuvent alors être estimés à partir des travaux sur les appariements. Ils sont de trois types :

a) des défauts de couverture et trous de collecte des bases administratives initiales

Des défauts de couverture et trous de collecte sont identifiés notamment en retrouvant des personnes en emploi selon l'enquête Emploi mais absentes des bases administratives :

- dans des secteurs d'activité clairement mal couverts par les bases administratives (aides familiaux, artistes, petits agriculteurs et salariés de particuliers employeurs) ;
- ou présents dans les bases administratives des années adjacentes (2011 ou 2013)

Le chiffrage correspondant à la contribution de l'ensemble de ces défauts de couverture ou trous de collecte, sur l'écart entre emploi des sources administratives et l'EEC, est estimé à -0,7 million (ligne (6)).

b) une sous-déclaration de l'emploi dans l'EEC

L'appariement permet d'identifier des désaccords entre les deux sources sur le caractère d'« emploi ». Certaines personnes ne sont jamais en emploi selon l'EEC alors qu'elles sont en emploi selon les DADS au moins une des quatre semaines de référence; cela vient probablement d'une sous-déclaration (totale) des « petits boulots » dans l'EEC, particulièrement par les étudiants et les retraités ; l'impact est estimé à 0,8 million sur l'emploi annuel moyen.

Pour les personnes en emploi dans l'année selon les deux sources, des désaccords existent sur le volume d'emploi dans l'année : l'arbitrage de ces désaccords conduit à ajouter 0,3 million d'emploi annuel moyen (sous déclaration partielle). Au total, la sous-déclaration est estimée à 1,1 million d'emplois (ligne (7)).

c) une sous-pondération de l'emploi dans l'EEC

Les travaux illustrent une probable sous-pondération de certaines personnes en emploi dans l'EEC ; c'est notamment le cas des jeunes de 20 à 29 ans. L'effet sur l'emploi annuel moyen serait de 0,9 million (ligne (8)).

L'ensemble de ces chiffrages repose sur la qualité de l'appariement et l'hypothèse de représentativité des individus appariés au regard de la variable d'intérêt, ainsi que d'arbitrages entre les deux sources d'informations, propres à ce travail. Cela n'exclut pas une incertitude sur les différents chiffrages, qui doivent donc être interprétés avec précaution, ni l'existence de facteurs non distingués dans cette analyse.

Introduction

Le dispositif d'estimation d'emploi à partir des bases administratives, les Estimations d'emploi, livre un niveau d'emploi supérieur d'environ 0,7 million d'emplois, sur champs comparables, à celui de l'enquête Emploi en continu (EEC). Pour expertiser cet écart, un appariement entre l'EEC et les bases administratives (en amont des Estimations d'emploi), a été réalisé pour les années 2011, 2012 et 2013. Ce document décrit cet appariement et en tire les enseignements pour les estimations d'emplois.

Les sources administratives utilisées sont, pour les salariés, les DADS ; pour les non-salariés, les bases « non-salariés » issues de données collectées par l'Acoss et par la MSA. La notion de « DADS » utilisée couvre l'ensemble des salariés : ils regroupent, en plus des DADS stricto sensu (celles déclarées par les entreprises sous la forme de déclaration annuelle de données sociales), des données provenant de sources spécifiques, notamment pour la fonction publique d'Etat et les Particuliers employeurs. L'expertise de l'écart se fait, pour les salariés, à partir de la base DADS diffusée. Par contre la base DADS utilisée pour l'appariement proprement dit se situe en amont de la chaîne de traitement : des individus n'appartenant pas aux bases de diffusion (car les périodes les concernant ne sont pas considérées comme de l'emploi) ont pu être appariés.

L'expertise de l'écart conduite ici se limite à la comparaison des sources. Elle porte sur une mesure de l'emploi - l'emploi annuel moyen - différente de celle des Estimations d'emploi - l'emploi au en fin d'année - et ne tient pas compte des corrections apportées à ces sources par le processus de production des Estimations d'emploi. Ainsi les trous de collecte des bases administratives détectés dans ce document sont, dans la mesure du possible, corrigés dans le processus conduisant à l'Estimation de l'emploi.

Constats et hypothèses avant l'appariement

Les Estimations d'emploi ont longtemps été calées sur le recensement de la population (RP), avec un niveau d'emploi assez proche de celui de l'EEC. La substitution d'un recensement en continu (chaque année, mais non exhaustif) au recensement périodique (exhaustif) a remis en cause ce calage et a motivé l'utilisation des sources administratives pour déterminer le niveau d'emploi. Cela s'est traduit, sur le millésime 2007, par une hausse en niveau de l'ordre de 0,6 million de personnes en emploi.

Deux facteurs principaux étaient alors évoqués pour expliquer cette rupture :

- La faiblesse du processus précédent, procédant par application successive d'indices d'évolution à partir du niveau calé sur le RP 1999 : une dérive, impossible à quantifier en l'absence de nouveau RP exhaustif, est possible.
- Une sous-estimation de l'emploi dans le RP 1999, avec notamment une possible sous-déclaration des « petits boulots », qui pourrait aussi concerner l'EEC, en raison du protocole de collecte par interrogation directe des personnes.

En 2013, avant le présent appariement, plusieurs éléments pouvaient laisser suspecter un biais à la hausse dans les estimations d'emploi à partir des DADS.

- Dans la déclaration DADS émise par l'entreprise, la période déclarée par l'employeur peut résulter de la fusion de plusieurs périodes rémunérées discontinues, ce qui conduit à sur-estimer le décompte des salariés à un instant T. L'estimation de l'emploi au 31/12 est en principe préservée de ce biais¹ mais il s'avère que les dates de début et de fin de période ne sont pas toujours fiables.
- Un rapport entre niveau d'emploi en équivalent temps plein (ETP) et niveau en personnes physiques (PP) trop faible pour assurer la cohérence avec le taux de temps partiel et la quotité moyenne du temps partiel publiés à partir de l'EEC².

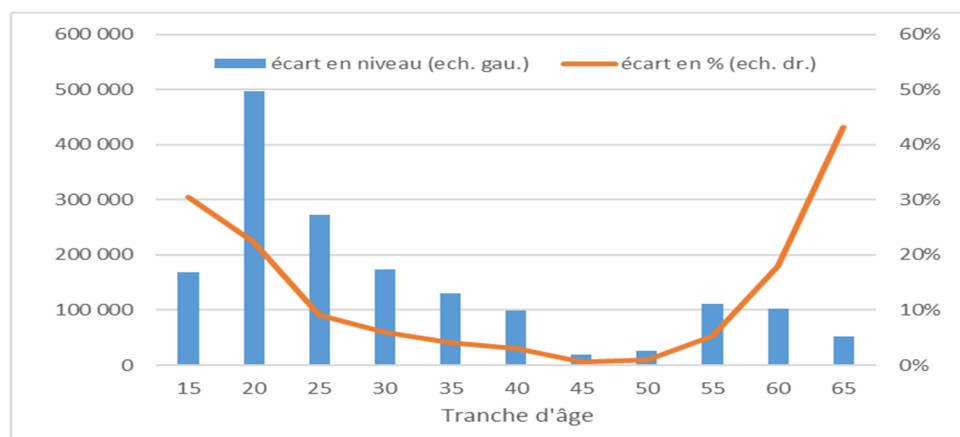
¹ Un salarié non rémunéré au 31/12 a, en principe, une période déclarée sur l'année n se terminant avant le 31/12. L'éventuelle période suivante, en n+1, se trouve dans la déclaration annuelle de l'année suivante et n'est donc pas fusionnée.

² L'emploi en PP et en ETP sont liés par la formule suivante : $ETP = (1 - p(1 - q)) PP$, avec p le taux d'emploi partiel et q la quotité moyenne des emplois à temps partiel. On sait que q est > 50% (58% selon l'EEC) et donc que $ETP/PP > 1 - p/2$. Les chiffres diffusés par la statistique d'entreprise à partir des DADS donnent un ratio $ETT/PP < 90\%$, ce qui implique un taux de temps partiel > 20%. Il n'est que de 14% sur le champ de la statistique d'entreprises en 2012 selon l'EEC.

Le constat d'une sur-extension des périodes pouvait alors étayer l'hypothèse que la faiblesse du ratio ETP/PP venait d'une sur-estimation du dénominateur, i.e. du niveau d'emploi estimé à partir des DADS. Ceci, combiné à la sous-déclaration des petits boulots dans l'EEC, permettait tenir un discours cohérent rendant compte de l'écart entre les deux niveaux d'emploi.

En 2012, une comparaison des effectifs par tranche d'âge indiquait que les écarts se concentraient aux extrêmes (graphique 1).

Graphique 1 : écart fin 2009 par tranche d'âge entre l'emploi selon les Estimations d'emploi et emploi selon l'EEC



Champ : métropole, hors frontaliers, fin 2009.

Sources : enquête Emploi, Estimations d'emploi

Lecture : fin 2009, il y a près de 500 000 personnes de 20 à 24 ans en emploi en moins, soit 20 %, dans l'EEC que dans les Estimations d'emploi

Dates de référence, champ, concepts d'emploi et éléments de comparaison

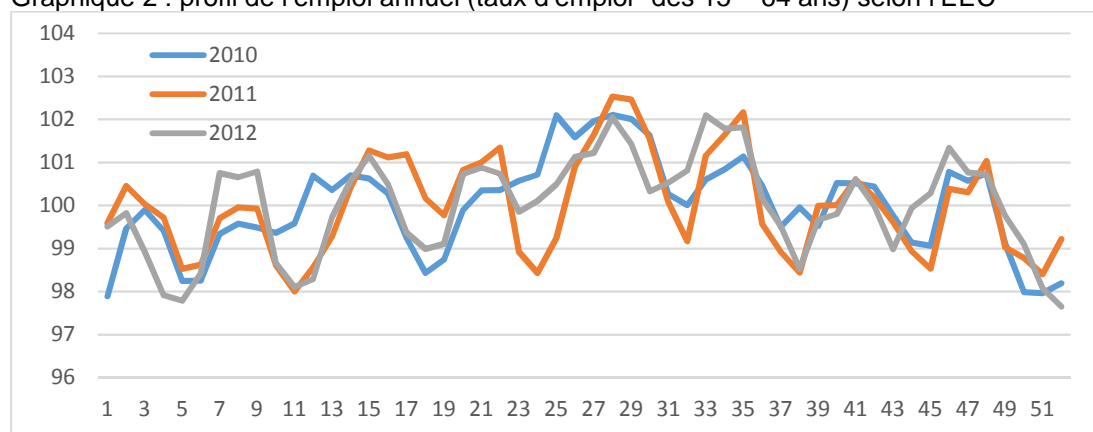
L'appariement ne permet pas de comparer directement l'emploi de l'EEC à celui des Estimations d'emploi. En effet ces dernières :

- Concernent l'emploi au 31/12 alors que l'EEC ne permet que de calculer un emploi annuel (ou trimestriel) moyen ;
- Résultent d'un processus de production et d'arbitrages en aval de la base DADS de diffusion et des bases non-salariés; il peut différer de ce qui peut être obtenu directement à partir de ces bases.

La date de mesure, le 31/12, ne saurait expliquer le surplus d'emploi des Estimations d'emploi par rapport à l'EEC : l'emploi la dernière semaine dans l'année est, d'après l'EEC, plutôt inférieur à l'emploi annuel moyen (graphique 2).

Le concept commun le plus pertinent, pour cette comparaison, est celui de **l'emploi annuel moyen au sens du BIT** ; en effet ramener les données de l'EEC à un concept d'emploi en fin d'année aurait considérablement réduit l'échantillon de cette dernière ; en outre l'EEC est calée et pondérée pour être utilisée en moyenne trimestrielle.

Graphique 2 : profil de l'emploi annuel (taux d'emploi* des 15 – 64 ans) selon l'EEC



Champ : France métropolitaine, personnes de 15 à 64 ans

Moyenne annuelle = 100. Données lissées (moyenne mobile d'ordre 3).

Lecture : en 2011, le taux d'emploi des 15 - 64 ans la dernière semaine de l'année (semaine 52) est, rapporté à la moyenne annuelle, de 99,2% (i.e. 0,8 point en dessous de la moyenne annuelle).

* le taux d'emploi est préféré à la mesure directe de l'emploi car l'EEC n'est pas faite pour une utilisation infra-trimestrielle, et a fortiori hebdomadaire, et un ratio est moins volatil qu'un volume.

Aucun chiffre d'emploi annuel moyen directement estimé à partir des sources administratives n'existe : dans le cadre de cette étude il faut donc le construire avant de le comparer à celui de l'EEC.

Le champ diffère sur deux points : les sources administratives mesurent l'emploi, salarié ou non salarié, rémunéré par les établissements situés en France, alors que l'EEC mesure l'emploi au sens du BIT des personnes habitant en France métropolitaine (jusqu'en 2012³), hors communautés. Il n'est pas possible de travailler sur un champ commun qui serait l'intersection des deux. En effet, si on peut isoler de l'EEC les frontaliers domiciliés en France travaillant à l'étranger (environ 0,4 million d'emplois), on ne peut pas faire de même, dans les bases administratives, pour les personnes en communauté (un peu plus de 0,2 million d'emplois, estimés à partir du Recensement de la population).

Le concept d'emploi diffère également : 173 000 emplois répertoriés dans les DADS sont, dans l'EEC, des situations de congés de longue durée. Ces situations sont exclues de la définition d'emploi au sens du Bureau international du travail. Comme il n'est pas possible de les isoler dans les DADS, on pourrait, pour comparer à concept équivalent, élargir dans l'EEC, le concept d'emploi aux congés de longue durée supérieurs à un an. Mais ces derniers, identifiés comme ceux ayant un emploi rémunéré mais n'ayant pas travaillé la semaine de référence sont pour moitié hors emploi des DADS. On ne peut donc pas, pour se ramener à un concept d'emploi plus proche de celui des DADS, les ajouter tous à l'emploi EEC : seuls les congés rémunérés aussi en emploi DADS seront considérés, pour les besoins de la comparaison, comme de l'emploi⁴.

Il s'agit de la seule différence de concept entre enquête Emploi et fichiers DADS de diffusion clairement identifiée grâce à l'appariement. On considère donc que, hors cette différence, tout emploi DADS est de l'emploi au sens du BIT, i.e. que les autres périodes DADS déclarées par les entreprises qui ne seraient pas de l'emploi au sens du BIT ont toutes été exclues au cours du processus de constitution des DADS.

Au total, les effets de champ et de concept se compensent globalement et apparaissent en cumulé relativement mineurs : ils ne permettent pas d'expliquer les 0,7 million d'écart.

³ L'extension de l'EEC aux DOM est postérieure.

⁴ L'enjeu ici est moins de déterminer un concept d'emploi un peu plus large que celui du BIT que de définir comment on traite les divergences entre sources : une semaine de référence où un salarié n'est pas en emploi dans les DADS sera traité comme une absence DADS s'il est en emploi EEC au sens du BIT, pas s'il est en congés longue durée.

Présentation de la démarche

L'appariement est présenté en **partie I**. Il porte sur des personnes, pas sur des postes : le champ est celui des personnes en emploi au moins une fois dans l'année (PEA).

Le concept d'emploi visé est celui d'emploi moyen dans l'année (EMA), seul concept comparable avec celui de l'EEC.

L'estimation des personnes en emploi est détaillée dans la **partie II**.

Le passage des PEA à l'EMA se fait par l'estimation de la durée moyenne en emploi dans l'année (DMA) des personnes en emploi : il fait l'objet de la **partie III**. DMA y est estimé par la moyenne de la présence en emploi sur les 4 semaines de référence (une par trimestre) de l'EEC, ce qui nécessite de restreindre l'échantillon aux répondants les 4 trimestres de l'année considérée.

D'où l'équation de base : $EMA = PEA * DMA$

Un avantage de cette procédure en deux étapes est de pouvoir s'affranchir dans la 1^{re} étape de l'incertitude sur les dates⁵ dans les bases administratives et de pouvoir ainsi pleinement bénéficier de leur avantage en termes d'exhaustivité (tout en utilisant l'EEC pour détecter des défauts de couverture ou des trous de collecte). Dans la 2^e étape, une fois le nombre de PEA établi en se calant surtout sur les bases administratives, et donc en s'affranchissant des problèmes liés aux pondérations de l'EEC, on peut confronter l'emploi des bases administratives à celui de l'EEC.

On dispose à l'issue de ce travail d'une base d'individus appariés avec, pour chacun, un poids et une durée moyenne en emploi dans l'année. À partir de cette base, la synthèse des enseignements de l'appariement est présentée en partie IV.

Des personnes en emploi selon les bases administratives mais jamais en emploi selon l'EEC

Alors que toute personne présente dans les bases administratives est, par définition, PEA, le champ de l'EEC couvre une population plus large, incluant les inactifs. En outre, certaines PEA peuvent ne pas être en emploi les semaines où elles sont enquêtées.

On appelle « JEMP » (pour « Jamais en EMPloi ») la population de l'EEC qui n'est pas en emploi lors des semaines de références. Cette population, définie empiriquement par « aucun trimestre en emploi au sens du BIT⁶ selon l'EEC » couvre conceptuellement trois cas, que l'on peut discerner dans une certaine mesure grâce à l'appariement :

- Ceux qui se seraient déclaré en emploi s'ils avaient été interrogés à d'autres semaines de référence : JEMP_C (C pour contingent, dépendance à la semaine de déclaration)
- Ceux qui sont en emploi (selon les bases administratives) au moins une des semaines de référence mais qui ne le déclarent pas dans l'EEC : JEMP_D (D pour déclaratif)
- Ceux qui ne sont véritablement jamais en emploi : JEMP_V (V pour « Vrai »).

L'appariement ne concerne, parmi les JEMP, que les deux premières catégories.

⁵ Au décalage de paie près : des personnes en emploi uniquement en décembre 2012 peuvent être présentes uniquement dans le fichier de validité 2013. L'impact est faible (cf. III.B.1).

⁶ C'est le critère retenu dans la partie II. Dans la partie III, une fois construit l'emploi DADS des semaines de référence, on réintègre dans les PEA2 les JEMP (au sens du II) en congés longue durée une semaine où ils sont en emploi selon les DADS.

Repondération et estimation des défauts d'exhaustivité des bases administratives

La stratégie générale consiste à calculer deux poids pour chaque individu apparié, l'un obtenu par calage de l'appariement sur les bases administratives, l'autre par calage sur l'EEC ; à comparer ensuite les effectifs par strate selon les deux pondérations.

Elle ne peut toutefois pas s'appliquer au nombre de personnes jamais en Emploi dans l'EEC (JEMP), qui ne peut être calé que sur les bases administratives. En effet, les JEMP_V, qui ne sont pas des personnes en emploi, constituent la majorité des JEMP de l'EEC et rien ne permettrait de les retrancher, à partir des seules données de l'EEC, pour caler sur les seuls JEMP_C et JEMP_D.

Pour les PEA selon les deux sources (PEA2) , qui peuvent, eux, être calés sur les 2 sources (après correction de l'attrition pour l'EEC, imputable à la restriction aux répondants les 4 trimestres de 2012), on constitue une typologie des PEA et la confrontation entre les deux calages permet de repérer un défaut de couverture des bases administratives pour certaines catégories (les aides familiaux, les agriculteurs, les aides familiaux et, dans une moindre mesure, les particuliers employeurs, cf. II.E).

À côté de ces défauts de couverture, ciblés sur certaines catégories, des trous de collecte plus ponctuels peuvent exister. Quelques précisions à propos de ces « trous de collecte », en ce qui concerne les DADS :

- La collecte des DADS concerne des postes (déclarés par des entreprises), pas des individus.
 - Cela affecte l'estimation de PEA uniquement si tous les postes d'un individu sont concernés (trou de collecte *total*) ;
 - Si seuls certains postes d'un individu sont concernés, il s'agit d'un trou de collecte *partiel* qui n'affecte pas l'estimation de PEA mais celle de DMA (traitée en III).
- On distingue deux types de trous de collecte :
 - Les trous de collecte proprement dits (absents des fichiers sources). On ne peut donc pas les détecter avec l'appariement 2012 mais on peut utiliser les appariements de 2011 et 2013 pour repérer des salariés... ;
 - ...présents dans les DADS et en emploi selon l'EEC une de ces deux années...
 - ...en emploi selon l'EEC en 2012...
 - ...absents des DADS 2012.
 - Des « trous de collecte » induits par les chaînes de traitement. Ces individus ont pu être appariés en 2012 (les bases administratives utilisées sont en amont des traitements de suppression des périodes considérées comme ne relevant pas de l'emploi) mais sont absents des DADS de diffusion.

On ne peut ajouter les trous de collecte à l'estimation des PEA que pour les catégories où on a choisi le calage sur les bases administratives (i.e. les catégories sans défaut de couverture détecté). En effet, s'il est légitime de compléter la source de calage, en l'occurrence les bases administratives, par ses défauts d'exhaustivité détectés, il ne le serait pas de compléter un calage sur une source, en l'occurrence l'EEC, déjà calée sur une source couvrant l'ensemble de la population, en l'occurrence le recensement.

Au total, le calage est, soit sur l'EEC, soit sur les bases administratives, selon les cas suivants :

Cas	Catégories où défaut de couverture des bases administratives (BA)	Autres catégories
PEA2	Calage EEC	Calage BA
JEMP	Calage BA	Calage BA
Trous de collecte bases administratives (détectés sur appariement amont ou autre année)	Non retenus	Calage EEC

Principales typologies utilisées

La comparaison doit tenir compte des différents cas de figure : il n'est par exemple pas possible de comparer de la même manière l'emploi des salariés, où les périodes d'emploi sont datées, et l'emploi des non-salariés, où l'on ne dispose que d'un indicateur, construit, d'activité au 31/12.

Parmi les différentes typologies utilisées, deux sont centrales dans l'élaboration des résultats, et leurs catégories seront souvent évoquées par leur acronyme.

Première typologie, celle des personnes en emploi (PEA) selon leur présence dans les sources :

- PEA2 : présent en emploi à la fois dans les bases administratives et l'EEC
- JEMP : présent en emploi dans les bases administratives mais jamais en emploi les semaines de référence dans l'EEC
- TCOL : présent (puisque apparié) dans les sources administratives en 2011, 2012 ou 2013 mais absent (en tant qu'emploi⁷) des sources diffusées en 2012

Deuxième typologie des PEA selon leur statut salarié ou non-salarié (selon les sources administratives) sur l'ensemble de l'année

- SAL : salarié seulement ou « pur » (pas d'activité non salariée répertoriée)
- PSA : principalement salarié (revenus salariaux de l'année > revenus d'activité non-salariée)
- PNS : principalement non-salarié (revenus salariaux < revenus d'activité non-salariée)
- NSA : non-salarié « pur » (pas d'activité salariée répertoriée)

Enfin, une typologie élaborée *ex post*, distingue les catégories qui, d'après l'appariement, sont mieux couvertes par l'EEC que par les bases administratives :

- NAA : non-salariés : agriculteurs, aides familiaux et artistes (calé sur l'EEC)
- NZZ : autres non-salariés (calé sur les bases « non-salariés »)
- PE : salariés des particuliers employeurs (calé sur l'EEC)
- SalhPe : autres salariés (calé sur les DADS)

Le choix du calage n'est possible que pour les PEA2. Par exemple, les relativement nombreux PE sans emploi déclaré dans l'EEC (JEMP) sont calés sur les DADS.

Une stratégie adaptée aux différents cas

Seuls les salariés « purs » (SAL), qui recouvrent 88 % des PEA, permettent une confrontation entre DADS et EEC pour l'estimation de la durée moyenne en emploi dans l'année (DMA, estimée par la moyenne de la présence en emploi les 4 semaines de référence).

Cela nécessite un travail sur les dates dans les DADS et l'élaboration d'indicateurs permettant d'arbitrer en cas de désaccord entre les deux sources sur l'emploi d'une personne donnée une semaine de référence donnée.

Pour les pluriactifs (désignant ici le cumul dans l'année, pas forcément au même moment, d'une activité salariée et d'une activité non-salariée), le DMA de l'EEC est, en moyenne, supérieur à celui des DADS. On prendra donc, sauf pour les JEMP, le DMA de l'EEC⁸.

Enfin, pour les NSA, seule l'EEC permet d'estimer un DMA (dans les bases « non-salariés », hors création ou cessation, un non-salarié est censé exercer son activité toute l'année). Pour les JEMP, on prend comme proxy de l'emploi annuel moyen (au niveau agrégé) l'activité au 31/12. En conséquence, les actifs au 31/12 sont pris comme s'ils étaient présents toute l'année; les autres sont pris comme s'ils étaient inactifs toute l'année et éliminés *ex post* des PEA. Ce choix est sans biais sous l'hypothèse que, pour ceux ayant une activité en 2012, la probabilité d'être en emploi à un moment donné de l'année est identique à celle d'être en activité au 31/12.

⁷ Ce qui nécessite d'être contrôlé, la présence dans les DADS peut être au titre de période de non-emploi comme le chômage indemnisé.

⁸ Construire un indicateur de présence en emploi synthétisant la présence dans les DADS et celle dans les bases « non-salariés » serait peu aisé, compte tenu du peu d'informations à ce sujet dans les bases « non-salariés » (suspicion d'activité au 31/12 ; date de création ou de cessation) ; de plus, PNS et PSA ne représentent que 3 % des PEA.

I - Description de l'appariement

A - Appariement avec les DADS

1 - Fichiers en entrée

Le fichier DADS 2012 utilisé comporte 72,5 millions d'observations portant sur 53,6 millions de postes (soit des couples (individu, SIRET)) différents et 32,1 millions d'individus différents (y compris ceux, sans emploi, au titre de leur indemnisation par Pôle emploi).

Il y a dans ce fichier constitué en amont des chaînes de traitement DADS plus d'individus et de postes que dans le fichier postes de diffusion. Ce point est important, à la fois pour l'explication de certaines incohérences (des non-salariés déclarés monoactifs dans les bases « non-salariés » suite à l'appariement avec le fichier postes peuvent, ici, être appariés au même individu EEC que des salariés DADS) et pour le diagnostic sur l'exhaustivité des DADS diffusées.

Le fichier EEC comporte 369 000 observations (couples (individu, trimestre)) portant sur 168 000⁹ individus de 15 ans ou plus différents.

Le champ de l'appariement porte sur l'ensemble des DADS (salariés et chômeurs indemnisés) mais pas sur l'ensemble de l'EEC, même si l'ensemble des observations de l'EEC sont utilisées : les inactifs ne sont pas censés être appariés, même si certains le sont en raison de la sous-déclaration de l'emploi dans l'EEC. À titre indicatif, 79 000 individus EEC sont au moins une fois salariés, 77 000 en excluant les frontaliers.

2 - Choix des identifiants

Faute de numéro identifiant dans l'enquête Emploi, l'appariement entre les DADS et l'EEC est plus difficile qu'un appariement entre sources administratives - où on peut apparier en général avec un numéro identifiant comme le NIR.

Pour constituer des identifiants d'appariement moins forts que le NIR, on dispose de deux blocs d'information :

- Un bloc « état-civil » :
 - Sexe
 - Prénom
 - Date de naissance
 - Département de naissance
- Un bloc de résidence
 - Commune de résidence
 - N° de voie

⁹ Il y a en fait 172 000 personnes d'au moins 15 ans dans l'EEC 2012. Une petite partie des personnes inactives de plus de 65 ans et plus n'ont pas été incluses dans le fichier produit spécifiquement pour l'appariement.

On teste un premier jeu d'identifiants « Id(i) » construits de la manière suivante :

Variables communes*	Date de naissance	N° de voie	Prénom		
			6 Premières lettres	2 Premières lettres	6 Dernières lettres
X	Complète		Id3		Id6
X	Complète	X	Id17	Id18	
X	Mois, année		Id8		

* Sexe, département de naissance, commune de résidence

Un premier critère de sélection des identifiants est la part d'identifiants uniques, i.e. qui ne renvoient qu'à un seul individu. Cette part est d'autant plus élevée que l'identifiant est précis : il vaut 99,5% pour Id3, contre 97,7 % pour Id8. Ce dernier identifiant n'est pas retenu.

Mais un identifiant comme Id8, en relâchant la contrainte de précision sur la date de naissance, permettait d'améliorer l'appariement des immigrés (département de naissance = '99') pour lesquels la date de naissance est moins souvent complète (ou, si elle est complète, commence plus souvent par « 01 »). On crée donc un nouvel identifiant, id19, en relâchant la contrainte sur la date de naissance du plus complet des 4 identifiants retenus (id17). Enfin, pour pallier la baisse des taux d'identifiant unique pour les prénoms les plus fréquents ou pour les prénoms dont les 6 premières lettres sont communes avec un autre prénom fréquent (ex. CHRISTophe et CHRISTian ou CLEMENce et CLEMENTine), on ajoute des identifiants utilisant les 10 premiers caractères du prénom.

On a finalement le choix entre les identifiants suivants :

Variables communes*	Date de naissance	N° de voie	Prénom			
			10 Premières lettres	6 Premières lettres	2 Premières lettres	6 Dernières lettres
X	Complète		IdB**	IdB6P		B6D
X	Complète	X	IdA	IdA6P	IdA2P	
X	Mois, année	X	IdC	IdC6P		

* Sexe, département de naissance, commune de résidence

** Le nom des identifiants est composé d'une lettre indiquant l'information mobilisée pour la date de naissance et l'adresse (A, B, C par ordre décroissant de qualité) et d'une indication sur l'information mobilisée sur le prénom (laissé à blanc pour l'information la plus complète, 10P).

L'identifiant le plus complet est IdA. On allège les contraintes sur le prénom en allant sur la droite, les contraintes sur l'adresse en allant vers le haut et les contraintes sur la date de naissance en allant vers le bas.

Les identifiants n'utilisent pas les informations liées au travail. Il peut cependant y avoir plusieurs identifiants par individu, par exemple suite à un déménagement ou à des différences d'orthographe sur le prénom. Il y a, selon les variables utilisées, entre 35 et 36 millions d'identifiants DADS différents (cf. tableau 1).

Les identifiants IdA6P et IdA2P utilisent le numéro de l'adresse et sont donc plus précis (moins de doublons) que les identifiants IdB6P et IdB6D. Par contre, ils appartiennent moins de monde car ce numéro est assez souvent non renseigné dans les DADS : 18,6 % des identifiants IdA6P sont à information incomplète (une ou plusieurs variables constituant l'identifiant sont non renseignées ou non valides). Les taux d'identifiant non unique du tableau 1 sont calculés uniquement sur les identifiants à information complète et sur la France métropolitaine.

Tableau 1 - Principaux indicateurs pour 4 des 8 identifiants

	Unicité des identifiants				Tx appariement*			
	IdB6P	IdB	IdA6P	IdC6P	B6P	B	A6P	C6P
Nombre d'identifiants différents (en milliers)	35 097	36 096	35 826	35 439				
% id. à information incomplète	3,6	3,6	18,6	18,6				
Taux d'identifiant non unique en fonction de la taille de la commune de résidence (nés en France)								
Moins de 10 000 identifiants DADS	0,27	0,25	0,23	0,27	24	22	22	22
De 10 000 à 100 000	0,41	0,36	0,22	0,32	20	19	18	19
Plus de 100 000	0,62	0,58	0,24	0,44	20	19	18	19
Taux d'identifiant non unique en fonction de la taille de la commune de résidence (immigrés)								
Moins de 10 000 identifiants DADS	1,82	1,68	1,60	1,71	18	17	17	18
De 10 000 à 100 000	2,36	2,18	1,92	2,08	15	14	14	14
Plus de 100 000	2,81	2,59	2,12	2,32	15	14	14	15

* en pour 10 000

Lecture : Il y a 35,1 millions d'identifiants idB6P différents, dont 3,6 % à information incomplète. Dans les communes à moins de 10 000 identifiants différents et parmi les identifiants à information complète, 0,27 % des identifiants de personnes nées en France renvoient à plusieurs individus (taux d'identifiant non unique) contre 1,82 % des identifiants de personnes immigrées. Pour 10 000 identifiants renseignés et uniques, toujours dans les communes à moins de 10 000 identifiants, et pour les personnes nées en France, il y a 24 personnes appariées à l'EEC.

Le taux d'identifiant non unique est nettement plus élevé pour les immigrés. Il croît aussi avec la taille de la commune de résidence. Il dépend aussi de la fréquence de la chaîne de caractère retenue pour le prénom. On construit un indicateur de risque de confusion, appelé « type de prénom » en combinant la chaîne de caractère retenue (ex CHRIST), le sexe (CHRISTian n'est pas confondu avec CHRISTine) et le fait d'être né ou non en France. On vérifie qu'avoir un « type » de prénom fréquent (au moins 200 000 occurrences dans la base des identifiants) augmente le risque d'identifiant non unique (tableau 2). Ce risque est plus sensible dans les grandes communes où la probabilité de rencontrer plusieurs occurrences du même type de prénom augmente. Le recours au numéro de la voie supprime quasiment le risque (colonnes pour l'identifiant A6P).

Tableau 2 - Impact du type de prénom sur le taux d'identifiant non unique

Taille de la commune (en milliers d'identifiants)	B6P			B			A6P			C6P		
	0-10	<100	100+	0-10	<100	100+	0-10	<100	100+	0-10	<100	100+
Type de prénom non fréquent	0,26	0,36	0,53	0,24	0,33	0,50	0,23	0,23	0,25	0,26	0,31	0,40
Type de prénom fréquent, seul	0,28	0,55	0,99	0,26	0,49	0,94	0,23	0,19	0,21	0,27	0,36	0,60

Lecture : Parmi les B6P à information complète des petites communes, le taux d'identifiant non unique (i.e. renvoyant à plusieurs individus) est de 0,28% quand le type de prénom est fréquent (plus de 200 000 occurrences dans la base des identifiants) et de 0,26% sinon

On ne retient, pour chacun des huit identifiants, que les individus appariés de manière unique (ou, ce qui revient au même, que les individus à identifiant unique dans chacune des deux sources). On élimine aussi de l'appariement final les individus appariés différemment selon l'identifiant (un individu A apparié à B avec l'identifiant x et apparié à C différent de B avec l'identifiant y est éliminé). On a au total 77 156 individus appariés, par l'un d'au moins des huit identifiants retenus, dont 74 806 avec au moins une période d'emploi DADS (pour les les 77 156-74 804 = 2 352 restants, certains ne sont dans les DADS qu'au titre de l'indemnisation chômage, d'autres ont été supprimés des DADS de diffusion par les chaînes de traitement car considérées comme enregistrés pour des périodes de non-emploi).

Parmi ces 74 804, 61 978 sont au moins une fois salarié dans l'EEC. Rapporté aux 76 803 répondants de l'enquête Emploi déclarant au moins un emploi salarié en France en 2012, le taux d'appariement est ainsi de 81 %. Pour les 19 % d'individus de l'EEC non appariés, il s'agit, pour la plupart, des défauts d'appariement : ils sont sans doute présents dans les deux bases ; mais certains sont absents des DADS. Ainsi, parmi les 2 352 appariés sans période d'emploi dans les DADS de diffusion, 227 sont

salariés dans l'EEC. Enfin, le concept de salarié de l'EEC étant différent de celui des bases administratives¹⁰, certains peuvent être appariés avec les bases non-salariés.

3 - Quel critère de qualité des appariements ?

L'appariement ne porte que sur les identifiants uniques. Si le champ des individus appariables de l'EEC était strictement inclus dans celui des DADS - ce qui serait approximativement le cas si on ne retenait pour l'appariement que les personnes en emploi salarié selon l'EEC, cela suffirait à garantir la qualité de l'appariement puisque les DADS sont quasi exhaustives sur ce champ. Mais on ne peut pas opérer une telle restriction : un des apports de l'appariement est justement de montrer que les deux sources diffèrent quant aux mesures de l'emploi. C'est d'ailleurs pour cela que les informations relatives au travail (commune de travail, secteur, SIRET) ne font pas partie des variables constitutives des identifiants. Un identifiant unique sur le champ DADS peut donc ne pas être unique sur le champ EEC. Cette possibilité est d'autant plus forte que la part de salariés dans la population considérée est faible (tranche d'âge notamment).

Disposer d'un indicateur de qualité est nécessaire pour apprécier les cas de discordance sur la situation d'emploi entre les DADS et l'EEC : certains d'entre eux pourraient en fait résulter d'une erreur d'appariement.

On propose dans ce chapitre plusieurs indices permettant de s'assurer de la qualité de l'appariement :

- cohérence du taux d'appariement non unique avec le degré de précision de l'indicateur ;
- cohérence sur le SIRET, le statut (salarié ou non) et l'emploi selon l'EEC ;
- cohérence entre l'appariement sur 2012 et les appariements sur 2011 et 2013 ;
- restriction nécessaire aux personnes suivies dès la première interrogation.

