

Direction des Statistiques Démographiques et Sociales

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

Document de travail



Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des Documents de Travail
de la
DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES

N°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

(DIVISION SALAIRES ET REVENUS D'ACTIVITÉ)

Document de travail

avril 2019

Remerciements : Ce document a bénéficié des nombreux conseils et remarques d'Elise Coudin, de Sophie Maillard ainsi que de Guilhem Théron et Fabien Guggemos. Nous tenons à leur exprimer toute notre reconnaissance. Toutefois, nous restons seuls responsables des erreurs et imprécisions restantes.

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 authors' views.

Les modèles de contrôle des effets de structure : comparaisons et application à une analyse de la disparité départementale des revenus non salariés moyens

Résumé

Ce document de travail présente la mise en œuvre de différentes méthodes de contrôle des effets de structure, dans le cadre d'une analyse des disparités de revenu d'activité des non-salariés entre les départements dont les principaux résultats ont été publiés dans l'Insee Première n° 1672 « Le revenu d'activité des non-salariés : plus élevé en moyenne dans les départements du nord que dans ceux du sud ».

Une cartographie par département des revenus d'activité des non-salariés fait apparaître une forte corrélation spatiale de ceux-ci : les revenus d'activité moyens sont relativement élevés dans le tiers nord de la France et faibles dans le tiers sud. Ces écarts de revenus d'activité entre les départements peuvent découler en partie des différences de structures par secteur d'activité et catégorie juridique du non-salarié. De même, les autres caractéristiques individuelles des non-salariés et l'environnement économique dans lequel s'exerce l'activité participent aux écarts de revenus entre départements. Afin de mettre en évidence ces potentiels effets de structure, plusieurs types de modélisation sont possibles. Certains sont usuellement utilisés en économie géographique – modèles structurels-résiduels descriptifs, économétriques et géographiques –, d'autres en économie du travail – modèle de Oaxaca-Blinder. Dans ce document, ces quatre modèles sont appliqués aux données relatives aux non-salariés et leurs apports et inconvénients sont présentés. La mise en œuvre d'un modèle structurel-résiduel économétrique et d'un modèle de décomposition de Oaxaca-Blinder donnent des résultats identiques. En revanche, les parties expliquées obtenues à partir de ces deux méthodes diffèrent légèrement de celles issues d'un modèle structurel-géographique. Enfin, quelques pistes pour améliorer la part expliquée de ces modèles sont exposées.

The econometric tools to control composition effects : comparisons and implementation for an analysis of disparities between French departments in average earned incomes of self-employed persons

Abstract

This document presents the implementation of different econometric tools to control composition effects, for an analysis of disparities between departments in earned incomes of self-employed persons. The main results of this analysis were published in Insee Première n° 1672 « Le revenu d'activité des non-salariés : plus élevé en moyenne dans les départements du nord que dans ceux du sud » (« Earned income in non-employees : higher on average in France's northern departments than in the southern »).

Mapping earned incomes of self-employed persons by departments shows a strong spatial correlation : the average earned incomes are relatively high in the northern third of France and low in the southern third. These earned income disparities between departments can partly result from structural differences by activity sector and legal category of self-employed. Similarly, other individual characteristics of self-employed as well as the economic environment of the activity contribute to earned income disparities between departments. In order to highlight these potential composition effects, several modeling tools are possible. Some are commonly used in economic geography – descriptive, econometric and geographic structural-residual models –, other are commonly used in labour economics – Oaxaca-Blinder model. In this document, these four models are applied to the self-employed database and their pros and cons are presented. The implementation of an econometric structural-residual model and an Oaxaca-Blinder decomposition model give identical results. On the other hand, the explained parts obtained from these two methods differ slightly from those obtained from a geographic structural model. Finally, some ways to improve the explained part of these models are presented.

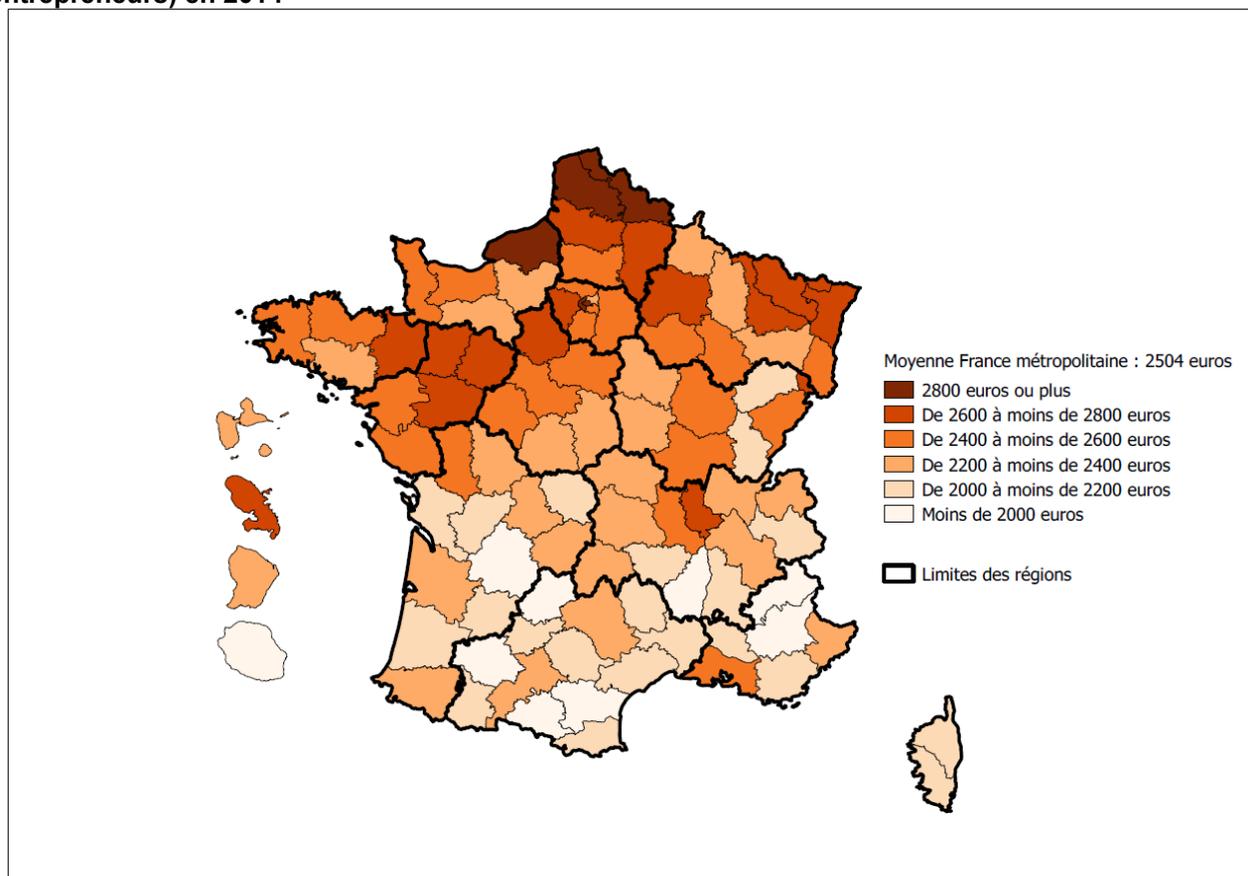
Table des matières

Introduction.....	3
1. Les données sur les non-salariés.....	4
1.1. Source, champ et concepts.....	4
1.1.1. Source des données.....	4
1.1.2. Champ de l'analyse.....	4
1.1.3. Définition du revenu d'activité.....	4
1.2. Méthodologie retenue.....	5
1.3. Les variables de structure retenues.....	5
1.3.1. Les grandes catégories de non-salariés.....	5
1.3.2. Secteur d'activité et catégorie juridique de non-salarié (croisement et regroupement des modalités).....	7
1.3.3. Les autres caractéristiques du non-salarié et de son activité.....	9
1.3.4. Une prise en compte partielle du volume de travail.....	10
1.3.5. La prise en compte de l'environnement économique.....	13
2. La variable d'intérêt : revenu moyen versus revenu médian ?.....	15
2.1. Impact de l'indicateur retenu sur les disparités de revenus d'activité, sur le champ de l'ensemble des non-salariés (y compris auto-entrepreneurs).....	15
2.2. Impact de l'indicateur retenu sur les disparités de revenus d'activité, sur le champ des non-salariés classiques à revenu non nul.....	18
2.3. Plusieurs facteurs expliquent la sensibilité du classement des départements à l'indicateur de revenu retenu.....	20
3. Contrôle des effets de structure : les modèles utilisés en économie géographique.....	24
3.1. Le Modèle Structurel-Résiduel descriptif.....	24
3.2. Le Modèle Structurel-Résiduel économétrique.....	25
3.3. Le Modèle Structurel-Géographique économétrique.....	26
4. Contrôle des effets de structure : les outils utilisés en économie du travail.....	27
4.1. Présentation du modèle de décomposition de Oaxaca-Blinder.....	27
4.2. Identification du modèle.....	29
4.3. Comparaison avec un modèle structurel-résiduel économétrique.....	31
4.3.1. Dans le cas où on utilise uniquement des variables structurelles catégorielles.....	31
4.3.2. Intérêts comparés de la décomposition de Oaxaca-Blinder et du modèle structurel-résiduel.....	32
5. Tableau de synthèse des différentes méthodes de contrôle des effets de structure.....	33
6. Application des différents modèles aux écarts de revenus d'activité non salariée entre départements.....	34
7. La transformation logarithmique des revenus.....	38
7.1. Impact du passage au logarithme sur la corrélation spatiale de la variable d'intérêt.....	38
7.2. Impact du passage au logarithme sur l'estimation des effets de structure.....	40
8. Les améliorations et extensions possibles.....	42
8.1. Les limites de l'indice de saisonnalité utilisé.....	42
8.2. La prise en compte de la pluriactivité.....	43
8.3. Effet de la concurrence ou effet de sélection.....	43
8.3.1. Essai de prise en compte de l'effet concurrence.....	45
8.3.2. Mesurer l'effet de sélection.....	48
Conclusion.....	48
Bibliographie.....	49
Annexes.....	50
Annexe 1 : résultats du modèle structurel-résiduel relatif à l'impact de la structure par grandes catégories de non-salariés sur les écarts de revenus d'activité entre départements.....	51
Annexe 2 : Indicateurs de revenu par département, pour l'ensemble des non-salariés.....	52
Annexe 3 : Principaux modèles de décomposition de Oaxaca-Blinder mis en œuvre.....	54
Annexe 4 : codes des régions.....	55
Annexe 5 : codes des départements.....	56

Introduction

Une cartographie par département des revenus d'activité des non-salariés fait apparaître une forte corrélation spatiale de ceux-ci : les revenus d'activité sont relativement plus élevés dans les départements du tiers nord de la France et faibles dans ceux du tiers sud (figure 1)¹. Une partie des disparités peuvent découler de différences de structure par secteur d'activité et type de non-salarié, mais aussi des autres caractéristiques individuelles des non-salariés. Enfin, l'environnement dans lequel s'exerce l'activité non salariée peut expliquer une partie des écarts de revenus résiduels.

Figure 1 : revenu d'activité mensuel moyen de l'ensemble des non-salariés (y compris auto-entrepreneurs) en 2014



Lecture : dans le département du Pas-de-Calais, le revenu d'activité mensuel moyen des non-salariés est supérieur à 2 800 euros en 2014.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014, hors agriculture et hors taxés d'office.

Source : Insee, base non-salariés 2014.

L'objectif de ce document de travail est triple : il décrit les méthodologies pouvant être mises en œuvre pour analyser le rôle des effets de structure sur les disparités de revenus d'activité entre départements ; il en mesure les avantages et inconvénients ; il explore des pistes d'amélioration à l'analyse.

Une fois le champ et les concepts de l'étude précisés, la première partie porte sur les variables de structure utilisées. L'analyse des écarts de revenus d'activité repose sur le contrôle des effets de structure : nous définirons l'indicateur synthétique retenu pour mesurer l'impact des variables structurelles utilisées. Dans une seconde partie, nous comparerons deux indicateurs d'intérêt : le revenu d'activité moyen et le revenu d'activité médian.

¹ Insee Première n° 1672, « le revenu d'activité des non-salariés : plus élevé en moyenne dans les départements du nord que dans ceux du sud », novembre 2017.

Différents outils de contrôles usuellement mis en œuvre en économie géographique seront présentés de façon théorique dans une troisième partie.

Dans une quatrième partie, nous décrivons le modèle de Oaxaca-Blinder, outil utilisé couramment en économie du travail.

Nous présenterons une synthèse comparative de ces modèles dans une cinquième partie, puis une application des méthodes les mieux adaptées à notre problématique d'analyse des disparités de revenus entre départements. Pour ce faire, on se limite à la prise en compte de deux groupes de variables de structure : le croisement « secteur d'activité X type de non-salarié » et l'ensemble des caractéristiques individuelles du non-salarié et de son activité.

Une septième partie décrit l'impact du passage du revenu au logarithme du revenu sur l'analyse des effets de structure, transformation nécessaire pour la validité des tests de significativité.

Enfin, nous terminons par la présentation de quelques améliorations et extensions possibles à cette analyse.

