

# Note sur le rapprochement des enquêtes SalSa 2009 et 2011

Martin CHEVALIER et Damien CARTRON – 24 janvier 2014

Originellement, les Enquêtes sur les salaires auprès des salariés (SalSa) 2009 et 2011 ont été conçues comme les deux « vagues » d'une même enquête. Cependant, les particularités de l'une et de l'autre rendent leur rapprochement particulièrement difficile, au point qu'il semble plus judicieux de les considérer comme deux millésimes distincts d'une même *série* d'enquêtes. Cette courte note revient sur les difficultés inhérentes au rapprochement de ces deux enquêtes et documente l'ensemble des opérations effectuées en ce sens.

## Difficultés posées par le rapprochement des enquêtes SalSa 2009 et 2011

Deux raisons ont conduit à restreindre le champ de l'enquête de 2009 aux seuls salariés présents dans les fichiers issus de l'exploitation des DADS, et donc à exclure de fait les salariés de la fonction publique d'État : d'une part, l'absence de base de sondage comparable aux DADS pour les salariés de la fonction publique d'État au moment de l'échantillonnage ; d'autre part, le surcoût de collecte, qui aurait excédé les montants prévus dans le financement ANR Corpus 2007. La mise en place du panel FGE dans la fonction publique d'État en 2009, ainsi que le financement par la Direction générale de l'administration et de la fonction publique (DGAFP) et le Centre pour la recherche économique et ses applications (CEPREMAP) ont permis l'extension en 2010-2011 de l'enquête menée en 2008-2009.

Cette seconde « vague » de l'enquête SalSa reprend ainsi la plupart des caractéristiques de l'enquête menée en 2008-2009, en particulier son échantillonnage (environ 5 000 salariés échantillonnés pour 3 000 répondants) et son questionnaire (à quelques ajustements près). Néanmoins, plusieurs éléments rendent le rapprochement direct de ces deux enquêtes délicat (Tableau 1) :

- **Période de collecte** : les deux enquêtes sont espacées de plus de deux ans, dans un contexte d'important ralentissement économique. Il semble difficile *a priori* d'écarter l'hypothèse que l'évolution du contexte et des perspectives économiques influence la perception et le jugement porté sur les salaires.

TABLEAU 1 – Synthèse des différences entre les enquêtes SalSa 2009 et 2011

Caractéristique	Enquête SalSa 2009	Enquête SalSa 2011
Période de collecte	Novembre 2008 – Janvier 2009	Janvier 2011 – Février 2011
Champ géographique	Dix régions + Essonne	France entière
Champ professionnel	Entreprises privées et publiques, fonctions publiques territoriale et hospitalière	Fonctions publiques d'État, territoriale et hospitalière
Base de sondage	Panel DADS 2006	Panels DADS et FGE 2008
Réseau d'enquêteur	INSEE	SOFRES
Passation	Face-à-face ou téléphone	Téléphone

- **Champ couvert** : les deux enquêtes portent sur des champs géographiques et professionnels différents. D'une part, l'enquête SalSa 2009 porte sur les salariés du privé et des fonctions publiques hospitalières et territoriales d'un sous-ensemble du territoire (dix régions et un département) ; d'autre part, l'enquête SalSa 2011 porte sur les salariés des trois fonctions publiques (y compris fonction publique d'État) et sur l'ensemble du territoire de

France métropolitaine. Dans ce cadre, il est particulièrement difficile de reconstituer une « population » clairement identifiées dont l'échantillon constitué par les individus interrogés au cours des deux enquêtes pourrait être considéré comme représentatif.

- **Modalités de collecte** : la collecte a été opérée dans un cas soit en face-à-face soit par téléphone par le réseau des enquêteurs de l'INSEE, et dans un autre par téléphone par les enquêteurs de la SOFRES. L'effet de l'évolution des modalités de collecte, plus difficile à appréhender, ne peut lui non plus être négligé *a priori*.

L'ensemble de ces éléments doit conduire l'utilisateur des données à être particulièrement prudent dans l'utilisation conjointe des fichiers des deux enquêtes : la variation simultanée de tous ces paramètres rend l'interprétation d'éventuelles différences entre les enquêtes particulièrement délicate.

## Modalités de rapprochement des enquêtes SalSa 2009 et 2011

Afin de permettre une exploitation commune des données de ces deux enquêtes, plusieurs opérations ont été néanmoins réalisées :

- d'une part, l'ensemble des variables et de leurs modalités ont été normalisées et recodées de façon homogène : relativement technique, cette opération a principalement consisté à uniformiser les noms et codage des variables et de leurs modalités entre les deux réseaux de collecte, notamment en permutant certains codages inversés dans le questionnaire SOFRES ;
- d'autre part, une pondération commune aux deux enquêtes est proposée.

La construction et l'utilisation d'une pondération commune aux deux enquêtes présupposent que les données collectées soient jugées comparables. Elles exigent donc l'acceptation d'hypothèses particulièrement fortes sur les données, dans la mesure où un grand nombre de biais peuvent contribuer à les rendre dissemblables (*cf. supra*). Il en va de la responsabilité de l'utilisateur des données de bien mesurer l'importance de ces hypothèses, de leur conséquences sur son analyse et des éventuelles limites à y apporter.