Le taux d'appariement non unique est bien cohérent avec le degré de précision de l'indicateur

On s'attend à ce que plus un indicateur est précis, plus l'appariement réalisé à partir de cet indicateur est fiable, le prix à payer étant un plus faible taux d'appariement. Si, pour tester cette hypothèse, on réalise l'appariement de manière asymétrique, en ne prenant des DADS que les identifiants uniques et en gardant tous les identifiants EEC, on obtient un nombre très limité d'individus DADS appariés à plusieurs individus EEC. La part, parmi les appariés, de ces appariés multiples, suit exactement le degré de précision de l'identifiant, minimale au centre et à gauche (date de naissance complète et prénom sur 10 caractères) (tableau 3).

Tableau 3 - Taux d'appariement non unique (en un pour 10 000)

Variables communes	Date de naissance	N° de voie	Prénom				
				10P	6P	2P	6D
X	Complète		B	7,14	7,42		7,09
X	Complète	X	A	2,98	3,21	3,60	
X	Mois, année	X	C	3,25	3,46		

Lecture : parmi les individus dont l'identifiant avec la date de naissance complète et les 10 premiers caractères du prénom est unique dans les DADS, pour 10 000 individus appariés, 2,98 sont appariés à plusieurs individus EEC.

Cohérence sur le SIRET, le statut (salarié ou non) et l'emploi selon l'EEC

Puisque le SIRET n'est pas utilisé dans la constitution des identifiants, on peut l'utiliser *ex post* comme critère de qualité de l'appariement.

Il ne s'agit pas de rejeter les appariements où le SIRET diffère, ce qui reviendrait d'ailleurs à utiliser le SIRET comme clé d'appariement, car d'une part le SIRET n'est pas toujours renseigné dans l'EEC et d'autre part une divergence de SIRET est toujours plausible (défaut de localisation dans les DADS, possibilité de désigner un même poste par plusieurs types de SIREN, par exemple, mais pas seulement, pour l'Intérim).

¹⁰ Par exemple, les dirigeants-salariés sont salariés selon les bases administratives (ils cotisent au régime général) et indépendants selon l'EEC (selon la logique de la PCS).

Quand on compare les identifiants selon ce critère, il ne semble pas déterminant au premier abord : il y a toujours, quelque soit l'identifiant utilisé, 2/3 de SIRET renseignés dans l'EEC et, parmi ceux à SIRET renseignés, 1/4 où aucun SIRET DADS n'est retrouvé parmi les au plus 4 (un par trimestre) SIRET EEC. Ceci n'est pas très surprenant car les appariements selon les différents identifiants ont en commun de nombreux couples (DADS, EEC). Il vaut mieux donc raisonner en différentiel.

On examine, à chaque étape, les nouveaux individus appariés grâce à l'introduction d'un nouvel identifiant. On commence par l'appariement le plus précis, celui avec IdA. Il faut aussi distinguer, parmi ceux qui ne sont pas appariés avec idA, ceux qui ne le sont pas parce que le numéro de voie n'est pas renseigné dans l'adresse DADS de ceux avec adresse DADS renseigné. L'appariement avec un autre identifiant est plus douteux dans ce dernier cas.

Tableau 4 - Cohérence avec l'EEC en fonction du type d'identifiant retenu (en %)

		Même siret (a)	autre siret (b)	Non salarié ou siret non renseigné	hors emploi	Total	%	a/(a+b)
Numéro adresse DADS renseigné								
1	idA	53	15	17	15	52 349	70	78
2	Autres identifiants avec voie	40	12	29	19	5 739	8	76
Identifiants sans voie								
3	voie EEC non numérique	52	13	19	16	1 936	3	80
4	voie EEC numérique	48	14	19	18	3 864	5	77
Num. adresse DADS non renseigné								
5	IdB	56	14	17	13	10 466	14	80
6	IdB6Pou IdB6D	35	12	34	20	452	1	75
Total		52	15	19	15	74 806	100	78

Lecture : parmi les 5 739 individus dont le numéro de l'adresse DADS est renseigné, non apparié en utilisant l'idA mais apparié en utilisant un autre identifiant comportant le numéro de la voie (idA6P, id2P, idC ou idC6P), 40 % ont un SIRET renseigné dans l'EEC qui se retrouve aussi dans au moins une des périodes DADS.

Le numéro de la voie est renseigné pour 85 % des individus appariés. L'appariement est alors nettement meilleur s'il est fait sur l'identifiant le plus précis, IdA (tableau 4). De même, quand le numéro de la voie n'est pas renseigné, l'appariement est meilleur sur l'identifiant le plus précis, IdB et la cohérence avec l'EEC, sur la situation d'emploi comme sur le SIRET, est comparable avec le cas avec numéro d'adresse (alors que le taux d'appariement unique est nettement moins bon, cf. supra).

Les cas où la cohérence est la moins bonne sont ceux où l'appariement est rejeté avec l'identifiant le plus précis mais accepté avec un autre identifiant possédant le même niveau d'information (A6P, A2P, C ou C6P dans le cas où le numéro est renseigné). Les cas où l'appariement a lieu sur des identifiants sans numéro alors que le numéro est renseigné sont dans une situation intermédiaire, avec une cohérence un peu meilleure, comme attendu, quand le non- appariement avec un identifiant à numéro vient de l'absence de numéro dans l'adresse récupérée dans l'EEC. On note que le taux de SIRET cohérent parmi les individus à SIRET renseigné dans l'EEC - dernière colonne du tableau - varie moins entre ces différents cas que le taux de hors emploi ou de SIRET non renseigné (si la personne s'est déclarée non-salariée, son SIRET est ici considéré comme non renseigné). Cela peut s'interpréter comme une conséquence de la quasi-exhaustivité des DADS sur le champ des salariés : un identifiant unique a peu de risque d'être apparié à deux salariés différents ; par contre un identifiant unique sur le champ des salariés peut ne pas l'être sur le champ complet.

On a ainsi trois niveaux de qualité :

1. Appariement avec des identifiants à date de naissance complète et sur les 10 premiers caractères du prénom : IdA, quand le numéro de voie est renseigné, IdB sinon (lignes 1 et 5 du tableau précédent)
2. Appariement avec des identifiants sans numéro de voie, alors que ce dernier est renseigné dans les DADS (lignes 3 et 4)
3. Appariement avec un identifiant avec numéro de voie, alors que non apparié avec idA (ligne 2) ou avec un identifiant sans numéro de voie, alors que non apparié avec idB (ligne 6)

Le premier niveau de qualité regroupe 84 % des appariés. Au sein de ce bloc, le critère de cohérence appliqué aux différents indicateurs (taille de la commune, fréquence du prénom, naissance en France) montre que la cohérence se dégrade pour les grosses communes (plus de 100 000 identifiants DADS) et les immigrés (tableau 5). On isole ces cas pour mieux cerner la catégorie des appariés à qualité la plus assurée.

Tableau 5 - Ventilation des appariés par niveau de qualité

	Même siret (a)	autre siret (b)	Non sal. ou siret nr	hors emploi	Total	%	a/(a+b)	
IdA (num renseigné) ou IdB (non renseigné)								
« petite » commune et né en Fr.	1A	55	15	17	14	52 951	71	79
Grosse commune ou immigré	1B	45	17	22	16	9 864	13	73
Num renseigné mais non utilisé	2	50	14	19	17	5 800	8	78
Num, hors IdA ou non num, hors IdB	3	39	12	29	19	6 191	8	76
Moyenne		52	15	19	15	74 806	100	78

Au total, parmi les salariés, le taux de SIRET cohérent parmi les individus à SIRET renseigné dans l'EEC est de 78 %.

Cohérence entre l'appariement sur 2012 et les appariements sur 2011 et 2013

L'appariement sur 2011 et 2013 fournit un test supplémentaire de la qualité de l'appariement. Parmi les 64 000 appariés 2012 qui sont aussi appariés en 2011 et/ou en 2013, 166, soit 0,26 %, sont appariés avec un salarié ayant un identifiant EEC différent de celui de 2012 (tableau 6). L'effet indicateur de qualité de l'identifiant semble assez faible, sauf pour les identifiants de type 2, où le taux d'incohérence passe à 0,60 %. Mais il se peut que l'incohérence provienne de l'appariement 2011 ou 2013. Les effectifs sont trop faibles, eu égard au faible nombre d'incohérence, pour croiser les effets d'indicateur de chaque année. Seul l'identifiant de type 1 le permet : quand l'identifiant est de type 1 aussi bien en 2013 qu'en 2012, le taux d'incohérence descend à 0,15%.

Tableau 6 : cohérence entre les appariements DADS 2011, 2012 et 2013

Indicateur qualité 2012	Nb appariés en 2011 et/ou 2013	Sur un identifiant EEC différent de celui de 2012	%
1A	45185	103	0,23
1B	8478	21	0,25
2	4525	27	0,60
3	5335	15	0,28
Total	63523	166	0,26

Champ : appariés DADS 2012

À noter qu'il n'est théoriquement pas impossible que, suite à un déménagement, une même personne soit interrogée 2 années de suite par l'EEC (avec alors un identifiant différent). L'approche symétrique (même identifiant EEC apparié deux années de suite) donne un taux moyen d'incohérence (entre les identifiants DADS, stables lors des déménagements) proche : 0,23 %.

Restriction nécessaire aux personnes suivies dès la première interrogation

L'enquête Emploi est un panel glissant sur 6 trimestres soumis, comme tous les panels, à l'attrition. Notamment, s'agissant d'un panel de logements, les personnes ayant déménagé quittent le panel et sont remplacées par les nouveaux entrants. Ainsi, parmi les 172 000 personnes d'au moins 15 ans répondant au moins une fois en 2012, 30 000 sont entrées après la première vague d'interrogation (elles sont repérées par un rang d'interrogation, RGI, inférieur au rang de la vague, RGA).

Si les personnes ajoutées ont le même profil que les personnes parties par attrition, ce qui permet d'avoir un échantillon non biaisé en coupe, elles ont un profil différent de celles restant tout au long du panel. Or la personne présente tout au long de l'année est comptée pour 1 dans l'appariement alors que la personne partie et son remplaçant dans le panel sont comptées pour 2 personnes différentes.

Un exemple pour illustrer ce biais :

- parmi l'ensemble des personnes appariées, il y a deux fois plus de personnes présentes au titre de la première vague d'interrogation (RGA=1) que de personnes entrées après la première vague (RGI < RGA).
- Inversement, parmi les personnes appariées ayant plus de 5 contrats CDD ou intérim dans l'année, il y a plus d'ajoutés (RGI<RGA) que de RGA = 1.

Cela biaise à la hausse le taux d'appariement des personnes plus mobiles (en fait, cela compense, et donc masque, la sous-représentation de cette population dans l'EEC).

Il vaut donc mieux, pour analyser le taux d'appariement du point de vue des DADS, se limiter à la première vague d'interrogation. Une fois contrôlé de l'âge (les jeunes sont sous représentés dans l'EEC), du sexe, de la naissance à l'étranger et de la taille de la commune, le nombre d'heures cumulées dans l'année ressort (plus que le nombre de contrats ; la distinction CDD/CDI ne ressort pas) : avoir moins de 300 heures diminue d'1/5^e la probabilité d'être apparié.

B - Description de l'appariement avec les bases non-salariés

La base « non-salariés » est en fait constituée de deux bases :

- Une base Acooss où l'on dispose de la dénomination (nom, prénom, métier) dont il faut extraire le prénom : 3 195 960 observations ;
- Une base MSA où l'on ne dispose pas du prénom : 473 809 observations.

Pour la localisation, on ne dispose, dans les deux bases, que de la commune de travail. Il n'est donc pas possible de construire un identifiant indépendant de l'activité : contrairement à l'appariement DADS, le repérage de la sous-déclaration EEC n'est pas possible. On en récupérera quand même certains en prenant, dans l'EEC, en cas de commune de travail non renseignée, la commune ou le département de résidence : 2 non-salariés sur 3 travaillent dans leur commune de résidence et 4 sur 5 dans leur département de résidence.

Autre difficulté : la population des non-salariés est relativement restreinte et l'unicité d'un identifiant offre moins de garantie quant à sa qualité que pour la base DADS. Ainsi, « Ac0 », avec le département de travail (cf. tableau 7), est un identifiant unique du fichier Acooss à 99,59% et s'est pourtant avéré de piètre qualité.

Ces difficultés conduisent à privilégier le SIREN¹¹. Les identifiants construits avec le SIREN ne conduisent pratiquement jamais à des appariements avec des salariés (selon l'EEC)¹².

Les identifiants testés sont décrits dans le tableau 7. La date de naissance est incluse dans tous les identifiants. Le sexe aussi, sauf pour l'identifiant Ac5 où le croisement SIREN et date de naissance semble suffisamment précis pour apparier, sans le prénom, les observations dont le sexe n'est pas renseigné. On ne dispose pas d'une variable spécifique pour le prénom. La variable « dénomination » commence par le nom puis le prénom. On retient pour le prénom le deuxième mot de la dénomination,

¹¹ On prend le SIREN personnel du non-salarié, mieux retrouvé dans l'EEC que celui de la variable SIREN_final ; le SIREN de la société gérée, SIREN_SG, est toutefois utilisé en complément : c'est celui que l'on s'attendrait à avoir dans l'EEC pour les gérants

¹² Ce n'est pas le cas avec un identifiant réduit au seul SIREN puisque les non-salariés peuvent être employeurs.

sauf lorsque cette dernière commence par 'DE', 'DOS' ou 'LE', auquel cas on prend le troisième. On teste aussi le mot suivant, au cas où le nom serait composé de 2 mots.

Tableau 7 : Composition des identifiants testés pour la base « non-salariés » Aross

	Pr10P	Pr6P	Pr2P	PrSuiv6P*	Aucun Pr
Commune	Ac9	Ac1	Ac3	Ac7	
Département + SIREN			Ac4		Ac5
Département + SIREN_SG			Ac10		Ac11
Commune + APE (3 carac.)		Ac12			Ac6
Département	Ac8	(Ac0)		Ac2	

* 6 premiers caractères du mot suivant le mot habituellement retenu comme prénom.

On apprécie l'appariement par rapport à la situation dans l'EEC des appariés : on s'attend à ce que les appariés soient non-salariés dans l'EEC. Ce critère n'est toutefois pas absolu. Des individus des bases non-salariés peuvent être appariés (hors erreur d'appariement) à des individus qui ne sont pas non-salariés selon l'EEC par suite :

1. de la différence de concept dans le partage salariat / non-salariat entre sources administratives (basé sur le mode de rémunération) et EEC (basé sur l'autonomie)
2. de la multiactivité salarié/ non-salarié, multi activité qui peut être sous déclarée dans l'EEC
3. de la non-déclaration d'une activité jugée marginale (pour ceux qui sont hors emploi).

On construit une variable de situation par rapport au non-salariat dans l'EEC avec les modalités suivantes :

1. Non salarié à la fois selon la déclaration et selon la codification EEC du statut
2. Non salarié selon la déclaration mais recodifié salarié (harmonisation avec la PCS)
3. Déclaré salarié et recodifié en non-salarié
4. Salarié avec activité secondaire de non-salarié
5. Salariés purs
6. Hors emploi

Pour les personnes enquêtées plusieurs fois dans l'année, on retient la situation la plus proche du non-salariat (la proximité suivant l'ordre des modalités indiquées supra)

Tableau 8 - Ventilation, pour chaque identifiant Aross, des appariés selon leur situation dans l'EEC

	Non salarié	Déclaré non-salarié	Reclassé non-salarié	Salariés multi actifs NS	Salariés purs	Hors emploi	Nb appariés
Ac1	78,23	3,33	0,66	1,22	7,20	9,35	5 582
Ac2	48,91	1,09	0,00	0,87	32,10	17,03	458
Ac3	77,73	3,32	0,68	1,22	7,38	9,66	6 053
Ac4	95,54	3,84	0,17	0,00	0,45	0,00	2 891
Ac5	95,63	3,69	0,16	0,00	0,52	0,00	3 091
Ac6	94,42	3,07	0,38	0,15	1,98	0,00	3 385
Ac7	80,58	1,65	0,00	0,83	8,68	8,26	242
Ac8	66,12	3,00	0,68	2,82	16,83	10,54	7 229
Ac9	78,33	3,24	0,67	1,24	7,23	9,27	5 488
Ac10	94,30	3,50	1,06	0,00	1,14	0,00	1 315
Ac11	94,42	3,48	1,01	0,00	1,09	0,00	1 380
Ac12	94,69	3,31	0,35	0,17	1,48	0,00	2 898
Total	81,90	3,27	0,56	1,07	7,00	6,19	
Individus EEC	5,61	0,28	0,46	0,33	45,98	47,33	

Les identifiants utilisant le SIREN (Ac4, Ac5, Ac10 et Ac11) conduisent logiquement à des appariés quasi exclusivement non-salariés selon l'EEC (tableau 8) : seuls ceux ayant déclaré dans l'EEC leur activité principale sous ce SIREN peuvent être appariés. On notera que les appariements utilisant la commune de travail et le code APE (Ac6 et Ac12) sont aussi ciblés : par construction pour l'exclusion des hors emploi (pas de code APE) ; un peu plus de salariés, par rapport aux identifiants avec SIREN, sont toutefois appariés. Les identifiants n'utilisant que le département de travail sont beaucoup moins ciblés : Ac2 est carrément à exclure ; Ac8 n'est pas non plus très bien ciblé mais c'est celui qui apparie le plus de multi actifs.

Pour construire l'échantillon, on retient les couples appariés avec les identifiants ci-dessus, à l'exception d'Ac2, ordonnés dans l'ordre du tableau ci-dessous. Les - rares - couples appariés différemment selon les identifiants (ex (x,y) avec l'identifiant a et (x,z) avec l'identifiant b) sont éliminés. Les premiers identifiants selon cet ordre, ceux les plus ciblés sur les non-salariés selon l'EEC - ont, par construction, un taux de non-salariés très élevé mais n'apparient aucun non-salarié principalement salarié (selon l'EEC), ce qui est source de biais. Les appariés supplémentaires apportés par les identifiants suivants sont nettement moins ciblés mais c'est logique puisque, parmi les couples appariés par ces identifiants, ne sont retenus que ceux non appariés par les identifiants les plus ciblés. Parmi les salariés purs ou les hors emploi appariés figurent sans doute à la fois des appariements corrects relevant de la différence de concept (ou de la sous-déclaration EEC) et des appariements incorrects. Il n'est bien entendu pas possible de les identifier individuellement. Les erreurs d'appariement sont sans doute conséquentes avec l'identifiant Ac8 qui n'utilise que le département mais le supprimer conduirait à se priver de plus de la moitié des multiactifs appariés. On ne retient, parmi les 1 522 couples appariés uniquement grâce à cet identifiant, que les 136 multiactifs (non-salariés principalement salariés), d'où la ligne totalb du tableau 9.

Tableau 9 - Ventilation, pour chaque identifiant Acooss, des nouveaux appariés lors de l'introduction de cet identifiant, selon leur situation dans l'EEC

	Non salarié	Déclaré non-salarié	Reclassé non-salarié	Salariés multi actifs NS	Salariés purs	Hors emploi	Nb appariés
Ac4	95,57	3,84	0,17		0,42		2 888
Ac5	97,00	1,50	0,00		1,50		200
Ac10	94,25	3,53	1,07		1,15		1 304
Ac11	97,10	2,90	0,00		0,00		69
Ac12	89,63	1,51	1,08	1,08	6,70		463
Ac6	79,13	1,74	1,74	0,00	17,39		115
Ac7	49,33	2,67	0,00	2,67	18,67	26,67	75
Ac9	37,55	2,95	0,96	4,04	21,82	32,67	1 558
Ac1	23,08	7,69	0,00	0,00	19,23	50,00	26
Ac3	27,67	3,14	0,63	3,14	26,42	38,99	159
Ac8	17,74	1,58	0,85	8,94	54,20	16,69	1 522
Total	5 698	250	55	211	1 307	858	8 379
Totalb en nb	5 378	219	41	206	435	529	6 808
Totalb en %	79	3	1	3	6	8	
Total EEC	9 409	470	777	552	77 101	79 359	
Tx appariement	57	47	5	37	1	1	

Lecture : 2 888 couples (Acooss, EEC) sont appariés avec l'identifiant Ac4 (2891 dans le tableau précédent mais 3 de ces couples sont appariés différemment avec d'autres identifiants), dont 0,42 % où l'apparié est salarié selon l'EEC. Parmi les individus non appariés avec Ac4, 200 le sont avec Ac5. Au total 8 379 couples différents sont appariés à l'aide des 11 identifiants retenus. Si on élimine les identifiants Ac1 et Ac3 et si on ne retient, parmi les couples appariés uniquement avec Ac8, que les salariés avec une activité secondaire non salariée, ce total revient à 6 808, dont 6 % de salariés purs selon l'EEC. Le taux d'appariement, rapporté à l'EEC, est alors de 57 % pour les non-salariés selon l'EEC et d'un peu moins de 1 % pour les salariés (taux d'appariement avec les bases non-salariés Acooss).

On note le très faible taux d'appariement des déclarés salariés recodifiés en non-salariés (5 %) au regard de celui des déclarés non-salariés recodifiés en salariés (47 %) : la déclaration spontanée est

plus proche du concept de salarié selon la logique de la déclaration sociale que de la logique « PCS » mise en œuvre lors du reclassement de salariés en non-salariés.

L'appariement sur le champ MSA est très fruste : le prénom n'est pas disponible et le SIREN n'est renseigné qu'une fois sur 2. On complète donc un identifiant avec SIREN, le seul fiable en l'absence de prénom par un identifiant composé du sexe, de la date de naissance, du département de naissance et de la commune de travail. 1 526 individus sont appariés de cette façon¹³ (tableau 10).

Tableau 10 - Ajout de l'appariement avec la base MSA

	Non salarié	Déclaré non-salarié	Reclassé non-salarié	Salariés multi actifs NS	Salariés purs	Hors emploi	Nb appariés
ACOSS	5 378	219	41	206	435	529	6 808
MSA	1 414	6	18	12	50	26	1 526
Total	6 792	225	59	218	485	555	8 334
Total en %	81	3	1	3	6	7	100
Total EEC	9 409	470	777	552	77 101	79 359	
Tx appariement	72	48	8	39	1	1	

C - Fusion des appariements

Note : on inclut dans cette section les individus DADS appariés alors qu'ils n'ont que des périodes d'indemnisation chômage.

L'échantillon total compte 83 567 individus dont 77 158 appariés avec les DADS, 8 334 appariés avec les bases « non-salariés » ; parmi eux, 1 925 sont appariés avec les deux sources.

Tableau 11 : taux d'appariement selon la situation dans l'enquête Emploi

Situation dans l'enquête Emploi	Ensemble de l'EEC		Répondants les 4 trimestres de 2012	
	Appariés	Taux (%)	Appariés	Taux (%)
Au moins une activité non-salariée, aucune activité salariée*	7 107	81	1809	82
A la fois activités non-salariée et salariée	868	87	280	88
Au moins une activité salariée, aucune activité non-salariée	61 797	81	16 123	83
Aide familial, aucune autre activité	108	20	23	16
En emploi, uniquement frontaliers	200	15	47	15
Jamais en emploi, allocations chômage	4 021	80	741	80
Déclaré chômage, sans allocations	2 091	40	314	32
Toujours inactif	7 373	11	1 516	9
Total	83 565		20 853	

* on utilise ici l'information sur la multiactivité de l'EEC

Le taux d'appariement, rapport entre le nombre d'appariés et le nombre de personnes susceptibles a priori d'être appariées dans l'enquête Emploi, est de 81 % +/- 2 points pour chaque catégorie d'actifs occupé, y compris les chômeurs indemnisés (tableau 11). Font exception les aides familiaux, non couverts par les bases « non-salariés ». Les multiactifs, qui peuvent être appariés aux DADS, et/ou aux

¹³ Après élimination des quelques appariés MSA déjà appariés avec les bases Acooss. La double affiliation est possible (on en recenserait environ 5 000) mais il a été jugé plus prudent, après examen visuel de ces quelques cas (moins de 10), de les exclure.

bases non-salariés ont logiquement un taux d'appariement à au moins une des deux bases supérieur (87%)¹⁴.

Parmi les 1925 appariés selon les deux sources, qui sont donc censées avoir une double activité, salariée et non-salariée, certaines situations semblent incohérentes, soit au regard de la multiactivité selon l'EEC, soit au regard de l'information sur la multiactivité dans les bases non-salariés qui ont une indicatrice de présence dans les DADS. La plupart de ces incohérences sont explicables :

- Monoactives selon l'EEC : il peut s'agir de sous-déclaration de la multiactivité dans l'EEC
- Pas d'activité salariée, selon les « bases non salariés » :
 - Postes DADS présents lors de l'appariement mais absents des fichiers DADS de diffusion (cf. section suivante)
 - Les postes (salariés) d'élus, ne sont pas retenus pour caractériser une activité salariée dans les bases non-salariés

77 cas (4 %) d'incohérence ne sont pas expliqués et semblent liés à la qualité des identifiants utilisés pour l'appariement (tableau 12).

Tableau 12 : taux d'incohérence non expliquée selon l'indicateur de qualité des identifiants utilisés

	1A	1B	2	3	Total
Appariés bases «non-salariés» avec SIREN	0,4	0,8	2,8	5,5	1,4
Autres appariés Acoss	3,4	7,6	10,0	6,6	5,1
Appariés MSA	9,9				9,4

Champ : individus ayant répondu aux 4 trimestres à l'EEC et appariés à la fois aux DADS et aux bases non-salariés

Lecture : parmi les appariés DADS grâce aux identifiants de type 1A (cette typologie renvoie au tableau 6 « Ventilation des appariés par niveau de qualité »), et appariés aussi aux bases « non-salariés », le taux d'incohérence est de 0,4 % si l'identifiant bases « non-salariés » utilise le SIREN et de 3,4 % s'il s'agit d'un autre identifiant construit à partir de la source Acoss.

Un autre cas peut sembler incohérent : l'appariement des frontaliers. En fait, parmi les 47 répondants quatre fois appariés et qui ont pour seul emploi un emploi frontalier, seuls 15 ont un poste dans les DADS de diffusion. Les autres sont soit appariés au titre de l'indemnisation chômage (ils se déclarent aussi au chômage au moins un trimestre dans l'EEC) ou au titre de périodes non retenues dans le fichier de diffusion.

D - Restriction de l'appariement DADS aux présents dans le fichier poste

Parmi les 77 156 appariés DADS, seuls 73 878 sont retrouvés dans le fichier postes de diffusion, postes d'indemnisation chômage inclus, soit un écart de 3 278 individus. La suite de l'étude (repondération, comparaison des situations d'emploi) portera uniquement sur les individus présents dans le fichier postes mais cette différence mérite examen. En effet, elle est le résultat du traitement DADS qui vise à ne retenir, parmi les déclarations DADS, que celles qui correspondent réellement à de l'emploi. L'appariement est l'occasion d'avoir un regard externe, celui de l'enquête Emploi, sur ce traitement quant à l'emploi des personnes concernées. On commence par établir une typologie des disparitions à partir d'un examen interne aux DADS afin de confronter ensuite la situation d'emploi EEC au regard de cette typologie.

¹⁴ On s'attendrait même à un taux supérieur (97%) si le taux d'appariement aux DADS et aux bases non-salariés était indépendante de la multiactivité. Cela s'explique par l'utilisation du SIRET pour l'appariement des non-salariés : si le poste principal dans l'EEC est salarié, alors l'appariement avec les bases non-salariés est le plus souvent rejeté (sauf en cas de succès avec l'identifiant Ac8, cf I.B) : le taux d'appariement des multiactifs avec les bases non-salariés est inférieur à celui des monoactifs (resp. 58% et 81%) alors que ce n'est pas le cas pour l'appariement avec les DADS, qui n'utilise pas le SIRET.

1 - Examen interne aux DADS des raisons de l'élimination des postes

Le fichier « postes », diffusé, est construit à partir d'un fichier intermédiaire de « périodes » (il peut y avoir plusieurs périodes par postes), non diffusé, lui-même construit à partir des fichiers sources. Le fichier utilisé pour l'appariement se situe, dans le processus de traitement, entre les fichiers sources et le fichier des périodes.

La moitié des 3 278 individus disparaissent entre l'extraction du fichier ayant servi à l'appariement et le fichier périodes définitif et l'autre moitié lors de la constitution du fichier postes à partir du fichier des périodes. Pour analyser ces disparitions, on se place au niveau poste (croisement entre un individu et un établissement) : un individu qui n'apparaît pas dans le fichier poste est un individu dont tous les postes ont été supprimés (sans que le niveau individu soit pris en compte pour la suppression de chaque poste).

Le fichier ayant servi à l'appariement comprend 53,6 millions de postes (couples (NIR, Siret), sans préjuger de la réalité d'un travail effectif dans l'établissement signalé par le Siret), le fichier « périodes » 50,6 millions et le fichier « postes » de diffusion n'en a que 47,6 millions. De même, les individus appariés ont 132 000 postes dans le fichier initial, dont 125 000 retrouvés dans le fichier périodes et 118 000 dans le fichier postes. Dans les deux cas, un peu plus de 10 % des postes sont supprimés par la chaîne de traitement.

Pour rendre compte du passage des périodes aux postes, la variable « type de rémunération » du fichier « périodes » est la plus discriminante (tableau 13).

Tableau 13 : présence dans le fichier « poste » en fonction du type de rémunération (en nombre de postes)

Type de rémunération*	Absents fichier poste (nombre)	Postes	Chômage indemnisé présent dans le fichier « poste » (nombre)	Total
A : Allocation	5 459	0	0	5 459
C : Chômage	1 447	0	8 973	10 420
I : Indemnités	102	0	0	102
P : Participation	125	0	0	125
S : Salaire	126	108 839	0	108 965
T : Traitement	20	16	0	36
Total	7 279	108 855	8 973	125 107

* La documentation sur cette variable s'est améliorée depuis les travaux sur l'appariement. D'où la possibilité d'ajouter des libellés aux modalités de la variable. Cette documentation confirme les conclusions de l'analyse ci-dessous. Ne sont retenus dans le fichier poste que ceux avec au moins une période de type S.

L'examen des catégories juridiques des établissements concernés permet d'interpréter certaines modalités de cette variable (tableau 14).

Tableau 14 : croisement du type de rémunération et de la catégorie juridique de l'établissement

Type Rémun.	CJ 8170 Régime d'assurance chômage	CJ 9220 Association déclarée	CJ 9260 Association de droit local	Autres CJ	Total
A	0	5 328	130	1	5 459
C	10 407	0	0	13	10 420
I	0	18	0	84	102
P	0	8	0	117	125
S	1	8 475	268	100 221	108 965
T	0	0	0	36	36
Total	10 408	13 829	398	100 472	125 107

Les rémunérations de type « C » correspondent aux versements des Assedic (CJ 8170). Ne sont conservées dans le fichier « poste » que les indemnités concernant des personnes ayant eu un poste salarié l'année N-1.

Les rémunérations de type « A » sont versées par des associations (CJ 92), principalement de type consulaire (code APE 9411Z : aucune période de type S pour ce code APE). Les rémunérations versées par les caisses de congés payés du bâtiment relèvent certainement de ce type : les postes correspondants sont supprimés par les chaînes de traitement DADS et leurs rémunérations ajoutées à celle du poste de BTP où le salarié travaille (ou un des postes s'il en a plusieurs).

Les rémunérations de type I, P et T sont peu fréquentes et non reprises (sauf en partie pour le type T) dans le fichier « poste », composé quasi exclusivement de périodes de type S¹⁵.

Pour les 7 000 postes supprimés entre l'extraction ayant servi à l'appariement et le fichier périodes définitif, on ne dispose que du Siret. Ces postes sont répartis sur 2 600 établissements différents. La distribution de la catégorie juridique et du code APE de ces établissements ne permet pas de voir dans ces caractéristiques la cause de l'élimination de ces postes. D'ailleurs ces établissements sont presque tous (2 500 sur 2 600) présents dans le fichier établissement de diffusion avec des postes salariés.

Au total, l'examen permet de retrouver dans certains cas les raisons de l'élimination de ces périodes : elles ne correspondraient pas à de l'emploi. On fait l'hypothèse qu'il en est de même pour les autres périodes éliminées.

2 - Parmi ces individus dont tous les postes DADS ont été supprimés, certains sont salariés selon l'EEC

On peut répartir les salariés dont tous les postes sont supprimés de la manière suivante

- a_Nper : aucune période dans le fichier périodes
- b_Per_C : que des périodes d'indemnisation chômage (type C) dans le fichier périodes
- c_Per_AC : que des périodes de type A ou C
- d_Per_Aut : autre raison

Les salariés présents dans le fichier « poste » sont quant à eux ventilés en :

- e_Pos_C : que des postes d'indemnisation chômage
- f_Pos_S : avec poste de salarié.

Nota : une personne n'ayant que des périodes d'indemnisation chômage est présente dans le fichier poste (cas e_Pos_C) si elle avait un poste salarié en N-1 et absente sinon (cas b_Per_C).

Tableau 15 - Modalité d'absence/présence dans le fichier postes et situation d'emploi selon l'EEC

	Appariés avec uniquement les DADS					Idem, répondants EEC aux 4 trimestres				
	Salariés	Mixtes	Non salariés	Hors emploi	Effectifs	Sal.	Mixtes	N Sal.	Hors E.	Eff.
a_Nper	19	0	5	76	1500	17	0	7	76	344
b_Per_C	8	0	2	90	1296	7	0	2	90	308
c_Per_AC	42	1	2	55	88					19
d_Per_Aut	16	1	3	80	99					20
e_Pos_C	12	0	2	86	1082	13	0	2	85	248
f_Pos_S	85	0	1	14	71166	89	1	1	10	17774

Champ : individus appariés avec les DADS et non appariés aux bases non-salariés

Lecture : 19 % des 1500 appariés absents du fichier périodes répondent au moins une fois être salariés à l'EEC. Parmi les 344 d'entre eux répondants aux 4 trimestres de l'EEC, 17 % sont salariés au moins un trimestre selon l'EEC.

On croise cette typologie avec la situation dans l'EEC en se limitant aux individus appariés uniquement avec les DADS (ceux appariés aux deux sources sont logiquement plus souvent en emploi selon l'EEC).

¹⁵ En fait, les seules périodes d'un autre type dont le poste est conservé sont celles relevant d'un poste ayant au moins une période de type S.

Seuls les individus éliminés parce qu'ils n'ont que des périodes de chômage et qu'ils n'étaient pas en emploi DADS l'année précédente (ligne b_Per_C) sont moins d'une fois sur dix en emploi selon l'EEC (tableau 15). On note que ceux qui sont au chômage depuis moins longtemps (ligne e_Pos_C : ils avaient un emploi DADS en 2011) sont un peu plus souvent en emploi selon l'EEC, ce qui est cohérent avec ce que l'on sait par ailleurs sur la relation entre ancienneté au chômage et accès à l'emploi. Ceux que l'on ne retrouve plus dès le fichier période (ligne a_Nper) sont plus d'une fois sur cinq en emploi et un peu moins d'une fois sur cinq en emploi salarié. On note que, même en se limitant à ceux qui ne sont pas appariés aux bases non-salariés, ils sont beaucoup plus souvent non-salariés (5 %) que les salariés présents dans le fichier postes (1 %). Rien ne permet de savoir s'il s'agit d'un défaut d'appariement ou d'une absence dans les bases « non-salariés ». Enfin, ceux qui sont éliminés au titre de périodes dont le type de rémunération est « A » (versées par des associations) sont peu nombreux mais assez souvent salariés selon l'EEC (ligne c_Per_AC).

Il se pourrait que, suite au décalage de paie, certains postes DADS créés en décembre 2012 soient absents du fichier postes 2012 (car intégrés au fichier postes 2013). Dans ce cas, la proportion de salariés selon l'EEC absents du fichier postes serait plus élevée au 4^e trimestre. Le tableau suivant montre que ce n'est pas le cas. C'est même nettement le contraire pour la ligne e_Pos_C (pour la ligne c_Per_AC aussi mais les effectifs sont très faibles), ce qui s'interprète aisément et montre que ces emplois EEC ne sont pas des artefacts liés à l'appariement : ceux de la ligne E étaient en emploi en 2011 et touchent des allocations en 2012 : leur proximité à l'emploi est donc plus forte en T1 (taux d'emploi de 14 % selon l'EEC) qu'en T4 (7 %) (tableau 16).

Tableau 16 - Taux d'emploi trimestriel par modalité d'absence/présence dans le fichier postes

	T1	T2	T3	T4
a_Nper	22	22	21	21
b_Per_C	7	9	8	7
c_Per_AC	51	53	45	43
d_Per_Aut	16	17	15	14
e_Pos_C	14	12	9	7
f_Pos_S	82	83	83	83

Au total, les emplois supprimés par les traitements DADS semblent le plus souvent justifiés (ils sont le plus souvent hors emploi dans l'EEC), mais il y aurait un certain nombre de suppressions à tort.