1. Les données sur les non-salariés

1.1. Source, champ et concepts

1.1.1. Source des données

La base non-salariés est issue de deux sources administratives, gérées par l'Agence centrale des organismes de sécurité sociale (Acoss) et par la Caisse centrale de la mutualité sociale agricole (CCMSA). Ces deux organismes collectent les cotisations sociales et la CSG-CRDS assises sur les rémunérations des non-salariés.

Les non-salariés regroupent l'ensemble des personnes affiliées à un régime de protection sociale des travailleurs non salariés. Sont concernés les auto-entrepreneurs (AE) et les non-salariés classiques. Ces derniers sont pour l'essentiel des entrepreneurs individuels (dont des professions libérales) ou des gérants majoritaires de sociétés à responsabilité limitée.

1.1.2. Champ de l'analyse

Dans un souci de cohérence avec l'Insee Première annuel sur les « revenus d'activité des non-salariés », l'analyse est restreinte aux informations provenant de l'Acoss, hors secteur agricole. Elle se limite aux non-salariés en activité au 31 décembre 2014. Tous les non-salariés en activité à cette date sont pris en compte, y compris ceux qui exercent une activité salariée (à titre principal ou non).

Le champ géographique est la France entière hors Mayotte. Pour certaines parties de l'analyse, le champ est réduit à la France métropolitaine pour des raisons de disponibilité des données.

1.1.3. Définition du revenu d'activité

Le revenu d'activité est la variable d'intérêt de l'analyse. Il correspond à la rémunération issue de l'activité non salariée, déduction faite des cotisations sociales payées dans l'année mais pas des contributions sociales (CSG, CRDS). Il est calculé à partir du revenu professionnel imposable, auquel sont réintégrés certains allègements fiscaux et cotisations sociales facultatives.

Il correspond :

- pour les auto-entrepreneurs, au chiffre d'affaires après abattement pour frais professionnels (dépendant du type d'activité exercée) ;
- pour les entrepreneurs individuels, au bénéfice net retiré de l'activité professionnelle ;
- pour les gérants majoritaires, à la rémunération, fixée par décision collective des associés, incluant une partie des dividendes perçus et, pour les gérants d'entreprises assujetties à l'impôt sur les sociétés, la déduction fiscale forfaitaire de 10 % pour frais professionnels.

En cas d'exercice déficitaire, le revenu d'activité est considéré comme nul. Il est calculé pour les non-salariés en activité au 31 décembre 2014, et rapporté à la durée d'affiliation dans l'année. Ce revenu peut correspondre à une activité à temps complet ou à temps partiel. Les non-salariés ne déclarant pas leur revenu font l'objet d'une taxation d'office par l'Acoss pour le recouvrement des cotisations. Ils ne sont pas pris en compte dans l'analyse des revenus d'activité.

1.2. Méthodologie retenue

Afin d'expliquer la disparité des revenus d'activité des non-salariés entre les départements, on essaie de contrôler les effets de structure et de mettre en évidence les écarts restant inexpliqués par ceux-ci. Sur une variable d'intérêt quantitative, les outils les plus couramment utilisés pour ce faire sont les modèles structurels-résiduels (descriptif et économétrique), le modèle structurel-géographique et la décomposition de Oaxaca-Blinder.

Quand on régresse le revenu sur les variables de structure, le R^2 ajusté du modèle est un indice de la pertinence globale de celles-ci, prises dans leur ensemble. De même, pour mesurer la pertinence des variables de structure utilisées dans un modèle de contrôle des effets de structure, on peut calculer un indicateur de gain relatif de variance des écarts de revenus entre départements. Il mesure la différence relative entre la variance des écarts de revenus entre départements et la variance résiduelle une fois les effets de structure contrôlés :

$$[(\text{variance des écarts restant inexpliqués} / \text{variance des écarts initiaux}) - 1] * 100$$

où les écarts restant inexpliqués sont issus d'un modèle structurel-résiduel, d'un modèle structurel-géographique ou d'un modèle de décomposition de Oaxaca-Blinder (cf. 3. Contrôle des effets de structure : les modèles utilisés en économie géographique et 4. Contrôle des effets de structure : les outils utilisés en économie du travail).

La possibilité de réaliser des tests de significativité des effets estimés implique l'utilisation de variables de distribution normale. C'est pourquoi l'analyse ne porte pas directement sur les écarts de revenu moyen, mais sur les écarts entre moyennes du logarithme du revenu. On fait alors l'hypothèse que les deux écarts sont fortement corrélés, hypothèse que l'on testera dans le cadre de la mise en œuvre du modèle de Oaxaca-Blinder (cf. 7. La transformation logarithmique des revenus).

1.3. Les variables de structure retenues

1.3.1. Les grandes catégories de non-salariés

Du fait des conditions d'accès au statut d'auto-entrepreneur (accessible si le chiffre d'affaires reste au-dessous de seuils fixés), leur revenu d'activité moyen est sensiblement plus faible que celui des autres non-salariés (i.e. les non-salariés « classiques », qui comprennent les entrepreneurs individuels et les gérants majoritaires) : 410 euros mensuels contre 3 260 euros mensuels en 2014.

Par ailleurs, environ 10 % des non-salariés classiques ont un revenu nul dans les fichiers de l'Acoss. Ce sont :

- des entrepreneurs individuels ayant réalisé un exercice déficitaire (dans ce cas l'Acoss fixe par convention leur revenu à zéro) ou dont le bénéfice a été nul ;
- des gérants majoritaires de sociétés qui ont choisi de ne rémunérer leur activité que par le biais des dividendes (pour un montant inférieur à 10 % du capital social de l'entreprise) ;
- des gérants majoritaires qui perçoivent une rémunération par le biais d'une autre société (liée à la première) dont ils sont salariés.

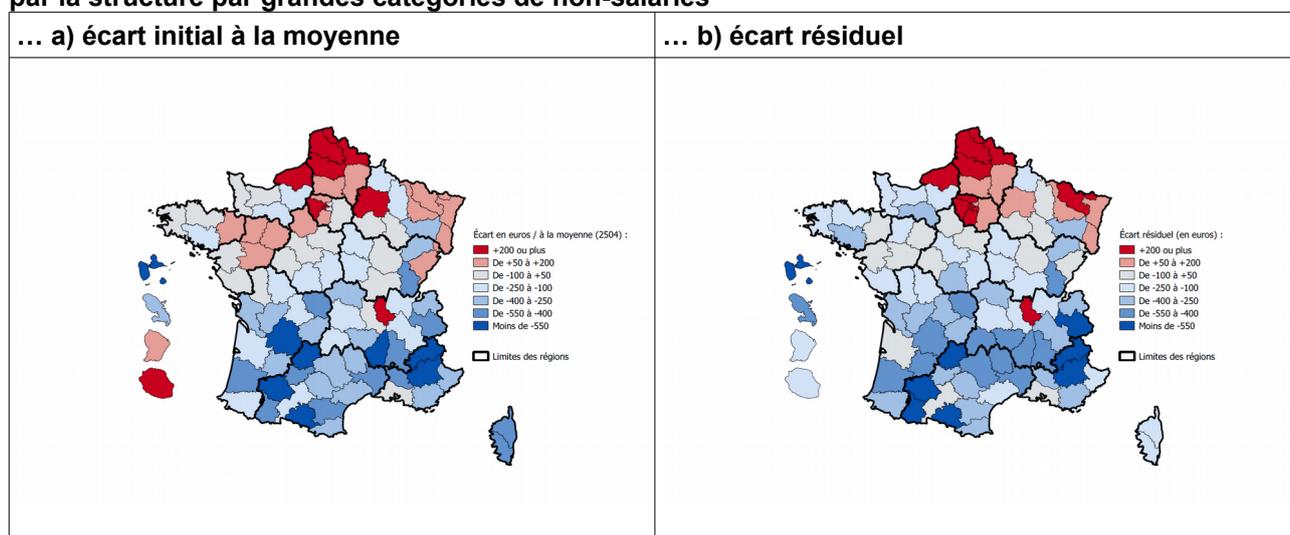
Les parts d'auto-entrepreneurs et de non-salariés classiques à revenu nul ont potentiellement un impact sur le niveau du revenu d'activité moyen observé dans un département.

L'effet de la structure « auto-entrepreneurs / non-salariés classiques à revenu nul / non-salariés classiques à revenu non nul » sur la disparité des revenus moyens départementaux est analysé à l'aide d'un modèle structurel-résiduel (cf. 3.2. Le Modèle Structurel-Résiduel économétrique). Pour ce modèle, la variable d'intérêt est la moyenne des revenus et non la moyenne des logarithmes de revenus, du fait de l'existence de revenus nuls (les tests de significativité ne sont donc pas valables puisque la variable d'intérêt n'est pas supposée normale). La France est la zone de référence et le département est la zone comparée.

Une fois contrôlée la structure par grandes catégories de non-salariés, la corrélation spatiale reste élevée : les revenus d'activités moyens sont toujours élevés au nord et faibles au sud (figure 2). L'indicateur de gain relatif de variance des écarts de revenus entre départements est légèrement supérieur à 100 % : neutraliser les effets de la structure par grandes catégories de non-salariés ne permet pas de réduire la variance des écarts (cf. Annexe 1).

On se concentre donc sur le champ des non-salariés classiques à revenu non nul. Une transformation logarithmique de la variable de revenu peut être alors réalisée.

Figure 2 : Écart entre le revenu d'activité mensuel moyen des départements et la moyenne française, pour l'ensemble des non-salariés (y compris auto-entrepreneurs) – écart initial et écart non expliqué par la structure par grandes catégories de non-salariés



Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

1.3.2. Secteur d'activité et catégorie juridique de non-salarié (croisement et regroupement des modalités)

Les non-salariés classiques (hors auto-entrepreneurs) peuvent être répartis en deux groupes, celui des gérants majoritaires de sociétés et celui des entrepreneurs individuels. Parmi ces derniers, on peut distinguer les professions libérales des autres entrepreneurs individuels.

On retient donc trois catégories juridiques de non-salariés :

- les entrepreneurs individuels hors professions libérales (notation EI) ;
- les professions libérales (notation PL) ;
- les gérants majoritaires de société (notation GM).

Remarque : la définition des professions libérales retenue dans l'étude ne repose que sur la catégorie juridique (CJ = 1500) et ne fait pas appel au secteur d'activité. Dans d'autres travaux, la définition retient une liste d'activités et de croisements « secteur d'activité X catégorie juridique ».

Le secteur d'activité des non-salariés est quant à lui agrégé en 8 modalités :

- 02 = Industrie (hors artisanat commercial)
- 03 = Construction
- 04 = Commerce et artisanat commercial
- 05 = Transports
- 07 = Services aux entreprises et services mixtes
- 13 = Services aux particuliers
- 20 = Santé humaine et action sociale
- 90 = Secteur indéterminé

Le secteur d'activité et la catégorie juridique du non-salarié sont les facteurs les plus « explicatifs » du niveau de revenu d'un non-salarié classique. Ces facteurs ne sont pas indépendants et leurs effets ne sont pas simplement additifs : par exemple, alors que les professions libérales ont une rémunération en moyenne plus élevée que les autres catégories de non-salariés dans les services aux entreprises, c'est l'inverse dans les services aux particuliers. Il est donc nécessaire de croiser les deux facteurs.

La variable obtenue en croisant le secteur d'activité et la catégorie du non-salarié compte 24 modalités. Certaines modalités de cette variable ont cependant des effectifs très faibles (voire nuls pour quelques départements). On ne pourra alors pas estimer les coefficients correspondants dans la décomposition de Oaxaca-Blinder. On procède donc à quelques regroupements :

- La part de professions libérales est faible (< 1,5 %) dans les secteurs 02, 03, 04, 05 et 90. On regroupe ces modalités en une seule (« XX_PL » = professions libérales hors services aux entreprises, services aux particuliers et santé). Pour les secteurs des services et de la santé, la part de professions libérales est suffisamment importante dans tous les départements.
- La part des entrepreneurs individuels est faible (1 %) dans le secteur de la santé. Le nombre d'entrepreneurs individuels dans ce secteur est inférieur à 10 non-salariés dans 16 départements, à 20 dans 48 départements. De même, la part des gérants-majoritaires est assez faible (4 %) dans le secteur de la santé et leur nombre est inférieur à 20 dans 7 départements. On regroupe ces deux modalités en une seule (« 20_YY » = non-salariés hors professions libérales du secteur de la santé).
- La part des entrepreneurs individuels dans les secteurs indéterminés n'est pas négligeable (8,5 %), mais les effectifs sont inférieurs à 20 dans 23 départements. On regroupe les entrepreneurs individuels et les gérants majoritaires ayant un secteur d'activité indéterminé en une seule modalité (« 90_YY » = non-salariés hors professions libérales des secteurs indéterminés). Les professions libérales dont le secteur d'activité est indéterminé ont en effet un revenu d'activité beaucoup plus élevé que les autres non-salariés ayant un secteur indéterminé (le revenu mensuel moyen est de 4630 euros pour les professions libérales, de 1820 euros pour les entrepreneurs individuels et de 2515 euros pour les gérants-majoritaires).