D'un point de vue technique, la construction d'une pondération commune aux enquêtes SalSa 2009 et 2011 ne peut reposer sur un échantillonnage probabiliste : par construction, les populations de référence des deux enquêtes diffèrent fortement (ne serait-ce que parce que l'une date de 2006 et l'autre de 2008), ce qui rend impossible la reconstitution de poids de sondage commun aux deux enquêtes qui soient susceptibles d'être redressés. En revanche, dès lors que l'on considère que pour être représentatif de certaines variables un échantillon doit présenter une distribution pondérée conforme à celle d'une population de référence, il est possible de construire une pondération qui ait du sens d'un point de vue statistique par calage sur marges. Comme souvent en pareil cas, la population de référence est issue de l'exploitation de l'Enquête Emploi (pour l'année 2010 ici), qui du fait de son échantillon important et de sa méthodologie rigoureuse permet d'obtenir une bonne estimation de la distribution des variables d'intérêt pour l'ensemble du champ couvert.

D'autre part, il convient de circonscrire précisément la population dont on souhaite rendre l'échantillon représentatif : dans la mesure où le champ géographique couvert par les deux enquêtes n'est pas le même, il semble déraisonnable de chercher à considérer comme représentatif de toute la France métropolitaine un échantillon qui pour moitié n'a été tiré que dans dix régions et un département. De ce fait, l'échantillon effectivement repondéré est restreint aux individus appartenant au champ géographique de l'enquête SalSa 2009, ce qui conduit à exclure environ la moitié des individus interrogés lors de l'enquête SalSa 2011. Des considérations techniques<sup>1</sup> conduisent à restreindre encore l'échantillon repondéré en excluant les individus résidant en Ile-de-France.

---

1. D'une part l'indisponibilité des données au niveau départemental dans les fichiers de l'Enquête Emploi en libre accès sur le site internet de l'INSEE et d'autre part l'incertitude sur le codage du département de résidence dans SalSa 2011 (*cf.* la note d'accompagnement de l'enquête).

L'ensemble de ces éléments conduit à repondérer un échantillon de 4 221 individus, soit 69,0 % des 6 117 personnes interrogées (2 973 issus de SalSa 2009 et 1 248 issus de SalSa 2011). Concrètement, il s'agit des salariés nés entre 1955 et 1992 et résidant dans les dix régions du champ de l'enquête SalSa 2009 hors Essonne<sup>2</sup>. Cet échantillon correspond à une population de 9 640 000 individus dans l'Enquête Emploi. Les variables de calage sont analogues à celles utilisées pour le redressement de l'enquête SalSa 2009 (*cf.* la note d'accompagnement de l'enquête) : le sexe, l'âge, le domaine d'emploi et le salaire. Sans surprise, le principal écart entre les distributions des variables de calage dans l'échantillon SalSa et dans l'Enquête Emploi concerne le domaine d'emploi (Tableau 2) : la fonction publique est clairement sur-représentée en raison de la composition de l'échantillon de l'enquête SalSa 2011.

TABLEAU 2 – Statistiques univariées des variables de calage

Variabiles	Modalité	Répondants SalSa	Enquête Emploi
Sexe	Homme	46,27 %	50,63 %
	Femme	53,73 %	49,37 %
Année de naissance	1976-1992	22,62 %	33,85 %
	1960-1975	49,54 %	43,69 %
	1944-1959	27,84 %	22,46 %
Domaine d'emploi	Entreprises privées ou publiques	54,42 %	76,56 %
	Fonction publique d'État	16,80 %	11,41 %
	Fonction publique territoriale	17,93 %	7,90 %
	Fonction publique hospitalière	10,85 %	4,13 %
Salaire	Pas de salaire déclaré	2,96 %	2,33 %
	Moins de 1 000 € par mois	10,26 %	15,67 %
	De 1 000 à 1 499 € par mois	34,61 %	34,54 %
	De 1 500 à 2 499 € par mois	39,23 %	36,13 %
	Plus de 2 500 € par mois	12,94 %	11,32 %

En pratique, le calcul de la pondération commune aux enquêtes SalSa 2009 et 2011 est opéré par déformation d'un vecteur de poids uniforme (de valeur initiale  $9\,640\,000 / 4\,221 = 2\,283,82$ ) à l'aide de la macro SAS CALMAR accessible depuis le site internet de l'INSEE. Les quatre méthodes proposées (linéaire, du *Raking ratio*, logistique et linéaire tronquée) sont utilisées : la méthode logistique est préférée, dans la mesure où elle conduit à un vecteur de poids présentant une faible dispersion (Tableau 3).

TABLEAU 3 – Caractéristiques du vecteur de poids POND0911

Nombre d'individus	4 221	Maximum	4 110,74
Somme des poids	9 640 000	D9	4 078,00
Moyenne	2 283,82	Q3	3 513,45
Ecart-type	1 264,39	Médiane	2 255,91
Rapport Q3/Q1	3,77	Q1	931,156
Rapport D9/D1	5,07	D1	804,71
Rapport Max/Min	5,14	Minimum	800,10

2. Picardie, Centre, Basse-Normandie, Lorraine, Alsace, Pays de la Loire, Midi-Pyrénées, Rhône-Alpes, Auvergne et Languedoc-Roussillon.