On peut tenter une estimation du nombre de salariés absents à tort du fichier « postes » en faisant l'hypothèse que la proportion de salariés EEC est la même pour les non-appariés. Sous cette hypothèse, il manquerait 380 000 personnes (tableau 17), dont 262 000 (lignes a_Nper et c-d) avaient des postes hors chômage supprimés à tort. Les autres (118 000) sont sans doute majoritairement absentes des sources (au moins au stade où a été extrait le fichier pour appariement) en tant que postes hors chômage¹⁶ mais repérées grâce à leur présence dans le fichier de Pôle emploi.

On proposera une autre estimation s'affranchissant de cette hypothèse après les repondérations.

Tableau 17 - estimation du nombre de personnes en emploi dans l'année manquantes dans le fichier postes

	Nb individus source (a)	Nb appariés (b)	(a)/(b)	Nb appariés salariés EEC (c)	(c) * (a)/(b)
a_Nper	1 282 277	1 641	781	288	225 043
b_Per_C	705 071	1 430	493	115	56 702
c - d	137 449	207	664	55	36 520
e_Pos_C	571 336	1 224	467	131	61 148
f_Pos_S	29 989 996	72 654	413		
	32 686 129*	77 156	424		379 413

Champ : France entière

* Sont comptés ici tous les individus, y compris ceux à NIR fictifs. Ce nombre est donc supérieur aux 32,1 millions ayant servi à l'appariement (qui s'effectue hors NIR fictifs).

¹⁶ Même si certains catégorisés ici comme n'ayant que des postes de chômage avaient des postes hors chômage avant leur suppression par les chaînes de traitement (c'est, selon les données de l'appariement, le cas de 40% de ceux de la ligne e_Pos_C).

II – Analyse du nombre de personnes en emploi au moins une fois dans l'année

L'appariement de l'enquête Emploi avec les DADS se fait uniquement sur des caractéristiques individuelles. Ce sont donc des individus qui sont appariés, non des postes. Cela conduit à une stratégie d'estimation du niveau d'emploi en deux étapes :

- d'abord l'estimation du nombre de personnes en emploi dans l'année (PEA)¹⁷ ;
- puis, pour ces personnes, calcul de l'emploi moyen dans l'année (en fait la moyenne de présence en emploi les 4 semaines d'interrogation de l'EEC).

Nota : dans cette partie on alterne l'utilisation de l'ensemble des appariés et l'utilisation des seuls répondants les 4 trimestres à l'EEC. L'acronyme P4T signale que l'on est sur cette restriction.

A - Population de référence dans les sources administratives : partage entre principalement salariés et principalement non-salariés

1 - Identification des sources de biais

Les sources administratives donnent, après traitement de la multiactivité salarié / non-salarié, qui est l'objet de ce chapitre A, une première estimation de personnes au moins une fois dans l'année (PEA). Cette estimation repose sur deux hypothèses :

- HA1 : exhaustivité des sources ;
- HA2 : chaque identifiant renvoie à une seule personne et chaque personne est identifiée par un seul identifiant

Le salarié est identifié dans les DADS d'origine par son NIR. Le NIR est remplacé par un identifiant anonymisé mais un code de redressement du NIR est disponible dans les bases utilisées. Cela permet de repérer 673 000 NIR incorrects¹⁸, dits aussi NIR fictifs, pour lesquels l'hypothèse HA2 est fautive : presque tous les NIR incorrects n'ont qu'un seul poste (contre 1,4 poste par NIR en moyenne) et il est quasiment certain que certains NIR incorrects renvoient à une personne déjà identifiée, au NIR correct ou non. Il est par contre impossible d'estimer l'ampleur du biais à la hausse sur le nombre de PEA ainsi créé. À titre indicatif, seuls 55 000 postes de ces NIR peuvent correspondre à un temps plein et 490 000 postes ont moins de 500 heures dans l'année. Diviser par 2 le poids de ces NIR (en leur attribuant un poids de 1/2, les autres ayant un poids de 1), revient à faire l'hypothèse que ces NIR ajoutent 336 000 personnes aux 28,6 millions de NIR corrects. Il y a ainsi, liée à cette hypothèse, une incertitude de l'ordre de 300 000 sur le nombre de PEA¹⁹.

L'hypothèse HA1 peut être mise en défaut de deux manières :

- Par défaut de couverture de certaines populations systématiquement absentes ou quasi absentes des sources administratives
- Par trou de « collecte » : certains individus seraient, certaines années, absents.

¹⁷ Il n'existe pas actuellement de publication sur le nombre de PEA. Jusqu'en 2014, il était mentionné, sur le seul champ des salariés, dans les publications sur le revenu salarial dans l'*Insee Référence* « France Portrait social ».

¹⁸ Les modalités du code de redressement du NIR commençant par S ou X ne sont jamais appariées. Cela représente 673 000 NIR.

¹⁹ Sachant qu'aucune des deux extrémités de la fourchette - ils ne rajoutent aucun PEA aux NIR corrects ; ils correspondent tous à des individus différents entre eux et différents de ceux avec NIR correct - n'est réaliste. De plus, l'impact sur l'incertitude en terme d'emploi est beaucoup plus faible car ces PEA à NIR fictif ont un faible taux d'emploi moyen dans l'année.

L'appariement permet de tester ces hypothèses en comparant avec l'EEC. Cette comparaison n'est toutefois pas directe : d'une part l'EEC n'est par construction pas exhaustive et d'autre part n'est connu que l'emploi pendant les 4 semaines de référence de l'année. De plus, le calage sur l'EEC n'a de sens que pour ceux qui sont au moins une fois en emploi (PEA) selon l'EEC. Les autres sont principalement des inactifs qui ne sont pas censés être retrouvés dans les bases administratives sur l'emploi. Seule une petite partie d'entre eux est appariée : ceux qui sont en emploi selon les sources administratives mais pas selon l'EEC, appelés ici pour cette dernière raison « JEMP » (pour Jamais en EMPloi).

Il faut donc adapter l'approche à chacun des cas suivants :

- PEA2 : les personnes en emploi dans les 2 sources, c'est à dire à la fois selon les bases administratives et selon l'EEC ; ils peuvent être calés (i.e. repondérés) sur chaque source. Cette double repondération permettra de détecter d'éventuels défauts de couverture (cf. E) ;
- JEMP : les JEMP sont PEA uniquement selon les bases administratives ; leur estimation se fera donc uniquement par calage sur celles-ci. Il conviendra toutefois de vérifier que la présélection des P4T ne conduise pas à les surestimer, les personnes ne changeant pas de situation étant moins sujettes à l'attrition (cf. B et D) ;
- TCOL : les trous de « collecte » qui comprennent aussi bien les trous de collecte proprement dits (absence des sources) que les trous de « collecte » induits par les traitements de production des fichiers de diffusion à partir des sources. Une estimation plancher de ces trous de collecte sera produite à partir de ceux directement observables à travers l'appariement et repondérés sur l'EEC (cf. F).

2 - Le partage entre principalement salariés et principalement non-salariés

L'estimation du nombre de personnes en emploi dans l'année (PEA) se fait à la fois par calage sur les sources administratives et par calage sur l'enquête Emploi, ce qui permet une première comparaison entre les deux sources et la détection d'éventuels défauts de couverture. Un même individu de l'EEC peut être apparié à la fois aux DADS et aux bases « non-salariés ». Pour pouvoir estimer un nombre de PEA sur les sources administratives sans double-compte, il faut donc l'affecter à une seule des deux sources. On s'appuie pour cela sur la notion d'activité principale (salariée ou non salariée) mise en œuvre dans les traitements de la chaîne « non-salariés » à partir de la comparaison entre les revenus d'activité non salariés et la somme des salaires.

Tableau 1 - ventilation des PEA selon leur activité principale

		Nombre PEA	Dont appariés	Taux 1/10 000	PEA pondérés***
NSA	Non salariés monoactifs (sans période DADS)	2 665 805	6 746	25	2 665 805
PNS	Non salariés pluriactifs principalement NS*	288 332	739	26	288 332
PSA	Non salariés pluriactifs principalement salariés*	715 632	1 771	25	715 632
SAL	Salariés monoactifs	28 302 709	70 366	25	27 965 725
	Total sans doubles comptes**	31 972 478	79 500	25	31 635 494

Champ : France métropolitaine

* En l'absence d'une table de passage entre les identifiants des bases « non-salariés » et les identifiants des DADS, un appariement a été effectué à partir d'autres informations. Environ 3 000 pluriactifs selon les bases « non-salariés » n'ont pas pu être rattachés à leurs postes DADS.

** La somme pour les appariés est de 79 622 car 122 monoactifs selon les bases « non-salariés » sont appariés avec les DADS. Ils seront par la suite comptabilisés uniquement parmi les salariés. Parmi ces 122, au moins 59 sont appariés à juste titre aux deux sources : ils n'ont que des postes d'élus et ne sont donc pas considérés comme salariés dans les bases non-salariés. On notera que ces non-salariés élus ne peuvent pas être repérés directement (i.e. hors appariement) dans les bases non-salariés. Ils sont donc comptés, dans la colonne nombre de PEA, à la fois en NSA et en SAL. Sous hypothèse d'un taux d'appariement de 1/400, cela représenterait environ 24 000 double-comptes (59 appariés * 400).

*** application d'un poids de 1/2 aux NIR incorrects

Il y a 29,3 millions de personnes ayant au moins un poste dans les DADS, 3,7 millions de personnes dans les bases « non-salariés » et 1,0 million de personnes présentes dans les deux sources, soit au total 32,0 millions de personnes différentes dans les sources administratives (tableau 1).

Le taux d'appariement est de 25 pour 10 000 (1 sur 400) dans les DADS comme dans les bases « non-salariés ». Mais, parmi les pluri-actifs, certains ne sont appariés qu'aux DADS et d'autres qu'aux bases « non-salariés ». Le calage des principalement salariés ne peut se faire que sur ceux appariés aux DADS et celui des principalement non-salariés sur ceux appariés aux bases « non-salariés ». Or, environ 20 % des salariés et des non-salariés de l'EEC ne sont pas appariés : un principalement salarié (selon les bases administratives) peut donc être apparié uniquement avec les bases « non-salariés » et un principalement non-salarié uniquement avec les DADS. Cela conduit à éliminer certains appariés et diminue par conséquent le taux d'appariement des pluriactifs (tableau 2). On note aussi que, comme l'appariement des bases « non-salariés » utilise la commune de travail ou le Siret, un pluriactif EEC qui se déclare salarié dans l'EEC a moins de chance d'être apparié aux bases « non-salariés » : on constate que beaucoup plus de pluriactifs ne sont appariés qu'aux seules DADS (932) plutôt qu'aux seules bases «non-salariés» (222).

Tableau 2 - De l'échantillon apparié total à l'échantillon retenu
(en gras : observations éliminées)

	Monoactifs	Pluriactifs principalement Non Salariés	Pluriactifs principalam ent salariés	Total
Appariés aux seules bases « non-salariés »	6 624	106	116	6 846
Appariés aux bases «non-salariés» et DADS (salarié fichier postes)	132	506	850	1 488
Appariés aux seules DADS (salarié fichier postes)	70 234	127	805	71 166
Total appariés	76 980	742	1 778	79 500
Total retenus	76 980	615	1 662	79 257
Taux d'appariement	25	21	23	25

B - Population de référence dans l'EEC : correction du biais d'attrition par calage du panel EEC

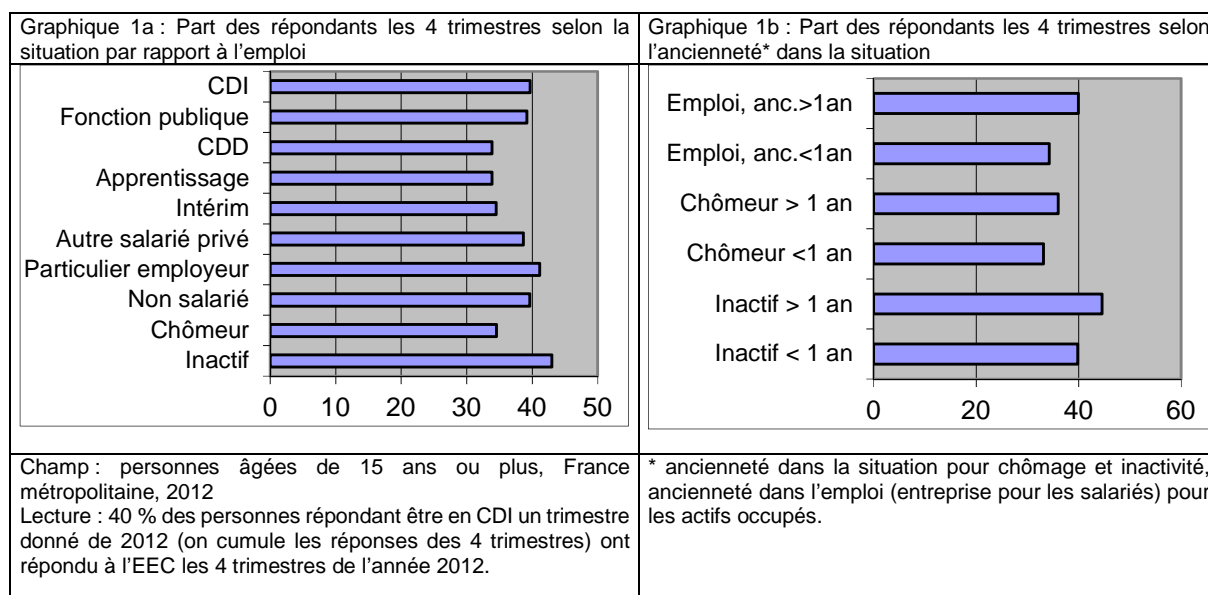
Comme l'appariement se fait sur des individus, le calage sur l'enquête Emploi doit se faire sur les répondants aux 4 trimestres de l'année. Cela permet notamment d'éviter les biais liés à la compensation de l'attrition (cf. I.A.3) et de mettre en œuvre la stratégie d'estimation de l'emploi en deux étapes :
Emploi moyen dans l'année (EMA) = (1) personnes en emploi dans l'année (PEA)

* (2) durée en moyenne en emploi dans l'année (DMA).

Il faut toutefois éviter le biais direct d'attrition : les personnes qui occupent les situations les moins stables sont plus concernées par l'attrition dans l'EEC utilisée en longitudinal, notamment parce que l'EEC suit des logements et non des individus.

La situation des répondants aux 4 trimestres d'une année est plus stable qu'en moyenne

Les répondants aux 4 trimestres (P4T) ne représentent, un trimestre donné, que 34 % des situations de chômage ou d'emploi en CDD et en intérim, contre 40 % des situations d'emploi en CDI ou non-salarié et 43 % des situations d'inactivité (graphique 1a). L'écart est encore plus net en comparant selon l'ancienneté dans la situation (graphique Xb).



En surestimant la part des situations stables, i.e. ceux en emploi les 4 trimestres, par rapport à ceux alternant emploi et non-emploi, on surestime la durée moyenne en emploi (DMA) et donc on sousestime, à emploi donné, le nombre de personnes en emploi dans l'année (PEA). Pour éviter cela la population des répondants les 4 trimestres est repondérée de manière²⁰ à ce qu'elle soit calée sur l'ensemble des répondants à l'EEC sur les variables suivantes :

- pour chacun des 3 premiers trimestres, activité (variable Acteu)
- pour le 4^e trimestre : sexe, âge en 5 tranches, lieu de naissance (France/étranger) et une variable de situation d'activité (inactivité, chômage, non salarié, CDI, administration, Particulier employeur, CDD, intérim, apprentissage, autre) distinguant pour chacune des 4 premières situations, par type d'ancienneté (< 1 an et >= 1 an).

Ce calage interne à l'EEC permet donc d'augmenter le nombre de PEA selon l'EEC, et, parmi ces derniers, la part des situations les moins stables (ceux qui ne sont pas un emploi au moins une fois) (tableau 3).

Tableau 3

	Pondération initiale	Après calage	Nombre après calage
Part des en emploi au moins une fois (PEA)	53,1	55,5	28 052 911
Parmi les PEA, part de non-emploi au moins une fois	15,4	17,0	4 770 660

Après ce calage, le nombre de PEA s'établit à 28,0 millions.

L'écart brut, avant toute utilisation de l'appariement, entre les bases administratives et l'EEC, concernant le nombre de PEA est donc de l'ordre de 4,0 millions²¹. Cet écart peut être dû à :

- de l'emploi hors semaines de référence (JEMP_C) ;
- de l'emploi non déclaré par l'enquête (JEMP_D) ;
- des différences de pondération parmi les personnes en emploi dans les deux sources (PEA2).

Pour quantifier ces différentes sources d'écart, il faut utiliser l'appariement en affectant des poids aux individus appariés.

²⁰ A l'aide de la macro CALMAR, cf <https://www.insee.fr/fr/information/2021902>

²¹ Soit l'écart entre 32,0 millions dans les bases administratives - avant baisse du poids des NIR incorrects - et 28,0 millions dans l'EEC, y compris les frontaliers, absents des bases administratives. Les effets NIR incorrects et différence de champ et de concept sont du même ordre de grandeur.

C - Le calage de l'échantillon apparié sur les bases administratives

Pour les salariés et principalement non-salariés, on effectue un calage sur marge sur les DADS en utilisant les caractéristiques socio-démographiques, les formes d'emploi, le volume de travail et le profil infra annuel de l'emploi (nombre de personnes en emploi à certaines semaines). On compare différentes méthodes de calage en mesurant l'écart du résultat obtenu au profil trimestriel de l'EEC.

Partant d'une EEC repondérée avec le calage interne, la méthode retenue est en deux nouvelles étapes :

1. Pour chacun des 4 trimestres, calage de l'ensemble des appariés du trimestre en partant des poids obtenus par calage interne à l'EEC (section précédente).
2. Calage des présents les 4 trimestres (P4T) en utilisant comme variables de calage supplémentaires, pour chaque trimestre, une variable de présence en emploi la semaine de référence. Cette variable à 6 modalités (croisement emploi DADS Oui/non et activité EEC emploi/chômage/inactivité) est construite à partir des résultats de l'étape précédente.

Pour les non-salariés, on ne dispose pas de profil infra-annuel dans les bases administratives. On utilise donc un calage plus simple, avec les caractéristiques socio-démographiques, la tranche de revenu et la catégorie juridique de l'entreprise, et on part de la pondération issue du calage interne à l'EEC.

Les détails de ce calage sont présentés en annexe 1.1.

D - Résultats – les personnes « Jamais en emploi dans l'EEC »

1 - Pour les salariés et principalement salariés

La notion de personnes jamais en emploi selon l'EEC (JEMP) – mais en emploi dans les bases administratives – mobilise pour chaque personne, l'information sur les 4 trimestres. Il n'y a donc aucun critère externe à l'échantillon des présents les 4 trimestres. On peut toutefois examiner comment la correction du biais d'attrition influe sur l'estimation du nombre de JEMP. Comme précédemment signalé, les JEMP regroupent en fait deux cas forts différents

- JEMP_D (déclaratif) : ils sont en emploi DADS au moins une des 4 semaines de référence. Il s'agit alors là bien d'un écart entre déclaration et source administrative ;
- JEMP_C (contingent) : ils sont en emploi DADS dans l'année, mais pas les semaines de référence.

Tableau 4 - Effet de la pondération sur la présence en emploi, dont JEMP

Code		Nb obs.	A : calage "de base"*	B : calage retenu	100 * B/A
JEMP	JEMP_C	655	1 233 537	1 215 525	98,5
	JEMP_D	1045	1 819 946	1 735 727	95,4
Emploi EEC 1 à 3 T		2 570	4 673 634	4 868 826	104,2
Emploi EEC 4 T		13 692	20 954 239	20 861 279	99,6
Total		17 962	28 681 357	28 681 357	100,0

* calage direct à partir des poids initiaux EEC

Note de lecture : 655 personnes en emploi dans les DADS ne sont pas en emploi dans l'EEC au moins une des 4 semaines de référence ; ils représentent 1 233 537 personnes en partant, pour le calage sur les DADS, des pondérations de l'EEC, et 1 215 525 en partant des pondérations EEC corrigées.

Les critères empiriques utilisés ici (emploi DADS les semaines de référence) ne captent qu'imparfaitement la différence conceptuelle entre déclaratif et contingent. En effet, d'une part on ne sait évidemment pas ce qu'ils auraient déclaré s'ils avaient été interrogés une semaine correspondant

à de l'emploi DADS ; d'autre part la qualité des dates déclarées dans les DADS peut brouiller cette distinction.

Comme attendu, l'attrition concerne surtout ceux dont la situation change en cours d'année : la correction augmente les poids de ceux dont le nombre de trimestres en emploi EEC est compris entre 1 et 3. La compensation s'opère sur les JEMP, surtout les JEMP_D (tableau 4).

Pour vérifier que le « phénomène JEMP » est robuste à la qualité de l'appariement, on utilise l'indicateur de qualité proposé en I.A.3. On ne retient comme « hors de doute » que les appariés prenant la valeur la plus haute de cet indicateur de qualité, soit 73 % des appariés finalement retenus, et comme potentiellement douteux les autres. On s'attend à trouver plus de JEMP parmi les appariés « douteux », la différence devant être plus marquée si on se restreint aux JEMP déclaratifs (désaccord entre EEC et les DADS sur l'emploi aux semaines de référence). C'est effectivement ce qui se passe entre 30 et 59 ans, où l'interprétation des JEMP est moins aisée (tableau 5). La qualité de l'identifiant joue (odd ratio, ou risque relatif, nettement supérieur à 1), ce qui indique qu'un certain nombre de JEMP dans cette tranche d'âge résultent d'un défaut d'appariement et doit donc conduire à de la prudence dans l'interprétation pour cette tranche d'âge. Le phénomène est cependant moins net si on se restreint aux personnes nées en France et il est possible que le comportement de réponse des immigrés, dont tous les identifiants sont considérés comme douteux, soit différent.

En tout cas, il n'y a pas d'effet qualité de l'identifiant pour les jeunes et les seniors, qui constituent près de 4 JEMP sur 5.

Tableau 5 : part des JEMP en fonction de la qualité de l'identifiant

	Part dans les			Tous JEMP			JEMP en emploi DADS		
	appariés	Tous JEMP	JEMP Décl.	Identifiant « douteux »	Bonne qualité	Odd ratio	Identifiant « douteux »	Bonne qualité	Odd ratio
15 à 29 ans	22%	61%	60%	27%	27%	1,0	20%	18%	1,1
<i>dont étudiant</i>				56%	53%	1,0	46%	43%	1,0
30 à 59 ans	73%	23%	22%	4,5%	2,5%	1,8	2,9%	1,3%	2,1
<i>hors de pays=99</i>				3,2%	2,5%	1,3	2,1%	1,3%	1,6
60 ans et +	6%	16%	18%	25%	27%	0,9	21%	20%	1,0

Note de lecture : alors qu'ils ne représentent que 22 % des appariés, les jeunes de 15 à 29 ans représentent 61 % de ceux qui ne sont jamais en emploi (JEMP) et 60 % des JEMP déclaratifs (en emploi DADS au moins une des 4 semaines de référence). 27 % des jeunes appariés sont JEMP, qu'ils soient appariés avec un identifiant de bonne qualité ou un de moins bonne qualité ; cette part étant la même, l'odd ratio (rapport des chances) est égal à 1. Toujours parmi les jeunes appariés, les JEMP déclaratifs sont un peu plus fréquent quand l'identifiant est de qualité douteuse (20 %) que lorsqu'il est de bonne qualité (18 %). L'odd ratio est alors de 1,1.

2 - Sur l'ensemble du champ

Les JEMP sont sans doute sous-estimés pour les non-salariés car l'appariement se fait en utilisant des identifiants liés au travail (commune de travail ou SIRET). Ne sont alors repérés que les JEMP travaillant dans leur commune ou département de résidence. On ne peut pas faire de distinction analogue à celle entre « en emploi DADS » et hors emploi les semaines de référence pour les salariés.

10 % des PEA selon les sources administratives, soit un peu plus de 3 millions de personnes, ne sont en emploi EEC à aucune des 4 semaines de référence (tableau 6). Cette part est particulièrement importante pour les CDD et l'intérim (un tiers). Si on se limite aux JEMP censés être en emploi DADS au moins une semaine de référence, cette part est aussi élevée pour l'apprentissage (16 %). Les non-salariés, à l'exception des autoentrepreneurs, semblent moins concernés (mais cette part est sous estimée). Le cas des auto entrepreneurs est délicat à interpréter car il est possible que certains soient présents dans les bases administratives sans être pour autant actifs l'année considérée. Cette question est traitée en III.D.

Tableau 6 - Part des personnes jamais en emploi (JEMP) dans les EEC, par catégorie

Code	Libelle	Total	Dont JEMP		% JEMP	
			Non emploi DADS 4T	Emploi DADS	Non emploi DADS 4T	Emploi DADS
Salariés						
APP	Apprentissage	507 950	3 226	79 067	1	16
CDD	CDD	3 183 242	514 439	560 427	16	18
CDI	CDI	14 729 652	140 310	318 133	1	2
INT	Intérim	1 305 895	229 942	198 529	18	15
MZZ	Privé, autre cas	1 707 558	100 138	177 617	6	10
PE	Particulier employeur	1 062 967	35 179	108 261	3	10
SI	Fonction publique hors CDD	5 123 018	65 366	169 155	1	3
SID	Fonction publique CDD	1 061 077	132 133	130 524	12	12
Non-salariés						
nsA	Auto entrepreneur	568 564	0	90 037		16
nsC	Artisan ou commerçant	588 249	0	13 322		2
nsL	Profession libérale	526 408	0	15 745		3
nsM	Agriculteur (MSA)	446 144	0	7 784		2
nsS	Associé gérant	680 188	0	21 864		3
nsZ	Autre cas	144 737	0	11 516		8
		31 635 649	1 220 733	1 901 982	4	6

Lecture : Parmi les 507 950 apprentis (situation principale de l'année) selon les DADS, 3 226 ne sont jamais en emploi DADS les semaines de référence (moins de 1 % de la catégorie) et 79 067 (16 %) sont jamais en emploi EEC mais en emploi DADS au moins une des semaines de référence.

E - Résultats - PEA2

On appelle « PEA2 » ceux qui sont en emploi dans l'année (PEA) selon les deux sources.

1 - Démarche

La stratégie consiste à calculer deux poids pour chaque individu apparié, l'un obtenu par calage de l'appariement sur les bases administratives, l'autre par calage sur l'EEC ; à comparer ensuite les effectifs par strate selon les deux pondérations. Le calage sur les bases administratives a été présenté en C ; le calage sur l'EEC consiste en un calage sur le panel repondéré en B (on part évidemment, pour les appariés, du poids EEC dans ce panel repondéré, dit aussi poids après calage interne).

Soient a_{i1} a_{i2} ... a_{ij} ... les individus appariés de la strate i ; p_{ij} leur poids selon la repondération bases administratives et q_{ij} leur poids par rapport à l'échantillon²² EEC. On a les estimations suivantes :

- $D_i = \sum (p_{ij})$ est l'effectif de la strate i dans les bases administratives
- $E_i = \sum (q_{ij})$ est l'estimation de l'effectif de la strate i selon l'EEC

Si X_i est le « vrai » effectif de la strate i , alors D_i en est un minorant (tout individu présent dans les bases administratives existe, au cas douteux près déjà évoqué des NIR incorrects). On ne peut a priori rien dire de tel pour E_i : la somme des E_i est calée sur la population globale et, pour un i donné, E_i peut être supérieur ou inférieur à X_i .

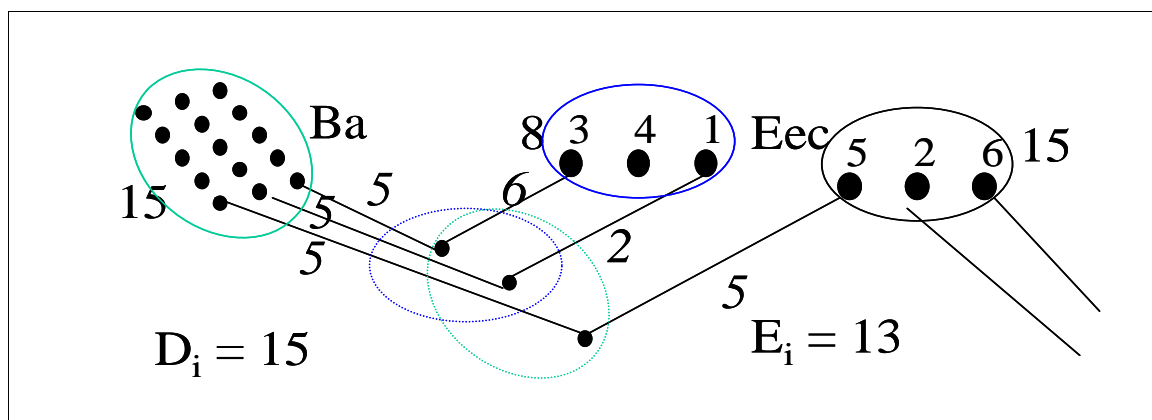
- Si $E_i < D_i$, on peut estimer que D_i représente l'estimation la plus prudente de X_i
- Si $E_i \gg D_i$, on peut estimer que l'on est en présence d'un trou de collecte dans les bases administratives.

Par la suite on analyse donc le rapport $R_i = E_i / D_i$

²² Il y a en fait trois niveaux de poids : ceux de l'EEC, ceux, calés sur l'EEC, des répondants sur les 4 trimestres de l'année et ceux de l'appariement.

Illustration par un exemple

Schéma 1:



Dans l'exemple illustré dans le schéma 1, et dont l'objectif est uniquement pédagogique, tous les individus d'une strate sont affectés du même poids ou du même coefficient multiplicatif par rapport au poids initial - alors qu'en pratique, à l'issue du calcul des pondérations (macro calmar), les poids de chaque individu peuvent différer.

Soit une strate i , par exemple CDD, de 15 individus dans les bases administratives, dont 3 appariés. Le poids de chacun est alors de 5. 2 de ces 3 appariés appartiennent à la strate homologue, CDD, de l'EEC et le troisième à une autre strate (s'il est par exemple multiactif, principalement salarié selon les bases administratives et principalement non-salarié selon l'EEC). La strate CDD de l'EEC a 3 individus, de poids global 8, dont deux appariés, de poids global 4. On multiplie donc les poids par $8/4=2$. Dans cet exemple tous les individus de la strate EEC à laquelle appartient le 3^e CDD (selon les bases administratives) apparié sont appariés : ils conservent donc leur poids EEC initial. Au total, les 3 appariés de la strate i ont un poids global de 15 (D_i) selon le calage sur les bases administratives et de 13 (E_i) selon le calage sur l'EEC. R_i vaut 87% (13/15). On en déduit, dans cet exemple fictif, que les personnes en CDD, selon les DADS, sont légèrement sous-pondérées dans l'EEC.

2 - Résultats par groupe

Une première comparaison en utilisant la typologie établie en A lors du partage salarié / non-salarié montre un net écart entre les monoactifs (au sens retenu ici) et les pluriactifs (tableau 7). Cet écart est particulièrement important parmi les non-salariés. Ce point, délicat à interpréter, est développé en annexe 1.2.

Tableau 7 - Comparaison des effectifs selon le calage - Typologie salarié/non-salarié selon les bases administratives

Code	Libelle	Nombre appariés	Effectifs pondérés		Ratio E/D
			Calage bases administratives (D)*	Calage EEC (E)	
Salariés		16 247	25 704 698	24 953 882	97
SAL	Salariés « purs »	15 871	25 038 743	24 340 568	97
PSA	Principalement salariés	376	665 955	613 314	92
Non-salariés		1 789	2 794 770	2 725 827	98
NSA	« purs »	1647	2 512 559	2 516 771	100
PNS	Principalement non-salariés	142	282 211	209 056	74
Total		18 036	28 499 469	27 679 709	97

Champ : Personnes en emploi dans les deux sources (PEA2)

Lecture : 15 871 appariés, salariés « purs » (i.e. sans activité non-salariée) sont en emploi selon les DADS ; ils représentent 24,3 millions de personnes en emploi selon le calage sur l'EEC, soit 97% de l'estimation obtenue par calage sur les bases administratives (25,0 millions).

* alors que les marges indiquées pour le calage EEC représentent l'ensemble de la catégorie correspondante dans l'EEC, celles indiquées pour le calage bases administratives sont inférieures à l'effectif correspondant dans les bases administratives : en effet, le calage se fait pour l'ensemble des appariés (JEMP et PEA2) et l'effectif indiqué ici est l'application aux seuls PEA2 des poids issus de ce calage.

La comparaison selon les types d'emploi indique une nette sous-pondération, dans l'EEC, des CDD et des auto entrepreneurs (tableau 8).

Tableau 8 - Comparaison des effectifs selon le calage selon la typologie d'emploi des bases administratives

Code	Libelle	Nombre appariés	Effectifs pondérés		Ratio E/D
			Calage bases administratives (D)	Calage EEC (E)	
Salariés		16 247	25 704 698	24 953 882	97
APP	Apprentissage	261	425 656	412 175	97
CDD	CDD	1088	2 095 627	1 929 790	92
CDI	CDI	9185	14 266 941	13 926 298	98
INT	Intérim	504	880 680	866 157	98
MZZ	Privé, autre cas	827	1 429 802	1 211 655	85
PE	Particulier employeur	609	919 820	952 310	104
SI	Fonction publique hors CDD	3268	4 886 676	4 826 153	99
SID	Fonction publique CDD	505	799 497	829 343	104
Non-salariés		1 789	2 794 770	2 725 827	98
nsA	Auto entrepreneur	229	474 378	403 499	85
nsC	Artisan ou commerçant	376	575 096	557 722	97
nsL	Profession libérale	308	504 954	463 938	92
nsM	Agriculteur (MSA)	401	441 796	515 707	117
nsS	Associé gérant	409	662 941	662 824	100
nsZ	Autre cas	66	135 605	122 137	90
Total		18 036	28 499 469	27 679 709	97

Lecture : voir tableau 7

L'utilisation de la typologie d'emploi construite à partir de l'EEC permet d'une part de confirmer, pour les salariés, la sous-pondération concernant les CDD et les CDI non en emploi les 4 trimestres²³ et, d'autre part, d'affiner le diagnostic pour les non-salariés : une fois tenu compte du défaut de couverture

²³ La ligne MZZ du tableau précédent, qui regroupe les cas où, dans les DADS, aucune forme d'emploi ne regroupe à elle seule plus de la moitié des heures du salarié, recouvre partiellement la ligne CDJ de la typologie EEC

des aides familiaux et des artistes, ainsi que d'une couverture imparfaite des agriculteurs, les autres non-salariés (plus des 4/5^e) seraient, eux aussi, sous pondérés dans l'EEC (tableau 9).

Tableau 9 - Comparaison des effectifs selon le calage selon la typologie emploi de l'EEC

Code	Libelle	Nombre appariés	Effectifs pondérés		Ratio E/D (%)
			Calage bases administratives (D)	Calage EEC (E)	
Salariés		16 167	25 587 291	24 764 900	97
ADM	Fonction publique	3725	5 658 835	5 607 649	99
APP	Apprentissage	307	519 046	496 534	96
CDD		1124	2 173 791	1 959 680	90
CDI	CDI, salarié les 4 T	8970	13 670 482	13 285 275	97
CDJ	Autre CDI	786	1 526 718	1 360 375	89
CXX	Autre salarié	163	276 149	306 227	111
INT	Intérim	416	731 301	678 815	93
PE	Particulier employeur	676	1 030 968	1 070 346	104
Non-salariés		1 869	2 912 178	2 914 809	100
NSA	Agriculture	374	405 380	458 735	113
NSF	Aide familial	21	33 544	151 590	452
NSR	Artiste	28	57 282	104 916	183
NSZ	Autre	1446	2 415 972	2 199 568	91
Total		18 036	28 499 469	27 679 709	97

Lecture : voir tableau 7

Nota : afin de se rapprocher du concept de non-salarié des bases administratives, les dirigeants salariés sont reclassés parmi les salariés.

Quelle pondération retenir pour la suite de ces travaux ? Une règle de prudence pose qu'il vaut mieux, en cas de doute, privilégier les groupes issus des bases administratives, fondés sur une présomption d'exhaustivité, aux groupes issus de l'EEC, dont l'estimation est moins robuste, surtout pour les faibles effectifs (cf. discussion sur les principalement non-salariés en annexe X). Cette règle peut se traduire opérationnellement de la manière suivante :

a/ retenir la pondération bases administratives dans tous les cas où $E/D < 1$ (ce qui signifie que l'EEC ne permet pas de remettre en cause la présomption d'exhaustivité des bases administratives) ;

b/ traiter au cas par cas les catégories où $E/D > 1$ (l'EEC met en doute l'exhaustivité des bases administratives).