Après regroupements, aucune modalité n'a un effectif nul au niveau départemental. On obtient un croisement secteur d'activité X catégorie du non-salarié en 18 modalités :

Modalités	Libellé	Effectifs	Revenu d'activité mensuel moyen (en €)	Minimum des effectifs départementaux
02_EI	Industrie – Entrepreneurs individuels (hors professions libérales)	27 770	1 970	50
02_GM	Industrie – Gérants majoritaires	33 030	3 790	29
03_EI	Construction – Entrepreneurs individuels (hors professions libérales)	113 470	2 190	160
03_GM	Construction – Gérants majoritaires	108 140	2 980	87
04_EI	Commerce – Entrepreneurs individuels (hors professions libérales)	142 680	2 310	259
04_GM	Commerce – Gérants majoritaires	160 320	3 220	106
05_EI	Transports – Entrepreneurs individuels (hors professions libérales)	32 770	1 630	33
05_GM	Transports – Gérants majoritaires	15 740	2 890	14
07_EI	Services aux entreprises – Entrepreneurs individuels (hors professions libérales)	45 890	4 450	45
07_GM	Services aux entreprises – Gérants majoritaires	129 820	4 610	57
13_EI	Services aux particuliers – Entrepreneurs individuels (hors professions libérales)	121 150	1 530	254
13_GM	Services aux particuliers – Gérants majoritaires	109 230	2 110	58
20_YY	Santé – <i>hors professions libérales</i>	20 640	8 320	22
90_YY	Secteur indéterminé – <i>hors professions libérales</i>	95 010	3 210	94
07_PL	Services aux entreprises – professions libérales	120 500	5 450	96
13_PL	Services aux particuliers – professions libérales	40 440	1 310	33
20_PL	Santé – professions libérales	351 200	5 450	352
XX_PL	Secteurs hors services et santé – professions libérales	2 850	3 620	2
	Ensemble	1 670 660	3 610	2 470

Note : Les effectifs sont calculés en tenant compte de la durée d'affiliation.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors auto-entrepreneurs et hors taxés d'office).

Source : Insee, base non-salariés 2014.

Le modèle de régression du logarithme de revenu sur le croisement secteur d'activité X catégorie juridique du non-salarié donne un R^2 de 0,151 (cf. Annexe 3, modèle M1).

1.3.3. Les autres caractéristiques du non-salarié et de son activité

Outre le revenu d'activité, la localisation de l'activité, le secteur d'activité détaillé et la catégorie juridique détaillée, les bases non-salariés fournissent de l'information sur les caractéristiques du non-salarié (sexe, âge...) et de son activité (ancienneté, emploi de salariés, occupation simultanée de poste(s) salarié(s), le cas échéant montant des salaires perçus...).

Les caractéristiques retenues pour l'analyse sont :

- **Le sexe**
Divers éléments expliquent l'effet associé au facteur sexe, dont les différences de volume de travail existant entre femmes et hommes.
- **L'âge**
La variable d'âge a été discrétisée en 6 modalités :
 - « moins de 30 ans »
 - « de 30 à moins de 40 ans »
 - « de 40 à moins de 50 ans » (modalité de référence)
 - « de 50 à moins de 60 ans »
 - « de 60 à moins de 70 ans »
 - « 70 ans et plus ».
- **L'ancienneté de l'activité non salariée**
L'ancienneté est calculée à partir de la date d'assujettissement associée au compte du non-salarié en cours durant l'année de validité de la base non-salariés (2014). Elle est discrétisée en sept modalités² :
 - « moins de 1 an »
 - « de 1 à moins de 2 ans »
 - « de 2 à moins de 3 ans »
 - « de 3 à moins de 6 ans »
 - « de 6 à moins de 11 ans » (modalité de référence)
 - « de 11 à moins de 21 ans »
 - « 21 ans ou plus ».
- **Le fait d'être employeur ou non**
Deux variables sur l'emploi éventuel de salariés par le non-salarié ont été définies. La première est dichotomique (employeur ou non.) La seconde précise le nombre de salariés employés. Les modalités de la variable détaillée sont alors :
 - « non employeur en fin d'année » (modalité de référence)
 - « de 1 à 4 salariés »
 - « de 5 à 9 salariés »
 - « de 10 à 19 salariés »
 - « 20 salariés ou plus »

Le gain relatif de variance des écarts de revenus est à peine meilleur quand la variable sur l'emploi de salariés est détaillée : avec un modèle de Oaxaca-Blinder prenant en compte le secteur d'activité X la catégorie du non-salarié et les autres caractéristiques du non-salarié et de son activité, le gain relatif de variance est de - 54 % avec la variable détaillée, contre - 52 % avec la variable dichotomique (cf. Annexe 3, modèle M2).

Dans la suite de l'analyse, on retient la variable dichotomique.

La prise en compte de l'ensemble de ces variables sur les caractéristiques du non-salarié et de son activité, en plus du secteur d'activité X la catégorie juridique du non-salarié, fait progresser le R² du modèle de régression du logarithme du revenu : il passe de 0,151 à 0,273 (cf. Annexe 3, modèles M1 et M2).

² La date d'assujettissement présente un défaut : elle a été réinitialisée au 01/01/2008 pour un certain nombre de non-salariés. Les résultats des 3 modalités d'ancienneté les plus importantes doivent donc être regardés avec prudence.

1.3.4. Une prise en compte partielle du volume de travail

Certaines informations influant fortement sur le niveau de revenu sont partiellement ou complètement manquantes dans les bases non-salariés : c'est notamment le cas du volume de travail effectué par le non-salarié.

Le volume de travail (en heures) effectué par le non-salarié peut s'écrire comme le produit de deux éléments : le nombre de jours travaillés dans l'année et le nombre d'heures effectué en moyenne journalière sur la période d'activité.

- Le nombre de jours travaillés dans l'année peut être inférieur à une année complète dans plusieurs cas de figures :
 - Si l'activité a débuté ou a cessé en cours d'année. Ce cas peut être directement pris en compte, car la date d'assujettissement et la date de cessation de l'activité sont connues dans les bases non-salariés.
 - Si l'activité non-salariée est non continue, voire saisonnière. Cet élément ne peut être directement pris en compte par le biais des bases non-salariés, les non-salariés restant assujettis durant leurs périodes d'inactivité hors-saison.

Certains non-salariés (et notamment ceux dont l'activité est saisonnière) complètent leur volume de travail annuel en tant que non-salarié par une activité salariée³ (on peut parler de pluriactivité séquentielle).

- La quantité de travail effectuée en moyenne journalière sur la période d'activité varie en fonction de divers facteurs : le sexe, la situation familiale, le niveau de la demande, des facteurs psychologiques tels que la motivation du non-salarié, un éventuel cumul avec une activité salariée qui obligerait le non-salarié à partager son temps de travail journalier disponible entre ses activités non salariée et salariée (on peut parler de pluriactivité simultanée).

Les bases non-salariés ne permettent pas de distinguer la pluriactivité séquentielle de la pluriactivité simultanée : elles permettent uniquement de repérer les non-salariés ayant une activité salariée dans l'année, et les dates de début et de fin de cette activité salariée.

Une partie seulement des éléments jouant sur le volume de travail peuvent être directement pris en compte dans le cadre de l'analyse des disparités de revenus entre départements :

- la durée d'assujettissement du compte du non-salarié dans l'année ;
- le sexe ;
- le cumul d'une activité salariée et non salariée au cours de l'année (pluriactivité).

De manière indirecte, on peut également prendre en compte **le caractère saisonnier ou non** de l'activité non salariée, en utilisant d'autres sources (les déclarations annuelles de données sociales – DADS – par exemple). On fait l'hypothèse (forte) que la saisonnalité d'une activité est similaire chez les non-salariés et les salariés. On approxime donc le caractère saisonnier ou non d'une activité non salariée, à l'aide du caractère saisonnier ou non de la même activité pour les salariés.

Pour ce faire, on utilise l'activité détaillée (sous-classes, classes ou groupes de la NAF), dans la mesure où les secteurs d'activité agrégés peuvent en effet regrouper des activités ayant une saisonnalité très différente.

Deux options se présentent pour appréhender la saisonnalité de l'activité : soit calculer un **indice « localisé »**, par exemple au niveau départemental, soit calculer un **indice national**.

La première solution permet de prendre en compte le fait que la saisonnalité d'une activité varie selon la localisation : par exemple une activité de service peu saisonnière au niveau national, comme la médecine, peut l'être dans des zones très touristiques. On peut supposer que l'indicateur localisé permet de mieux neutraliser, dans l'étape de contrôle des effets de structures liés aux caractéristiques du non-salarié et de son activité, l'impact du volume de travail des non-salariés sur les disparités de revenus entre départements. Cependant, la spécificité territoriale de la saisonnalité reflète une caractéristique de l'environnement économique dans lequel s'exerce l'activité non salariée : la touristicité de la zone (cf. 1.3.5. La prise en compte de l'environnement économique). Intégrer cette dimension dans les caractéristiques du non-salarié

³ À noter que l'activité salariée peut être l'activité principale de la personne : celle qui lui apporte le revenu le plus élevé sur l'année.

et de son activité reviendrait à minimiser une partie des effets des caractéristiques de l'environnement économique. De plus, la construction d'un indicateur localisé repose sur des croisements pouvant comporter peu d'observations et donc aboutir à un indicateur moins robuste.

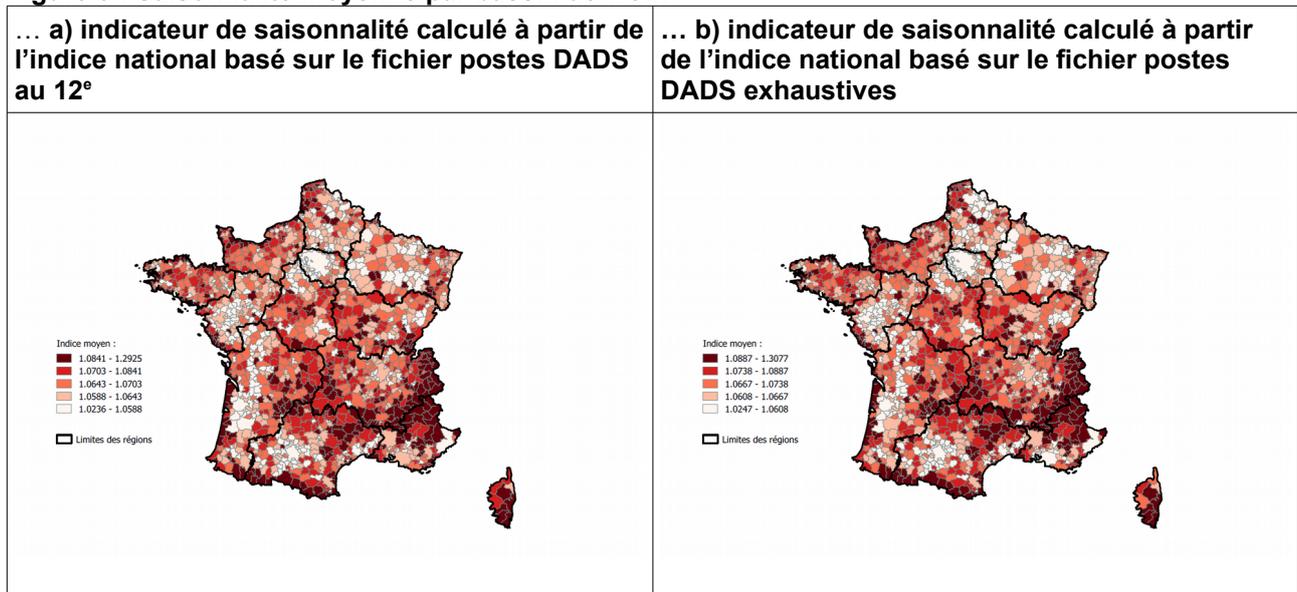
La seconde solution consiste à construire un indice au niveau national de la façon suivante :

- à partir des DADS, on établit pour chaque activité détaillée la courbe journalière d'emploi (nombre de postes actifs chaque jour de l'année). L'indice est égal au rapport du 99^e décile sur le 1^e quartile des effectifs journaliers.
Cette courbe journalière d'emploi est construite à partir des fichiers postes des DADS 2014 (fichier postes au 12^e). Elle ne se limite pas aux seuls postes non annexes, étant donné qu'on s'intéresse à la saisonnalité qui se traduit parfois par un recours à des contrats de très courte durée. Les postes annexes d'une durée de 10 jours ou plus (et pour lesquels le nombre d'heures travaillées est supérieur à 40) ont donc été intégrés au champ. Le rajout de postes annexes ne modifie que très peu la valeur de l'indice de saisonnalité obtenu : le coefficient de corrélation entre l'indicateur calculé avec ajout de postes annexes et celui calculé sans ajout est de 0,99, que l'on pondère ou non les activités par leurs effectifs salariés.
- On associe à chaque non-salarié l'indice de saisonnalité de son activité ainsi obtenu.
- L'indice de saisonnalité est ensuite discrétisé en quatre classes, déterminées selon la distribution⁴ de celui-ci :
 - les « activités de faible ou moyenne saisonnalité » dont l'indicateur est inférieur à la médiane ;
 - les « activités d'assez forte saisonnalité » dont l'indicateur est compris entre la médiane et le 3^e quartile ;
 - les « activités de forte saisonnalité » dont l'indicateur est compris entre le 3^e quartile et le 9^e décile ;
 - les « activités de très forte saisonnalité » dont l'indicateur est supérieur ou égal au 9^e décile.