Pour les salariés, sont concernés par des ratios $E/D > 1$:

- Les particuliers employeur (PE). $E/D = 1,04$ aussi bien à partir de la caractérisation des PE dans les bases administratives qu'à partir de leur caractérisation dans l'EEC (la différence en niveau entre ces deux caractérisations vient de personnes qui ne proviennent pas majoritairement - en termes de nombre d'heures - de la source PE selon les DADS et se déclarent PE dans l'EEC). Comme les effectifs sont conséquents et que l'exception PE, par rapport aux autres salariés, est facilement interprétable car probablement imputable à la source (il s'agit d'une source spécifique ajoutée à la source DADS pour constituer les DADS « grand format »), on retiendra la pondération EEC.
- Les autres salariés (CXX), qui constituent une catégorie « solde » difficilement interprétable (la moitié sont en CDI selon les bases administratives) et à relativement faibles effectifs. Il est donc plus prudent de ne pas en tenir compte.

On retient donc la typologie construite à partir des bases administratives et, à l'exception des PE, la pondération calée sur ces bases administratives.

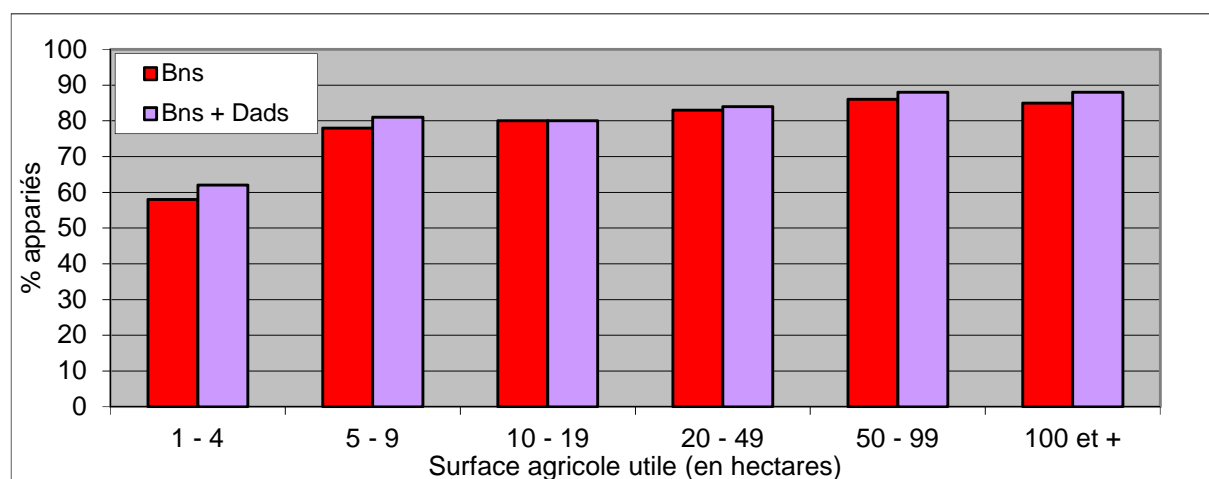
Pour les non-salariés, sont concernés :

- Les agriculteurs, aussi bien à partir de la caractérisation par la source (MSA) que par le déclaratif. Sont exclus des bases de diffusion les cotisants solidaires, i.e. les agriculteurs en

deçà d'un certain seuil (variable selon les départements) de surface. On constate effectivement que le taux d'appariement est plus faible pour les petits exploitants (graphique X).

- Les aides familiaux, identifiables uniquement dans l'EEC puisque n'étant pas personnellement rémunérés il n'y a pas de source administrative les concernant. Ils sont donc rarement appariés : en fait uniquement en cas de multiactivité ou de divergence entre déclaratif et administratif.
- Les artistes, identifiables eux aussi uniquement dans l'EEC. La couverture des artistes par les bases « non-salariés » est partielle : elle ne couvre que certains modes de déclaration des droits d'auteurs. Le service statistique du ministère de la culture (Deps) « évalue à plus de 30 000 le nombre d'artistes auteurs absents de la base Non-salariés » (Guyon, 2015). L'écart constaté ici est du même ordre de grandeur.

Graphique 2 : taux d'appariement des agriculteurs en fonction de la taille de l'exploitation



Champ : individus de l'EEC 2012 pour lesquels la surface d'exploitation est renseignée au moins 1 trimestre (872 observations)

Lecture : 58 % des agriculteurs ayant une exploitation de moins de 5 hectares (s.a.u.) sont appariés aux bases Non-salariés, 62 % sont appariés aux bases non-salariés et/ou aux DADS.

On retient donc provisoirement, pour les non-salariés, la typologie construite à partir de l'EEC et, pour les trois catégories susmentionnées, la pondération EEC (et la pondération bases administratives pour les autres non-salariés).

L'utilisation d'une typologie bases administratives pour les salariés et d'une typologie EEC pour les non-salariés pose de nouveau la question du partage entre salariés et non-salariés. Deux options sont possibles :

- privilegier le partage entre salariés et non-salariés des bases administratives et n'appliquer la typologie EEC qu'aux non-salariés selon celles-ci. Cette procédure a l'avantage de la simplicité mais conduit à diminuer à tort les catégories de non-salariés mal couvertes par les bases administratives : ainsi les artistes se retrouvent majoritairement classés parmi les salariés puisque seule cette partie de leur revenu est connue et leur effectif, en tant que non-salariés passe de 105 000 à 39 000 (tableau X) ;
- identifier les catégories de non-salariés mal couvertes par les bases administratives (aides familiaux et artistes) à partir de l'EEC et, pour le reste, appliquer le partage salariés non-salariés des bases administratives. C'est la solution retenue (tableau 10).

Au total, ces arbitrages conduisent à retenir 239 000 personnes en emplois de plus que par rapport au calage sur les bases administratives.

L'écart avec le calage sur l'EEC (27 680 000) est bien plus important : près de 1,1 million. Cet écart inclut un effet de champ : les actifs des communautés sont dans les bases administratives mais pas dans le champ de l'EEC. Cet effet de champ étant estimé, d'après le recensement de la population, à 0,2 million, il reste un écart de 0,9 million qui peut être interprété comme l'effet d'une sous-pondération des personnes en emploi, surtout les jeunes, par rapport aux chômeurs et inactifs.

Tableau 10 : synthèse sur les personnes en emplois dans les deux sources (PEA2)

Code	Libelle	Option a			Option b	
		Ventilation appariés	Calage bases administratives	Calage arbitré	Ventilation appariés	Calage retenu
Salariés		16 247	25 718 972	25 745 860	16 216	25 685 173
APP	Apprentissage	261	423 628	423 628	261	423 628
CDD	CDD	1 088	2 104 676	2 104 676	1 080	2 087 199
CDI	CDI	9 185	14 269 499	14 269 499	9 176	14 255 102
INT	Intérim	504	880 334	880 334	500	871 722
MZZ	Privé, autre cas	827	1 432 127	1 432 127	822	1 425 044
PE	Particulier employeur	609	921 340	948 228	606	937 000
SI	Fonction publique hors CDD	3 268	4 885 881	4 885 881	3 267	4 885 087
SID	Fonction publique CDD	505	801 487	801 487	504	800 391
Non-salariés		1 789	2 794 134	2 943 245	1 820	3 066 851
NSA	Agriculture	366	392 663	447 266	374	456 191
NSF	Aide familial	15	23 282	101 354	21	151 391
NSR	Artiste	11	22 977	39 413	28	103 810
NSZ	Autre	1 397	2 355 212	2 355 212	1 397	2 355 459
Total		18 036	28 513 106	28 689 105	18 036	28 752 024

Lecture : 9185 appariés en emploi dans les deux sources sont en CDI selon les bases administratives et représentent 14 269 499 personnes ; 9 (9 185 - 9 176) d'entre eux sont aides familiaux ou artistes selon l'EEC ; ils ne sont donc pas considérés comme salariés en CDI dans la typologie retenue ; les 9 176 restant représentent, toujours selon la pondération calée sur les bases administratives, 14 255 102 personnes en emplois.

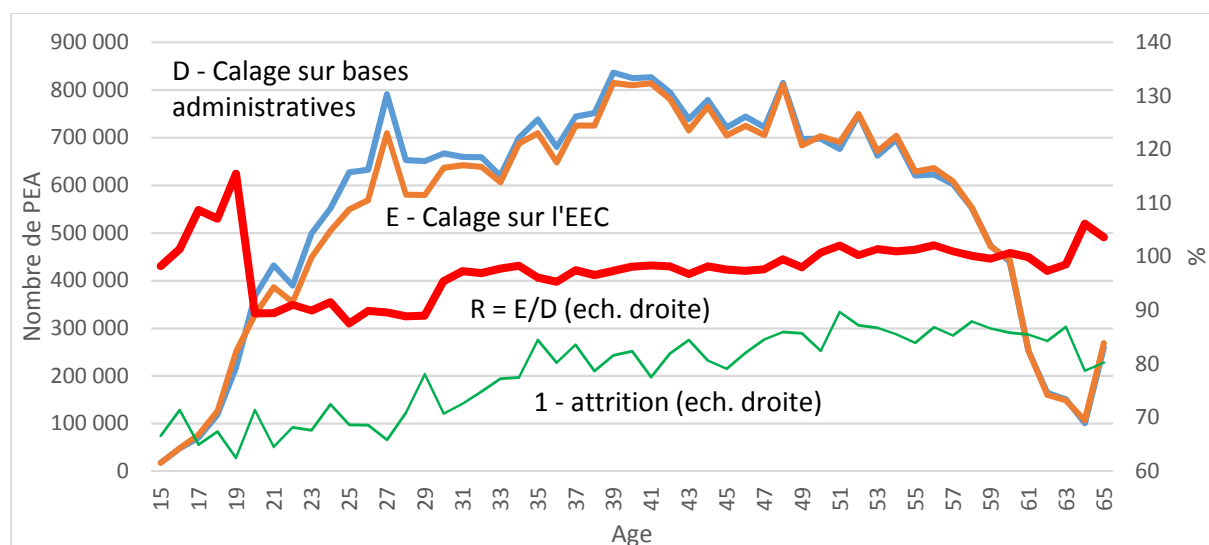
3 - Par âge : une sous-pondération des jeunes en emploi dans l'EEC

On sait qu'il y a moins d'emploi de jeunes dans l'EEC et on a déjà repéré en II.D une des raisons : la sous-déclaration de l'emploi dans l'EEC (le « phénomène JEMP »). Dans cette partie, on s'intéresse uniquement aux personnes en emploi dans les deux types de sources (PEA2), c'est à dire à l'emploi hors JEMP.

En comparant pour les PEA2, les effectifs calculés avec la pondération bases administratives et les effectifs calculés avec la pondération EEC, on obtient un profil marqué selon l'âge (graphique 3) : les jeunes en emploi sont nettement sous-pondérés dans l'EEC. Ce résultat demeure si on exclut les petits boulots (en se limitant aux horaires annuels « normaux », côté DADS, ou aux CDI et fonction publique, côté EEC). Une partie peut provenir des communautés, absentes de l'EEC, et qui concernent surtout, pour l'emploi, les jeunes (d'après le Recensement de la Population). On reviendra sur le profil par âge, en tenant compte des communautés en III.G.

On peut supposer que l'essentiel vient d'une surestimation des jeunes hors emploi, au détriment des jeunes en emploi, plus difficiles à joindre. L'attrition, qui peut capter la difficulté à joindre les personnes et donc à avoir un échantillon représentatif, est aussi plus forte sur les jeunes. Mais le profil par âge de l'attrition montre un plancher entre 15 et 20 ans puis une amélioration progressive jusqu'à 35 ans alors que le profil du rapport des pondérations indique un trou pour les deux tranches d'âge de 20 - 24 ans et 25 - 29 ans. Avant 20 ans, ce rapport, qui repose sur de faibles effectifs, est assez heurté mais globalement supérieur à 1. Après 30 ans, il remonte progressivement pour atteindre 1 sur la tranche des 50 - 59 ans (en fait, un peu moins de 1 pour les hommes et un peu plus de 1 pour les femmes).

Graphique 3 : Comparaison, pour les personnes en emploi au moins une fois dans les deux sources, du calage sur les bases administratives et de celui sur l'EEC, par âge



Champ : personnes en emploi au moins une fois sur l'année dans les deux sources (PEA2)

Lecture : les personnes de 25 ans en emploi au moins une fois sur l'année selon l'EEC sont 549 000 selon la pondération EEC, soit 88 % de la pondération des bases administratives (627 000) appliquée aux mêmes individus. À cet âge, 69 % des personnes enquêtées au 1^{er} trimestre et en emploi sont enquêtées sur les 4 trimestres (attrition).

Nota : le profil heurté des deux courbes d'effectifs vient du calcul à partir des PEA2 appariés, avec un nombre d'observations compris entre 200 et 300 pour chaque âge compris entre 20 et 30. Le calage sur les sources se fait par tranche d'âge (15 - 19 ; 20 - 24 ; 25 - 29 ; 30 - 39 ; 40 - 49 ; 50 - 59 ; 60+).

F - Ajout des trous de « collecte » des bases administratives

Nota : on appelle ici « trou de collecte », toute absence des fichiers de diffusion de données individuelles, que cette absence soit constatée dans les fichiers initiaux (trous de collecte proprement dits) ou soit la conséquence de l'épuration effectuée par les chaînes de traitement.

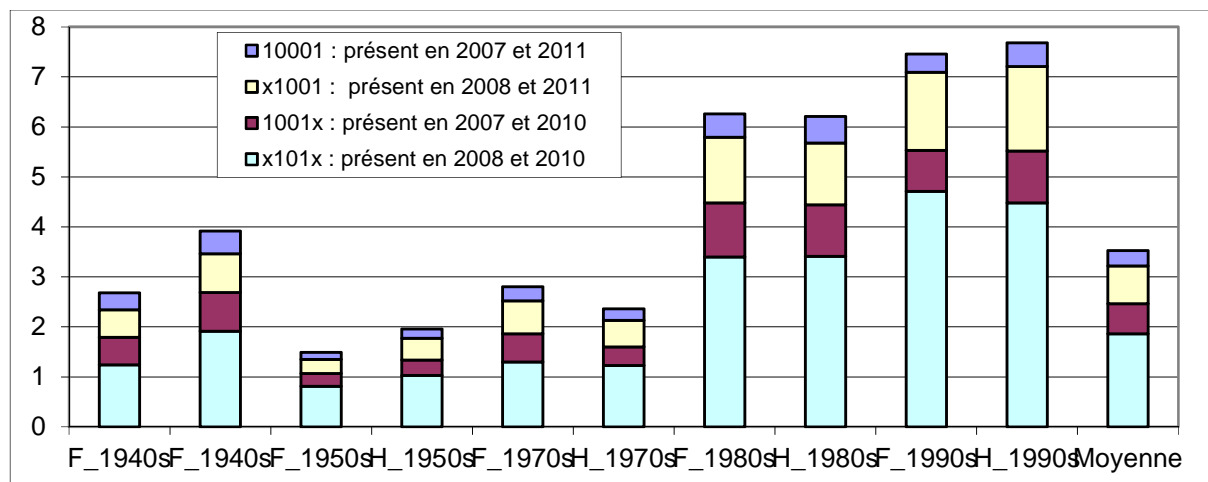
La collecte des DADS porte sur des postes, non des individus. Concernant un individu, un trou de collecte peut être partiel (certains de ses postes manquent, mais pas tous) ou complet (il n'a aucun poste dans le fichier de diffusion).

Une première estimation, indépendante de l'appariement, est interne aux DADS et utilise la dimension panel. Une exploitation du panel au 1/12e montre que plus de 3 % des salariés présents dans les DADS (« classiques ») en 2008 (ou 2007) et en 2010 (ou 2011) ne le sont pas en 2009 et que les trous « simples » (« 101 » : présents en n-1 et n+1, absents en n) représentent 1,8% +/- 0,1 point des présents que ce soit en 2008, 2009 ou 2010 (1^{ère} année où le comptage peut se faire sur les DADS « grand format ») (graphique 4).

En retenant l'ordre de grandeur de 2 %, cela fait 600 000 personnes. Cette estimation présente à la fois un biais à la hausse (des allers-retours entre salariat et inactivité ou travail indépendant peuvent se traduire par des « faux » trous de collecte) et à la baisse (défauts de couverture partiel : certaines populations seraient absentes ou sous-représentées).

Ne pouvant estimer les trous de collectes complets à partir des taux d'appariement, trop corrélés aux catégories permettant de supputer des trous, on choisit une estimation a minima en se limitant aux individus observés en début d'appariement 2012 ou appariés en 2011, en emploi EEC en 2012, et absents du fichier postes de diffusion en 2012.

Graphique 4 : absences 2009 (quand présence avant et après) par sexe et année de naissance



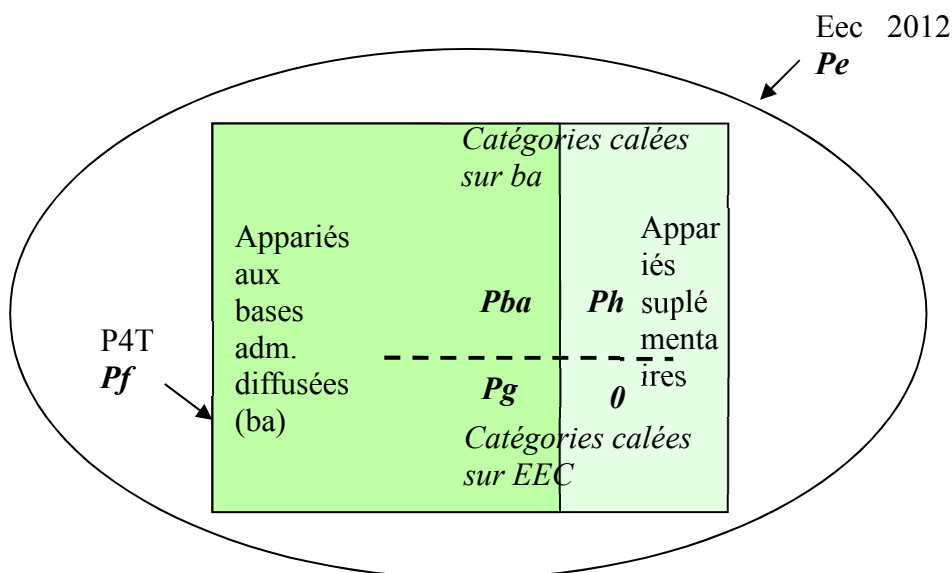
Lecture : 1,5% des femmes nées dans les années 1950 et présentes dans le panel Dads en 2008 (ou 2007) et en 2010 (ou 2011) sont absentes en 2009, dont 0,8% présentes en 2008 et 2010, 0,3% présentes en 2008 et 2011, 0,3% présentes en 2007 et 2010 et 0,1% présentes uniquement en 2007 et 2001.

Rappelons que nous travaillons ici sur les bases administratives de diffusion des données individuelles. Certaines corrections sont apportées, à partir de ces bases, dans le processus conduisant aux Estimations d'emploi.

Pour repérer les trous de collecte 2012 et en donner une estimation avec une pondération cohérente avec celle adoptée dans le reste de l'étude, on complète l'appariement retenu sur le champ des personnes présentes les 4 trimestres de 2012 (P4T 2012) par :

- Les appariés absents du champ DADS de diffusion;
- Les appariés en 2011 et/ou 2013 pour ceux enquêtés ces années.

Schéma 2 : différentes sous populations de l'univers des répondants à l'EEC et pondérations associées



Lecture du schéma : la pondération calée sur l'EEC s'effectue à plusieurs niveaux : d'abord calcul des poids P_f par calage des P4T sur l'EEC 2012 (avec les poids de l'EEC divisés par 4 : P_e) ; puis un calage des différents échantillons d'appariés sur les P4T, ce qui donne des poids P_g pour l'échantillon apparié sur les bases administratives de diffusion et un poids P_h ($P_h < P_g$) pour l'échantillon plus large incluant

les appariés supplémentaires. Les trous de collecte identifiés ne sont ajoutés que pour les catégories où la pondération issue de bases administratives (*Pba*) a été privilégiée (3^e condition infra).

Les P4T 2012 étant un sous-champ des enquêtés de rang 1 à 3 en 2012 T1, ceux de rang 1 et 2 sont encore enquêtés en début 2013 et ceux de rang 2 et 3 ont été enquêtés en fin 2011.

Parmi les appariés supplémentaires, sont retenus comme trou de collecte ceux réunissant les trois conditions suivantes :

- PEA selon l'EEC en 2012 ;
- Absents des bases administratives de diffusion 2012 : certains ajoutés au titre de l'appariement 2011 ou 2013 peuvent en effet être dans les bases administratives de diffusion 2012 sans avoir été appariés en raison par exemple de variables d'identification de moins bonne qualité en 2012 ;
- Appartient à une catégorie calée sur les bases administratives.

La 3^e condition sert à éviter les double-comptes : cette démarche ne peut s'appliquer qu'aux catégories où la pondération issue des bases administratives (*Pba*) a été privilégiée ; là où c'est la pondération EEC (*Pg*) qui a été privilégiée (particuliers employeurs, agriculteurs...), i.e. là où la couverture des bases administratives est absente ou partielle, les bases administratives ne servent pas de socle à l'estimation des PEA et il serait absurde de compléter cette estimation par des individus présents dans les bases administratives dans d'autres fichiers que le fichier de diffusion 2012. Cela conduit à retenir pour présenter les résultats dans ce chapitre la typologie suivante :

	Calés sur les bases administratives	Calés sur l'EEC
Salariés selon l'EEC*	Salariés hors PE (ShPE)	Particuliers Employeurs
Non salariés selon l'EEC	autres non-salariés	NSA + NSF + NSR**

* puisque l'on ne connaît leur situation en 2012 que via l'EEC

** NSA : Agriculteurs ; NSF : aides familiaux ; NSR : artistes

Pour les catégories ShPE et NSZ on peut additionner les observations absentes des sources administratives et calées sur l'EEC avec un poids *Ph* aux observations de l'appariement 2012 sur les sources administratives et calées sur ces sources avec un poids *Pba*.

Tableau 11 : Ajouts à l'appariement 2012

	En emploi dans les bases administratives 2012			Pas en emploi dans les bases administratives 2012				Total	% ajout
				Multi actifs	Non salarié s	Salariés	Sous total		
	Appariés 2012 absents bases administratives diffusion	En emploi EEC 2011 et/ou 2013	Pas en emploi EEC						
Aides familiaux, agriculteurs, artistes	35	502	71	103	49	39	15	711	
Autres non salariés	149	1 826	161	68	30	34	4	2 204	3,0
Particuliers employeurs	24	13	1 013	20	5	12	3	1 069	
Autres salariés	263	139	23 000	297	163	115	19	23 700	1,3
Total	472	2 480	24 240	487	246	201	41	27 680	

Champ : PEA 2012 selon l'EEC (pondération Ph)

Lecture : en 2012, selon l'EEC, il y a 23 700 000 salariés hors Particuliers employeurs, en emploi au moins une fois dans l'année. Parmi eux, 297 000 ne sont pas en emploi dans les bases administratives de diffusion 2012, dont 163 000 appariés avec les fichiers 2012 en amont des fichiers de diffusion et 134 000 (115 000 + 19 000) absents de ces fichiers amont mais appariés en 2011 et/ou 2013. On ne retient pas les 19 000 qui n'étaient pas en emploi selon l'EEC l'année de leur appariement. On retient

donc au total, pour les salariés hors PE, 278 000 (163 000 + 115 000) personnes en emploi au titre des trous de collecte DADS, soit 1,3 % de l'effectif EEC total.

L'estimation retenue, de 342 000 (somme des 4 cellules en gras du tableau 11) est contrôlée de l'activité salariée dans l'EEC en 2012. Elle corrige ainsi le biais à la hausse mais pas celui à la baisse : il s'agit donc d'une estimation plancher.

Des compléments sur le devenir en 2013 des appariés en 2012 sont présentés en annexe 1.3.

G - Synthèse sur les PEA

On peut séparer la population des résidents en France (hors communautés), couverte par l'EEC, en « personnes en emploi dans l'année » (PEA) et jamais en emploi dans l'enquête Emploi (JEMP). La quantification, *via* calage de l'échantillon apparié sur une source de référence, nécessite de travailler à un niveau plus fin. Il est en général préférable de caler sur les bases administratives, réputées exhaustives, qui procurent des marges de calage plus robustes mais pour les catégories où la couverture par les bases administratives est déficiente (catégories identifiées par la confrontation des deux calages pour les PEA2), il est préférable de caler sur l'EEC. Le tableau suivant résume l'arbitrage adopté dans chacun des cas de figure :

Situation EEC	Catégorie*	Présence bases administratives de diffusion 2012	
		Oui	Non
PEA	Non- salariés ; agriculteurs, artistes, aides familiaux Salariés : particuliers employeurs	PEA2_A : Poids EEC	Non ajoutés
	Autres non-salariés et autres salariés	PEA2_B : Poids bases administratives	TCOL : Poids EEC
JEMP		JEMP : Poids bases administratives	Ne sont pas PEA

* On repère d'abord les NAA (agriculteurs, artistes, aides familiaux) dans l'EEC ; pour les autres, la catégorie est déterminée à partir des bases administratives.

On obtient au final 32,2 millions de personnes en emploi dans l'année (PEA), soit 0,6 million de plus que dans les bases administratives et 3,5 millions de plus que dans l'EEC (tableau 12).

En relatif, la plus forte correction, par rapport aux bases administratives, concerne les catégories de non-salariés mal couvertes (+ 40 %) ; en absolu, elle concerne les trous de collecte pour les salariés hors PE, estimés à 278 000. Alors que les non-salariés ont peu de JEMP, l'autre catégorie où la pondération EEC est privilégiée, les salariés des PE, en ont beaucoup : 13 % des PE selon les bases administratives ne sont jamais en emploi selon l'EEC. Il y a ainsi plus de salariés des PE dans les bases administratives que dans l'EEC, même si l'analyse de PEA2 conduit à estimer que, pour les personnes se déclarant PE dans l'EEC, la source PE les sous-estime un peu (4 %).

Cette estimation de 32,2 millions peut sembler élevée. Elle part toutefois d'un socle déjà élevé : il y a 32,0 millions d'identifiants différents dans les bases administratives de diffusion. Si on accepte l'hypothèse que les postes conservés dans les DADS à l'issue des chaînes de traitement correspondent bien à de l'emploi et que les chaînes de traitement non-salariés ne conservent bien que les non-salariés actifs²⁴, la seule marge d'incertitude sur ce socle concerne les 0,7 million de NIR incorrects (et le décalage de paie : mais cela se compense d'une année sur l'autre et l'impact est marginal : on en donnera une estimation au III). Ces NIR incorrects ne peuvent pas être appariés (même si l'appariement n'utilise pas le NIR) et l'appariement n'est donc d'aucun secours pour savoir combien de PEA supplémentaires apportent réellement ces NIR incorrects (par rapport aux 31,3 millions de PEA à NIR correct). On a fait le choix de les prendre avec une pondération de 1/2, ce qui laisse une incertitude de +/- 0,3 million sur les 31,6 millions du calage sur bases administratives ainsi obtenu. Comme ces PEA à NIR incorrect sont en moyenne moins souvent en emploi, l'ampleur de l'incertitude en termes d'emploi est plus faible.

²⁴ On traite au III.D des indicateurs d'activité pour les non-salariés. 48 000 non-salariés JEMP sont alors exclus des PEA.

Tableau 12 - estimation du nombre de personnes en emplois dans l'année (PEA)

En milliers

	Calage bases administratives (A)	Calage EEC (B)	Calage final				Autre lecture du total
			PEA2	JEMP* (C)	TCOL (D)	Total	
Aides familiaux, agriculteurs, artistes	546	760	760	4	0	764	= B + C
Autres non salariés	2 463	2 092	2 307	156**	63	2 527**	= A + D
Particuliers employeurs	1 057	937	937	142	0	1 079	= B + C
Autres salariés	27 569	23 895	24 748	2 821	278	27 847	= A + D
Total	31 636	27 684	28 752	3 123	342	32 216**	

* pour les non-salariés, les bases administratives ne permettent pas de repérer les artistes et les aides familiaux ; les non-salariés JEMP issus de la source MSA sont classés en NAA et ceux de la source Acoess en NSZ ;

** parmi ces 156 000 JEMP NSZ, 48 000 seront finalement considérés comme inactifs à la suite de l'examen en partie III. Le nombre de PEA NSZ sera donc alors de 2 479 000 et le nombre total de PEA de 32 169 000.

Par rapport à ce calage, un peu plus de 0,2 million sont ajoutés par calage sur l'EEC pour les catégories de non-salariés mal couvertes par les bases « non-salariés ». Pour chacune de ces catégories de non-salariés, des informations externes à l'appariement confirment l'existence d'un défaut de couverture des bases « non-salariés ». La précision de l'ampleur de cet ajout est celle de l'EEC sur des populations de taille relativement restreinte.

Un peu plus de 0,3 million sont ajoutés au titre de trous dans les bases administratives. De nombreux indices montrent que de tels trous existent. On a retenu ici une estimation prudente puisque les PEA ajoutés sont ou ont été dans les sources administratives (en amont des bases de diffusion ou l'année antérieure) et sont en emploi selon l'EEC, tout en évitant les double-comptes en partant d'une base commune (tout apparié retenu avant ajout et tout ajout appartient à l'EEC 2012 restreinte aux répondants les 4 trimestres) et en se limitant aux catégories non déjà calées sur l'EEC. Reste, pour ces ajouts, un doute sur l'effet qualité de l'appariement, que le nombre très faible d'observations ne permet pas de tester. On a par prudence éliminé ceux qui étaient en incohérence sur la situation d'emploi l'année de l'appariement (si un individu est apparié en 2013 aux bases administratives et n'a pas d'emploi observé dans l'EEC en 2013 alors on ne le retient pas).

En fait, on ne dispose pas de mesure de référence déjà publiée pour le concept de PEA. Si la mesure proposée ici peut paraître élevée, c'est par comparaison au niveau d'emploi publié. A noter qu'une partie de l'écart, celle due à la différence de concept, provient de la part des parcours d'emploi discontinus (de nombreux PEA ne sont pas en emploi de manière continue sur toute l'année).

L'enseignement le plus important de cette partie concerne toutefois l'EEC avec l'identification de deux facteurs affectant l'estimation de l'emploi des jeunes :

- La sous-déclaration de l'emploi : plus de 3 millions de personnes jamais en emploi dans l'EEC se retrouvent au moins une fois dans les bases administratives en 2012 (JEMP) ; parmi les 2,9 millions de JEMP salariés, 1,2 million en emploi DADS lors d'au moins une des quatre semaines de référence. Les emplois concernés étant le plus souvent des « petits boulots », l'effet induit sur l'emploi en moyenne annuelle est bien plus faible : une estimation est proposée en partie III.
- La sous-pondération des PEA, par rapport aux chômeurs et inactifs, qui concerne quasi exclusivement les jeunes : cela conduirait à une sous-estimation de près de 1,2 million de PEA : aux 0,9 million de PEA ajoutés lors du calage de certaines catégories sur les bases administratives au II.E s'ajoutent les 0,3 million de PEA aux bases administratives au II.F pour ces mêmes catégories.

III – Passage à l'estimation de l'emploi moyen annuel

A - Démarche, concept d'emploi et préliminaires

On rappelle la décomposition qui structure ce travail : $EMA = PEA * DMA$.

Cette partie est consacrée à l'estimation, pour chaque observation, de la durée moyenne en emploi sur l'année (DMA). Comme PEA est déjà estimé en partie II, les résultats seront présentés ici plus souvent en termes d'emploi moyen annuel (EMA).

On estime la durée moyenne en emploi sur l'année (DMA) par la moyenne de l'emploi des 4 semaines de référence de l'EEC. Comme l'enquête Emploi est collectée en continu sur toutes les semaines de l'année, il s'agit un estimateur sans biais de l'emploi annuel moyen qui serait calculé comme la moyenne de l'emploi, au sens du BIT, des 52 semaines de l'année.

1 - Un traitement spécifique pour les salariés purs

La comparaison entre la DMA de l'EEC et celle des sources administratives n'est possible que pour les DADS, le concept d'emploi dans les bases non-salariés se limitant à la présomption d'activité.

Se pose alors la question des pluriactifs. Alors que l'emploi des DADS est supérieur à l'emploi EEC pour les salariés purs, même en excluant les JEMP (Jamais en EMPloi selon l'EEC), c'est l'inverse pour les pluriactifs (tableau 1).

Tableau 1 : comparaison de l'emploi en moyenne sur l'année, EEC et DADS

		Nb observations	Emploi (en milliers)		Emploi des PEA2*	
			EEC	DADS**	EEC	DADS
SAL	Salariés purs	17 564	22 791	23 905	22 662	22 971
PSA	Principalement salariés	398	614	572	601	551
PNS	Principalement non-salariés	145	254	126	254	126

Champ : Hors non-salariés purs et hors ajouts pour trou de collecte

* Ici : PEA2 = en emploi EEC et en emploi DADS au moins une des 4 semaines de référence

** Emploi moyen des 4 semaines de référence, après traitement des dates explicité au III - B.

Cela conduit à retenir l'emploi EEC pour les pluriactifs hors JEMP, faute de comparaison possible avec les sources administratives (tableau 2).

Tableau 2 - Emploi retenu selon les cas

		PEA2	JEMP
SAL	Salariés purs		Arbitrage (cf. III - C)
PSA	Principalement salariés	Emploi EEC	Emploi DADS
PNS	Princip. Non-salariés	Emploi EEC	Emploi bases «non-salariés» : toujours en emploi*
SNA	Non-salariés purs	Emploi EEC	Emploi bases «non-salariés» : toujours en emploi*

* après élimination de certains en fonction des critères d'activité des bases «non-salariés» (cf. III - D)

Ceci introduit une asymétrie de traitement entre salariés purs et salariés principalement salariés (PSA), alors que la part des revenus d'activité non-salariée dans les PSA est le plus souvent très faible, mais l'enjeu se révèle insignifiant : puisque, comme on le verra au III - C, l'emploi EEC est considéré comme de l'emploi, la seule différence porte sur le traitement des absences EEC (en emploi une semaine donnée selon les DADS et pas selon l'EEC) ; or les absences EEC ne concernent que 12 000 emplois

des PSA. De plus, pour compenser, alors que pour les salariés purs l'emploi des JEMP est arbitré, l'ensemble de l'emploi DADS des PSA JEMP (21 000 emplois) est retenu.

2 - Une différence de concept : les absences de longue durée

Pour comparer la situation dans l'EEC à l'emploi DADS, il faut aller plus loin que la seule ventilation selon la variable d'activité au sens du BIT : en emploi, au chômage et inactif. On décompose les situations d'emploi au sens du BIT en 4 modalités :

- Salarié au sens strict : à la fois en déclaré et en affecté selon la variable STATUT
- Statut salarié ou non-salarié indécis : différence entre déclaration et affectation ou dirigeant salarié
- Non-salarié au sens strict
- Frontalier (les 3 autres modalités ne concernent que les non-frontaliers)

Et les situations de non-emploi au sens du BIT en 4 modalités

- En emploi rémunéré mais en absence et/ou congés de longue durée (variable RASB - raison d'absence - renseignée) : ne sont pas considérées, en 2012, comme de l'emploi au sens du BIT les absences pour congés maladie de plus d'un an ou pour congé parental ou de formation non rémunérée de plus de 3 mois.
- A trouvé un emploi mais qui commencera plus tard
- Au chômage au sens du BIT et dans aucun des deux cas précédents
- En inactivité et dans aucun des deux cas isolés.

Parmi les situations d'emploi selon l'EEC, on vérifie que le travail frontalier (résident en France travaillant à l'étranger) est le fait de personnes peu souvent appariées, les DADS concernant l'emploi rémunéré par des établissements situés en France (tableau 3). Parmi les situations de non-emploi selon l'EEC, les deux cas isolés se distinguent. Avoir un emploi plus tard est le fait de personnes le plus souvent appariées (76 %) ce qui est normal s'il s'agit d'un emploi salarié et que la promesse d'embauche se concrétise, mais assez peu souvent, pour les personnes appariées, en emploi DADS la semaine correspondante (29 %). Les personnes en emploi rémunéré mais non classées en emploi au sens du BIT pour cause de congés de longue durée (ligne eCon) sont un peu moins souvent appariées (60 %) mais plus souvent en emploi DADS (69%) la semaine correspondante quand elles sont appariées. En effet, les DADS recensent l'emploi rémunéré et non la présence effective en emploi. Il s'agit là d'une différence de concept.