Le calcul d'un indice de saisonnalité national permet l'utilisation, dans un souci de simplification, du fichier postes des DADS au 12^e. À noter que l'utilisation des fichiers postes régionaux exhaustifs donne des résultats très proches (figure 3) : le coefficient de corrélation entre l'indice calculé à partir des DADS au 12^e et celui calculé à partir des DADS exhaustives est de 0,99 (avec pondération des activités par les effectifs salariés).

⁴ Les quantiles sont déterminés sans pondérer les activités par leur poids dans la population des non-salariés.

Figure 3 : Saisonnalité moyenne par bassin de vie...



Note : l'indice de saisonnalité calculé au niveau national s'appuie à gauche sur le fichier postes des DADS au 12^e (validité 2014), à droite sur les fichiers postes exhaustifs régionaux des DADS 2014 (départementaux pour l'Île-de-France).

Pour chaque bassin de vie, l'indicateur de saisonnalité est la moyenne (pondérée par la durée d'assujettissement dans l'année) des indices de saisonnalité des non-salariés y exerçant leur activité.

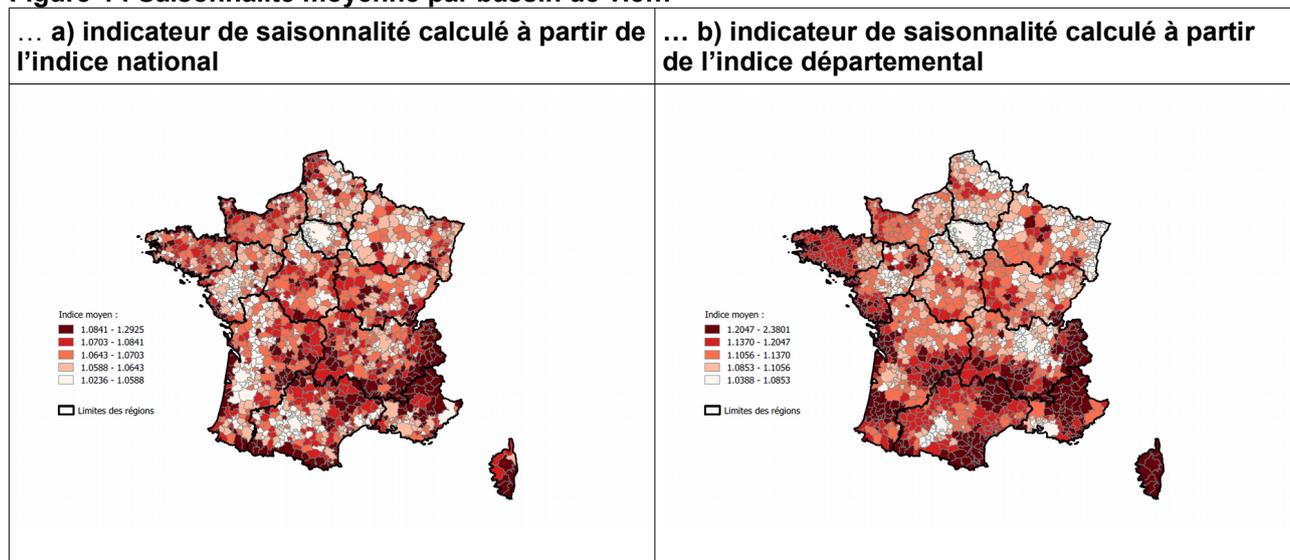
Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors auto-entrepreneurs, hors taxés d'office).

Source : Insee, base non-salariés 2014.

La prise en compte de la saisonnalité de l'activité (utilisation de l'indice de saisonnalité national), en plus des autres caractéristiques du non-salarié et de son activité fait légèrement progresser le R^2 ajusté du modèle de régression du logarithme du revenu, qui passe de 0,273 à 0,288 (cf. Annexe 3, modèles M2 et M3).

Construire un indice de saisonnalité localisé (calcul pour chaque croisement activité X département) donne des résultats sensiblement différents de celui construit au niveau national (figure 4). Cependant, il ne permet pas d'améliorer le R^2 ajusté de la régression du logarithme du revenu : en prenant en compte les croisements secteur d'activité X catégorie du non-salarié et les autres caractéristiques du non-salarié et de son activité, le R^2 ajusté est de 0,282 si on s'appuie sur l'indice de saisonnalité localisé, contre 0,288 si on s'appuie sur l'indice de saisonnalité national (cf. Annexe 3, modèle M2). On retient donc l'indice de saisonnalité national.

Figure 4 : Saisonnalité moyenne par bassin de vie...



Note : l'indice de saisonnalité calculé au niveau national s'appuie sur le fichier postes des DADS au 12^e (validité 2014). L'indice départemental s'appuie sur les fichiers postes exhaustifs régionaux des DADS 2014 (départementaux pour l'Île-de-France).

Pour chaque bassin de vie, l'indicateur de saisonnalité est la moyenne (pondérée par la durée d'assujettissement dans l'année) des indices de saisonnalité des non-salariés y exerçant leur activité.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors auto-entrepreneurs, hors taxés d'office).

Source : Insee, base non-salariés 2014.

1.3.5. La prise en compte de l'environnement économique

Au-delà des caractéristiques propres au non-salarié et à son activité, l'environnement économique dans lequel il exerce cette activité influe également sur le niveau de son revenu.

Pour mesurer l'apport de variables d'environnement économique au pouvoir « descriptif » d'un modèle économétrique, un premier test a été réalisé en ajoutant un effet fixe départemental aux variables relatives aux caractéristiques du non-salarié et de son activité. Avec cet ajout, le R^2 ajusté de la régression du logarithme de revenu sur les variables de structure passe de 0,288 à 0,293, modèle mis en œuvre pour la France métropolitaine (cf. Annexe 3, modèles M3 et M6). En ajoutant un effet fixe par bassin de vie plutôt que par département, le R^2 ajusté n'augmente pas davantage : il reste à 0,293 (cf. Annexe 3, modèle M7). En intégrant dans le modèle des variables décrivant l'environnement économique des bassins de vie ou des départements, le R^2 ajusté des modèles de régression mis en œuvre ne pourra donc pas dépasser cette valeur (0,2934). Le pouvoir explicatif de ces variables restera donc relativement limité.

Reflétant le dynamisme du marché local, les variables d'environnement économique testées décrivent la densité de population, le niveau de vie, la touristicité, le taux de chômage ou la pyramide des âges. Pour pouvoir être intégrées dans un modèle de décomposition de Oaxaca-Blinder sur les écarts de revenus entre les départements et la France métropolitaine prise dans son ensemble, le niveau géographique de ces variables doit être infra-départemental. Le niveau géographique retenu est le bassin de vie, excepté pour le taux de chômage : non disponible à ce niveau, on utilise les taux de chômage par zone d'emploi (pour les définitions du bassin de vie et de la zone d'emploi, voir <https://www.insee.fr/fr/metadonnees/definition/c2060> et <https://www.insee.fr/fr/metadonnees/definition/c1361>).

Lorsqu'elles sont quantitatives, les variables introduites dans le modèle sont soit discrétisées, soit standardisées. Dans le cas où elles sont discrétisées, un problème méthodologique peut émerger : pour que la décomposition de Oaxaca-Blinder soit réalisable, il faut que toutes les modalités de toutes les variables explicatives soient représentées au niveau du département. Or, ce n'est pas le cas par exemple pour la densité de population : dans les départements très urbanisés, les modalités correspondant à une faible densité de population ne sont généralement pas représentées. De plus, Paris est un département particulier où seule existe la modalité associée à la densité la plus forte. Dans ce cas extrême, le coefficient de cette modalité ne peut pas être estimé du fait de la présence d'une constante dans le modèle.

Ce problème peut toutefois être contourné, car la contrainte d'une représentation de toutes les modalités ne s'impose que pour le calcul de la décomposition par facteur de la partie inexpliquée des écarts de revenus (cf. 4. Contrôle des effets de structure : les outils utilisés en économie du travail). Si l'on se contente de l'estimation de la partie inexpliquée totale, alors il est possible de la calculer par différence : partie inexpliquée = écart – partie expliquée. C'est la solution qui a donc été retenue pour mettre en œuvre les modèles de Oaxaca-Blinder intégrant des variables d'environnement économique.

Les variables d'environnement économique retenues sont :

- **La densité de la population** (source : Recensement de la population)
La densité de population par bassin de vie varie de moins de 10 habitants au km² dans certains territoires ruraux, à 2 700 habitants pour celui de Paris (ce bassin s'étend largement au-delà du département de Paris, d'où une densité très inférieure à celle de la ville de Paris stricto sensu). Même après standardisation, l'amplitude de la variable reste très forte. De plus la relation non-linéaire entre le logarithme du revenu et la densité de population conduit à intégrer la densité au carré parmi les régresseurs, l'amplitude de cette dernière variable étant encore plus importante. C'est pourquoi ces deux variables ont été remplacées par la densité de population discrétisée en six classes, construites selon la distribution de la densité pour l'ensemble des bassins de vie :
 - densité inférieure au premier décile ;
 - densité comprise entre le premier décile et le premier quartile ;
 - densité comprise entre le premier quartile et la médiane ;
 - densité comprise entre la médiane et le troisième quartile ;
 - densité comprise entre le troisième quartile et le 9^e décile ;
 - densité supérieure au 9^e décile.
- **Le niveau de vie** (source : Fichier Localisé Social et Fiscal – Filosofi – 2013)
La variable retenue est le revenu moyen par Unité de Consommation. Cette variable a été standardisée.
- **La touristicité** (Sources : Insee en partenariat avec la DGE et les partenaires territoriaux, Kit tourisme capacité d'hébergement données 2014)
La touristicité d'un territoire revêt divers aspects : tourisme d'affaires / tourisme de loisirs, hébergement en chambres d'hôtel, en campings, dans les Autres Hébergements Collectifs de Tourisme (AHCT : résidences de tourisme, villages de vacances, auberges de jeunesse, ...), en résidences secondaires, ...
Deux variables ont été testées : la densité hôtelière (nombre de chambres d'hôtel au km²) et la densité de campings (nombre d'emplacements au km²). Le choix a été fait de ne pas construire un indicateur synthétique de touristicité, afin de mieux conserver la diversité des profils touristiques. Les deux variables ont été discrétisées selon leur distribution :
 - densité inférieure au premier quartile ;
 - densité comprise entre le premier quartile et la médiane ;
 - densité comprise entre la médiane et le troisième quartile ;
 - densité comprise entre le troisième quartile et le 9^e décile ;
 - densité supérieure au 9^e décile.

In fine, seule la variable de densité hôtelière a été conservée, les coefficients associés à la densité de campings étant difficiles à interpréter (alternativement positifs et négatifs d'une modalité à l'autre). Une limite de la densité hôtelière est qu'elle ne permet pas de distinguer les zones où le tourisme d'affaires prédomine de celles où le tourisme de loisirs prédomine.
- **Le taux de chômage** (source : Insee, taux de chômage localisés)
La variable retenue pour décrire le marché du travail est le taux de chômage au sens du Bureau International du Travail par zone d'emploi.
- **La part de personnes âgées** (source : Recensement de la population)
Afin de prendre en compte la spécificité des territoires ayant une forte proportion de personnes âgées, la part des seniors a été introduite dans les modèles. Plus précisément, la variable retenue est la part des 65 ans et plus dans la population totale.

La prise en compte de l'ensemble de ces variables d'environnement économique fait à peine progresser le R^2 ajusté du modèle de régression du logarithme du revenu, qui passe de 0,288 à 0,290 (cf. Annexe 3, modèles M3 et M8).

2. La variable d'intérêt : revenu moyen versus revenu médian ?

L'indicateur de revenu d'activité retenu dans le cadre de l'Insee Première n° 1672, « Le revenu d'activité des non-salariés : plus élevé que la moyenne dans les départements du nord que dans ceux du sud », est le revenu mensuel moyen. C'est également l'indicateur principal de l'Insee Première annuel sur les revenus d'activité des non-salariés.

Toutefois, une des limites de la moyenne est sa sensibilité aux valeurs extrêmes, en particulier, dans le cas des revenus, aux valeurs les plus élevées : le revenu médian est ainsi nettement inférieur au revenu moyen (figure 5). Il pourrait être intéressant d'analyser les écarts de revenus sur les niveaux médians, plutôt que sur les niveaux moyens. Choisir le revenu médian comme indicateur d'intérêt a un impact non négligeable sur la répartition spatiale des écarts de revenus à la médiane nationale, si on s'intéresse à l'ensemble des non-salariés. L'impact est plus limité si on restreint l'analyse des disparités de revenus d'activité entre départements aux non-salariés classiques à revenu non nul.