Tableau 3 : appariement et emploi DADS selon la situation dans l'EEC

Situation EEC		Nb** (milliers)	% appariés	% sal. pur	Sal. pur : % emp_DAD S	Emp_DAD S sal pur	Emp_DAD S total**
aSal	Salariés "strict"	22 249	83	98	98	21 906	22 396
bSNa	Sal / Non-Sal	413	75	54	97	216	241
cNsa	Non-salariés "strict"	2 639	76	6	72	128	273
dFro	Frontalier	370	8	65	59	16	17
eCon	Absence longue durée	345	60	89	68	173	176
fFut	Emploi plus tard	300	76	90	29	82	85
hCho	Chômage	2 540	49	95	22	373	384
ilna	Inactivité	21 724	11	93	33	1 011	1 035
Total		50 580				23 905	24 608

Champ : PE4T, situation trimestrielle

Lecture : En 2012, en moyenne, 345 000 personnes ne sont pas en emploi au sens du BIT tout en ayant un emploi rémunéré (longue absence, ligne eCon). 60 % de ces situations sont le fait de personnes appariées aux bases administratives, dont 89 % sont salariés purs. Parmi les appariés salariés purs, 68 % sont en emploi DADS la semaine où ils sont enquêtés et dans cette situation (eCon). Au total, 173 000 emplois DADS de salariés purs correspondent à cette situation dans l'EEC, 176 000 si on élargit à l'ensemble des salariés DADS.

** Le nombre de situations dans l'EEC utilise la pondération des PE4T calée sur l'EEC alors que le nombre d'emploi DADS utilise la pondération établie au II, qui donne un nombre plus élevé de PEA.

Cependant, seule la moitié des 345 000 situations d'absence de longue durée dans l'EEC correspondent à de l'emploi DADS (176 000 emplois DADS mis en regard avec l'appariement). On ne peut donc pas, dans les comparaisons, considérer toutes les absences de longue durée comme de l'emploi rémunéré au sens des DADS. On gardera donc l'emploi EEC au sens du BIT quand on compte l'emploi EEC, tout en tenant compte des absences de longue durée dans la comparaison avec l'emploi DADS ce qui conduit à considérer certains JEMP comme en emploi au sens des DADS dans l'EEC (cf. cellules signalées par un * dans le tableau 4).

3 - Emploi DADS : moyenne annuelle vs moyenne des semaines de référence

La séparation entre PEA2 et JEMP est tributaire des semaines de référence. À ce titre, il faut aussi catégoriser les DADS en fonction de leur emploi DADS les semaines de référence (ce n'est possible qu'après traitement des dates dans les DADS ; on anticipe ici les résultats du III - B), ce qui fait apparaître des nouvelles catégories - par exemple les 239 000 PEA2 qui ne sont pas en emploi DADS les semaines de référence - délicates à traiter (tableau 4). Cette séparation introduit aussi un biais de sélection dans le calcul de l'emploi moyen des DADS : l'emploi moyen hors JEMP est plus élevé calculé sur les semaines de référence (car ils sont en emploi EEC, et donc le plus souvent en emploi DADS, au moins l'une de ces semaines) que sur l'ensemble des semaines. On vérifie que, sur l'ensemble des cellules, l'emploi moyen calculé sur les 4 semaines de référence (réparties aléatoirement sur toute l'année) est très proche de celui calculé sur les 52 semaines de l'année : 23 905 milliers vs 23 852, soit un écart de 0,2%.

Tableau 4: Emploi des salariés purs, en fonction de leur situation annuelle

En milliers

	Emploi DADS (semaine réf.)	Nb PEA		Nb emplois EEC (BIT)		Nb emplois DADS moy. sur sem. ref.		Nb emplois DADS moy. sur 52 sem.	
		Emploi EEC		Emploi EEC		Emploi EEC		Emploi EEC	
		OUI	NON	OUI	NON	OUI	NON	OUI	NON
PEA2	OUI	24 925		22 662		22 926		22 820	
	NON	239		129				24	
JEMP	OUI	92*	1 662	0*		978		44	868
	NON		1 133				0		97
Total		28 051		22 791		23 905		23 853	

Champ : salariés purs

Lecture : 24,9 millions de PEA sont en emploi au moins une des 4 semaines de référence, à la fois selon l'EEC et selon les DADS. Leur emploi moyen ces 4 semaines est de 22,7 millions selon l'EEC et de 22,9 millions selon les DADS. Leur emploi moyen calculé pour les 52 semaines à partir des DADS est de 22,8 millions.

** 92 000 PEA sont classées en JEMP car sans emploi au sens du BIT dans l'EEC mais sont au moins un trimestre à la fois en emploi DADS et en absence longue durée EEC*

On traitera ensemble les PEA2 et les JEMP dans l'examen des incohérences entre EEC et DADS pour les salariés purs (cf. III - C).

B - La question des dates et des fusions de période dans les DADS

Deux particularités des DADS peuvent susciter un doute sur la validité des dates observées : le décalage de paie et la fusion des périodes.

Le décalage de paie consiste à payer le mois m les heures effectuées le mois m-1. Pour des déclarations annuelles, cela implique que certaines périodes de décembre ne seront présentes que dans le fichier de l'année n+1 et, a contrario, que le fichier de l'année n contient des périodes de décembre n-1.

La fusion des périodes s'effectue à deux niveaux :

- Au niveau de la déclaration quand, par exemple pour le travail intermittent ou l'intérim, l'entreprise déclare sous une seule période continue des périodes de travail discontinues ;
- Au niveau des chaînes de traitement DADS : il y a une seule observation par poste de travail dans le fichier « postes » ; un deuxième jeu de dates de début et date de fin permet de conserver l'information pour les postes à 2 périodes (sur l'échantillon non pondéré, 6 % des postes ont 3 périodes ou plus).

Comme on dispose du fichier des périodes, seul le premier niveau de fusion nous concerne ici. On vérifiera l'impact du second niveau en comparant les incohérences avec l'EEC selon que l'on part du fichier « postes » ou du fichier « périodes ».

1 - Le décalage de paie

Le décalage de paie a plusieurs impacts potentiels :

- Sur le nombre de PEA : certaines personnes en emploi uniquement en décembre peuvent être absentes du fichier 2012 (leur paie de décembre est dans le fichier 2013) ; réciproquement, certaines personnes en emploi en décembre 2011 mais pas en 2012 peuvent être dans le fichier 2012.
On vérifie que ces différentes sources de biais se compensent.
- Sur les dates de présence et donc sur la cohérence entre EEC et DADS pour les PEA2.
En prenant comme critère la minimisation des écarts - en termes de présences en emploi - avec l'EEC, on constate qu'il vaut mieux, pour la confrontation au cas par cas, partir des dates déclarées par les entreprises, plutôt que des dates corrigées par les chaînes de traitement (corrigées en fonction de critères statistiques pour assurer la cohérence avec les salaires)²⁵.

On adopte le traitement suivant :

- Ajouter les périodes du fichier 2013 concernant le mois de décembre 2012
- En l'absence de périodes dans le fichier 2013, prolonger les périodes se terminant au 30 novembre 2012 jusqu'au 31 décembre dans les cas suivants
 - Période débutant en 2011 (signe de décalage de paie)
 - Période débutant en 2012 et décalage de paie déclaré par l'entreprise

Des compléments sur le décalage de paie sont proposés en annexe 1.4

2 - Durée, nombre d'heures et densité

On traite ici de la question de la fusion des périodes à travers la cohérence entre durée de la période, nombre d'heures déclarées et quotité de travail. Une fusion de période devrait se traduire par une incohérence entre ces 3 paramètres.

²⁵ Cette comparaison entre dates déclarées et dates corrigées, détaillée en annexe, couvre non seulement la question du décalage de paie mais aussi celle des trop faibles densités (cf. infra)

Nota : sauf indication contraire, les chiffres donnés dans cette partie sont ceux, non pondérés, de l'ensemble des appariés, soit 73 000 personnes, 110 000 postes et 156 000 périodes

Le contrôle de **cohérence interne** repose essentiellement sur la notion de densité, le nombre d'heures divisé par la durée. On s'attend à une densité au moins égale à 5 (35 /7) pour les temps pleins. Pour les temps partiels, on normalise la densité en la divisant par la quotité (comprise entre 0 et 1).

Sur les 156 000 périodes utilisées,

- 82 000 sont déclarées à temps plein, dont 75 000 d'au moins une semaine. Parmi ces dernières, 23 % ont une densité trop faible (< 5)
- 33 000 sont déclarées à temps partiel, dont 27 000 d'au moins une semaine. Parmi ces dernières, 20 % ont une densité normalisée trop faible (< 5)
- 3 000 sont déclarées atypiques (intermittent, à domicile,...) avec quotité de temps partiel déclarée.
- 38 000 ne sont ni déclarées à temps plein, ni à temps partiel (pas de quotité) : elles sont majoritairement de très faible densité. La totalité des salariés des particuliers employeurs (14 000 périodes) sont dans ce cas de figure.

Le contrôle de **cohérence externe** mobilise le **taux d'absence EEC (TAE)** : % d'absence (en emploi) EEC les semaines où le salarié est en emploi selon les DADS. Un TAE élevé peut être le signe

1. D'une trop faible densité.
2. D'une sous-déclaration des petits boulots

Le second effet s'observe sur les périodes courtes : le TAE est de 1/3 pour les périodes de moins d'un mois et de 1/5 pour les périodes de 1 à 3 mois.

Pour observer le premier effet, on peut examiner les périodes couvrant toute l'année (où l'effet petit boulot peut être exclu, sauf très faible quotité) ; le TAE est alors de 1 % quand la densité normalisée est cohérente (entre 5 et 6,5) et de 6 à 7 % sinon.

L'imputation de la quotité, quand elle n'est pas renseignée dans les DADS, permet une comparaison avec l'EEC : il y a nettement plus de temps partiel selon les DADS que selon l'EEC, avec un écart concentré sur les plus faibles quotités où le soupçon d'intermittence est fort.

Le TAE est bien corrélé à la densité normalisée calculée avec les quotités imputées. Mais pour départager entre effet « fusion des périodes » (emploi DADS à tort) et effet sous-déclaration des petits boulots (absence EEC à tort), il faut considérer la situation sur plusieurs trimestres (partie suivante).

Des compléments sur le traitement des dates sont proposés en annexe 1.5

C - Traitement des salariés purs

Parmi les environ 23 millions d'emplois de salariés purs, il y a pour environ 2 M d'entre eux désaccord entre la situation selon l'EEC et la situation selon les Dads (tableau 5)

Tableau 5 : désaccord entre sources sur l'emploi des salariés purs une semaine donnée

Emplois, en milliers		En emploi EEC			Total emploi Dads	
		Oui	Non, congés rémunérés longs	Non, autres	Emploi rémunéré	Concept BIT*
En emploi Dads	Oui	22 266	173	1 466	29 905	23 732
	Non	525				
Total emploi EEC		22 791				

Champ salariés purs (présents dans les Dads, pas dans les bases non-salariés)

* hors congés rémunérés de longue durée

Un désaccord sur la situation d'emploi peut relever soit d'une erreur d'appariement (on se réfère en fait à deux personnes différentes), soit d'une incohérence sur la situation d'emploi d'une personne bien identifiée.

D'un point de vue logique, une incohérence sur la situation d'emploi d'une personne bien identifiée relève forcément d'un des quatre cas suivants :

Absence EEC

- Présence à tort dans les DADS (ex. suite à fusion de périodes)
- Absence à tort dans l'EEC (ex. sous déclaration des petits boulots)

Absence DADS

- Présence à tort dans l'EEC
- Absence à tort dans les DADS (ex. trou de collecte)

Seul le cas de présence à tort dans l'EEC n'a à ce stade pas encore été examiné. Rejeter cette possibilité permettrait d'adopter une stratégie d'estimation. En effet, si toute situation déclarée comme de l'emploi dans l'EEC est effectivement de l'emploi, alors toute absence DADS est une absence à tort. On pourrait alors considérer l'emploi EEC comme socle de l'emploi et arbitrer (entre emploi et non-emploi) sur les seules absences EEC. Il peut y avoir bien entendu des absences EEC dues à des erreurs d'appariement. Les erreurs d'appariement pouvant être assimilées à du bruit (sans biais a priori sur le niveau d'emploi, en supposant l'absence de lien entre erreur d'appariement et situation d'emploi), il faut juste s'assurer que celles qui sont comptées en emploi au titre des absences DADS ont leur contrepartie comptée en non-emploi du côté des absences EEC.

On commencera donc par justifier cette hypothèse – tout emploi EEC est effectivement de l'emploi - (1), ce qui permettra ensuite de se concentrer sur les absences EEC et de passer en revue les facteurs pouvant expliquer de telles absences, d'abord pour les PEA (2), puis pour les JEMP (3). Ces différents facteurs seront alors mobilisés pour un exercice d'affectation (emploi ou non) des absences EEC (4).

1 - Tout emploi déclaré dans l'EEC peut être supposé comme étant de l'emploi

En dehors des éventuels défauts d'appariement, deux facteurs pourraient augmenter la présence en emploi à tort dans l'EEC :

- L'individu enquêté n'est pas celui qui répond à l'enquête (le répondant est une personne « proxy » du ménage)
- La situation du trimestre précédent est reportée à tort le trimestre suivant (ce qui devrait se traduire par un effet rang d'interrogation)

Il semble y avoir un effet proxy, aussi bien pour les absences DADS (0,6 point) que pour les absences EEC (0,7 point, cf. tableau 6).

Tableau 6 : taux de discordance en fonction du répondant à l'enquête (en %)

	Taux d'absence DADS	Taux d'absence EEC
Répondant = enquêté	2,1	2,6
Répondant = proxy	2,7	3,3
Total	2,3	2,8

Champ : PEA2 salariés purs

L'effet proxy disparaît toutefois toutes choses égales par ailleurs (tenant compte de l'effet de sélection lié à l'appariement : un individu EEC apparié aux DADS a au moins une période d'emploi DADS) : les proxy sont plus fréquents pour les intérimaires et les apprentis et c'est plus l'effet forme d'emploi que l'effet proxy qui explique les absences DADS.

L'effet rang est délicat à analyser car l'échantillon panélisé utilisé ici n'est pas neutre : le rang 1 n'existe qu'au 1^{er} trimestre et le rang 6 qu'au 4^e trimestre. La modélisation des absences DADS ne montre aucun effet rang.

La contrepartie de l'hypothèse que tout emploi déclaré dans l'EEC est un emploi est l'acceptation que les absences DADS proviennent soit de décalages de dates dans les DADS, soit de trous de collecte. La présence de trous de collectes a déjà été documentée lors de la construction de l'appariement. Il s'agissait alors de personnes non présentes dans le fichier de diffusion 2012 mais présentes dans les DADS soit en amont du fichier de diffusion, soit en 2011 ou 2013, et en emploi EEC en 2012. Comme les DADS sont émises par les employeurs, l'absence d'une personne signifie l'absence de tous ses postes. On doit donc s'attendre aussi à des absences partielles, portant sur un ou plusieurs postes, mais pas tous, de la personne.

L'analyse des transitions le confirme. En effet, lorsqu'une personne est en emploi en T à la fois selon les DADS et l'EEC et en emploi en T+1 selon l'EEC, le taux d'absence DADS est 4 à 5 fois plus élevé quand un changement d'emploi est identifié à travers l'EEC (3,8 %) que quand aucun changement n'est identifié (0,8 %) (tableau 7). Seul un trou de collecte DADS portant sur le poste de T+1 permet d'expliquer un tel écart.

Tableau 7 : taux d'absence DADS en fonction de la transition EEC

Transition	Emploi sans changement	Changement emploi	Non emploi -> emploi	Total
Toute situation DADS en T				
Nb (milliers)	15 510	601	790	17 140
Tx abs. DADS	1,7	6,0	7,7	2,3
En emploi DADS et EEC en T				
Nb (milliers)	15 300	545		
Tx abs. DADS	0,8	3,8		

Champ : Transitions T -> T+1 des PEA2 en emploi EEC salarié en T+1

Certes les changements d'emploi étant très rares selon l'EEC, la majorité des absences DADS concernent des transitions sans changement d'emploi. Mais les transitions d'emploi en emploi sont difficiles à repérer dans l'EEC (on a du construire une variable spécifique) et sous estimées : sur 10 changements de postes repérés dans les DADS (repérage restrictif : le salarié est en emploi la semaine de référence de T et celle de T+1 mais aucun poste du salarié n'est présent à la fois en T et en T+1), 4 sont identifiés aussi comme changement de poste par la variable construite, 2 correspondent à une situation de non-emploi EEC un des deux trimestres et 4 à de l'emploi continu sans changement repéré : si le statut (emploi vs non-emploi, module A du questionnaire) n'est pas affecté par le report de la situation de T à T+1, la situation en emploi (module B) l'est.

2 - Absences EEC : des facteurs explicatifs significatifs mais trop ciblés

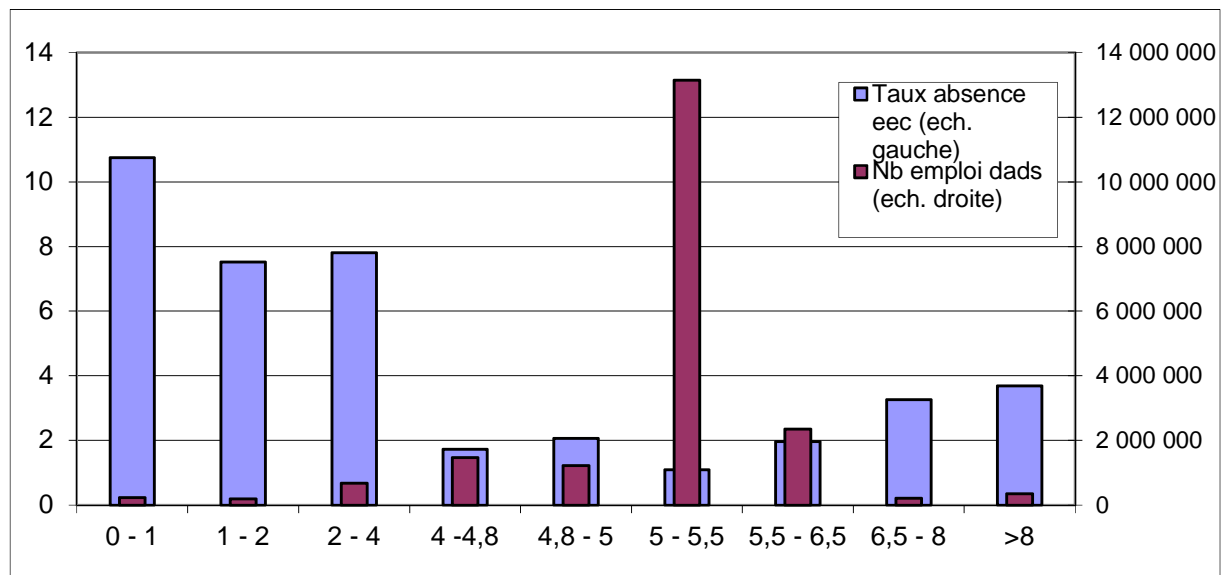
L'objectif de cette partie est de caractériser les facteurs permettant d'expliquer et interpréter les absences EEC, en vue de la procédure d'affectation de ces absences (II E.4). Chacun des facteurs examinés est à la fois très discriminant et très ciblé (seule une faible part de la population est concernée par ce facteur) et n'explique donc qu'une partie des TAE.

On reconstitue ici un fichier « postes » par agrégation des périodes en calculant pour chaque semaine un indicateur de présence et de quotité. Le poste principal est, pour une semaine donnée, celui de la plus grande quotité.

a - effet densité

On a déjà montré, pour les postes à une seule période en cours la semaine de référence, qu'une faible densité (nombre d'heures divisé par la durée, voir III.B.2) des périodes DADS augmentait le risque d'absence EEC. Après agrégation des périodes, on confirme cet effet mais on constate aussi que la plupart des emplois se situent dans les périodes de bonne densité. Ainsi, les densités comprises entre 4,8 et 6,5 regroupent 92 % des 20 millions d'emplois pour lesquels une densité corrigée de la quotité est calculable. Ils ont un taux d'absence EEC (TAE) inférieur à 2 % mais représentent tout de même 68 % des absences EEC (graphique1) : on ne peut expliquer les absences EEC par la seule densité.

Graphique 1 : taux d'absence EEC et nombre d'emplois DADS en fonction de la densité



Champ : PEA2 à temps plein ou à quotité renseignée, postes de plus d'une semaine (20 millions d'emplois)

Lecture : les postes de densité comprise entre 1 et 2 représentent 190 000 emplois, dont 7,5 % ne sont pas de l'emploi selon l'EEC

Ce critère de la densité tenant compte de la quotité, n'est calculable que lorsque que la quotité est déclarée. En effet, quand la quotité est imputée, on ne peut pas distinguer entre densité et quotité, et donc entre effet faible densité imputable aux DADS et effet faible quotité (« petit boulot ») imputable à l'EEC. Or le taux d'absence DADS est plus élevé quand la quotité n'est pas disponible : il est, pour les PEA2, de 8 % quand la quotité est imputée à partir du nombre d'heures contre 2 % quand la quotité est déclarée.

La comparaison des quotités entre sources a montré que la quotité imputée était souvent inférieure à celle de l'EEC (et donc que la densité était sans doute trop faible) mais, pour rendre compte des absences EEC, cette comparaison n'est possible que s'il y a au moins un autre trimestre où la personne est en emploi à la fois dans les DADS, sur le même poste que le trimestre où il y absence EEC, et dans l'EEC. Ceci conduit à s'intéresser aux transitions.

b - transitions DADS

On examine d'abord le cas le plus simple, et le plus fréquent, celui où le salarié est, selon les DADS, trois trimestres consécutifs (T-1, T et T+1)- sur le même poste principal. Il est aussi en emploi EEC en T-1 et/ou en T+1. Dans ces conditions, l'effet « petit boulot » sur l'emploi EEC en T est quasi inexistant et on peut mesurer un effet qualité de la quotité quasi pur. On compare pour cela la quotité DADS et EEC sur T-1 ou T. Si la quotité DADS est nettement inférieure à celle de l'EEC, alors il y a un risque que la durée de la période DADS soit trop longue (quotité = nombre d'heures / durée). Et on observe effectivement, dans ce cas, un TAE de 8 % alors qu'il n'est que de 1 % en l'absence de divergence sur la quotité.

Hors divergence, le TAE monte, pour les PEA2, à 23 % quand la personne en emploi DADS en T n'est pas en emploi DADS en T-1 et change de poste en T+1 (tableau 8, ligne NC). La divergence sur la quotité accentue l'effet de la transition puisque l'on cumule les absences liées à de la sous- déclaration EEC et celles liées à la faible densité DADS. Ainsi lors du passage du non-emploi DADS à l'emploi, avec poursuite sur le même poste le trimestre suivant, le TAE est de 9 % en l'absence de divergence sur la quotité en T+1 et de 17 % en cas de divergence (ligne NP).

Tableau 8 : taux d'absence EEC (TAE) en T en fonction du type de transition DADS et de la cohérence de quotité sur un autre trimestre

*		PEA2		Ensemble des salariés purs	
		Pas divergence	Divergence qt	Pas divergence (y.c. JEMP)	Divergence qt (possible que pour PEA2)
PP	Poste principal en t-1 et t+1	1	8	3	8
CP	Nouveau poste princ.	4	,	8	,
PC	Chgt poste en t+1	6	11	13	11
CX	Chgt poste t-1 -> t	6	,	14	,
PS	Autre poste princ en t+1	9	,	20	,
NP	Pas en emploi t-1 ou t+1	9	17	24	17
SS	Autre poste pr en t-1 et t+1	17	,	25	,
NC	Non emploi t-1, chgt t+1	23	,	45	,
NN	Emploi ni en t-1 ni en t+1	9	,	52	,

* code en 2 caractères : le 1^{er}, situation en t-1, le 2nd, en t+1 ; la situation est celle du poste principal en t.

P : poste principal ; S : poste secondaire ; C : en emploi, mais pas sur le poste de t ; N : non-emploi ; X : effet de bord (pour T1 et T4, on ne connaît qu'une des deux transitions)

Certains regroupements ont été effectués : ex PP = PP + XP + PX ;

L'utilisation des transitions montre aussi le biais qu'entraîne l'estimation sur le seul champ de PEA2 : on s'attendrait à ce que le TAE soit le plus fort pour les personnes en emploi isolé (en T mais ni en T-1, ni en T+1, ligne NN) mais ce n'est pas le cas, suite à un biais de sélection : les PEA2 sont par construction en emploi au moins une fois dans l'EEC et les absences EEC correspondant au cas NN sont le plus souvent des JEMP. Quand on réintègre les JEMP, le TAE progresse comme attendu. Il faudra tenir compte de ces différentes difficultés dans l'approche globale.

Enfin, là encore, c'est le cas de continuité sur le même poste principal, qui n'explique pas les absences EEC, qui est de loin le plus fréquent.

c - Présence d'un Siret commun

Pour les PEA2, un indicateur construit en comparant le Siret de l'EEC, quand il est renseigné, à celui des DADS, permet d'isoler d'une part un effet qualité de l'appariement - doute quand aucun Siret de l'EEC n'est retrouvé dans les DADS, ligne bNCO du tableau 9 - et d'autre part un indicateur de forte suspicion d'absence DADS - quand un Siret commun existe avec les DADS, donc très peu de doute sur la qualité de l'appariement, mais pas celui des DADS le trimestre considéré, ligne dCNT. Dans ce dernier cas, le TAE s'élève à 14%.

Tableau 9 : TAE en fonction de l'accord entre DADS et EEC sur le Siret

		Emploi DADS	TAE (%)	Nb absences EEC
aPAS	Aucun Siret dans l'EEC	3 268	3,1	100
bNCO	Siret dans l'EEC, aucun en commun avec DADS	3 866	2,4	92
cCNT	Siret commun avec DADS, pas celui du T	412	13,8	57
dSIC	Le Siret du trimestre DADS est dans l'EEC	13 952	0,9	132
ztot	Total	21 497	1,8	381

Champ : PEA2, hors intérim (pas de Siret utilisateur dans les DADS) et salariés des PE (pas de Siret)

Alors que l'effet indicateur de qualité de l'appariement présenté en I.A.3 joue surtout sur la part de PEA2 sans SIRET renseigné de l'EEC il a peu d'impact sur la part des cas où le Siret EEC n'est pas celui des DADS, part qui devrait augmenter si la qualité de l'appariement était moins bonne. Or cette part n'augmente que dans le cas AB, indicateur de bonne qualité mais appliqué aux grosses communes et aux immigrés (tableau 10).

Tableau 10 : indicateur de qualité de l'appariement et indicateur d'accord sur le Siret

Qid	aPAS	bNCO	cCNT	dSIC	Total*
A	13,6	17,5	1,4	67,5	11 138
AB	18,4	21,6	1,4	58,6	1 872
B	18,2	18,0	1,0	62,8	876
C	19,0	17,6	0,9	62,5	1 017
Total*	2 210	2 691	202	9 800	14 903

Champ : PEA2, hors intérim et salariés des PE

* nombre, non pondéré, d'appariés PEA2 avec au moins un poste hors intérim et PE

3 - Les Jemp sont-ils de l'emploi ?

On se restreint ici aux PEA salariés purs qui restent JEMP même en tenant compte des périodes de congés rémunérés ou d'emploi frontalier, soit 2,77 millions de JEMP, dont 1,65 million en emploi DADS au moins une des 4 semaines de référence.

On rappelle la distinction faite en introduction entre :

- JEMP_C (contingents) : ils ne sont pas en emploi DADS les semaines de référence (ils sont dans la colonne EhSref du tableau ci-dessous)
- JEMP_D (déclaratifs) : ils sont en emploi DADS les semaines de référence (colonne E_Sref)

À propos des JEMP, plusieurs doutes doivent être levés

- a. Cela relève du bruit de l'appariement
- b. C'est de l'emploi, mais il y a erreur sur les dates
- c. Ce n'est pas vraiment de l'emploi

a. Le doute relatif à l'appariement a déjà été examiné au II : les JEMP sont trop ciblés sur les étudiants et les retraités pour relever du bruit et ce n'est qu'aux âges intermédiaires, soit pour 1 JEMP sur 5, que l'on peut distinguer un effet de moindre qualité de l'identifiant.

b. L'effet dates peut être examiné à partir du cas symétrique : si un certain nombre de personnes sont, les semaines de référence, en emploi DADS mais pas en emploi EEC à cause d'une erreur sur les dates, on doit observer un nombre équivalent de personnes en emploi EEC mais pas en emploi DADS les semaines de référence : 255 000 PEA sont dans ce cas (tableau11). On peut tenter de capter le cas symétrique en se servant du calendrier de l'EEC ainsi que d'informations complémentaires comme l'ancienneté au chômage contrôlée par la date de dernier emploi. On a ainsi 515 000 PEA qui, selon l'EEC, sont en emploi uniquement hors des semaines de référence. Or 250 000 d'entre eux sont en emploi DADS au moins une des semaines de référence, ce qui laisse soupçonner une erreur sur les dates. La symétrie semble presque parfaite (250 000 et 255 000) quant au nombre de PEA concernés mais n'est plus vérifiée lors de la traduction en nombre d'emplois : en effet, les PEA en emploi DADS uniquement hors semaine de référence sont, en moyenne en emploi une semaine sur dix selon les DADS alors que ceux qui ne sont pas en emploi EEC les semaines de référence le sont, qu'ils soient détectés comme étant en emploi EEC hors semaine de référence ou non, une semaine sur deux. Une interprétation plausible est que les décalages de date ne rendent compte que partiellement des deux cas : il y aurait des trous de collecte DADS ou de la sous-déclaration EEC ou des erreurs d'appariement²⁶ parmi les 505 000 PEA en emploi EEC (en incluant le calendrier) pour lesquels une seule des deux sources les déclare en emploi au moins une semaine de référence. L'emploi des PEA en emploi DADS uniquement hors semaine de référence et en emploi EEC au moins une des semaines de référence, soit 26 000 emplois, serait alors un majorant du rôle des erreurs de date dans l'emploi des JEMP.

²⁶ Un indice : parmi les 1 651 000 JEMP_D, ceux qui sont en emploi EEC hors semaine de référence représentent 30 % des 30 - 49 ans, ceux pour lesquels la qualité de l'appariement a un impact, contre 15 % des 15 - 29 ans.

Tableau11 : Comparaison DADS EEC en fonction de l'emploi les semaines de référence

EEC DADS	PEA			Emploi DADS**			Emploi EEC***		
	En emploi sem. de référence		Total	E_Sref	EhSref	Total	E_Sref	EhSref	Total
	Oui (e_Sref)	Non (EhSref)							
PEA2:E_Sref*	25 027	255	25 283	22 895	26	22 921	22 965	141	23 106
Jemp : EhSref	250	265	515	117	29	146	44	42	86
Jemp : autres	1 401	852	2 253	720	65	785			
Total	26 679	1 372	28 051	23 732	121	23 852	23 010	183	23 192

Champ : PEA salariés purs.

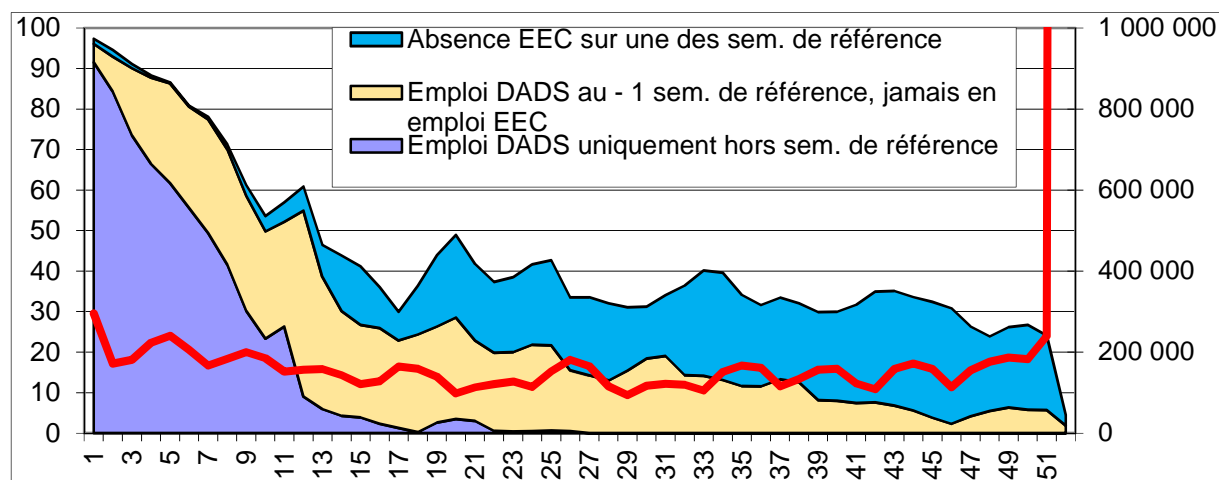
* y compris absences de longue durée en emploi DADS

** emploi calculé en nombre de semaine moyen d'emploi DADS (l'utilisation de l'emploi moyen des 4 semaines de référence donnerait un emploi nul dans la colonne EhSref).

*** emploi calculé en nombre de mois moyen d'emploi (à partir du calendrier eec) Lecture : Parmi les salariés purs (sources administratives) jamais en emploi les semaines de référence selon l'EEC, 515 000 se déclarent au moins une fois en emploi dans l'année (calendrier de l'EEC). Parmi ces derniers, 250 000 sont en emploi Dads au moins une des semaines de référence et 265 000 uniquement en dehors des semaines de référence. Ces derniers représentent un emploi moyen annuel de 29 000 dans les sources administratives et de 44 000 selon le calendrier EEC.

c. Le doute sur le statut (emploi ou non) des emplois DADS des Jemp est plus délicat à traiter. Rappelons d'abord que les chaînes de traitement DADS sont conçues pour éliminer les périodes qui ne sont pas de l'emploi. Ces chaînes ne sont sans doute pas parfaites et si, comme on a pu le constater, elles éliminent parfois à tort des périodes qui sont sans doute de l'emploi selon l'EEC, l'inverse est aussi possible. Le concept d'emploi des DADS (emploi rémunéré) n'est pas celui de l'EEC (emploi au sens du BIT) mais la seule traduction identifiée de cette différence de concept (les congés de longue durée) est déjà, à ce stade de l'étude, prise en compte en amont de la construction de la catégorie JEMP. Si la voie négative (identifier les emplois DADS qui ne seraient pas des emplois) est impraticable, on peut poursuivre sur la voie positive (identifier pourquoi une période d'emploi DADS n'est pas déclarée comme telle dans l'EEC) déjà empruntée dans la discussion sur les PEA : ces emplois devraient être surtout des petits boulots de personnes (étudiants...) ne se considérant pas principalement en emploi.

Graphique 2 : part des Jemp ou absence EEC en fonction du nb de semaines en emploi dans les DADS



Champ : PEA, salariés purs

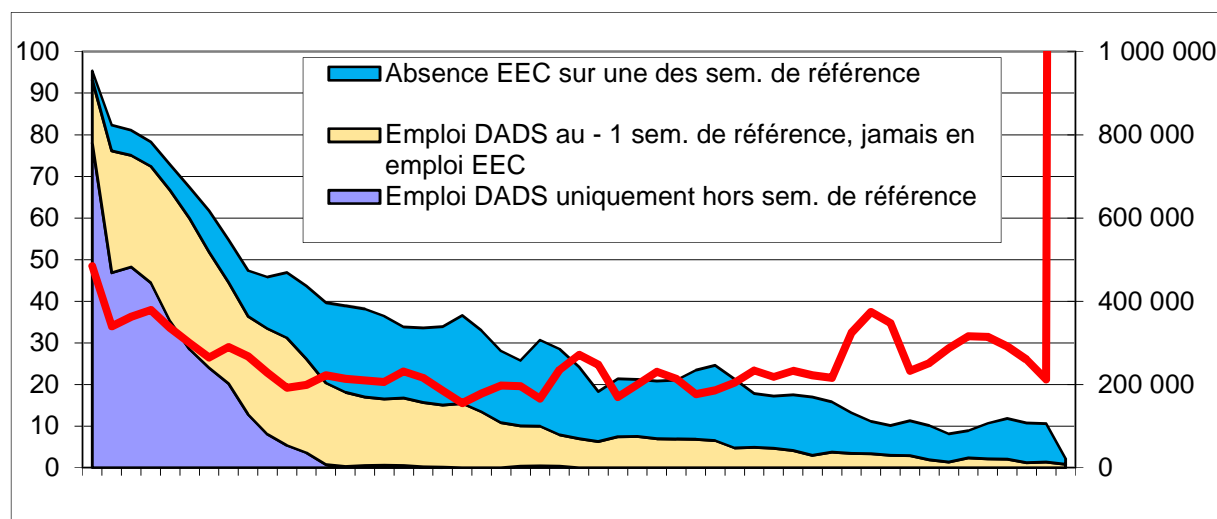
Lecture : 30 % des 200 000 PEA (courbe en gras) en emploi 9 semaines selon les DADS le sont en dehors des semaines de référence ; 28 % le sont pendant au moins une des semaines de référence selon les DADS mais pas selon l'EEC ; 3% sont des PEA2 avec au moins une semaine de référence en emploi DADS sans être en emploi EEC ; les 39% restant sont toujours, les semaines de référence, en emploi EEC quand ils sont en emploi DADS

La plupart des personnes en emploi moins de 6 semaines selon les DADS le sont hors des semaines de référence (graphique 2) : il n'est donc pas possible de savoir si elles se seraient déclarées en emploi EEC. Cette part devient négligeable après 12 semaines car il est rare (voire impossible s'il s'agit d'une seule période d'emploi) d'être en emploi plus de 12 semaines sans l'être au moins une des 4 semaines de références (12 semaines est l'intervalle séparant 2 semaines de référence).

Ce sont les JEMP_D, là où aucun emploi EEC n'est déclaré alors que la personne est, selon les DADS, en emploi au moins une des semaines de référence, qui soulèvent le plus de question quant à la réalité de l'emploi. Ce doute est d'autant plus fort que l'emploi selon les DADS est long. Si la part de JEMP_D est maximale (46 %) à 12 semaines, elle reste non négligeable (15 %) à 26 semaines.

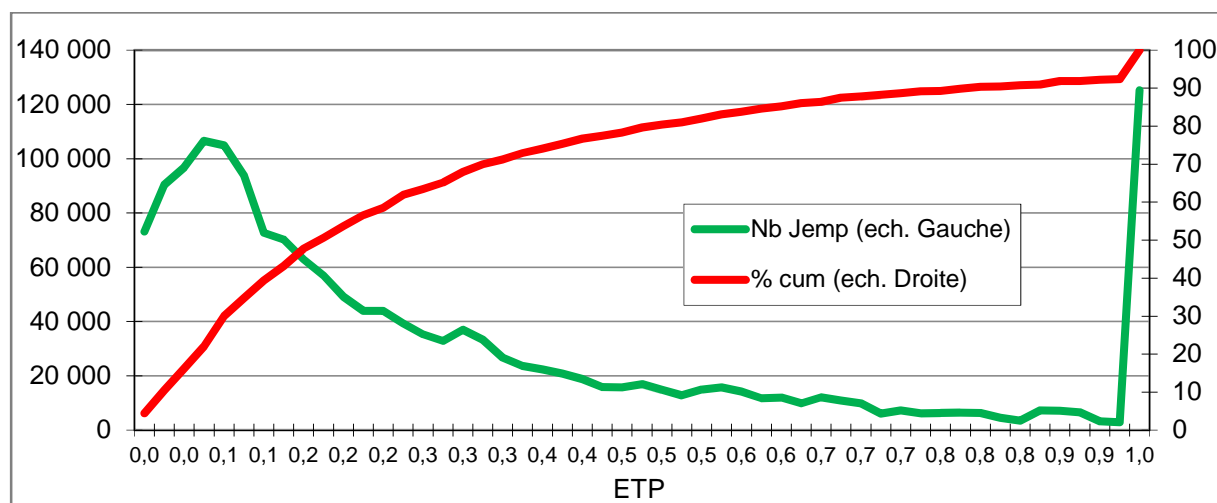
Les JEMP présents dans les DADS plus d'une semaine sur deux peuvent soit l'être à tort (problème de densité dans les DADS) soit ne pas être déclarés suite à une très faible quotité. C'est pourquoi l'etp est un critère préférable au nombre de semaines. Effectivement, la part de JEMP parmi les PEA DADS d'un demi etp n'est que de 7 % (graphique 3). 80% des JEMP ont moins d'un demi etp selon les DADS (graphique4).

Graphique 3 : nombre de JEMP (et part cumulée) en fonction de l'etp DADS.



Champ : PEA, salariés purs

Graphique 4 : nombre de JEMP (et part cumulée) en fonction de l'ETP DADS



Champ : PEA, salariés purs

Lecture : il y a 39 000 JEMP travaillant dans l'année 0,25 ETP (chaque point représente l'équivalent d'une semaine à temps plein ; le 13^e point représente donc 13/52 = 0,25 etp). 62% des JEMP travaillent moins de 0,25 ETP.

Il y a sans doute quelques erreurs d'appariement dans les JEMP_D à etp élevé : à partir de 0,7 etp, le taux d'apparié avec un identifiant à qualité plus douteuse augmente, ainsi que celui des 50 - 59 ans (tableau12). Mais même en considérant qu'au delà de 0,7 etp un JEMP sur cinq est du à une erreur d'appariement, cela ne concernerait que 50 000. On vérifiera au chapitre suivant que la procédure d'affectation des absences EEC mise en œuvre conduit bien à ne pas considérer ces 50 000 individus comme de l'emploi.

Tableau 12 : caractérisation des JEMP_D selon leur etp DADS

Etp	Nb Jemp_D	Type identifiant		Tranche d'âge			
		A - AB	B - C	15 - 29	30 - 49	50 - 59	60 +
0,0	268	79	21	55	13	7	26
0,1	445	81	19	75	9	3	13
0,2	262	77	23	69	10	7	15
0,3	180	84	16	61	17	9	14
0,4	112	86	14	63	14	5	18
0,5 - 0,6	140	91	9	73	8	8	12
0,7 - 0,9	111	76	24	60	10	21	9
1,0	133	66	34	53	12	20	15
Total	1 651	91	19	65	6	5	8

Champ : PEA pur salarié 'Jemp_D'

Lecture : 268 JEMP_D (jamais en emploi EEC, en emploi au moins une des semaines de référence dans les DADS) font moins de 0,1 etp selon les DADS. 79% d'entre eux ont un identifiant de très bonne qualité (type A ou AB) et 21% de moins bonne qualité (type B ou C). 55% d'entre eux ont entre 15 et 29 ans.

On est donc assuré que la part des JEMP liée aux erreurs d'appariement et aux erreurs sur les dates dans les DADS est restreinte. La plupart des JEMP sont des jeunes ou des seniors occupant des petits boulots (un tiers réalisent moins d'un dixième d'etp, 80% moins d'un demi etp) qu'ils ne déclarent pas à l'EEC. La prochaine étape a pour but d'estimer le volume d'emploi correspondant.

4 – Affectation des absences EEC

Nota : Il faut considérer ce qui suit comme un exercice, qui va au delà de ce que permet de faire avec assurance l'appariement

On a vu divers facteurs qui permettent d'expliquer les absences EEC observées soit par un problème de densité (présence à tort dans les DADS), soit par un problème de sous-déclaration de petits boulots (absence à tort dans l'EEC). Ces facteurs sont assez discriminants mais ne s'appliquent pas à toutes les absences EEC. Enfin, quelques critères de qualité de l'appariement permettent de soupçonner une erreur d'appariement.

Si on veut proposer une estimation du nombre d'emplois parmi les absences EEC, il faut une procédure qui permette de traiter toutes les absences EEC. Deux démarches ont été testées. Une trace des frontières successives au sein de l'échantillon (d'abord entre PEA2 et JEMP puis, au sein des PEA2, entre intérim, PE et autres, et ainsi de suite), et traite de manière différenciée chaque cas. On cherche alors à estimer pour chaque cellule un taux de conversion des absences EEC en emploi. L'autre, présentée ici, cherche à modéliser les absences EEC en tenant compte des différents facteurs examinés.

$$y_i = a + \sum_c c_c 1_{i,c} + \sum_j d_j 1_{i,j} + \sum_k e_k 1_{i,k} + \sum_m f_m 1_{i,m} + \varepsilon_i$$

On a 3 groupes de facteurs (en plus du bloc de contrôle : âge, sexe, forme d'emploi...) : ceux imputables aux DADS (présence à tort), D, ceux imputables à l'EEC (absence à tort), E, et ceux imputables aux doutes sur l'appariement ou à la qualité des données, F.

On somme pour chaque observation (= un PEA * trimestre en emploi DADS) les différents facteurs pour chaque groupe afin d'obtenir un score: $D_i = \sum_j d_j 1_{i,j}$; $E_i = \sum_k e_k 1_{i,k}$; $F_i = \sum_m f_m 1_{m,j}$ et on attribue à l'observation le groupe (D, E ou F) ayant le score (Di, Ei, Fi) le plus élevé

Pour le bloc D (présence à tort dans les DADS) :

- Quand la quotité est imputée (fonction linéaire estimée séparément pour chaque type d'imputation (par les heures, par la quotité du poste, par le salaire)) est faible ce peut être le signe d'une trop faible densité et donc de problèmes de date
- Quand la quotité est déclarée (tqt='A'), on peut calculer une densité corrigée de la quotité : $d = \text{nb d'heures} / (\text{durée en jours} * \text{quotité})$. Si la densité est inférieure à 5, il y a un risque de présence à tort : on le modélise par une fonction linéaire de cette densité (il est d'autant plus fort qu'elle est faible)
- Pour les postes de plus d'un trimestre, quand un écart de quotité avec l'EEC est observé un autre trimestre

Pour le bloc E (absence à tort de l'EEC)

- Étudiant (avec un effet modérateur de l'apprentissage)
- Retraité (identifié par la situation un an avant ou par le module ad hoc 2012)
- Quand quotité déclarée (tqt = 'A'), fonction linéaire de la quotité (la faible quotité augmente le risque de non déclaration)
- Quand au moins un Siret des DADS est retrouvé dans l'EEC mais pas celui du trimestre considéré (hors intérim et PE)
- Pour les postes de moins d'un trimestre, fonction linéaire de la durée du poste

Pour le bloc F (risque d'incohérence liée à une erreur d'appariement ou à forte incertitude)

- Identifiant de type C - non appariés avec les 10 premières lettres du prénom, appariés avec une version plus courte du prénom - pour ceux nés en France
- Identifiant de type B – appariement uniquement avec les identifiants sans numéro de voie alors que ce dernier est renseigné dans les DADS - pour les immigrés
- Quand aucun Siret commun entre l'EEC (avec Siret renseigné) et DADS (hors intérim et PE)
- Aucun élément pour calculer la quotité (tqt='Z') : forte incertitude sur la nature de l'emploi
- La chaîne de traitement DADS a corrigé les dates ou le nombre d'heure ou la variable CPF (temps plein/ temps partiel) : signe d'incertitude sur la présence effective à une date donnée

Se pose la question du champ sur lequel s'effectue cette modélisation. L'appliquer aux seuls PEA2 est source de biais : le biais de sélection sur les PEA2, déjà problématique avec l'autre méthode (on calcule l'emploi moyen DADS sur 4 semaines pour des personnes en emploi DADS au moins une des quatre semaines), devient économétriquement rhédibitoire : les personnes en emploi DADS un seul trimestre sont forcément, s'ils sont PEA2, en emploi EEC ce trimestre (sauf les cas très rare où ils sont en emploi EEC uniquement un autre trimestre). Les coefficients associés aux personnes en emploi un seul trimestre sont donc opposés à ce que l'on aurait sans ce biais.

On ne peut donc pas réaliser l'estimation sur les seuls PEA2. Mais on se heurte alors à une autre difficulté : les comportements des Jemp, notamment des étudiants et des retraités, sont très spécifiques : il faudrait introduire des indicatrices partout, ce qui reconduirait au problème précédent (une estimation de deux modèles, dont un limité aux PEA2). On adopte une solution de compromis en estimant d'abord le modèle hors étudiants et retraités (mais avec les autres Jemp), et en utilisant les coefficients issus de ce premier modèle ; on ajoute ensuite les étudiants et les retraités²⁷ uniquement pour obtenir les coefficients des indicatrices associées (au fait d'être étudiant ou retraité).

On constate tout d'abord que si ces critères, pris isolément, ne rendent compte que d'une petite partie des absences EEC, ils permettent, pris ensembles, de catégoriser la quasi-totalité de ces absences. Ainsi, parmi les PEA2, ceux qui ne répondent à aucun critère cité ci-dessus représentent plus de la

²⁷ Dans l'EEC (avant la refonte de 2013), les retraités ne sont pas identifiés quand ils sont en emploi (la question est posée uniquement à ceux hors-emploi). On complète l'information disponible par celle du module ad-hoc 2012 portant sur le passage de l'emploi à la retraite.

moitié des emplois DADS et moins d'une absence EEC sur 11 (tableau 13). Ceux qui font le plus haut score avec les critères du bloc E représentent moins d'un emploi sur 5 et plus d'une absence EEC sur 2.

Si les absences EEC du bloc E doivent être considérées comme de l'emploi et celles du bloc D comme du non-emploi, les absences EEC des autres blocs restent incertaines. Il faut certainement compter celles du bloc F comme du non-emploi car cela permet de compenser les cas similaires dans les absences DADS (par exemple suite à erreur d'appariement) qui sont toutes comptées comme de l'emploi. Par prudence, on compte aussi celles du bloc Z (aucun critère d'affectation) comme du non-emploi.

Au total, 296 000 des 564 000 absences EEC des PEA2 sont considérées comme de l'emploi, soit un peu plus d'une sur deux, et 736 000 emplois DADS des Jemp (pour les Jemp tous les emplois DADS sont des absences EEC) sur 902 000 sont considérées comme de l'emploi.

Tableau 13 : résultats de l'arbitrage sur les absences EEC

		Bloc D : non emploi	Bloc E : emploi	Bloc F : non emploi	Bloc Z : non emploi	Total	Dont non emploi
Total		3 213	5 332	3 068	12 119	23 732	434
Hors Jemp	Emplois DADS	3 173	4 596	2 989	12 072	22 830	
	TAE	3,7	6,4	3,3	0,4	2,5	
	AbsEEC	119	296	99	50	564	268
Jemp	Emplois DADS = AbsEEC	40	736	79	47	902	166
Intérim		238	340	149	60	787	103
Hors Jemp	Emplois DADS	228	248	140	59	675	
	TAE	24,8	25,2	16,0	6,1	21,5	
	AbsEEC	56	62	22	4	145	82
Jemp	Emplois DADS = AbsEEC	10	92	9	1	112	20
PE		84	137	669	0	890	44
Hors Jemp	Emplois DADS	84	88	651	0	823	
	TAE	5,3	13,4	3,4	.	4,6	
	AbsEEC	4	12	22	0	38	26
Jemp	Emplois DADS = AbsEEC	0	49	18	0	67	18
Autres		2 891	4 855	2 250	12 059	22 055	287
Hors Jemp	Emplois DADS	2 861	4 260	2 198	12 013	21 332	
	TAE	2,0	5,2	2,5	0,4	1,8	
	AbsEEC	58	222	55	46	381	159
Jemp	Emplois DADS = AbsEEC	30	595	52	46	723	128

Champ : PEA salariés purs, en emploi Dads (concept BIT) la semaine de référence

Lecture : En 2012, les salariés purs ont eu en moyenne 894 000 postes principaux en tant que salariés des particuliers employeurs (PE), dont 827 000 de la part des salariés ayant aussi déclaré au moins un trimestre en emploi dans l'EEC (hors jemp). Parmi ces 827 000 postes, 84 000 sont considérés comme présentant des risques de présence à tort dans les DADS pour la semaine de référence considérée (colonne D) et donc arbitrés comme du non-emploi en cas d'absence EEC. Le taux d'absence EEC étant, pour cette catégorie, de 5,3 %, cela concerne un peu plus de 4 000 postes. Au total, sur les 38 000 absence EEC des postes PE hors jemp, 26 000 sont arbitrés comme du non-emploi (somme des colonnes D, F et Z).

Ces résultats sont assez proches de ceux obtenus avec l'approche alternative citée en début de section. On vérifie pour les Jemp que les cas douteux des âges intermédiaires sont moins souvent retenus comme de l'emploi (tableau 14), surtout quand l'etp est supérieur à 1/2.

Tableau 14 : Emplois Jemp retenus par tranche d'âge

	15 - 29	30 - 39	40 - 49	50 - 59	60 +	Total
Nb emplois DADS	519	48	47	95	193	902
% retenu	90	44	38	56	92	82

Champ : emploi DADS des jemp salariés purs

Correction supplémentaire pour l'intérim

Pour réaliser l'étude sur la segmentation et les statuts d'emploi²⁸, l'emploi a été ventilé par statut d'emploi. Le travail de comparaison des sources (et d'arbitrage) sur les statuts d'emploi n'est pas reproduit dans ce document. Toutefois le cas de l'intérim a un impact sur l'estimation globale du niveau d'emploi et est donc considéré ici.

La méthode d'arbitrage sur l'emploi, présentée au III.C.4, conduisait à retirer du niveau d'emploi 103 000 emplois déclarés en Intérim dans les DADS. Au total, après ajout de 37 000 absences DADS considérées comme de l'emploi en intérim dans l'EEC et arbitrage sur les statuts, l'estimation de l'emploi en intérim s'élevait à 711 000, soit nettement au-dessus des niveaux habituellement retenus. En 2012, l'emploi moyen annuel intérimaire est de 536 000, selon les estimations trimestrielles d'emploi, et de 598 000 selon l'Acoss.

Comme l'intérim est le statut d'emploi dont la mesure du niveau d'emploi est la plus délicate, celle où le décalage de paie et les incertitudes sur les dates effectives d'emploi sont les plus importantes, celle, on l'a vu, où les taux d'absences EEC sont les plus élevés (21 %), il n'est pas absurde d'envisager une méthode spécifique d'estimation.

L'idée est de renoncer à arbitrer sur la présence en emploi pour chaque semaine de référence et de calculer directement l'emploi moyen en intérim à partir de l'etp, a priori plus fiable, corrigé pour tenir compte du temps partiel (quand la densité est cohérente avec la quotité déclarée). Il faut toutefois tenir compte des autres emplois de la personne. Le tableau suivant explicite la démarche sur quelques exemples :

Poids	Intérim			Autres		Intérim		Emploi avant		Emploi après	
	Quot	Etp	Nb T	Etp	Nb T	Etpn	Etpn/Q	Int	Aut	Int	Aut
p	a	b	c	d	e	f	g	h	k	m*	n
						b/(c+d)	f/a	p*c/4	p*e/4	p*g*	p*e/4
1000	1	1	4	0	0	1	1	1000	0	1000	0
1000	1	0.6	4	0	0	0.6	0.6	1000	0	600	0
1000	0.5	0.2	2	0.6	2	0.2	0.4	500	500	400	500
1000	0.5	0.4	1	0.8	3	0.33	0.67	250	750	250	750
1000	0.5	0.4	1	0.8	2	0.33	0.67	250	500	500	500

* $m = p \cdot g$ dans les cas simples, $p \cdot \min(g, 1 - e/4)$ en toute généralité. Les deux derniers exemples illustrent cette contrainte sur le nombre de trimestres hors intérim.

Cette procédure permet de réduire de 74 000 l'emploi intérimaire, en plus de la réduction de 103 000 déjà effectuée avec l'arbitrage décrit ci-dessus. Cela porte l'estimation à 637 000, soit 40 000 au dessus de celle de l'Acoss.

29 000 de ces emplois intérimaires étaient le fait de personnes en emploi, un autre trimestre, dans l'EEC : cela porte à 267 000 (296 000 – 29 000) l'emploi ajouté aux PEA2 de l'EE ; 45 000 étaient JEMP : cela porte à 691 000 (736 000 – 45 000) l'emploi ajouté aux JEMP.

Soit un ajout total de 958 000 alors que 564 000 absence EEC et 902 000 emplois des JEMP, soit 1 466 000 emplois DADS étaient soumis à arbitrage. 508 000 emplois DADS ont donc été arbitrés négativement. Cela est du même ordre de grandeur que les 525 000. Au total, pour les salariés purs,

²⁸ Claude Picart, « Trois segments pour mieux décrire le marché du travail », Insee référence Emploi, chômage et revenus du travail, édition 2017.

et hors ajouts liés à la pondération et aux trous de collecte, l'effet net sur l'emploi DADS de ces arbitrages est de + 16 000 emplois.

D - Traitement des non-salariés

Pour les non-salariés, l'absence d'information sur le profil d'emploi intra-annuel dans les bases administratives conduit, pour les PEA2, à privilégier l'emploi selon l'EEC, sachant que, hors agriculteurs, aides familiaux, et artistes, le nombre de non-salariés en emploi dans l'année est calé sur les bases administratives. 9 non-salariés purs sur 10 sont en emploi les 4 trimestres selon l'EEC et le ratio DMA (emploi annuel moyen/ PEA) est de 95%.

Pour statuer sur les JEMP où, par construction, l'EEC ne peut servir à estimer l'emploi, on construit des indicatrices d'activité à partir des bases « non-salariés » et on vérifie que, pour les actifs, la part de JEMP décroît nettement en fonction du revenu d'activité, ce qui est l'indice qu'ils relèvent de la sous-déclaration de petits boulots. On exclut par contre du champ des PEA les JEMP qui ne sont pas actifs au 31/12, soit 48 000 des 140 000 JEMP non-salariés purs.

Au total, les non-salariés et les principalement non-salariés représentent près de 2,9 millions d'emplois (hors trous de collecte).

Des compléments sur le le traitement des non-salariés sont proposés en annexe 1.6

IV – Enseignements de l'appariement

A - Retour sur la démarche

À l'issue du travail de confrontation des sources, on dispose d'un échantillon d'individus appariés avec, pour chacun, un poids arbitré (en partie II) et un emploi annuel moyen arbitré (en partie III).

En notant, pour une observation i :

- pe_i son poids après calage sur l'EEC
- pb_i son poids après calage sur les bases administratives
- pf_i son poids final, après arbitrage
- eme_i son emploi annuel moyen selon l'EEC
- emb_i son emploi annuel moyen selon les bases administratives
- emf_i son emploi annuel moyen après arbitrage

On peut décomposer, pour chaque source, l'écart entre la contribution d'une observation à l'emploi arbitré - $pf_i * emf_i$ - et sa contribution à la source, par exemple $pe_i * eme_i$ pour l'EEC de la manière suivante :

$$pf_i * emf_i - pe_i * eme_i = \underbrace{(pf_i - pe_i) eme_i}_{\text{Effet pondération}} + \underbrace{pf_i (emf_i - eme_i)}_{\text{Effet déclaration}}$$

L'interprétation en termes de pondération et de déclaration vaut pour l'EEC, y compris

- pour ceux jamais en emploi (JEMP) : effet déclaration ($eme_i = 0$)
- pour ceux ajoutés au titre des trous de collecte des bases administratives (TCOL) : effet pondération ($pe_i = 0^{29}$; $emf_i = eme_i$)

Pour les bases administratives, où le terme « effet pondération » est moins approprié, on précisera l'interprétation de cette décomposition en C.

Cette estimation des emplois à ajouter aux sources est une construction, qui repose sur un certain nombre d'hypothèses dont on rappelle ici les principales :

- Toute personne présente dans les sources administratives (hors périodes d'indemnisation chômage) est PEA : on suppose que les chaînes de traitement de ces sources ont éliminé ce qui n'était pas de l'emploi.
- Tout emploi au sens du BIT de l'EEC est un emploi (la réciproque est fautive, cf. les JEMP). Les éventuelles erreurs (par exemple déclaration par un proxy) sont de l'ordre d'un bruit sans biais (cf. III.C.1).
- Les éventuels défauts d'appariement n'ont pas d'impact sur la représentativité des personnes appariées au vu de la variable d'intérêt.

En plus de ces hypothèses structurantes, certaines sources de biais peuvent être identifiées :

- Le cas des NIR fictifs (cf. II.A.1). Doivent-ils être comptés pour un (un individu pointé par un NIR fictif n'est pas présent sous un autre NIR) ou pour zéro ? La décision, arbitraire, de leur affecter un poids de 1/2, introduit une incertitude. Cette incertitude, non négligeable sur le nombre de PEA, est moindre sur l'emploi (inférieure à 100 000) du fait de leur faible emploi moyen.

²⁹ Les trous de collecte ne sont ajoutés qu'aux catégories calées sur les bases administratives (cf. II.F). Soit, pour une catégorie donnée, E le nombre de PEA selon l'EEC, B selon les bases administratives et A le nombre de PEA ajoutés. Le nombre de PEA retenu est : $R = B + A$; l'écart avec l'EEC est $R - E = (B - E) + (A - 0)$. B - E est l'effet du calage sur les bases administratives.

- L'estimation des trous de collecte est une estimation plancher : ne sont repérés que les absents 2012 présents en 2011 ou en 2013.
- L'arbitrage, pour les PEA2 SAL, sur les absences EEC (cf. III.C.4) reste un exercice entaché d'une marge d'erreur d'environ 100 000 EMA.

On présentera les résultats en distinguant salariés et non-salariés. La frontière entre ces deux catégories dépend des sources : les salariés dirigeants, présents dans les DADS (donc considérés comme salariés au sens des sources administratives) sont considérés comme indépendants (au sens de l'autonomie de décision) dans l'enquête Emploi. On privilégie les sources administratives, sauf pour les catégories mal couvertes par ces dernières (tableau 1).

Tableau 1 : Affectation à une catégorie en fonction de la catégorie de chacune des sources

Bases administratives		Enquête Emploi	NAA	NSZ	PE	ShPE
NAA	Aides familiaux, agriculteurs, artistes		NAA	NAA		
NSZ	Autres non-salariés			NSZ		
PE	Salariés des particuliers employeurs			PE		
ShPE	Autres salariés			ShPE		

Cette typologie, utilisée pour la pondération (partie II) ne recouvre pas parfaitement celle utilisée pour l'estimation de l'emploi annuel moyen (partie III). En effet, la frontière pertinente pour cette estimation passe entre les salariés purs (présents dans les DADS, absents des bases non-salariés), où une confrontation avec l'enquête Emploi est possible, et les autres (présents dans les bases non-salariés) pour lesquels l'absence de profil infra annuel de l'emploi dans les bases non-salariés conduit à privilégier l'enquête Emploi. Le partage entre principalement salariés et principalement non-salariés de cette seconde typologie respecte toutefois celui entre salariés et non-salariés de la première, sauf pour les non-salariés distingués dans la catégorie NAA (tableau 2). En effet les aides familiaux, non couverts par les bases non-salariés, ne peuvent être appariés (via les DADS) que s'ils sont pluriactifs. Ils sont donc classés comme non-salariés selon la première typologie (NAA, d'après l'EEC) et salariés selon la seconde.

Tableau 2 : calage (dernière colonne) et méthode d'estimation de l'emploi moyen selon les cas.

Situation dans les bases administratives Regroupements pour pondération (calage)	Non salariés	Princip. non salariés	Princip. salariés	Salariés	Trous collecte	Calage
Aides familiaux, agricult., artistes	EEC (sauf JEMP)		EEC	(EEC)*		EEC
Autres non-salariés					EEC	BA
Salariés des particuliers emp.			EEC (sauf JEMP)	Arbitrage yc JEMP		EEC
Autres salariés					EEC	BA

* : On considère que dans cette case les bases administratives ne nous donnent que la partie salariée de l'activité. Il vaut donc mieux prendre l'emploi moyen de l'enquête Emploi.

Lecture : Pour les autres salariés (i.e. hors salariés des particuliers employeurs), l'emploi annuel moyen est celui de l'EEC pour les principalement salariés (hors JEMP) et ceux ajoutés au titre des trous de collecte et par arbitrage entre emploi DADS et emploi EEC pour les salariés purs. Le calage de ceux présents dans les deux sources se fait sur les bases administratives.

Tableau 3 : ventilation des PEA selon les deux typologies

En milliers

	Non salariés	Princip. non salariés	Princip. salariés	Salariés	Trous collecte	Total
Aides familiaux, agricult., artistes	591	49	21	103		764
Autres non-salariés	2 179	236			63	2 479
Salariés des particuliers emp.			26	1 053		1 079
Autres salariés			674	26 895	278	27 847
Total	2 770	286	721	28 050	342	32 169

B – Enseignements sur l'enquête Emploi

Sur le champ retenu (hors frontaliers), il y a, en 2012, 27,7 millions de personnes en emploi dans l'année (PEA) ; leur emploi moyen est de 25,4 millions (tableau 4).

Tableau 4 : PEA et emploi dans l'enquête Emploi, ventilés entre salariés et non-salariés : situation selon l'enquête Emploi (en ligne) et situation retenue dans l'échantillon (en colonne)

	PEA (en milliers)			Emplois (en milliers)		
	Non-salarié	Salarié	Total	PEA non-salarié	PEA salarié	Total
Non-salarié*	2 693	222	2 915	2 567	201	2 768
Salarié	159	24 611	24 769	133	22 464	22 598
Total	2 852	24 832	27 684	2 700	22 666	25 366

* on rappelle que, pour réduire les différences de concept, les dirigeants salariés, classés parmi les indépendants dans l'EEC, sont ici classés comme salariés.

Au total, on ajoute 4,5 millions de PEA à l'enquête Emploi, dont 3,0 millions pour la seule non-déclaration d'emploi de la part de ceux qui ne sont jamais en emploi (JEMP) et près de 2,2 millions d'emplois (tableau 5). Ce chiffre inclut un effet de champ propre à l'EEC, la non-interrogation des ménages en communauté, dont l'impact sur l'emploi est estimé à un peu moins de 200 000. Hors cet effet de champ, il y aurait de l'ordre de 2,0 millions d'emplois à ajouter à l'EEC.

Tableau 5 : Ajouts à l'enquête Emploi

		PEA initial	Ajout PEA	Emploi Initial	Ajout d'emploi	
					Effet pondération	Effet déclaration
Non salariés	Jamais en emploi		98	.	0	98
	PEA selon l'EEC	2 852	230	2 700	201	0
	« Trous » bases adm.		63	.	55	0
Salariés	Jamais en emploi		2 936*	.	0	712
	PEA selon l'EEC	24 832	879	22 666	628	267
	« Trous » bases adm.		278	.	211	0
Total		27 684	4 485	25 366	1 094	1 077
Correction pour communautés					- 200	
					894	1 077

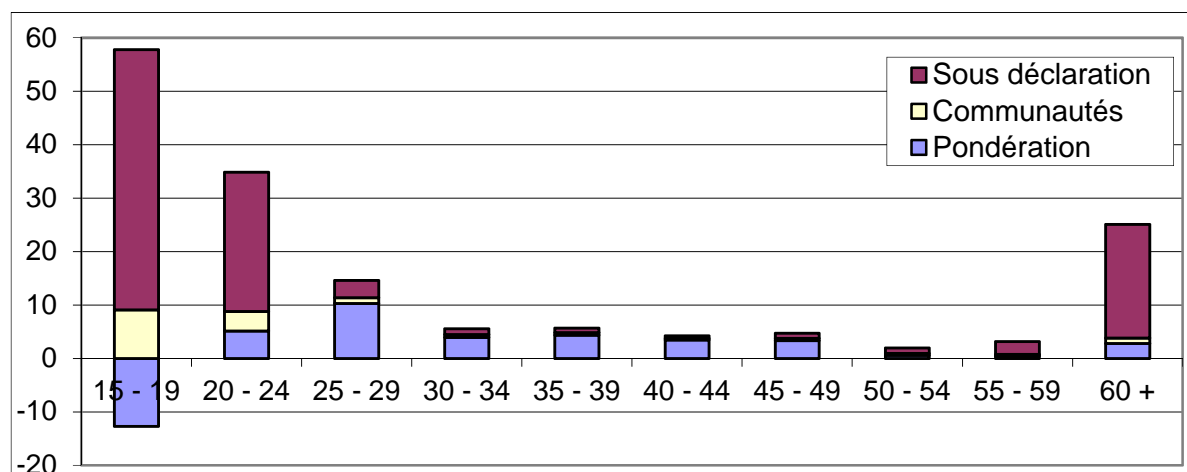
* y compris les PEA DADS qui ne sont pas en emploi BIT pour cause de congés longue durée (différence de concept) : ils ne contribuent pas aux ajouts d'emploi (où l'on se place à concept commun) mais n'en sont pas moins absents d'un comptage des PEA à partir de la seule EEC

Les effets de pondération - $(p_{fi} - p_{ei}) em_{ei}$ dans la décomposition proposée en A – résultent de l'application au travail du II (estimation des poids) de l'emploi moyen selon l'EEC. Il provient entièrement du calage sur une source exogène (les bases administratives enrichies d'une estimation de leurs trous de collecte) même si la démarche adoptée conduit à distinguer l'effet direct du calage (repondération) sur les bases administratives (201 000 + 628 000 = 829 000 emplois) de l'effet des trous de collecte (ajout de 266 000 emplois : application aux 342 000 PEA ajoutés au II.F de leur emploi moyen EEC). C'est à l'effet de pondération que doit retranché l'effet de champ (communautés).

Les effets de déclaration - $p_{fi} (em_{fi} - em_{ei})$ - sont, pour les salariés, issus de la procédure d'affectation des absences EEC du III.C.4. On notera que ceux ne déclarant jamais d'emploi, soit 2,9 millions de PEA, représentent 712 000 emplois : leur durée moyenne en emploi dans l'année est de moins de 3 mois. Pour les non-salariés au moins une fois en emploi selon l'EEC, l'emploi est celui de l'EEC, d'où l'impossibilité d'estimer un effet déclaration. L'emploi des autres non-salariés (JEMP) a été contrôlé par une indicatrice d'activité au 31/12 (qui vaut, en moyenne, pour l'emploi annuel moyen) qui a conduit à rejeter en III.D 48 000 des 140 000 PEA non salariés jamais en emploi selon l'EEC estimés au II.D.

Ces effets de sous-déclaration sont concentrés aux extrémités de la pyramide des âges des personnes en emploi : jeunes de 15 à 29 ans et seniors de 60 ans et plus (graphique 1). Les jeunes concentrent aussi les effets de pondération et les effets de champ (exclusion des communautés).

Graphique 1 : nombre d'emplois « ajoutables » au niveau initial de l'EEC, pour 100 emplois d'une classe d'âge donnée



Lecture : pour 100 emplois selon l'EEC dans la tranche d'âge des 20 à 24 ans, la repondération brute en ajoute 9, dont 4 peut être attribué à l'emploi dans les communautés et le solde, 5, à une sous-pondération des jeunes de cette tranche d'âge en emploi. Le traitement de la sous-déclaration en ajoute 26 : au total, 35 emplois sont ajoutés. Nota : l'emploi des communautés par âge a été estimé séparément (source RP) et séparé de l'effet pondération (sans ce retraitement, l'effet pondération apparaît presque aussi fort pour les 20 - 24 ans que pour les 25 - 29 ans).

C – Enseignements sur les sources administratives

Sur le champ retenu (hors frontaliers), il y a, en 2012, 31,6 millions de personnes en emploi dans l'année (PEA) ; leur emploi moyen est de 27,1 millions (tableau 6).

Tableau 6 : PEA et emploi dans les bases administratives ventilés entre salariés et non-salariés : situation selon les bases administratives (en ligne) et situation retenue dans l'échantillon (en colonne)

	PEA (en milliers)			Emplois (en milliers)		
	Non-salarié	Salarié	Total	Non-salarié	Salarié	Total
Non salarié	55	28 626	28 681	38	24 367	24 404
Salarié	2 907	.	2 907	2 648	.	2 648
Total	2 962	28 626	31 588	2 686	24 367	27 053

Au total, on ajoute 0,6 millionde PEA aux bases administratives : 0,3 million pour les non-salariés, principalement au titre des défauts de couverture (aides familiaux, artistes, agriculteurs) ; 0,3 million pour les salariés, principalement au titre des trous de collecte (tableau 7a). Pour l'emploi, il faut d'abord tenir compte des congés (ou absences) de longue durée rémunérés qui sont inclus dans l'emploi mesuré par les sources administratives mais pas dans l'emploi au sens du BIT dans l'enquête Emploi : 173 000 emplois sont concernés.

La décomposition et l'interprétation des écarts est plus complexe que pour l'enquête Emploi. En effet, alors que dans l'EEC l'unité d'observation est l'individu (en tant qu'appartenant à un ménage) et que la distinction entre un effet pondération et un effet déclaration coule de source, l'unité d'observation des DADS est le poste. Les trous de collecte peuvent aussi bien conduire à l'ajout d'individus si tous les postes d'un même individu sont absents (trous de collecte de la partie II), ou à la correction à la hausse de l'emploi moyen annuel d'un individu si seuls certains postes sont absents (absences DADS de la partie III).