Figure 5 : Revenu d'activité mensuel des non-salariés

	Revenu moyen (en euros)	Revenu médian (en euros)
Ensemble des non-salariés (y compris auto-entrepreneurs)	2 510	1 200
Non-salariés classiques	3 260	1 950
Non-salariés classiques à revenu non nul	3 610	2 230

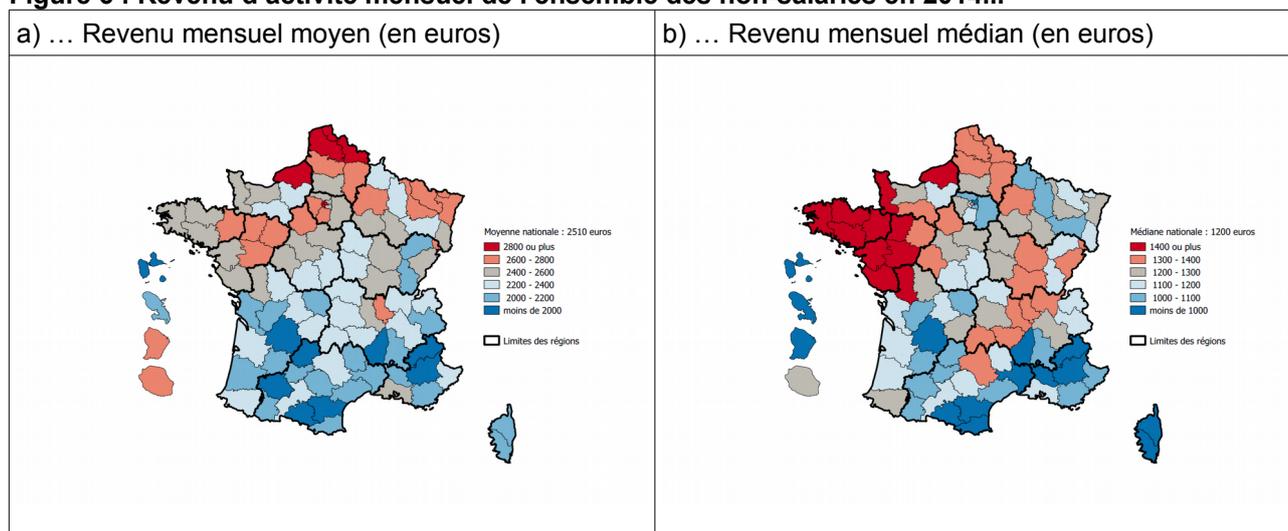
Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

2.1. Impact de l'indicateur retenu sur les disparités de revenus d'activité, sur le champ de l'ensemble des non-salariés (y compris auto-entrepreneurs)

Quand l'analyse porte sur l'ensemble des non-salariés (y compris les auto-entrepreneurs), la distribution géographique des écarts de revenus médians à la médiane française diffère sensiblement de celle des écarts de revenus moyens à la moyenne française : en particulier les revenus d'activité médians les plus élevés ne se concentrent pas au nord de l'hexagone mais en Bretagne et dans les Pays-de-Loire (figure 6). En revanche, les revenus médians départementaux les plus bas sont bien localisés dans le sud comme pour les revenus d'activité moyens.

Figure 6 : Revenu d'activité mensuel de l'ensemble des non-salariés en 2014...



Lecture : dans le département du Pas-de-Calais, le revenu d'activité mensuel moyen est supérieur à 2800 € en 2014. La moitié des non-salariés de ce département gagne plus de 1 400 € par mois.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

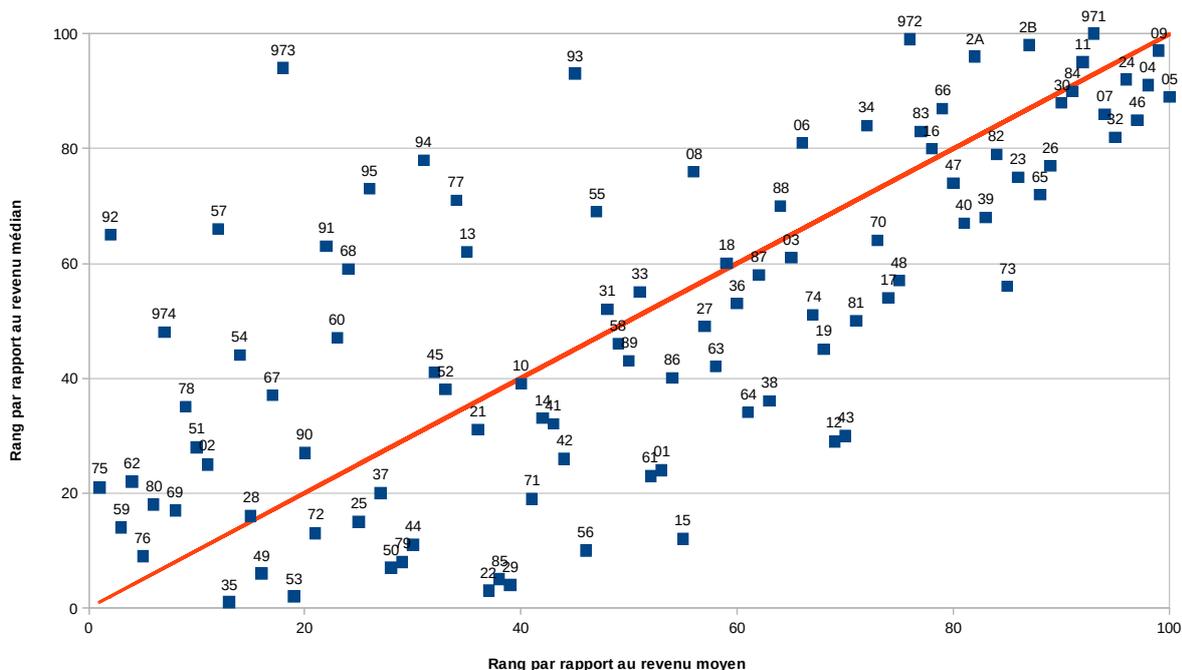
Le revenu d'activité médian est toujours plus faible que le revenu d'activité moyen. Cependant, l'écart entre ces deux indicateurs est très variable selon les départements (cf. Annexe 2). Ainsi leur classement suivant le revenu médian diffère de celui réalisé à partir du revenu moyen (figure 7). Le coefficient de Spearman mesurant la corrélation entre les deux classements n'est que de 0,68.

Certains départements perdent des places quand on les classe selon leur revenu d'activité médian plutôt que selon le revenu d'activité moyen. C'est le cas des DOM, notamment la Guyane (département 973) et la Réunion (département 974) : ces territoires ont un revenu moyen assez important et un revenu médian modéré ou faible. En France métropolitaine, le classement de la plupart des départements d'Île-de-France (région 11) se dégrade en retenant le revenu d'activité médian. Il en est de même pour une bonne part de ceux du Grand-Est (région 44).

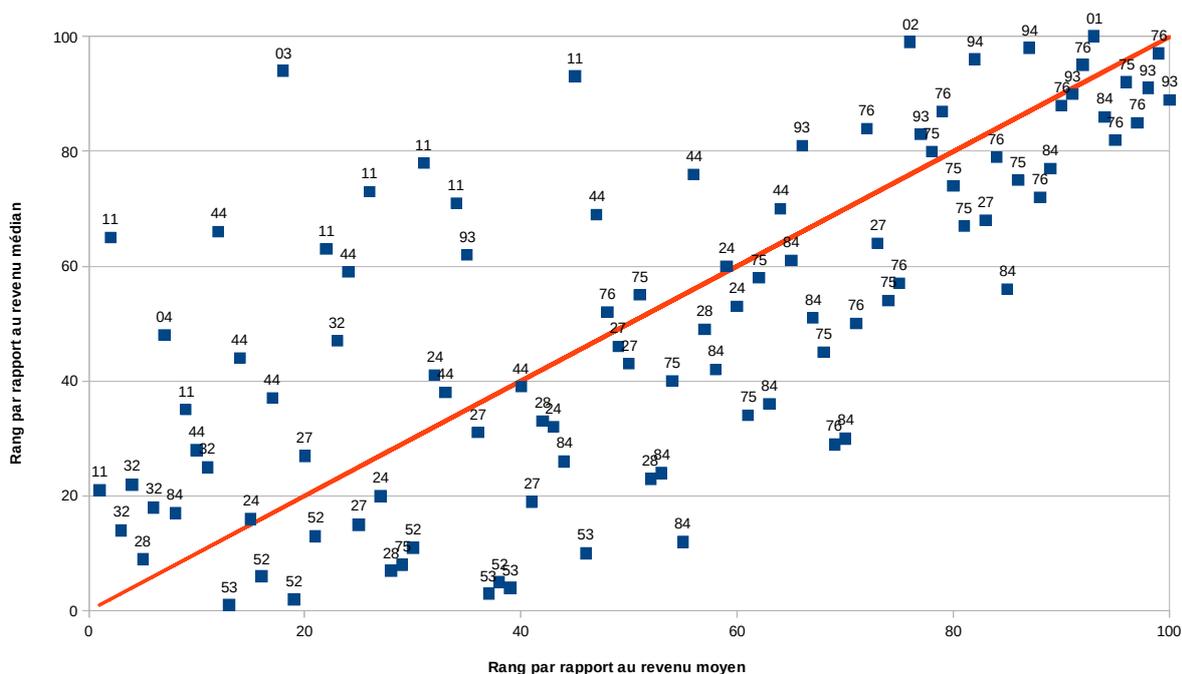
À l'inverse, d'autres départements ont un meilleur classement selon le revenu médian que selon le revenu moyen, comme ceux de Bretagne (région 53) ou des Pays-de-Loire (région 52). Une partie des départements des régions de la moitié sud de la France, Occitanie (région 76), Nouvelle-Aquitaine (région 75) et surtout Auvergne – Rhône-Alpes (région 84), sont aussi dans ce cas.

Figure 7 : Impact de l'indicateur de revenu retenu sur le classement des départements, pour l'ensemble des non-salariés ...

a) ... Points étiquetés selon le numéro du département (cf. Annexe 5)



b) ... Points étiquetés selon la région d'appartenance (cf. Annexe 4)



Note de lecture : au-dessus de la droite en rouge, le classement des départements selon le revenu médian est moins bon (c'est-à-dire plus élevé) que selon le revenu moyen. Au-dessous il est meilleur selon le revenu médian que selon le revenu moyen.

Par exemple, la Guyane (département 973, région 03) occupe le 18^e rang dans le classement selon le revenu moyen, et le 94^e rang selon le revenu médian.

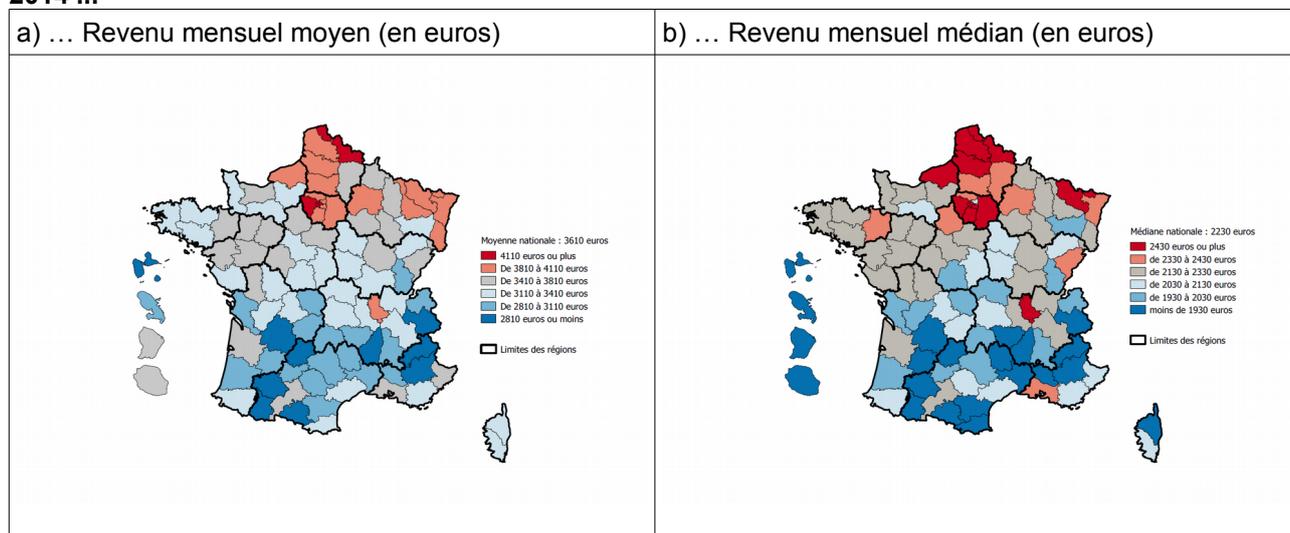
Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

2.2. Impact de l'indicateur retenu sur les disparités de revenus d'activité, sur le champ des non-salariés classiques à revenu non nul

Si on restreint l'analyse aux non-salariés classiques à revenu non nul, la distribution géographique des écarts de revenus d'activité médians à la médiane nationale ne diffère plus fondamentalement de celle des écarts de revenus moyens à la moyenne nationale (figure 8). selon les deux indicateurs, les écarts sont positifs au nord de l'hexagone et négatifs au sud. Dans le sud, la plupart des départements ayant une grande agglomération (Gironde, Haute-Garonne et Bouches-du-Rhône) ont un revenu d'activité plus fort que leurs voisins, quel que soit l'indicateur retenu (moyenne ou médiane).

Figure 8 : Revenu d'activité mensuel des non-salariés classiques ayant déclaré un revenu non nul en 2014 ...



Lecture : dans le département du Pas-de-Calais, le revenu d'activité mensuel moyen est compris entre 3 810 € et 4 110 € en 2014. La moitié des non-salariés classiques à revenu non nul de ce département gagne plus de 2 430 € par mois.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs)

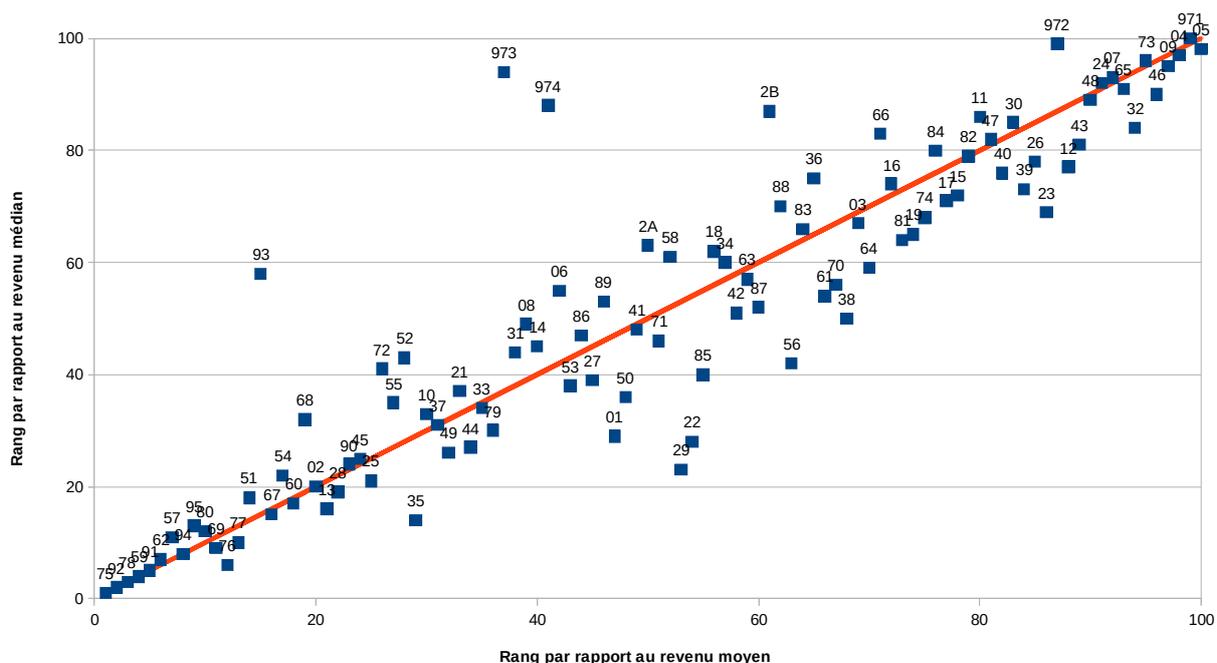
Source : Insee, base non-salariés 2014

Sur ce champ, quand on passe du revenu moyen au revenu médian, les changements de rang sont modérés : le coefficient de Spearman mesurant la corrélation entre les deux classements est de 0,91, relativement proche de 1.

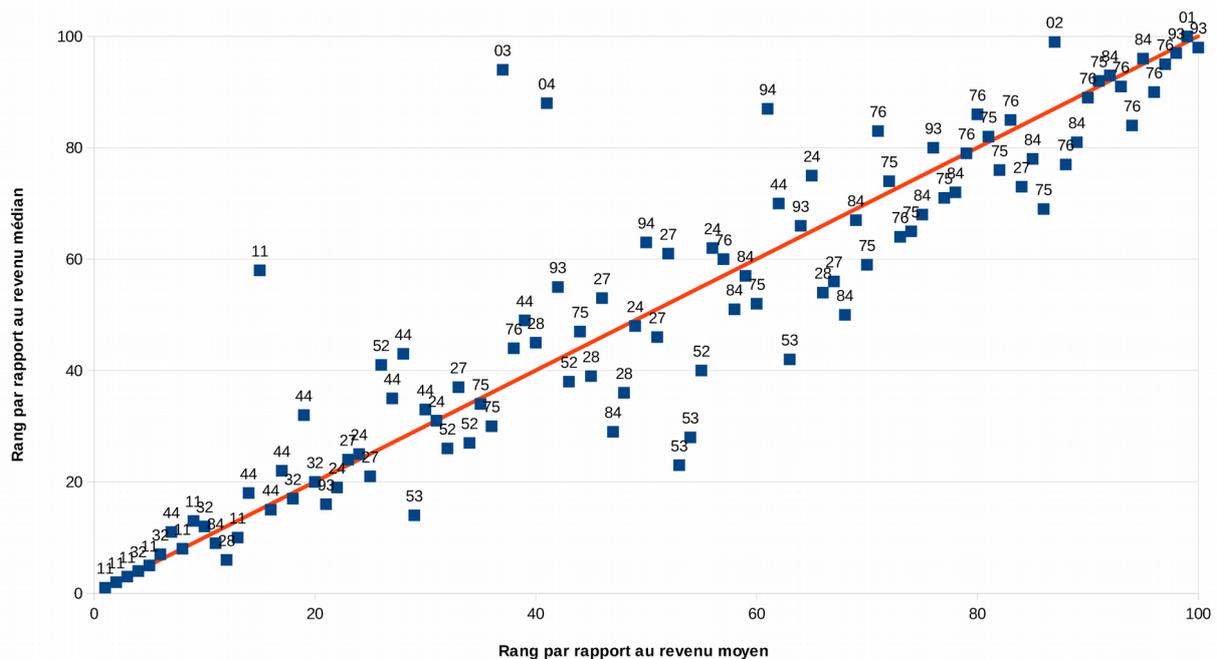
Cependant, quelques départements se distinguent toujours par des classements différant fortement selon l'indicateur (médiane ou moyenne) retenu (figure 9). Ainsi la Guyane et la Réunion ou la Seine-Saint-Denis perdent là aussi des places et les départements de Bretagne en gagnent.

Figure 9 : Impact de l'indicateur de revenu retenu sur le classement des départements, pour les non-salariés classiques à revenu non nul ...

a) ... Points étiquetés selon le numéro du département (cf. Annexe 5)



b) ... Points étiquetés selon la région d'appartenance (cf. Annexe 4)



Note de lecture : au-dessus de la droite en rouge, le classement des départements selon le revenu médian est moins bon (c'est-à-dire plus élevé) que selon le revenu moyen. Au-dessous il est meilleur selon le revenu médian que selon le revenu moyen.

Par exemple, la Guyane (département 973, région 03) occupe le 37^e rang dans le classement selon le revenu moyen, et le 94^e rang selon le revenu médian.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

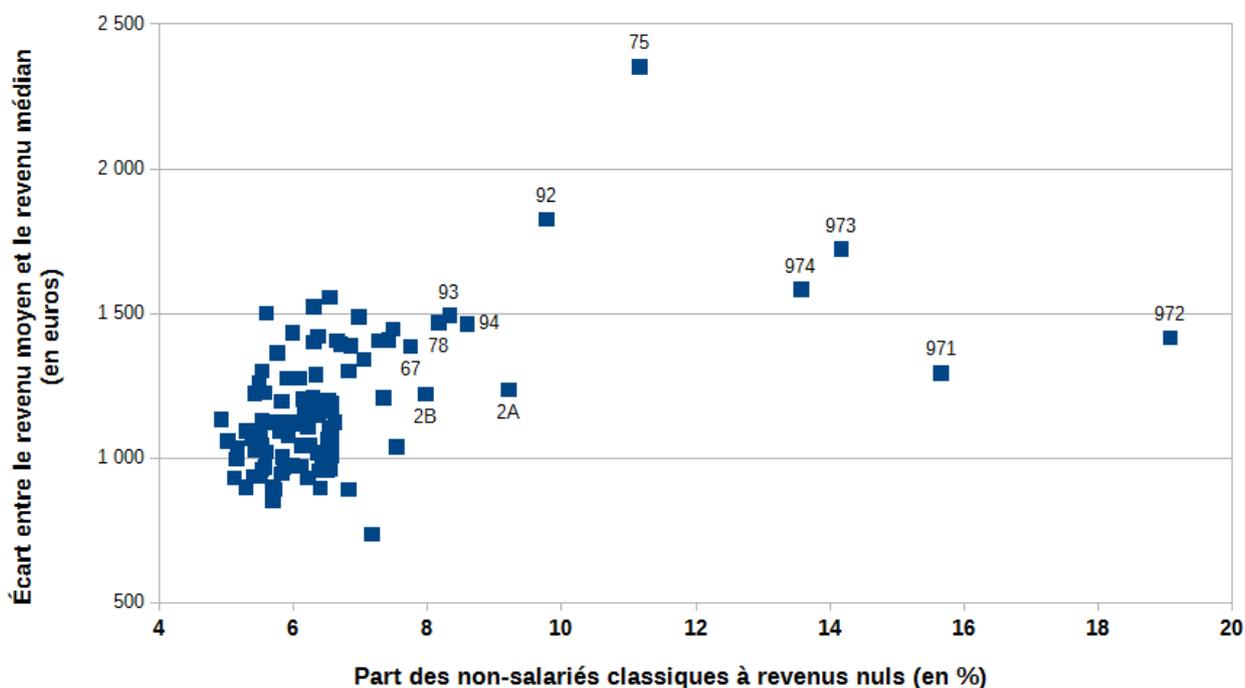
Source : Insee, base non-salariés 2014.

2.3. Plusieurs facteurs expliquent la sensibilité du classement des départements à l'indicateur de revenu retenu

L'impact de l'indicateur de revenu d'activité retenu sur le classement des départements découle de différents paramètres, comme la variabilité de la part des non-salariés classiques à revenu nul ou, dans une moindre mesure, la distribution des revenus d'activité parmi les non-salariés classiques à revenu non nul. En revanche, l'écart entre le revenu moyen et le revenu médian n'est pas significativement corrélé avec la part d'auto-entrepreneurs (coefficient de corrélation de 0,02, avec ou sans les DOM). D'autres éléments nous amènent à écarter les auto-entrepreneurs de l'analyse des disparités de revenus entre départements (cf. 1.3.1. Les grandes catégories de non-salariés).

L'écart entre le revenu moyen et le revenu médian est positivement corrélé avec la proportion de non-salariés classiques à revenu nul : le coefficient de corrélation est de 0,51 avec les DOM et de 0,64 sans les DOM (figure 10). Ainsi, les départements ayant un écart élevé ont en général une proportion importante de non-salariés classiques à revenu nul (cf. Annexe 2).

Figure 10 : Écart entre le revenu moyen et le revenu médian départemental, en fonction de la part de non-salariés classiques à revenu nul



Note de lecture : En Guadeloupe (département 971), le revenu d'activité mensuel médian est inférieur de 1 290 € au revenu d'activité mensuel moyen. 15,7 % des non-salariés de ce département sont des non-salariés classiques à revenu nul.

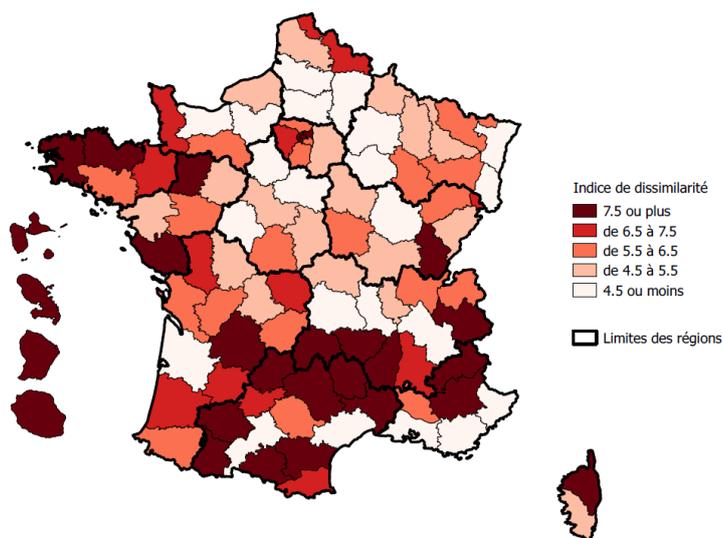
Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

La distribution des revenus parmi les non-salariés classiques à revenu non nul influe aussi sur l'écart entre le revenu moyen et le revenu médian : elle est assez variable selon les régions. La distribution des revenus dans les DOM, dans la plupart des départements de Bretagne, dans les départements alpins et dans de nombreux départements du sud diffère sensiblement de celle observée sur l'ensemble de la France.

Un indice de dissimilarité de la distribution des revenus, mesurant la proportion de non-salariés qu'il faudrait réaffecter entre les différentes tranches de revenus pour obtenir une distribution des revenus identique à celle observée en France, permet d'apprécier les écarts de distributions entre départements (figure 11).