Tableau 7-a : ajouts aux bases administratives

		PEA initial	Ajout PEA	Emploi Initial	Con gés	Emploi « BIT »
Non salariés	Jamais en emploi	98	0	98		98
	PEA selon l'EEC	2 864	218	2 589		2 589
	« Trous » bases adm.	0	63	0		0
Salariés	Jamais en emploi*	2 845	0	924		924
	PEA selon l'EEC	25 781	21	23 442	-173	23 269
	« Trous » bases adm.	0	278	0		0
Total		31 588	581	27 053	-173	26 880

* Les JEMP s'entendent ici hors frontaliers et hors congés longue durée, d'où une différence avec le tableau 5

On peut distinguer :

- Effets de couverture : non-couverture (aides familiaux) ou couverture partielle (artistes) ou couverture défectueuse liée à une source spécifique (agriculteurs, particuliers employeurs).
- Trous de collecte : contrairement aux effets de couverture, ils ne sont pas liés à une sous-population particulière. Il s'agit de trous traités par l'ajout de PEA si tous les postes d'un même individu sont concernés (lignes « trous ») ou, sinon, par ajout d'emploi à des personnes au moins une fois en emploi selon les bases administratives (lignes PEA³⁰).
- Présence à tort, détectées notamment par une trop faible densité. C'est la principale composante de la colonne « suppression emplois ».
- Bruit lié à des décalages de dates entre les DADS et l'EEC ou à des défauts d'appariement. Si ces bruits sont sans biais, ils se traduisent par des volumes équivalents d'ajouts et de suppressions.

Tableau 7b: ajouts aux bases administratives

		Emploi « BIT »	Effet couverture	Ajout emplois Brut	Suppression emplois	Ajout emplois Net
Non salariés	Jamais en emploi	98	0	0	0	0
	PEA selon l'EEC	2 589	179	257	-123	134
	« Trous » bases adm.	0	0	55	0	55
Salariés	Jamais en emploi	924	0	0	-212	-212
	PEA selon l'EEC	23 269	32	562	-303	259
	« Trous » bases adm.	0	0	211	0	211
Total		26 880	211	1 085	-638	447
<i>Décomposition selon situation dans les bases administratives</i>						
Trous de collecte		0	0	266	0	266
Non salariés purs	Jamais en emploi	93	0	0	0	0
	PEA selon l'EEC	2 332	138	172	-106	66
Princ. non salariés	Jamais en emploi	5	0	0	0	0
	PEA selon l'EEC	219	1	50	-15	35
Principal. salariés	Jamais en emploi	21	0	0	0	0
	PEA selon l'EEC	544	6	72	-8	64
Salariés purs	Jamais en emploi	902	0	0	-212	-212
	PEA selon l'EEC	22 764	66	525	-297	228
Total		26 880	211	1 085	-638	447

L'impact des effets de couverture est estimé à 210 000 emplois, principalement de l'emploi non salarié (tableau 7b). Cette estimation résulte de l'application de l'emploi EEC aux 239 000 personnes en emploi ajoutées aux bases administratives pour les catégories calées sur l'EEC au II.E.2.

³⁰ Les PEA des bases administratives sont le cumul des lignes « jamais en emploi » et « PEA selon l'EEC » mais l'ajout d'emplois se fait par référence à l'EEC et on ne peut donc pas ajouter de l'emploi aux JEMP.

L'effet net des ajouts d'emplois est estimé à 447 000, solde de 1 085 000 ajouts et 638 000 suppressions. 638 000 représente une estimation plafond des présences à tort dans les DADS. Les ajouts sont la somme de :

- 266 000 emplois liés aux trous de collecte totaux : application aux 342 000 PEA ajoutés au II.F de leur emploi moyen EEC.
- 525 000 emplois des salariés purs, en emploi selon l'EEC une semaine où ils ne sont pas en emploi selon les DADS, tout emploi EEC étant par hypothèse considéré comme de l'emploi (III.C.1)
- 72 000 emplois EEC des principalement salariés, non en emploi dans les DADS (on ne sait pas construire un emploi moyen à partir des sources administratives pour les pluri actifs)
- 222 000 emplois EEC des non-salariés inactifs au 31/12 selon les bases administratives (on a retenu l'indicatrice d'activité au 31/12 comme proxy de l'emploi moyen au III.D)

Les suppressions sont la somme de :

- 212 000 emplois DADS des salariés purs jamais en emploi selon l'EEC, considérés comme étant en emploi à tort dans les DADS suite à la procédure d'arbitrage du III.C.4
- 297 000 emplois des salariés purs, en emploi selon les DADS et non selon l'EEC la semaine de référence, considérés comme étant en emploi à tort dans les DADS suite à la procédure d'arbitrage du III.C.4
- 8 000 emplois de principalement salariés non en emploi selon l'EEC
- 121 000 emplois de non-salariés en emploi au 31/12 selon les bases administratives (et donc comptés pour 1 toute l'année) : ils ne sont pas en emploi toute l'année selon l'EEC.

Pour les pluri actifs, l'écart net de 99 000 emplois à ajouter aux bases administratives résulte clairement pour partie du choix de ne retenir qu'une seule des deux sources – les DADS pour les principalement salariés et les bases non-salariés pour les principalement non-salariés – pour l'emploi selon les bases administratives. Ce choix majore ce que l'on doit ajouter aux bases administratives et minore sans doute ce qu'on ajoute à l'EEC.

D – Synthèse : écart entre EEC et estimations annuelles d'emploi

Les estimations annuelles d'emploi partent des sources administratives expertisées ici. Mais, alors qu'à des fins de comparaison avec l'EEC, nous avons construit à partir de ces sources une notion d'emploi annuel moyen, les estimations annuelles d'emploi concernent l'emploi au 31/12.

Les traitements pour l'élaboration des Estimations d'emploi visent à déterminer pour chaque personne en emploi au 31/12 son statut (salarié ou non) et son activité principale. Des doutes sur l'indicatrice DADS de présence au 31/12, notamment à cause des décalages de paie, conduisent à supprimer des postes. En outre une cale à la baisse est nécessaire pour assurer la cohérence en évolution, ie pour assurer une cohérence avec les estimations d'emploi antérieures à 2012.

Au total, l'écart entre l'emploi moyen annuel moyen calculé à partir des sources administratives 27 053 000 – et l'estimation d'emploi au 31/12/2012 – 26 453 000, soit 600 000, est la somme des trois facteurs suivants :

- Passage de l'emploi annuel moyen à l'emploi au 31/12
- Traitements de constitution du fichier
- Calage pour assurer la cohérence en évolution

Cet écart s'articule avec les autres sources d'écarts analysées dans ce document pour proposer une interprétation globale de l'écart entre l'emploi publié, pour 2012, à partir de l'enquête Emploi – 25 754 000³¹ – et celui des Estimations d'emploi (tableau 8).

³¹ Source : « Une photographie du marché du travail en 2012 », Insee Première n°1466

Tableau 8

	2012	
Enquête Emploi	25 754	(a) Moyenne sur l'année, France métropolitaine
Estimation emploi	26 453	(b) Estimation fin d'année, France métropolitaine
Ecart à expliquer	699	(b) - (a)
Différences de champ	-193	
Communautés Transfrontalières	200	Source : Recensement de la population
	-393	Source : enquête Emploi
Différences de concept d'emploi	173	
Emploi rémunérés non BIT	173	Congés longue maladie notamment, source : appariement
Traitements des sources administratives initiales	-600	Passage aux Estimations d'emploi + effet (fin vs moyenne annuelle)
Défauts de couverture et trous de collecte des bases administratives	-658	Source : appariement
Sous-déclaration (totale ou partielle) dans l'EEC	1077	Source : appariement
Sous-pondération dans l'EEC	894	Source : appariement
Autres	6	Par solde

Annexes

Annexe 1 – Compléments (détails) au document

1 – Complément au II.C : détails sur le calage de l'échantillon apparié

Le calage se fait en utilisant la macro Calmar2. On teste pour apprécier la robustesse des résultats différentes variantes :

- selon les poids initiaux des individus appariés : 1, pe (poids dans l'EEC), pe* (poids dans l'EEC après calage des P4T, cf. II.B.) ;
- selon les variables utilisées pour le calage ;
- selon la méthode de calage : linéaire, raking ratio (proposées par la macro Calmar)

Le calage ne peut se faire qu'à partir des individus de poids non nul de l'EEC : 23 individus supplémentaires de l'appariement ont un poids nul. L'échantillon utilisable est ainsi de 19 843 (tableau 1). Parmi ces derniers, environ 50 ont un poids nul certains trimestres. Pour éviter une attrition supplémentaire, on travaille avec le poids moyen sur l'année.

Tableau 1 - de l'échantillon retenu à l'échantillon final sur les 4 trimestres

	Échantillon retenu	Présents les 4 T	Avec un poids non nul
Champ DADS	71 947	17 894	17 962
Champ Bases non-salariés	7 310	1 882	1881
Total	79 257	19 866	19 843

L'échantillon retenu est utilisé pour calculer des marges trimestrielles afin d'améliorer le calage de l'échantillon final.

Pour les salariés et principalement salariés, les variables de calage utilisées sont d'abord calculées à partir des DADS :

- sexe
- tranche d'âge
- naissance à l'étranger (oui/non)
- tranche de taille de la commune
- tranche de nombre d'heures rémunérées dans l'année
- typologie de caractérisation de l'emploi (tableau)
- salarié pur ou multiactif principalement salarié
- nombre de semaines en emploi dans l'année (comme variable numérique)

Tableau 2 - Typologie de caractérisation de l'emploi : quelques statistiques descriptives

Employeur (source)	Type de contrat			NIR incorrects	Parmi les NIR corrects			
					%	Principalement Salariés (%)	Femmes (%)	<30 ans (%)
Privé (DADS)	CDI	CDI	14 811 917	1,1	2,7	43,9	19,4	25,9
	Apprentissage	APP	514 517	2,6	0,7	33,7	99,6	0,0
	CDD	CDD	3 292 089	6,6	3,5	53,8	58,6	11,3
	Intérim	INT	1 340 110	5,1	3,4	28,8	51,6	9,5
	Mixte	MZZ	1 750 150	4,9	2,1	39,4	29,6	30,4
Particulier employeur	PE		1 108 852	8,3	1,8	90,3	10,4	46,4
Fonction publique (SIASP)	CDD	SID	1 073 300	2,3	3,4	65,8	47,7	15,9
	Autre	SI	5 127 406	0,2	1,4	60,3	13,1	34,6
Total			29 018 341	2,3	2,5	49,3	26,7	25,3

Champ : salariés purs (SAL) et principalement salariés (PSA), France métropolitaine

Lecture : Dans la source Dads, on dénombre 3 292 089 CDD parmi les salariés purs et principalement salariés. 2,6% d'entre eux ont un NIR incorrect. 3,5% de ces salariés en CDD à NIR incorrect sont principalement salariés, 53,8% sont des femmes.

Nota bene : les salariés avec plusieurs postes sont classés dans la catégorie où plus de la moitié des heures sont réalisées. Pour le privé, si aucune forme d'emploi renseignée ne forme plus de la moitié des heures, le salarié est placé dans la catégorie « Mixte ».

Afin de capter le profil infra annuel, on cale aussi sur le nombre de personnes en emploi les semaines 2, 10, 20, 30 et 50³².

Tableau 3 - Taux d'appariement par catégorie (en un pour 10 000)

Catégorie (cf tableau 2)	Appariement EEC, poids > 0*	Présents au T1	Présents les 4T (P4T)	Nb de P4T
CDI	26	15	6,5	9 454
APP	27	16	6,3	315
CDD	24	14	5,5	1 696
INT	24	14	5,8	734
MZZ	23	14	5,8	972
PE	25	15	6,8	695
SID	26	15	6,4	667
SI	26	16	6,7	3 429
Total	25	15	6,3	17 962

Champ : personnes en emploi dont le NIR est correct

Lecture : Sur 10 000 personnes en CDI dans les DADS, 26 sont appariés avec un poids non nul dans l'EEC. 15 sont appariés à un répondant du T1, 6,5 à un répondant les 4 trimestres. Soit, pour les répondants les 4 trimestres, 9 454 individus appariés.

* Parmi les 79 257 appariés retenus (cf. A), 273 ont un poids nul dans l'EEC

Taux rapporté aux PEA avec NIR OK.

Le calage se fait sur les DADS. On peut comparer les différentes méthodes de calage par leur effet sur le profil trimestriel selon l'EEC. On utilise comme critère la part en emploi (pemp), la part au chômage (pch) et la part des formes particulières d'emploi³³ parmi les personnes en emploi (pfpe). Calculées uniquement sur les individus appariés aux DADS, ces parts n'ont pas de signification économique, sauf celle des FPE puisqu'elle est rapportée aux personnes en emploi selon l'EEC.

Le profil de référence (colonne A du tableau 4) est celui obtenu en faisant un calage distinct pour chaque trimestre avec l'ensemble des individus appariés pour ce trimestre. On évite ainsi le biais d'attrition (mais ce calage avec des individus différents chaque trimestre ne permet pas d'estimer un nombre de PEA ; c'est pourquoi on ne peut pas s'en contenter).

Le calage spontané est celui utilisant comme poids de départ des individus présents les 4 trimestres le poids de l'EEC (colonne B). Il s'écarte assez sensiblement du profil de référence avec notamment une sous-estimation des chômeurs et des salariés en FPE. Le critère de la somme des carrés des écarts (écarts calculés sur chacune des 9 cellules) donne une SCE de 2,20.

Tableau 4 - profils trimestriels pour différentes méthodes de calage

trimestre	A			B			C			D		
	Par trimestre - Poids EEC - raking ratio			P4T - poids EEC raking ratio			P4T - poids calage interne - raking ratio			Idem + calage sur profil trimestriel		
	pemp	pch	pfpe	pemp	pch	pfpe	pemp	pch	pfpe	pemp	pch	pfpe
1	80,64	7,01	13,57	80,75	6,66	12,90	80,43	7,06	13,17	80,57	7,00	13,27
2	81,75	6,16	14,40	81,45	5,73	13,93	81,54	6,10	14,39	81,63	6,14	14,43
3	81,87	6,83	14,56	81,61	6,26	14,52	81,90	6,79	15,09	81,78	6,81	15,01
4	81,29	7,48	14,10	81,12	6,74	13,71	81,29	7,50	14,04	81,19	7,45	13,96
SCE*	0			2,20			0,54			0,35		

*Somme des carrés des écarts avec les cellules correspondantes du calage de référence par trimestre

Lecture : 7,01% des appariés du premier trimestre sont au chômage, en prenant les poids obtenus par calage sur les DADS en donnant à la macro Calmar les poids de l'EEC comme poids initiaux et en utilisant la méthode du raking ratio. En partant des mêmes poids initiaux et de la même méthode mais en limitant l'échantillon apparié aux P4T, on obtient que 6,66% des P4T sont au chômage au premier trimestre. En partant des poids obtenus par calage interne du panel EEC, cette part remonte à 7,06%.

Partir des poids obtenus après calage interne à l'EEC (étape présentée supra en B) permet de se rapprocher nettement du profil de référence : la SCE est divisée par 4 (colonne C). Ce résultat n'est pas surprenant puisque le calage interne visait lui aussi à réduire le biais d'attrition en utilisant les profils trimestriels.

Le meilleur ajustement selon le critère SCE est celui obtenu en partant des poids issus du calage interne à l'EEC et en ajoutant comme variables de calage, pour chaque trimestre, une variable de présence en emploi la semaine de référence (6 modalités par croisement emploi DADS oui/non et activité EEC, variable actu à 3 modalités) à partir de l'ensemble des appariés du trimestre (on évite le biais d'attrition) calés sur les DADS (colonne D).

³² On compte les présents sur des périodes de 3 semaines centrées sur ces semaines (exemple présents en S9, S10 ou S11 pour l'indicateur nbsem10) à partir du fichier des périodes et après retraitement spécifique des décalages de paie. On reviendra en partie III.B sur le traitement des dates.

³³ Les formes particulières d'emploi (FPE) sont définies ici comme les salariés qui ne sont ni en CDI ni titulaires de la fonction publique.

Pour les non-salariés, on ne dispose pas de profil infra annuel. On utilise donc un calage plus simple, avec les variables suivantes :

- sexe
- tranche d'âge
- naissance à l'étranger (oui/non)
- tranche de taille de la commune
- tranche de revenu
- statut (basé sur la catégorie juridique)
- non-salarié pur ou multiactif principalement non-salarié

On part de la pondération issue du calage interne à l'EEC, ce qui permet de corriger, ainsi que l'on vient de le constater pour l'appariement avec les DADS, une bonne partie de l'effet d'attrition.

2 – Complément au II.E.2 : une sous-pondération des multiactifs principalement non-salariés (PNS) dans l'EEC ? Faut-il caler sur la multiactivité ?

La population des PNS est délicate à traiter : ils sont sous-appariés dans les bases « non-salariés » (car peuvent se déclarer comme salariés dans l'EEC) et sur-appariés dans l'EEC (car il leur suffit d'être apparié à l'une des deux sources, bases «non-salariés» ou DADS, pour apparaître comme appariés). De plus, les effectifs concernés sont faibles dans l'EEC.

Salariés comme non-salariés sont mieux appariés aux bases administratives quand ils sont « purs » (mono activité au sens retenu ici). Le taux d'appariement décroît plus nettement en cas de multi activité pour les non-salariés car l'appariement avec les bases «non-salariés» est plus dépendant des données sur l'établissement et le lieu de travail et donc du fait que l'enquêté se déclare comme non-salarié dans l'EEC : seuls 12 pour 10 000 des principalement salariés sont appariés aux bases non-salariés contre 25% des non-salariés purs (tableau 1).

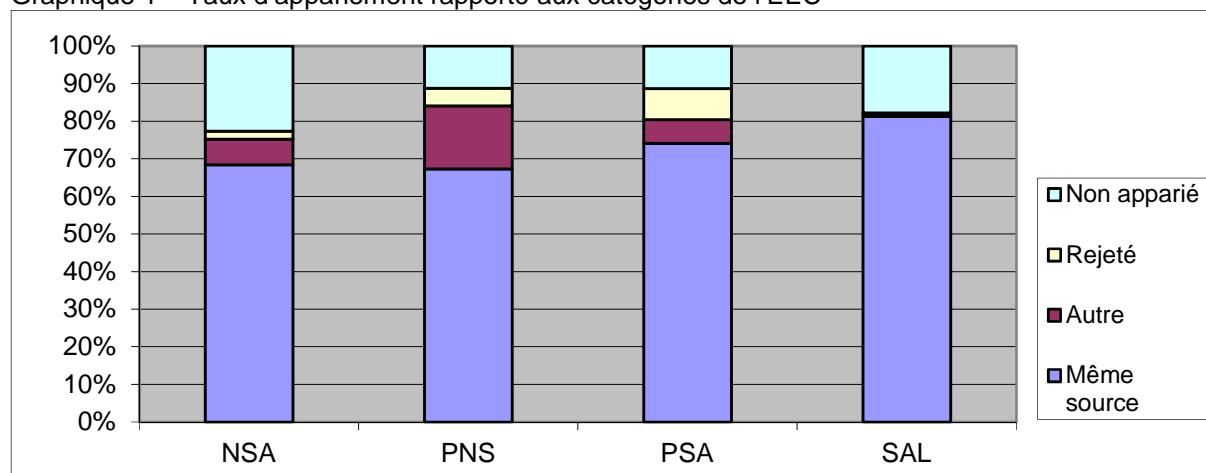
Tableau 1 – Appariement aux sources administratives

	Population (P)	Bns (B)	DADS (D)	Retenus(R)	dont P4T (A)	B/P	D/P	R/P	A/P	A/R
	En nombre					En 1/10000				En %
NSA	2 665 805	6 751		6 691	1 736	25	0	25	7	26
PNS	288 332	612	506	608	145	21	18	21	5	24
PSA	715 632	850	1 655	1 649	398	12	23	23	6	24
SAL	28 302 709		70 364	70 036	17 564	0	25	25	6	25
	31 972 478	8 213	72 525	78 984	19 843	3	23	25	6	25

Lecture : 612 des 288 332 principalement non-salariés sont appariés aux bases non-salariés, soit 21 pour 10 000 et 506 aux DADS, soit 18 pour 10 000. 608 sont finalement retenus, soit 21 pour 10 000, dont 145 répondant les 4 trimestres, soit 5 pour 10 000.

Au contraire, c'est pour les pluriactifs que le taux primaire d'appariement à l'EEC est le plus élevé (graphique 1).

Graphique 1 – Taux d'appariement rapporté aux catégories de l'EEC



Lecture : le taux d'appariement primaire des PNS selon l'EEC – on tient compte de la multi activité déclarée chaque trimestre et du nombre de trimestres où chaque statut est principal ou secondaire - est de 89% ; suite au rejet d'un certains nombre d'observations pour incohérence (exemple : apparié avec les seules bases «non-salariés» alors qu'il est, selon les bases, principalement salarié), ce taux descend à 84%. Il serait de 68% si on se limitait aux seuls appariés aux bns.

Ce sur appariement devrait disparaître dans l'échantillon finalement retenu puisque chaque multi actif doit être apparié avec la base correspondant à son statut principal (c'est comme si on constituait une base unique sans double compte). C'est bien le cas pour principalement salariés (PSA) mais pas pour les principalement non-salariés (PNS), dont une part assez importante est appariée avec les seuls DADS.

La population des PNS est donc sous appariée du point de vue des bases «non-salariés» et sur appariée du point de vue de l'EEC. Ceci « explique » mécaniquement qu'un PNS apparié représente nettement plus d'individus bases «non-salariés» que d'individus EEC (la prise en compte des poids EEC modifie peu ce constat) et qu'ils apparaissent donc comme sous représentés dans l'EEC.

Mais plusieurs bémols doivent être apportés à une interprétation en termes de sous-représentation. D'abord, le taux d'appariement EEC limité aux bases «non-salariés» n'est pas plus fort pour les PNS que pour les NSA. La différence vient des appariés aux seules DADS. La détermination du statut principal ne coïncide pas d'une source à l'autre (tableau 2). Ensuite, la population des PNS est très réduite : 107 dans l'EEC (présents les 4 trimestres), dont 90 appariés.

Tableau 2 – Croisement du statut selon les bases administratives (en ligne) et selon l'EEC (en colonne)
En nombre, non pondéré, d'appariés

	NSA	PNS	PSA	SAL	JEMP	Total
NSA	1594	14	13	34	81	1736
PNS	88	34	9	12	2	145
PSA	42	27	100	207	22	398
SAL	168	15	55	15648	1678	17564
Total	1892	90	177	15901	1783	19843

Le calage sur des populations aussi réduites n'est sans doute pas très robuste, c'est pourquoi on présente des variantes où ces variables de statut ne sont pas utilisées dans le calage (tableau 3).

Tableau 3 – Variantes de calage

		pdba_sns (A)	pdba_hsns (B)	pdEEC_sns (C)	pdEEC_hsns (D)	C/A	C/B	D/B	D/A
Statut selon les ba									
Total	18 036	28 499 469	28 495 177	27 679 709	27 679 709	97	97	97	97
SAL	15 871	25 038 743	25 065 966	24 348 564	24 340 568	97	97	97	97
PSA	376	665 955	637 193	599 846	613 314	90	94	96	92
NSA	1 647	2 512 559	2 564 798	2 527 478	2 516 771	101	99	98	100
PNS	142	282 211	227 220	203 821	209 056	72	90	92	74
Statut selon l'EEC									
Total	18 036	28 499 469	28 495 177	27 679 709	27 679 709	97	97	97	97
SA	15 882	25 113 634	25 119 704	24 305 949	24 291 459	97	97	97	97
PS	173	303 685	293 462	282 550	298 249	93	96	102	98
NS	1 891	2 930 340	2 943 645	2 966 565	2 949 064	101	101	100	101
PN	90	151 810	138 367	124 644	140 937	82	90	102	93

Lecture : les 142 appariés en emploi selon l'EEC (PEA2) et principalement salariés (PNS) selon les ba, représentent, dans les bns, 282 000 personnes en utilisant le statut (sns) comme variable de calage et 227 000 sans l'utiliser. Ils représentent, dans l'EEC 203 000 personnes en utilisant le statut et 209 sans l'utiliser. Les ratios population EEC/ population bases administratives varient, selon les poids utilisés, entre 72 et 92.

Champ : PEA2

Ne pas utiliser les variables de statut dans le calage diminue, comme attendu, l'écart entre les estimations EEC et les estimations bases administratives pour les PNS. Ce ratio reste toutefois toujours inférieur à celui obtenu pour les autres populations : une sous estimation de cette population particulière n'est pas à exclure. A noter que les multi actifs (PNS et PSA) appariés sont plus sensible à l'attrition EEC (colonne A/R du tableau 1). La sensibilité à l'attrition est souvent le signe d'une population plus difficile à capter avec l'EEC. L'utilisation du statut selon l'EEC donne des résultats beaucoup plus chaotiques. C'est sans doute dû à la mauvaise qualité de la variable de calage, calculée à partir de 107 individus EEC contre 288 332 dans les bases «non-salariés» (y compris JEMP selon l'EEC).

Comme, de plus, la variable de statut selon les bases administratives est une variable discriminante pour la suite des travaux (ceux qui ne sont pas salariés « purs ») sont systématiquement moins souvent

en emploi dans les DADS que dans l'EEC : l'estimation de leur emploi moyen ne peut donc être basée que sur l'EEC), le calage finalement retenu intègre cette variable. Par contre, la variable de statut selon l'EEC est à la fois peu robuste et non utilisée dans la suite des travaux : elle n'est pas retenue.

Le ratio retenu et présenté dans le corps du texte est donc celui de la dernière colonne du tableau 4 (D/A).

3 – Complément au II.F : devenir en 2013 des appariés 2012

250 000 salariés des DADS de diffusion 2012, et encore salariés en 2013 selon l'EEC, sont absents des Dads de diffusion 2013 (tableau 1). La moitié sont absents des fichiers d'origine (au moins au stade où a été effectué l'extraction pour l'appariement) et l'autre moitié éliminés par les chaînes de traitement DADS.

Tableau 1 : devenir en 2013 des salariés (principalement ou non) en 2012

	Absents sources adm.	DADS diffusé	DADS, non diffusé	Bns, uniquement	Bns et DADS non dif.	Total
Absents EEC 2013	105	2 013	72	6	7	2 202
N Non-salariés	12	367	15	77	34	509
P Mixtes	0	214	0	3	2	219
S Salariés	121	21 710	122	0	2	21 960
Y Frontaliers	3	49	2	0	0	53
Z Non emploi	809	3 158	532	6	18	4 523
<i>Dont Jemp 2012</i>	<i>492</i>	<i>1 574</i>	<i>151</i>	<i>0</i>	<i>4</i>	<i>2 221</i>
Total	1 049	27 520	743	91	63	29 470
<i>Dont Jemp 2012</i>	<i>570</i>	<i>2 232</i>	<i>156</i>	<i>2</i>	<i>8</i>	<i>2 969</i>

Champ : appariés P4T de rang < 6 en 2012T4 (i.e. enquêtables en 2013) salariés purs, principalement salariés et principalement non-salariés

Lecture : environ 2,0 millions³⁴ de salariés à NIR correct présents dans les fichiers DADS de diffusion en 2012 ne le sont pas dans ceux de 2013. 806 000 (743 000 + 63 000) d'entre eux sont retrouvés dans les fichiers ayant servi à l'appariement mais n'ont pas été retenus dans les fichiers de diffusion. Parmi ces 806 000, 550 000 ne sont pas en emploi en 2013 selon l'EEC (dont 155 000 n'étaient déjà pas en emploi selon l'EEC en 2012 : les JEMP). Si on se limite à ceux qui sont salariés en 2013 selon l'EEC, 250 000 (21 960 – 21 710) sont absents des fichiers de diffusion DADS 2013 (alors qu'ils étaient dans celui de 2012),

Ce chiffre de 250 000 constitue une estimation plancher des trous de collecte totaux (tous les postes d'un salarié sont absents) des DADS. Ne sont en effet pas comptés dans cette estimation :

- La part en emploi des 190 000 absents à la fois de l'EEC 2013 et des DADS diffusées
- La part en emploi des 650 000 JEMP 2012 absents des DADS diffusées (et pour lesquels l'information EEC ne suffit pas à confirmer le non-emploi)
- Ceux absents à la fois en 2012 et 2013

Tableau 2 : comparaison de la distribution du revenu salarial annuel entre présumés trous de collecte et NIR incorrects

	Présents DADS 2012, absents DADS 2013 (ou 2011) et salariés EEC 2013 (ou 2011)	NIR incorrect en 2012 (DADS exhaustives)
1 € à 1 K€	12	36
1 à 4 K€	19	29
4 à 12 K€	19	20
12 à 18 K€	21	8
18 K€ et +	29	7

Lecture : 12% des trous de collecte 2013 présumés, contre 36% des NIR incorrect ont, en 2012, un revenu salarial compris entre 1 et 1000 €.

³⁴ Chiffre tiré du tableau (29,5 - 27,5) ; A partir des DADS exhaustives, il y a près de 3 M de salariés de 2012 absents en 2013. La différence est principalement due aux NIR fictifs. Sur les NIR corrects (ind_nir='C') le taux de disparition est de 7%, conforme au tableau établi à partir de l'appariement.

Certains de ces trous de collecte (au niveau du NIR) pourraient être présents en 2013 sous un NIR incorrect. La comparaison de la distribution du revenu salarial montre que la population à NIR incorrect est fort différente de celle visée ici (tableau 2). Certes le revenu salarial des NIR incorrects est parfois incomplet (si le même individu est présent sous deux NIR différents). Mais il faudrait alors supposer que ceux à NIR correct en 2012, salariés en 2012 et 2013 selon l'EEC, soient présents dans les DADS 2013 sous plusieurs NIR, tous incorrects.

Parmi les près de 1 million de salariés qui ont aussi une activité non salariée, à titre principal ou non, ceux qui sont absents des DADS de diffusion 2013 sont presque tous présents dans les bases non-salariées (150 m). Les absences des bases administratives en 2013 concernent donc presque uniquement, parmi les salariés, les salariés purs (tableau 3).

Tableau 3 : devenir en 2013 des salariés purs en 2012

		Absents ba	DADS diffusé	DADS, non diffusé	Bns*, uniquement	Bns et DADS non dif.	Total
	Absents EEC 2013	105	1 950	69	0	0	2 123
N	Non-salariés	11	231	15	0	0	256
P	Mixtes	0	84	0	0	0	84
S	Salariés	121	21 320	118	0	0	21 560
Y	Frontaliers	3	44	2	0	0	49
Z	Non emploi	809	3 087	532	0	0	4 428
	Total	1 047	26 720	736	0	0	28 500

Champ : appariés P4T de rang < 6 en 2012T4 (i.e. enquêtés en 2013) salariés purs

* On ne peut pas, avec les bases disponibles, faire le lien entre les salariés purs en 2012 (pas d'identifiant non-salarié) et les non-salariés purs en 2013. Ce défaut de suivi ne concerne qu'une population sans doute réduite : en témoigne la proportion de non-salariés selon l'EEC dans les absences des bases administrative (1,0%) pas plus élevé que la moyenne (0,9%). Cela est confirmé par l'utilisation de l'appariement 2013 : on récupère, parmi les salariés purs appariés aux DADS en 2012, 8 (16 m en pondéré) non-salariés purs appariés aux bases «non-salariés» en 2013

Pour les non-salariés, les ordres de grandeur sont similaires : 37 000, soit 10 %, sont absents des bases administratives en 2013 alors qu'ils sont non-salariés selon l'EEC (tableau 4).

Tableau 4 – devenir en 2013 des non-salariés en 2012

		Absents ba	Salariés et non-salariés	Non-salariés purs	Salariés purs*	Appariés 2013 à salariés*	Total
	Absents EEC 2013	33	53	157	36	0	279
N	Non-salariés	37	232	2 269	6	21	2 565
P	Mixtes	0	127	13	3	2	144
S	Salariés	0	253	37	145	7	443
Y	Frontaliers	3	0	4	5	0	12
Z	Non emploi	51	67	169	23	13	323
	Total	124	732	2 649	218	43	3 766

Champ : appariés P4T de rang < 6 en 2012T4 (i.e. enquêtés en 2013) non-salariés purs, principalement salariés et principalement non-salariés

* les salariés purs 2013 sont directement identifiables pour les non-salariés 2012 ayant aussi une activité salariée ; pour les non-salariés purs en 2012, seul l'appariement permet d'identifier ceux devenus salariés purs.

4 – Complément au III.B.1 : détail de l'analyse du décalage de paie (et de la correction des dates)

L'analyse du décalage de paie nécessite le recours au fichier des périodes : c'est là que figure la variable déclarée par les entreprises, selon les modalités suivantes :

- 01 - sans décalage de paie
- 02 - adoption du décalage de paie
- 03 - décalage de paie constant
- 04 - suppression du décalage de paie
- 05 - activité en décalage de paie

La modalité 05 correspond à des périodes en décalage de paie incluses dans le mois de décembre et dont les éléments de salaire seront déclarés l'année suivante : elles sont quasiment toutes (97%) supprimées lors du passage au fichier postes.

a) Impact sur le nombre de PEA

À l'impact attendu du décalage de paie s'ajoute l'effet du traitement des périodes commençant et se terminant en décembre n-1. Les dates dans les DADS sont converties en nombres (dat) prenant leur valeur de 1 à 360 : $\text{dat} = (\text{numéro du mois} - 1) * 12 + \text{jour dans le mois (plafonné à 30)}$. Ce traitement ignore l'année. Les périodes commençant en décembre n-1 et se poursuivant en n sont alors caractérisées par une valeur dat de début postérieure à celle de fin : elles sont corrigées en retranchant 361 à la date de début (le 1^{er} décembre n-1 devient ainsi - 30). Avec cet algorithme, les périodes commençant et se terminant en décembre n-1 ne sont pas corrigées (car la valeur de la date de début est inférieure à celle de la date de fin). Cela n'a pas de conséquences pour le traitement des salaires (objectif des DADS au moment où ont été élaborées les chaînes de traitement) mais en a pour l'emploi au 31/12, notamment pour les effectifs des établissements. Pour le nombre de PEA, les personnes n'ayant, dans le fichier 2012, que des périodes se terminant en 2011 sont considérées en emploi DADS 2012 au seul examen du fichier poste. Symétriquement, certaines personnes en emploi DADS fin 2012 sont uniquement dans le fichier DADS 2013.

On peut corriger cette anomalie en utilisant les variables "année du début de période" et "année de fin de période" du fichier périodes (elles ne sont pas reprises dans le fichier postes).

L'effet de ce décalage de paie est estimé ici à partir d'un échantillon au 1/100^e des DADS (périodes et postes) 2012 et 2013 (y compris DOM). On ne considère que les périodes des postes retenus dans les fichiers DADS de diffusion. 650 000 postes de 2012 ont au moins une période se terminant en 2011. Cela concerne 620 000 PEA. L'effet net est de 350 000 sur les postes (ils n'ont aucune autre période en 2012), 60 000 sur les PEA et 175 000 sur l'emploi au 31/12³⁵. L'effet sur les PEA est plus que compensé par la prise en compte des périodes 2012 des DADS 2013 : + 96 000. L'écart correspond grosso modo à des trous de collecte : on récupère sur 2013 non seulement les petits boulots de décembre 2012 mais aussi des postes en décalage de paie absents en 2012 ; 30 000 PEA sur les 96 000 étaient dans le fichier postes 2011.

b) Choix entre dates déclarées et dates corrigées

On justifie ci-après le choix de travailler à partir des dates déclarées (pas uniquement pour le décalage de paie).

Avant de corriger les dates pour tenir compte du décalage de paie, il faut choisir à quelles dates doivent s'appliquer les corrections : aux dates déclarées par l'entreprise ou aux dates corrigées par les chaînes de traitement DADS - 5% des postes ont leurs dates corrigées³⁶, indicatrice IR_dates. Plusieurs arguments plaident en faveur du retour aux dates déclarées :

- Les chaînes de traitement corrigent les dates en fonction de critères statistiques (notamment pour assurer la cohérence avec les salaires) sans doute globalement satisfaisants mais qui, au cas par cas (ce qui nous intéresse ici dans la confrontation avec l'EEC) peuvent s'avérer peu pertinents.
- Le caractère asymétrique des corrections : en cas de période jugée trop longue, on décale la date de début de période, pas la date de fin.
- La disponibilité des variables pour effectuer les tests de cohérence interne à la période, cohérence entre la durée (en nombre de jours), la quotité (temps plein / temps partiel) et le nombre d'heures. Souvent la variable nbheur est mise à zéro dans le fichier des périodes alors qu'au moins une des deux variables déclarées - nombre d'heures ou heures travaillées - est renseignée.
- Ex post, il apparaît que la cohérence externe (avec l'EEC) est mieux assurée en partant des dates déclarées (tableau 1)

³⁵ Soit beaucoup moins que le nombre de personnes concernées dont la période 2011 se termine la dernière semaine de décembre : l'impact principal du décalage de paie se trouve sans doute dans l'estimation de la multi activité au 31/12.

³⁶ C'est pratiquement toujours la date de début qui est corrigée, une fois sur deux pour l'avancer, une fois sur deux pour la reculer.

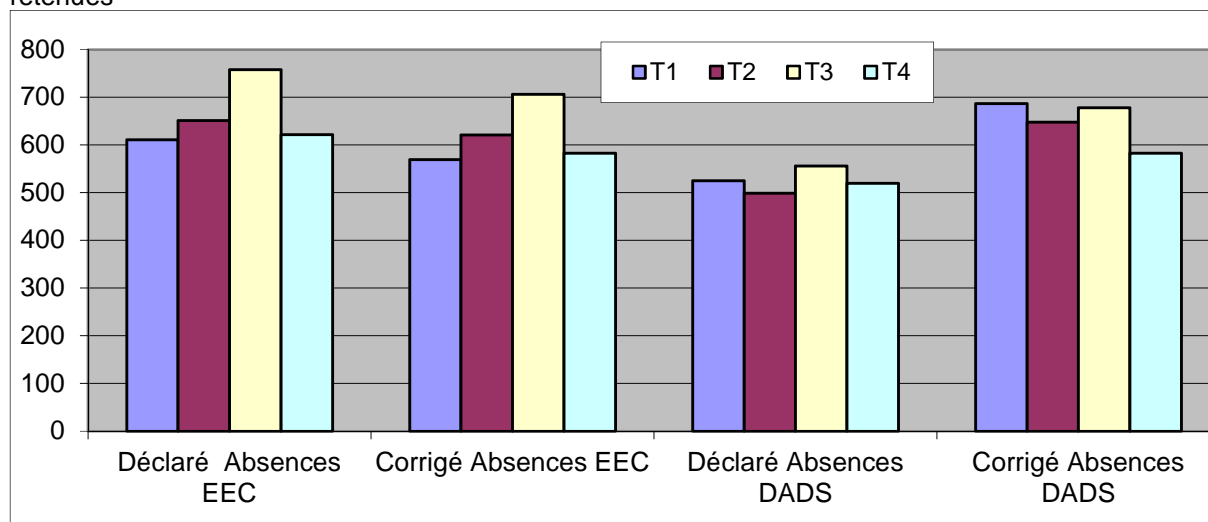
Tableau 1 : écarts entre sources selon les variables dates utilisées (en milliers)

	Dates déclarées	Dates corrigées
Absences EEC	673	632
Absences DADS	526	671
Total	1 199	1 303

*Champ : purs salariés PEA2, moyennes des 3 premiers trimestres (pour éviter l'effet du décalage de paie)
Lecture : En prenant les dates déclarées du fichier des périodes, il y a en moyenne sur les trois premiers trimestres 673 000 absences EEC (en emploi la semaine de référence selon les DADS mais pas selon l'EEC) et 526 000 absences DADS.*

La correction des dates diminue les absences EEC et augmente les absences DADS, ce qui signifie que ce sont essentiellement les corrections par raccourcissement des périodes (en reculant la date de début) qui ont un impact sur la cohérence externe. L'augmentation des absences DADS est de loin supérieure à la réduction des absences EEC, ce qui signifie que le recul de la date de début est le plus souvent effectué à tort.

Graphique 1 : profil trimestriel des absences EEC et des absences DADS en fonction des dates retenues



Champ : salariés purs PEA2

Lecture : au T1 2012, il y a un peu plus de 611 000 absences EEC (pas d'emploi EEC alors que emploi selon les DADS) par rapport à l'emploi DADS identifié à partir des dates déclarées et 569 000 à partir des dates corrigées ; il y a 525 000 absences DADS (pas d'emploi DADS alors que emploi selon l'EEC) avec les dates déclarées et 687 000 avec les dates corrigées

Le profil trimestriel des écarts confirme cette interprétation (graphique 1) : celui des absences EEC présente un pic marqué au T3 (petits boulots d'été) visible aussi bien avec les dates déclarées qu'avec les dates corrigées alors que celui des absences DADS est nettement différent selon que l'on prend les dates déclarées (profil assez plat avec un léger pic en T3) ou les dates corrigées (profil décroissant avec un creux marqué au T4, résultat de l'asymétrie des corrections de dates, qui ne portent que sur les dates de début).

c) traitement du décalage de paie

Pour la confrontation des sources, il importe de corriger le mois de décembre. Plusieurs possibilités

- Utiliser la variable ind_3112 du fichier poste pour prolonger jusqu'au 31/12 les postes s'arrêtant avant. C'est la seule possibilité avec le fichier « postes ». Trois inconvénients :
 - l'inclusion pour décembre 2012 de postes de décembre 2011, le plus souvent avec une indicatrice de présence au 31/12 positive
 - l'absence des postes de décembre 2012 inclus dans le fichier 2013
 - le côté un peu boîte noire de cette indicatrice
- L'utilisation du fichier « périodes » 2012. Cela permet d'éliminer les postes de 2011. On utilise la variable de décalage de paie pour prolonger jusqu'au 31/12 les périodes s'arrêtant au 30/11. Les périodes de décembre 2012 du fichier 2013 restent toutefois manquantes.

- L'utilisation du fichier « périodes » 2013 en complément du fichier « périodes » 2012. C'est le seul moyen de remédier aux 3 inconvénients du fichier postes.

Voici les principaux points de l'algorithme de correction des dates implémenté ici pour la dernière possibilité :

- Champ des périodes prises en compte
 - On ne garde que les périodes de postes retenus dans les fichiers postes de diffusion
 - On élimine les périodes dont l'année de fin de validité est 2011
 - On ajoute, à partir du fichier de 2013, les périodes de 2013 dont l'année de début de validité est 2012 et dont le poste est retenu dans le fichier poste 2012 ou dans celui de 2013
- Traitement du décalage de paie
 - Les périodes appartenant à des postes pour lesquels on a ajouté une période provenant du fichier 2013 ne sont pas concernées par ce qui suit ;
 - Les périodes se terminant au 30 novembre sont prolongées jusqu'au 31 décembre dans les cas suivants :
 - Périodes débutant en 2011
 - Périodes débutant en 2012 et avec la variable de décalage de paie à '02' ou '03'

Les deux dernières possibilités peuvent être testées sur l'ensemble des postes du fichier période ou sur le fichier période restreint aux postes présents dans le fichier postes.

Tableau 2 : écarts entre sources au T4 selon le mode de traitement du décalage de paie (en milliers)

	Fichier postes	Périodes des postes retenus		Toutes périodes (hors chômage)		
		Sans 2013	Avec 2013	Sans 2013	Avec 2013	2013 sans cor*
Absences EEC	630	589	622	624	654	600
Absences DADS	578	556	520	525	491	1 165
Total	1 208	1 145	1 142	1 149	1 145	1 765

Champ : purs salariés PEA2, 4^e trimestre

Lecture : En prenant les dates déclarées du fichier des périodes 2012, hors celles de dont l'année de fin de validité est 2011, il y a 589 m absences EEC (en emploi la semaine de référence selon les DADS mais pas selon l'EEC) et 556 m absences DADS.

* ajout des périodes 2012 du fichier 2013 mais sans corriger les périodes du fichier 2012 en décalage de paie

Le passage aux périodes sur le champ du fichier postes diminue le nombre des absences EEC de 41 000 (tableau 2) : c'est l'impact à la fois de la fusion des périodes dans le fichier postes (malgré l'utilisation des variables datdeb2 et datfin1) et de suppression des périodes de décembre 2011 dans le fichier des périodes (cf. supra). Il diminue aussi les absences DADS, sans doute suite à l'utilisation des dates déclarées (fichier périodes, alors que seules les dates corrigées sont dans le fichier postes). L'intégration des périodes 2012 déclarées en 2013 (3^{ème} colonne) diminue logiquement les absences DADS mais augmente d'un montant presque équivalent les absences EEC : sans doute parce que ce sont des petits boulots qui sont ajoutés (les périodes de plus d'un mois en décalage de paie sont déjà corrigées du décalage de paie dans le traitement du fichier périodes de 2012).

La prise en compte de l'ensemble des périodes réduit encore les absences DADS mais au prix d'une augmentation des absences EEC à peu près équivalente. A noter que 2013 ne suffit pas à compenser le décalage de paie (dernière colonne): il faut en plus corriger les périodes du fichier 2012.

On retient les dates calculées à partir du fichier périodes enrichi de 2013 (3^e colonne du tableau, correspondant à l'algorithme détaillé supra).

5 – Complément au III.B.2 : détail de l'analyse sur la durée, le nombre d'heures et la densité

Calcul du nombre d'heures et des durées

- Dans le fichier périodes, le nombre d'heures déclaré est parfois un nombre d'heures rémunérées, parfois un nombre d'heures travaillées. On prend par défaut le nombre d'heures rémunérées. S'il est nul ou non renseigné, on prend alors le nombre d'heures travaillées, que l'on multiplie par 1,1 (pour tenir compte des congés et se ramener ainsi au concept des heures rémunérées).

- Pour les périodes commençant en 2011 ou se terminant en 2013, on les cale sur 2012 en mettant la date de début (resp. de fin) au 1^{er} janvier (resp. 31 décembre) et on calcule la durée sur l'année civile en appliquant une règle de trois.

a - Cohérence interne

Le contrôle de cohérence interne repose essentiellement sur la notion de densité, le nombre d'heures divisé par la durée. On s'attend à une densité au moins égale à 5 (35 / 7) pour les temps plein. Pour les temps partiels, on normalise la densité en la divisant par la quotité (comprise entre 0 et 1). Pour caractériser la modalité d'exercice de l'emploi, on mobilise trois variables

- CIPDZ³⁷ : C pour complet, I pour intermittent, P pour partiel, D pour domicile et Z pour autre. Cette variable n'est renseignée que pour 62% des périodes. Elle est absente pour les particuliers employeur (PE) et la fonction publique d'état (FPE).
- Temps partiel : quotité
- Caractéristique du contrat. Plus détaillée que CIPDZ, elle n'est pas renseignée pour les PE mais l'est pour la FPE (avec toutefois la seule distinction entre temps plein et temps partiel).

On construit une variable ttp reprenant CIPDZ (avec I et D regroupés sous la modalité atypique), complétée par la caractéristique du contrat. Comme aucune de ces deux variables n'est renseignée pour les PE, on crée une modalité spécifique pour les isoler. La variable de quotité est cohérente, à 1% près, avec la distinction temps plein/ temps partiel et permet d'affecter une petite partie des atypiques en temps partiel. Si on juge plus prudent de ne pas considérer que les atypiques sont à temps plein quand la quotité n'est pas renseignée, il reste 38 000 périodes (somme des 3 dernières lignes du tableau 1), soit près d'une sur quatre, pour lesquelles le contrôle de cohérence interne entre la densité et la quotité n'est pas possible.

Tableau 1 : nombre de périodes par type

Selon la variable CIPDZ retravaillée (ttp, cf. texte)		Selon la variable quotité		
		Temps plein ou NR	Temps partiel	Total
aTpl	Temps plein	82 359	393	82 752
bTpa	Temps partiel	453	33 003	33 456
cAty	Atypique (intermittent, à domicile, autre)	13 160	2 917	16 077
dPE	Salarié des particuliers employeurs	13 617	0	13 617
zZzz	Aucune information	10 061	143	10 204
Total		119 650	36 456	156 106

Champ : toutes périodes (des postes diffusés) des personnes appariées

On calcule la densité normalisée par la quotité ($ds_q = n\text{heur}/\text{durée} * 1/\text{quotité}$) pour les périodes de plus d'une semaine. Les périodes à quotité non renseignée sont mises à temps plein. Pour conserver l'information liée à la confrontation entre CIPDZ et la quotité, on crée les nouvelles modalités suivantes :

- Atp pour les atypiques dont la quotité est renseignée (donc, à temps partiel)
- Inc (incohérence) pour les temps plein ou temps partiel CIPDZ quand la quotité est incohérente avec CIPDZ

On constate des densités trop faibles (<5) pour une part non négligeable (23%) des temps plein (tableau 2). Il y a aussi des densités trop élevées (>6.5), notamment pour les temps partiel (11%), signe d'une incohérence entre quotité déclarée et nombre d'heures. Enfin, les atypiques (hors ceux qui ont une quotité, classés en cAtp), les PE et les non renseignés (zZzz) ont le plus souvent une faible densité, ce qui justifie de ne pas les considérer par défaut à temps plein.

Tableau 2 : ventilation des types de période par densité (en %)

ttp		Heures NR	Densité normalisée				Total
			< 4	4 à 5	5 à 6.5	> 6.5	
aTpl	Temps plein	3	9	14	73	2	75 362
bTpa	Temps partiel	2	7	13	67	11	26 792
cAtp	Atypique, quotité renseignée	5	8	9	71	7	2 440
cAty	Atypique, quotité NR	29	45	7	18	2	9 962
dPE	Salarié des PE	0	89	7	4	1	13 175
gInc	Incohérence	7	36	40	17	1	573
zZzz	Aucune information	28	52	7	13	1	8 713
Total		7 868	29 702	16 612	77 851	4 984	137 017

Champ : Périodes d'au moins une semaine (des postes diffusés) des personnes appariées

³⁷ Variable du fichier périodes. Dans le fichier postes CIPDZ figure, après redressement, sous le nom de CPF. En cas de désaccord entre CIPDZ et CPF, l'EEC est plus souvent en accord avec CIPDZ.

b - Effet sur la cohérence externe (présence en emploi)

Une densité trop faible est le signe, soit d'un trop faible nombre d'heures déclarées, soit d'une durée trop longue, quand, par exemple, plusieurs courtes périodes de travail sont déclarées en une seule période, ou quand les périodes sont déclarées par défaut comme allant du 1^{er} janvier au 31 décembre. En cas de durée trop longue, le salarié ne travaille pas dans l'entreprise pendant toute la période : s'il est interrogé une semaine de cette période où il ne travaille pas dans cette entreprise, on observera alors, soit, s'il ne travaille pas au même moment ailleurs, une incohérence entre l'emploi DADS (oui) et l'emploi EEC (non), ce que l'on appelle ici une absence EEC, soit de la multi activité à tort (non traitée ici).

L'utilisation des périodes dans l'analyse de la cohérence externe est délicate dans les cas où plusieurs périodes coexistent (sur un ou plusieurs postes) la même semaine de référence. Le traitement de ces cas (un quart des observations) est nécessaire pour l'estimation de l'emploi mais comme il ne s'agit ici que de mettre en évidence l'impact de la densité sur les absences EEC, on se limitera aux cas à une seule période (l'arbitrage au III.C.4 couvre l'ensemble des cas).

La faible densité n'est pas le seul facteur pouvant expliquer les absences EEC. La sous-déclaration des petits boulots en est un autre : un tiers des périodes d'un mois maximum, un cinquième des périodes de 1 mois à un trimestre, ne sont pas déclarées. L'effet de la densité ne se fait sentir que sur les périodes de plus d'un trimestre. Pour les périodes couvrant toute l'année, le taux d'absence EEC est de 1% quand la densité est normale et de 6 à 7 % sinon (tableau 3).

Tableau 3 : situation dans l'EEC en fonction de la durée de la période DADS et de la densité

Durée période	Toute l'année		91 J à 1 an		31 à 90 J	1 à 30 J	Total
	OK	Non	OK	Non			
Densité normalisée (OK de 5 à 6.5)							
Emploi EEC BIT	98,6	92,8	91,6	83,6	78,8	65,2	127 264
Congés EEC	0,3	0,9	0,6	1,1	0,7	1,0	791
Emploi EEC calendrier	0,0	0,1	0,4	0,9	1,0	1,8	280
Absence EEC	1,1	6,2	7,4	14,4	19,5	32,0	7 152
Total	67 658	37 246	11 596	12 971	4 169	1 847	135 487

Champ : appariés * semaine de référence où emploi DADS sur une seule période.

Nota : dans ce tableau, contrairement aux précédents, une même période peut être comptée plusieurs fois (si, pendant son emploi sur cette période, le salarié répond à plusieurs interrogations EEC).

Lecture : quand un salarié est interrogé une semaine où il est en emploi DADS sur une période DADS de plus d'un trimestre et de moins d'une année et de densité normale, il est : dans 91,6 % des cas en emploi EEC au sens du BIT ; sinon, dans 0,6 % des cas, en congés de longue durée ; sinon, dans 0,4 % des cas, en emploi selon le calendrier de l'EEC ; sinon, dans 7,4 % des cas, en absence EEC, quel que soit le critère d'emploi utilisé.

Si on se concentre sur les périodes couvrant toute l'année (deux cas sur trois dans le tableau 3), on peut détailler l'effet de la densité selon le type de période. C'est quand les heures ne sont pas renseignées que le taux d'absence EEC est le plus élevé (tableau 4). Sinon, elles décroissent avec la densité pour remonter aux densités trop élevées qui signalent une anomalie. Le taux est très faible, autour de 1%, pour les périodes de densité normale, sauf pour les temps partiel. Pour ces derniers intervient un effet petit boulot pour les faibles quotités : le taux d'absence EEC est de 13% pour les quotités de moins d'un jour par semaine et de 9% pour celles d'un jour à moins d'un mi-temps (en cas de densité cohérente, chiffres hors tableau). On notera que les salariés des PE à un seul employeur ont, pour chaque strate de densité, le plus faible taux d'absence EEC mais ils ont presque tous une très faible densité (pour les PE la quotité n'est pas renseignée et est mise ici à 1 : la densité vaut donc pour la quotité).

Tableau 4 : taux d'absence EEC par type de période et densité

		Heures NR	Densité normalisée				Total
			< 4	4 à 5	5 à 6.5	> 6.5	
aTpl	Temps plein	7,2	7,3	1,0	0,7	0,8	1,1
bTpa	Temps partiel	8,0	8,8	4,5	3,2	7,6	4,0
cAtp	Atypique, quotité renseignée	4,3		14,5	1,3		3,9
cAty	Atypique, quotité NR	2,0	16,9	0,3	1,0		3,9
dPE	Salarié des PE		3,7	0,3	0,0	2,4	3,3
gInc	Incohérence		16,6	1,9			13,0
zZzz	Aucune information	31,4	21,5	6,1	2,1	0,0	25,6
Total		14,3	6,1	1,6	1,1	4,4	2,9

Champ : appariés * semaine de référence où emploi DADS sur une seule période couvrant toute l'année.

c - Quotité et cohérence externe

La quotité n'est pas toujours renseignée. On l'attribue ou on l'impute en fonction du cas de figure :

A	Tps plein ou quot renseignée	pour les temps plein q = 1 et pour les temps partiels q = quot
H	Imputé via nbHeures	q inconnu, nbheures connu : q = min (nbheures/(dur*35/7), 1)
P	Imputé via densité du Poste	Si nbheures est inconnu, on calcule q à partir de la densité moyenne du poste
S	Imputé via Salaires	Quand les heures sont inconnues au niveau poste comme au niveau période, q = min (salaire / (10 * dur * 35 / 7), 1)
Z	Forcé à 0.5	ni heures, ni salaire connu : qt = 0.5 (cas très rare)

Les quotités imputées sont majoritairement inférieures à un mi-temps (tableau 5). Cela peut correspondre aussi bien à du véritable mi-temps qu'à de l'intermittence. La confrontation avec l'EEC permet de le vérifier.

Tableau 5 : quotités par mode d'affectation (en %)

	< 0,2	< 0,5	0,5	< 0,8	0,8	0,9	1-	1	Total
A Tps plein ou quot renseignée	5	7	2	7	4	2	1	72	104 848
H Imputé via nbHeures	48	17	2	10	2	3	5	14	26 873
P Imputé via densité du Poste	34	7	1	9	2	6	7	34	4 314
S Imputé via Salaires	40	34	1	6	2	1	2	14	947
Z Forcé à 0.5	0	0	100	0	0	0	0	0	35
	19	11		10				80	
Total	616	967	3 165	458	4 281	3 775	2 985	770	137 017

Champ : Périodes d'au moins une semaine (des postes diffusés) des personnes appariées

Il y a nettement plus de temps plein selon l'EEC (106 m) que selon les DADS (87 m, tableau 6). L'excès de temps partiel dans les DADS vient surtout des plus faibles quotités où le soupçon d'intermittence est fort. Moins important en volume, mais très significatif, est l'excès des quotités strictement comprises entre 0.8 et 1.

Tableau 6 : comparaison des quotités EEC (en ligne) et des quotités DADS déclarées ou imputées

	< 0.5	0.5	< 0.8	0.8	0.9	1-	1	Total
NR	399	8	42	5	6	17	95	572
< 0.5	6 395	187	331	47	35	49	324	7 368
0.5	2 222	1 197	929	50	47	51	480	4 976
0.5 - 0.8	3 369	324	3 026	333	256	112	559	7 979
0.8	850	72	730	2 295	680	206	426	5 259
0.8 - 1	745	40	189	176	763	345	218	2 476
1	12 367	799	3 709	904	1 698	2 026	84 988	106 491
Total	26 347	2 627	8 956	3 810	3 485	2 806	87 090	135 121

Champ : appariés * semaine de référence où emploi DADS sur une seule période de plus d'une semaine.

L'excès de périodes de faible quotité est surtout sensible pour les quotités imputées à partir du nombre d'heures déclarées, là où le soupçon d'intermittence est le plus fort (les tableaux correspondants ne sont pas reproduits).

Le taux d'absence EEC dépend plus de la quotité DADS que du mode d'établissement de cette quotité, à l'exception de l'imputation via les salaires (tableau 7). Ceci justifie l'utilisation de la densité corrigée de la quotité, y compris quand la quotité est imputée, pour l'arbitrage sur les absences EEC.

Tableau 7 : taux d'absence EEC en fonction de la quotité et de son mode d'affectation (en %)

	< 0,2	< 0,5	0,5	< 0,8	0,8	0,9	1-	1	Total
A Tps plein ou quot renseignée	18	15	5	5	1	3	3	3	4
H Imputé via nbHeures	9	11	7	6	2	3	2	4	8
P Imputé via densité du Poste	16	23	,	13	9	8	3	1	8
S Imputé via Salaires	43	48	,	,	,	,	,	27	42
Total	13	16	6	6	2	3	3	3	14

Champ : appariés * semaine de référence où emploi DADS sur une seule période de plus d'une semaine.

6 – Complément au III.D : détails du traitement des non-salariés

1 - Indicateurs d'activité

On dispose d'une indicatrice d'activité au 31/12 et d'une variable de revenu. Les revenus pouvant être négatifs, on retient comme en inactivité les individus à revenus nuls ou non renseignés. La variable ainsi construite, *actr*, comporte 3 modalités :

- A : actif et revenu renseigné non nul
- B : actif et revenu non renseigné ou nul
- Z : inactif

Ce ne sont toutefois pas ces variables qui sont utilisées pour les publications. Le revenu n'est pas mobilisé, sauf pour les autoentrepreneurs. On reconstitue une variable de diffusion, en lui donnant la valeur 0 (non-diffusion) pour les cas suivants :

- Les taxés d'office à supprimer des statistiques (variable *eviction_to*)
- Les auto-entrepreneurs inactifs selon l'indicatrice d'activité
- L'indicateur d'activité selon l'Acoss (variable *CODETAT*)

De plus, pour l'analyse de l'appariement, l'activité au 31/12 n'est pas la seule pertinente. On dispose de la variable *DATEDISP* (non renseignée pour la MSA) permettant de récupérer les cessations de 2012 (potentiellement appariables avec l'EEC 2012).

Presque tous les non-salariés inactifs au 31/12 sont cessés en cours d'année et donc à retenir pour l'appariement (tableau 1).

Tableau 1 : critères d'activité

		Diffusé	Cessés en cours d'année	Non diffusé	Total	Tx app.
A	Actif et revenu non nul	2 306	0	25	2 332	26
B	Actif et revenu nul	271	0	90	362	23
Z	Inactif au 31/12	14	237	11	261	19
Total		2 592	237	126	2 954	
Taux d'appariement (1/10000)		26	19	16		25

Champ : non-salariés (NSA) et principalement non-salariés (PNS)

Le taux d'appariement est fonction croissante de la confiance quant à l'activité. Cela ne va pas de soi. Trois raisons peuvent être invoquées :

- Les identifiants incluent des variables (commune ou SIREN) liées à l'activité. Mais cela ne justifie pas une telle différence dans la mesure où les non salariés travaillent dans leur majorité dans leur commune de résidence
- Des effets de structure : sur représentation de la MSA parmi les revenus nuls, des auto entrepreneurs parmi les inactifs et les non-diffusés.
- Les personnes à la marge de l'emploi sont sous représentées dans l'EEC

Ce dernier point est illustré par le plus faible taux d'appariement des non-salariés purs à revenu renseigné positif et inférieur à 5 K€ (tableau 2).

Tableau 2 : taux d'appariement par tranche de revenu

Tranche de revenu	Négatif	Nul	< 1K€	1 - 5	5 - 10	10 - 50	50 K€ +	Total
Nombre (en milliers)	28	250	105	223	248	1 086	383	2 328
Taux d'appariement (1/10000)	34	25	23	24	27	27	27	26

Champ : non-salariés purs (NSA) actifs au 31/12 et diffusés

2 - Emploi des JEMP

L'effet revenu sur la déclaration en emploi dans l'EEC est très net et significatif, malgré le faible nombre d'observations quand on se restreint aux appariés présents les 4 trimestres dans l'EEC (tableau 3).

Tableau 3 : taux de JEMP par tranche de revenu

Tranche de revenu	Négatif	Nul	< 1K€	1 - 5	5 - 10	10 - 50	50 K€ +	Total
Nombre d'observations	21	168	64	152	161	762	270	1 601
Taux de JEMP (%)	0	11	45	12	3	1	0	4

Champ : non-salariés purs (NSA) actifs au 31/12 et diffusés, appariés à l'EEC P4T

Il semble donc justifié de considérer les JEMP non-salariés comme de l'emploi quand les indicateurs d'activité des bases non-salariés sont positifs, cas auxquels se limite le tableau X ci-dessus. Dans les autres cas, la part de JEMP est plus élevée (tableau 4), sans que l'on puisse l'attribuer à un effet revenu (le revenu est souvent non renseigné et, quand il l'est, il est plutôt élevé).

Tableau 4 : indicateur d'activité synthétique et JEMP

		Nb observations	% JEMP	Nb JEMP pondéré (milliers)
A	Actif au 31/12, diffusé, revenu non nul	1 430	3	69
B	Actif au 31/12, diffusé, revenu nul	171	9	24
C	Cessé en cours d'année	94	17	35
Z	Autres cas	41	17	13
Total		1 736	5	140

Champ : non-salariés purs (NSA) appariés à l'EEC P4T (pas de JEMP pour les PNS)

Au total, parmi les 140 000 JEMP non-salariés, 92 000 sont conservés et 48 000 sont exclus du champ des PEA.

3 - emploi des PEA2

Comme la seule source sur le profil annuel de l'emploi des non-salariés est l'EEC, l'emploi moyen des PEA2 sera le nombre moyen de trimestres en emploi selon l'EEC. On vérifie que l'emploi moyen annuel par PEA est plus faible pour les non-salariés cessés au cours de l'année 2012 (83%) ou sans indice fort d'activité (82%) (tableau X). Mais, même pour ces cas, la majorité des non-salariés sont en emploi EEC les 4 trimestres. On peut aussi, comme proxy de l'emploi moyen annuel selon les bases non-salariés, ne retenir que les PEA2 actifs au 31/12 (dernière colonne du tableau). L'emploi non salarié selon les bases « non-salariés » est alors, à pondération identique, 100 000 en deçà de celui selon l'EEC (dont 62 000 pour les seuls principalement non salariés, signe qu'il faudrait aussi prendre en compte l'emploi salarié).

Tableau 5 : nombre de trimestre en emploi et emploi moyen selon les cas

		Nb trimestres en emploi EEC (%)				Nb PEA2 (milliers) (a)	Nb emplois (b)	100 * b/a	Nb PEA2 actifs 31/12
		1T	2T	3T	4T				
A	Actif, rev non nul	2	3	3	92	2 213	2 126	96	2 213
B	Actif, rev nul	3	7	7	83	257	238	93	257
C	Cessé en 2012	14	10	9	68	149	124	83	0
Z	Autre cas	17	6	9	68	59	48	82	0
<i>Sous total NSA</i>		3	3	4	89	2 678	2 535	95	2 470
PNS	Principalem. NS	9	3	5	83	281	254	90	219
Total		4	3	4	88	2 958	2 789	94	2 689

Champ : PEA2 non-salariés

Au total, en ajoutant les 98 000 JEMP (92 000 pour les non-salariés purs et 5 000 pour les principalement non-salariés), les non-salariés et principalement non-salariés représentent près de 2,9 millions d'emplois. (hors trous de collecte)

Annexe 2 – Glossaire des sigles

Nb : les sigles spécifiques à cette étude (ex PEA) sont en maigre, les autres (ex DADS) en gras

ACOSS	Agence Centrale des Organismes de Sécurité Sociale
BA MSA)	Bases administratives (DADS + non-salariés relevant de l'ACOSS + non-salariés relevant de la MSA)
BIT	Bureau International du Travail
DADS	Données Annuelles de Déclarations Sociales
DMA	Durée Moyenne en emploi dans l'Année des PEA
EEC	Enquête Emploi en Continu
EMA	Emploi Moyen dans l'Année
ETP	(volume d'emploi en) Equivalent Temps Plein
JEMP	Jamais en EMPloi dans l'EEC (à aucune des 4 interrogations de l'année)
MSA	Mutualité Sociale Agricole
PCS	(nomenclature des) Professions et Catégories Socio-professionnelles
PE	(salariés des) Particuliers Employeurs
PEA	Personne en Emploi au moins une fois dans l'Année
PEA2	Personnes en Emploi dans l'Année à la fois selon les sources administratives et selon l'EEC
PP	(volume d'emploi en) nombre de Personnes Physiques
RP	Recensement de la Population

Série des documents de travail de la DSDS

- F1905** : Le revenu des ménages selon la comptabilité nationale et selon l'Enquête Revenus Fiscaux et Sociaux (ERFS) - Jérôme Accardo
- F1904** : Appariement entre l'enquête Emploi et le fichier Historique de Pôle emploi sur la période 2012-2017 - Méthode et premiers résultats - Alexis Hameau, Sylvain Larrieu, Anis Marrakchi, Alexis Montaut, Christophe Dixte, Yohan Coder et Sophie Hamman
- F1903-G2019/04** : L'économie numérique fausse-t-elle le partage volume-prix du PIB ?
Lorraine Aeberhardt, Florian Hatier, Marie Leclair, Benoît Pentinat et Jean-Denis Zafar
- F1902** : Les modèles de contrôle des effets de structure : comparaisons et application à une analyse de la disparité départementale des revenus non salariés moyens - Christophe Bertran et Laurianne Salembier
- F1901 - G2019/01** : Effets de moyen terme d'une hausse de TVA sur le niveau de vie et les inégalités : une approche par microsimulation - Mathias André et Anne-Lise Biotteau
- F1808** : La construction d'un indice des prix des logements neufs - Thomas Balcone
- F1807** : Estimer les effectifs de couples de personnes de même sexe au recensement : expérimentation d'une solution de validation du sexe par le prénom - Élisabeth Algava et Sébastien Hallépée
- F1806** : L'effet d'une variation du montant de certains transferts du système socio-fiscal sur le niveau de vie : résultats sur 2016 à partir du modèle de microsimulation Ines (Cahier de variantes) - Maëlle Fontaine et Michaël Sicsic
- F1805** : Family, Firms and the Gender Wage Gap in France - Elise Coudin, Sophie Maillard et Maxime Tô
- F1804** : Mode de collecte et questionnaire, quels impacts sur les indicateurs européens de l'enquête Emploi ? - Klara Vinceneux
How do the collection mode and questionnaire used affect the European indicators in the French Labour Force Survey ? - Klara Vinceneux
- F1803** : Trois versions du taux d'effort en matière de logement - Pascal Godefroy
- F1802** : Heterogeneous exposure to labor earnings risk - Pierre Pora et Lionel Wilner
- F1801** : L'espérance de vie par niveau de vie Méthode et principaux résultats - Nathalie Blanpain
- F1708** : Les données fiscales de l'EDP : une nouvelle source d'informations sur les couples et les familles ? - Vianney Costemalle
- F1707** : La situation matrimoniale dans le recensement : impact de la refonte du questionnaire de 2015. - Guillemette Buisson
- F1706** : Indices de prix à la consommation - Patrick Sillard
- F1705** : Effet d'un choc d'inflation sur le revenu disponible et ses composantes deux ans après : une approche par microsimulation - Anne-Lise Biotteau et Maëlle Fontaine

- F1704** : Scanner data and quality adjustment - Isabelle Léonard, Patrick Sillard, Gaëtan Varlet et Jean-Paul Zoyem
- F1703** : Les structures familiales en France : comparaison entre le recensement, l'enquête famille et logements et l'enquête emploi - Guillemette Buisson et Aude Lapinte
- F1702** : Projections de la population active à l'horizon 2070 - Malik Koubi et Anis Marrakchi
- F1701** : Les taux marginaux effectifs de prélèvement pour les personnes en emploi en France en 2014 - Juliette Fourcot et Michaël Sicsic
- F1606** : Projections de population 2013-2070 pour la France : méthode et principaux résultats - Nathalie Blanpain et Guillemette Buisson
- F1605** : Les durées passées en famille monoparentale - Méthode d'estimation des durées et résultats - Vianney Costemalle
- F1604** : ESeG = European Socio economic Groups - Nomenclature socio-économique européenne - Monique Meron, Michel Amar, Charline Babet, Milan Bouchet-Valat, Fanny Bugeja-Bloch, François Gleizes, Frédéric Lebaron, Cédric Huguée, Étienne Penissat et Alexis Spire
- F1603** : Catégorie sociale d'après les déclarations annuelles de données sociales et catégorie sociale d'après le recensement : quels effets sur les espérances de vie par catégorie sociale ? Comparaison entre les déclarations annuelles de données sociales et les recensements de la population. Comparaison de méthodes d'estimation des espérances de vie - Vianney Costemalle
- F1602** : L'espérance de vie par catégorie sociale et par diplôme - Méthode et principaux résultats - Nathalie Blanpain
- F1601** : Échantillonnage des agglomérations de l'IPC pour la base 2015 - Laurence Jaluzot et Patrick Sillard
- F1508** : Worker-firm matching and the family pay gap: Evidence from linked employer-employee data - Lionel Wilner
- F1507** : Effet des nouvelles mesures sociales et fiscales sur le niveau de vie des ménages : méthodologie de chiffrage avec le modèle de microsimulation Ines - Mathias André, Marie-Cécile Cazenave, Maëlle Fontaine, Juliette Fourcot et Antoine Sireyjol
- F1506** : Nowcasting du taux de pauvreté par la micro-simulation - Maëlle Fontaine et Juliette Fourcot
- F1505/376-501** : Bilan du projet EDP++ - division Camap et division Enquêtes et études démographiques
- F1504** : Contrôles des rémunérations dans les déclarations annuelles de données sociales (DADS) - Une analyse exploratoire pour améliorer la détection des points atypiques - Claire Jacod
- F1503** : Précision de l'enquête Patrimoine 2010 - Pierre Lamarche et Laurianne Salembier
- F1502** : Pourquoi l'indicateur de pauvreté en conditions de vie baisse malgré la crise économique ouverte en 2008 ? Jean-Louis Pan ké shon
- F1501** : Évolution de la population de la France entre 1981 et 2011 : contributions de la fécondité, de la mortalité, du solde migratoire et de la structure de la pyramide des âges - Catherine Beaumel et Pascale Breuil-Genier

- F1410** : "Personal network" and retirement: Is retirement bad for friendship and good for family relationships ? Anne Laferrère
- F1409** : Retraités mais pas en retrait : La retraite pousse-t-elle à de nouvelles activités ? Anne Laferrère
- F1407** : Production "aval" de l'enquête emploi en continu EEC2 2013 - 20XX - Fabien Guggemos
- F1406 bis** : La constitution de l'échantillon démographique permanent de 1968 à 2012 - Stéphane Jugnot
- F1405 (tome 1)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne Daguet
- F1405 (tome 2)** : Hommes et femmes vivant en couple en 2009, 1999 et aux recensements précédents - Fabienne Daguet
- F1404** : L'addition est-elle moins salée ? La réponse des prix à la baisse de TVA dans la restauration en France - Quentin Lafféter et Patrick Sillard
- F1403** : Estimer les flux d'entrées sur le territoire à partir des enquêtes annuelles de recensement - Chantal Brutel
- F1402** : Une rotation de la main d'œuvre presque quintuplée en 30 ans : plus qu'un essor des formes particulières d'emploi, un profond changement de leur usage - Claude Picart
- F1401** : Calculs statistiques de stock et de flux sur la révision électorale 2012-2013 - Christelle Rieg