Figure 11 : Indice de dissimilarité de la distribution des revenus des non-salariés classiques à revenu non nul (référence = France)



Note : La variable de revenu est découpée en tranches de 200 euros. L'indice de dissimilarité mesure la proportion de non-salariés qu'il faudrait réaffecter entre les tranches de revenus pour obtenir une distribution des revenus identique à celle observée pour la France.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

La distribution du revenu d'activité s'approche d'une distribution log-normale dans l'ensemble des régions. Le revenu médian est donc toujours plus faible que le revenu moyen (pour rappel, lorsque la distribution est log-normale, l'inégalité suivante est vérifiée : Mode < Médiane < Moyenne). Or, pour ce type de distribution, la différence entre la médiane et la moyenne est d'autant plus forte, pour une moyenne donnée, que l'écart-type est important.

De fait, l'écart-type du revenu d'activité est très élevé en Île-de-France, en Guyane ou à la Réunion, régions pour lesquelles l'écart entre revenu médian et revenu moyen est important (figure 12). Dans les départements d'outre-mer, ce résultat s'explique avant tout par une surreprésentation des non-salariés classiques à faible revenu, mais aussi par celle des non-salariés à revenu élevé (figure 13). En Île-de-France, seule joue la surreprésentation des non-salariés classiques à revenu élevé : 28,5 % d'entre eux ont un revenu d'activité supérieur ou égal à 5 000 euros par mois, contre 20,5 % en France (figure 14).

À l'inverse, en Bretagne où l'écart entre le revenu médian et le revenu moyen est faible, l'écart-type du revenu d'activité est très modéré. En effet, la part des non-salariés classiques à faible revenu est moindre qu'à l'échelon national.

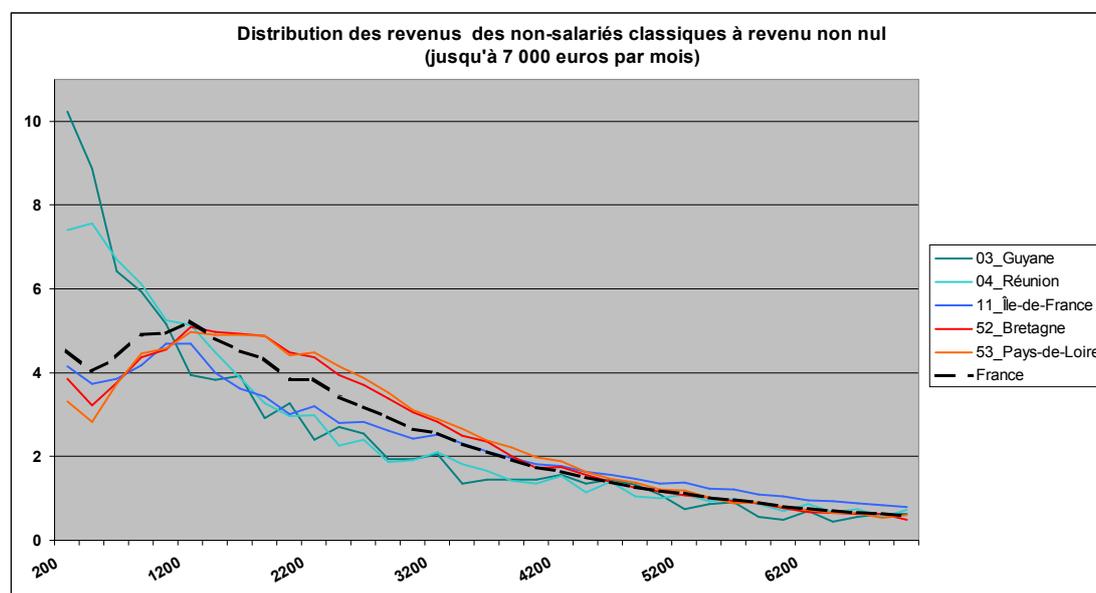
Figure 12 : Écart-type du revenu d'activité des non-salariés classiques à revenu non nul

Région	Écart revenu médian - revenu moyen	Écart-type du revenu
11 - Île-de-France	-1930	8330
03 - Guyane	-1760	5520
04 - Réunion	-1620	4660
02 - Martinique	-1570	4180
32 - Hauts-de-France	-1500	4970
44 - Grand-Est	-1450	4770
01 - Guadeloupe	-1440	4050
France	-1380	5190
94 - Corse	-1350	4290
28 - Normandie	-1270	4390
24 - Centre - Val de Loire	-1270	4330
93 - Provence - Alpes - Côte d'Azur	-1240	4230
27 - Bourgogne - Franche-Comté	-1200	4170
52 - Pays de la Loire	-1190	4160
84 - Auvergne - Rhône-Alpes	-1160	4130
76 - Occitanie	-1150	3950
75 - Nouvelle Aquitaine	-1140	3970
53 - Bretagne	-1040	3870

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Figure 13 : Distribution des revenus des non-salariés classiques à revenu non nul



Note : graphique tronqué à 7 000 euros mensuels. La distribution est approchée par un histogramme, avec des classes correspondant à des tranches de revenus d'amplitude de 200 euros (l'ordonnée correspond à la part en % de la classe dans l'ensemble de la population étudiée).

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Figure 14 : Part des non-salariés classiques (hors revenus nuls) à faibles ou à forts revenus, selon les régions

Région	Moins de 1 000 euros		5 000 euros ou plus		20 000 euros ou plus	
	Part	Rang	Part	Rang	Part	Rang
01 - Guadeloupe	47,1	1	14,7	17	1,0	14
02 - Martinique	41,3	2	18,2	12	1,0	12
03 - Guyane	36,6	3	19,4	7	1,8	2
04 - Réunion	33,0	4	21,3	4	1,4	5
11 - Île-de-France	20,6	15	28,2	1	2,7	1
24 - Centre - Val de Loire	21,7	12	19,5	6	1,1	8
27 - Bourgogne - Franche-Comté	22,2	11	18,5	11	1,1	9
28 - Normandie	20,9	14	20,5	5	1,1	7
32 - Hauts-de-France	21,0	13	25,0	2	1,6	3
44 - Grand-Est	22,4	10	22,2	3	1,5	4
52 - Pays de la Loire	19,8	16	18,8	10	1,0	11
53 - Bretagne	18,9	17	18,1	13	0,7	17
75 - Nouvelle Aquitaine	23,4	9	17,2	15	0,8	16
76 - Occitanie	25,5	6	16,6	16	0,9	15
84 - Auvergne - Rhône-Alpes	23,5	8	17,5	14	1,0	10
93 - Provence - Alpes - Côte d'Azur	24,7	7	19,3	8	1,0	13
94 - Corse	27,6	5	19,0	9	1,1	6
France	22,8		20,5		1,3	

Note de lecture : En Guadeloupe, 47,1 % des non-salariés classiques à revenu non nul ont un revenu inférieur à 1000 euros, ce qui la porte au 1^{er} rang lorsqu'on classe les régions selon ce critère.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

La variabilité des proportions de non-salariés à revenu nul et celle des dispersions de revenus expliquent la sensibilité du classement des départements à l'indicateur de revenu retenu, sur un champ incluant l'ensemble des non-salariés. Sur le champ restreint des non-salariés classiques à revenu non nul, seul le deuxième aspect joue et les effets sont moins prégnants. Le cœur de l'étude reposant sur la population des non-salariés classiques à revenu non nul, choisir comme indicateur de revenu le revenu moyen ou le revenu médian ne change donc pas fondamentalement le résultat principal de l'analyse : le revenu d'activité (moyen ou médian) est plus élevé au nord qu'au sud.

Une des extensions existantes à la méthode de décomposition de Oaxaca-Blinder s'appuie sur les quantiles de la variable d'intérêt au lieu de la moyenne. Développer un modèle de décomposition basé sur le revenu médian (ou d'autres quantiles ou statistiques comme des mesures de dispersion) serait donc néanmoins possible (cf. Maillard S., Boutchenik B. : « Méthodes économétriques de décomposition des inégalités – de la théorie à la pratique », Document de travail, Insee, 2018).

3. Contrôle des effets de structure : les modèles utilisés en économie géographique

En économie géographique, les outils classiques de contrôle des effets de structure, pour une variable d'intérêt quantitative continue, sont les modèles **structurels-résiduels** (« descriptif » et « économétrique ») et les modèles **structurels-géographiques**.

Les différents modèles sont présentés ici dans le cadre de notre problématique : les écarts de revenus entre un département et la France découlent-ils, en partie ou totalement, de différences de structure par secteur d'activité X catégorie de non-salarié, par caractéristiques des non-salariés et de leur activité ou par environnement économique dans lequel s'exerce l'activité. On veut pouvoir analyser les effets de structure pour chacun des 96 départements métropolitains⁵.

Comme variable d'intérêt, on utilise ici, dans un souci d'améliorer la spécification de la forme fonctionnelle, le logarithme du revenu d'activité et non le revenu d'activité (ce qui implique de se restreindre au champ des non-salariés à revenu positif). En fait les modèles de contrôle des effets de structure utilisés en économie géographique (qui découlent d'égalités comptables), de même que le modèle de décomposition de Oaxaca-Blinder utilisé en économie du travail, n'impliquent pas que l'hypothèse de normalité soit vérifiée. Celle-ci n'est nécessaire que pour l'inférence mise en œuvre habituellement, encore que cette hypothèse puisse être relâchée en recourant à des méthodes telles que le Bootstrap.

3.1. Le Modèle Structurel-Résiduel descriptif

Le modèle structurel-résiduel descriptif ne permet d'utiliser qu'une seule variable de structure. Cette variable doit être catégorielle.

→ Si on veut prendre en compte plusieurs variables structurelles, il est nécessaire de les croiser. Cette possibilité reste limitée : aucun croisement ne doit avoir un effectif nul.

→ Si on veut utiliser une variable structurelle quantitative, il est nécessaire de la discrétiser.

Dans le cadre de notre problématique, la variable d'intérêt est le logarithme du revenu d'activité et la variable de structure peut être le croisement « secteur d'activité X catégorie de non-salarié ».

L'écart entre la moyenne des logarithmes du revenu d'un département d et celle de la France peut s'écrire :

$$(R_d - R_F) = \sum_i ((P_{di} * R_{di}) - (P_{Fi} * R_{Fi})) = \sum_i (P_{di} - P_{Fi}) * R_{Fi} + \sum_i P_{di} * (R_{di} - R_{Fi})$$

où :

P_{zi} est le poids de la catégorie i (modalité i de la variable de structure) dans la zone z (le département d ou la France F), c'est-à-dire la proportion de non-salariés de la zone z appartenant à la catégorie i .

R_{zi} est la moyenne des logarithmes du revenu des non-salariés de catégorie i dans la zone z

R_z est la moyenne du logarithme du revenu des non-salariés dans la zone z

Le premier terme de cette décomposition correspond à la partie expliquée (effet structurel) de cet écart :

$$\sum_i (P_{di} - P_{Fi}) * R_{Fi} = \sum_i (P_{di} - P_{Fi}) * (R_{Fi} - R_F)$$

Le second terme ($\sum_i P_{di} * (R_{di} - R_{Fi})$) correspond à la partie inexpliquée (effet résiduel).

⁵ Certaines variables représentant l'environnement économique dans lequel s'exerce l'activité ne sont pas disponibles pour les DOM. C'est pourquoi on se limite aux départements métropolitains.

En mettant en œuvre cette décomposition, on peut calculer la contribution de chaque modalité de la variable de structure à l'effet structurel (partie expliquée de l'écart) ainsi qu'à l'effet résiduel (partie inexpliquée). Ce modèle ne permet pas d'estimer directement des intervalles de confiance pour les écarts ainsi mesurés. Ceux-ci peuvent être cependant estimés au moyen de méthodes de Bootstrap.

3.2. Le Modèle Structurel-Résiduel économétrique

Le modèle structurel-résiduel économétrique fait appel à une analyse de la variance. Il permet d'utiliser plusieurs variables de structure. Celles-ci doivent être catégorielles.

→ Si l'on veut utiliser une variable structurelle quantitative, il est nécessaire de la discrétiser.

Les données de base utilisées sont soit des données individuelles, soit des données agrégées correspondant à des regroupements d'individus. Les résultats obtenus sont identiques, car les variables de structure sont catégorielles. Dans le cas de données agrégées, l'analyse de la variance est pondérée par le poids des regroupements considérés.

Dans le cadre de notre problématique, on analyse la variance du logarithme du revenu, sur le champ de la France métropolitaine. On utilise pour variables de structure les croisements secteur X catégorie du non-salarié et les autres caractéristiques du non-salarié et de son activité.

Soient les variables de structure $X_i, i=1...K$.

On notera P_{zij} le poids de la modalité j de la variable X_i dans la zone z (département d ou France F).

On a donc $\sum_j (P_{zij}) = 1$, pour une variable X_i et une zone z données.

La partie expliquée (effet structurel) de l'écart entre la moyenne des logarithmes du revenu d'un département d et celle de la France peut s'écrire :

partie expliquée⁶ = $\sum_i \sum_j (P_{dij} * \hat{\beta}'_{ij})$ avec $\hat{\beta}'_{ij} = \hat{\beta}_{ij} - \sum_j (P_{Fij} * \hat{\beta}_{ij})$
 où $\hat{\beta}_{ij}$ est le coefficient estimé de la modalité j de la variable X_i dans l'analyse de la variance.
 (On a donc imposé les contraintes : $\forall i, \sum_j (P_{Fij} * \hat{\beta}'_{ij}) = 0$, ce qui correspond à prendre la France comme référence).

On a : $\hat{\beta}'_{ij} = (R_{Fij} - R_{F..})$

où :

R_{Fij} est la moyenne des logarithmes du revenu des non-salariés de catégorie j pour la variable i en France

$R_{F..}$ est la moyenne des logarithmes du revenu des non-salariés en France

La partie inexpliquée (effet résiduel), qui est la moyenne des résidus des non-salariés du département dans l'analyse de la variance, peut aussi se calculer par différence entre l'écart observé et la partie expliquée.

⁶ Afin de faciliter la comparaison avec le modèle de décomposition de Oaxaca-Blinder, on utilise les notations traditionnelles de la régression (et non celles de l'analyse de la variance).

À noter que les parties expliquée et inexpliquée (effets structurel et résiduel) issues d'un modèle structurel-résiduel « économétrique » à une seule variable de structure sont identiques à celles calculées dans le cadre d'un modèle structurel-résiduel « descriptif » sur cette même variable de structure.

En effet, avec un modèle structurel-résiduel « économétrique » à une seule variable, la partie expliquée peut s'écrire :

$$\sum_i P_{di} * \hat{\beta}'_{ij} = \sum_i P_{di} * (R_{Fi} - R_F)$$

où $\hat{\beta}_i$ est le coefficient estimé de la modalité i de la variable X (on reprend ici les notations correspondant à celles utilisées dans la présentation du modèle structurel-résiduel descriptif).

La partie expliquée dans le modèle structurel-résiduel descriptif est égale à :

$$\begin{aligned} & \sum_i (P_{di} - P_{Fi}) * (R_{Fi} - R_F) \\ &= \sum_i P_{di} * (R_{Fi} - R_F) - \sum_i P_{Fi} * R_{Fi} + \sum_i P_{Fi} * R_F \\ &= \sum_i P_{di} * (R_{Fi} - R_F) \\ & \text{car : } \sum_i P_{Fi} * R_{Fi} = R_F \cdot \sum_i P_{Fi} = R_F \cdot R_F \text{ et } \sum_i P_{Fi} * R_F = R_F \cdot \sum_i P_{Fi} = R_F \cdot R_F \end{aligned}$$

On retrouve donc bien la même expression de la partie expliquée dans les deux modèles.

En mettant en œuvre un modèle structurel-résiduel « économétrique », on peut calculer la contribution de chaque modalité de chaque variable de structure (et donc la contribution de chaque variable de structure) à l'effet structurel (partie expliquée), mais pas à l'effet résiduel (part inexpliquée).

3.3. Le Modèle Structurel-Géographique économétrique

La mise en œuvre d'un modèle structurel-géographique est proche de celle d'un modèle structurel-résiduel économétrique : les contraintes portant sur les variables de structure et sur la variable d'intérêt sont les mêmes. On introduit simplement un facteur géographique parmi les variables explicatives.

Dans le cas de notre problématique, le facteur géographique introduit est le département d'activité.

On impose, pour toutes les variables de structure, y compris pour le facteur géographique, les contraintes :

$$\forall i, \sum_j (P_{Fij} * \hat{\beta}'_{ij}) = 0$$

On suppose qu'il y a $(K - 1)$ variables structurelles (notées X_1 à X_{K-1}) et on note X_K la variable correspondant au facteur géographique.

On a alors :

$$\text{partie expliquée} = \sum_{i=1}^{K-1} \sum_j (P_{dij} * \hat{\beta}'_{ij}) \text{ avec } \hat{\beta}'_{ij} = \hat{\beta}_{ij} - \sum_j (P_{Fij} * \hat{\beta}_{ij})$$

où $\hat{\beta}_{ij}$ est le coefficient estimé de la modalité j de la variable X_i dans l'analyse de la variance.

Pour la variable correspondant au facteur géographique, P_{dKj} est égal à 1 si la modalité j correspond au département d , à 0 sinon. On a donc :

$$\text{Effet géographique} = \sum_j (P_{dKj} * \hat{\beta}'_{Kj}) = \hat{\beta}'_{Kj} \text{ , soit le coefficient de la modalité } j \text{ correspondant au département } d.$$

Du fait de l'intégration d'une variable supplémentaire (le facteur géographique), la partie expliquée diffère de celle obtenue en mettant en œuvre un modèle résiduel-structurel ou un modèle de décomposition de Oaxaca-Blinder (cf. 4. Contrôle des effets de structure : les outils utilisés en économie du travail).

L'hypothèse sous-jacente au modèle structurel-géographique est qu'il existe un effet géographique identique pour tous les individus d'une même zone (dans notre problématique, le département). Sous cette hypothèse, on peut décomposer la partie inexpliquée en un effet géographique et une partie résiduelle. Cette dernière correspond à la moyenne des résidus, dans l'analyse de la variance, des individus du département. Toutefois, avec l'intégration d'un effet fixe géographique, les moyennes des résidus par zone géographique sont nulles. La partie inexpliquée est donc simplement égale à l'effet géographique.

Comme pour le modèle structurel-résiduel « économétrique », on peut calculer la contribution de chaque modalité de chaque variable de structure à l'effet structurel (la partie expliquée), et donc la contribution de chaque variable de structure. On ne peut pas calculer les contributions à la partie inexpliquée.

4. Contrôle des effets de structure : les outils utilisés en économie du travail

En économie du travail, un outil classique de contrôle des effets de structure, pour une variable d'intérêt quantitative continue, est le modèle de décomposition de Oaxaca-Blinder.

Ce modèle est présenté ci-après dans le cas de la décomposition d'un écart de moyennes.

4.1. Présentation du modèle de décomposition de Oaxaca-Blinder

Avec le modèle de décomposition de Oaxaca-Blinder, comme avec les modèles de contrôle des effets de structure utilisés en économie géographique, on utilise pour variable d'intérêt le logarithme du revenu d'activité et non le revenu d'activité, dans un souci d'amélioration de la forme fonctionnelle (ce qui implique de se restreindre au champ des non-salariés à revenu positif).

Le modèle permet d'utiliser plusieurs variables de structure. Celles-ci peuvent être catégorielles ou continues.

Les données de base utilisées sont des données individuelles. Si les variables de structure sont toutes catégorielles, on peut également utiliser des données agrégées.

Dans le cadre de notre problématique, on doit, pour chaque département, procéder à deux régressions du logarithme du revenu d'activité sur les variables de structure :

- une régression sur la population des non-salariés du département ;
- une régression sur la population des non-salariés de la France métropolitaine, hors ceux du département (= la population de référence).

Les deux populations respectent bien la condition de groupes mutuellement exclusifs (« Mutually Exclusive Groups »). On notera d'ailleurs que la population de référence est différente pour chaque département. Ainsi, pour les 96 départements métropolitains, il est nécessaire de réaliser 192 régressions.

On peut toutefois poser l'hypothèse que le poids d'un département dans l'ensemble de la France est relativement faible et donc faire le choix de ne pas exclure de la population de référence les non-salariés du département⁷. On suppose que cette non-exclusion ne modifie pas de façon sensible les résultats de la régression sur la population de référence. Sous cette hypothèse, on limite le nombre de régressions à 97 (une sur chaque département et une sur l'ensemble de la France métropolitaine).

⁷ Cette hypothèse a été posée dans le cadre de l'Insee Première n° 1672 sur « les revenus d'activité des non-salariés : plus élevés en moyenne dans les départements du nord que dans ceux du sud ».

Pour un département d , on a :

$$\begin{aligned} \bar{Y}_d - \bar{Y}_F &= (\hat{\beta}_{d0} - \hat{\beta}_{F0}) + \sum_i ((\hat{\beta}_{di} - \hat{\beta}_{Fi}) * \bar{X}_{di}) && \text{(partie inexpliquée)} \\ &+ \sum_i ((\bar{X}_{di} - \bar{X}_{Fi}) * \hat{\beta}_{Fi}) && \text{(partie expliquée)} \end{aligned}$$

où :

\bar{Y}_z correspond à la moyenne des logarithmes du revenu d'activité dans la zone z (département d ou France métropolitaine hors ce département)

\bar{X}_{zi} correspond à la moyenne de la variable explicative X_i dans la zone z

$\hat{\beta}_{zi}$ le coefficient estimé de la variable X_i dans la régression réalisée sur les non-salariés de la zone z

$\hat{\beta}_{z0}$ la constante estimée dans la régression réalisée sur les non-salariés de la zone z

Dans le cas de variables de structure catégorielles, les régresseurs sont les variables dichotomiques, qu'on peut noter X_{ij} , associées aux modalités j des variables catégorielles X_i . \bar{X}_{zij} est alors égal à P_{zij} , le poids de la modalité j de la variable X_i dans la zone z .

On a donc $\sum_j (P_{zij}) = 1$, pour une variable X_i et une zone z données.

La décomposition devient :

$$\begin{aligned} \bar{Y}_d - \bar{Y}_F &= (\hat{\beta}_{d0} - \hat{\beta}_{F0}) + \sum_i \sum_j ((\hat{\beta}_{dij} - \hat{\beta}_{Fij}) * P_{dij}) && \text{(partie inexpliquée)} \\ &+ \sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) && \text{(partie expliquée)} \end{aligned}$$

En mettant en œuvre un modèle de décomposition de Oaxaca-Blinder, on peut calculer la contribution de chaque modalité de chaque variable de structure (et donc la contribution de chaque variable de structure) à la partie expliquée ET à la partie inexpliquée.

Dans le cas de variables catégorielles :

- La contribution de la modalité j de la variable i à la partie expliquée est égale à :

$$(P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}$$

- La contribution de la variable i à la partie expliquée est égale à :

$$\sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij})$$

- La contribution de la modalité j de la variable i à la partie inexpliquée est égale à :

$$(\hat{\beta}_{dij} - \hat{\beta}_{Fij}) * P_{dij}$$

- La contribution de la variable i à la partie inexpliquée est égale à :

$$\sum_j ((\hat{\beta}_{dij} - \hat{\beta}_{Fij}) * P_{dij})$$

4.2. Identification du modèle

Avec des variables de structure catégorielles, se pose le problème de l'identification du modèle : il faut imposer des contraintes sur les coefficients. Par exemple, on peut :

- omettre une modalité pour chaque variable catégorielle
- imposer la nullité de la somme des coefficients des modalités de chaque variable catégorielle.

Modifier la contrainte imposée aux coefficients des modalités d'une variable catégorielle X_i pour la régression dans la zone géographique z revient à effectuer une translation de l'ensemble de ces coefficients par une même quantité, c_{zi} , ne dépendant que de la variable X_i et de la zone z .

Considérons donc $\hat{\beta}_{zij}$ et $\hat{\beta}''_{zij}$ ($z = d$ ou F), les coefficients associés à deux jeux de contraintes distincts.

Alors : $\hat{\beta}''_{zij} = \hat{\beta}_{zij} + c_{zi}$

La partie expliquée et la partie inexpliquée ne dépendent pas de la contrainte :

Avec le second jeu de contraintes, la partie expliquée = $\sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}''_{Fij})$

$$\begin{aligned}
 &= \sum_i \sum_j ((P_{dij} - P_{Fij}) * (\hat{\beta}_{Fij} + c_{Fi})) \\
 &= \sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) + \sum_i \sum_j ((P_{dij} - P_{Fij}) * c_{Fi}) \\
 &= \sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) + \sum_i (c_{Fi} \sum_j P_{dij}) - \sum_i (c_{Fi} \sum_j P_{Fij}) \\
 &= \sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) \\
 &\text{car } \sum_j P_{dij} = \sum_j P_{Fij} = 1
 \end{aligned}$$

La partie expliquée ne dépend donc pas du choix des contraintes appliquées aux coefficients. C'est également le cas pour la partie inexpliquée, puisqu'elle peut s'obtenir comme différence entre les écarts moyens observés et la partie expliquée.

Les contributions des modalités des variables catégorielles à la partie expliquée sont impactées par le choix des contraintes imposées aux coefficients :

La contribution de la modalité j de la variable i à la partie expliquée est égale à : $(P_{dij} - P_{Fij}) * \hat{\beta}''_{Fij} = (P_{dij} - P_{Fij}) * (\hat{\beta}_{Fij} + c_{Fi}) \neq (P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}$

Pour une variable catégorielle X_i , donnée, les contributions des modalités à la partie expliquée ne s'apprécient que relativement les unes par rapport aux autres.

Les contributions des variables catégorielles à la partie expliquée ne dépendent pas des contraintes appliquées aux coefficients :

La contribution de la variable i à la partie expliquée est égale à :

$$\begin{aligned}
 & \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) \\
 &= \sum_j ((P_{dij} - P_{Fij}) * (\hat{\beta}_{Fij} + c_{Fi})) \\
 &= \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) + c_{Fi} \sum_j P_{dij} - c_{Fi} \sum_j P_{Fij} \\
 &= \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) \\
 &\text{car } \sum_j P_{dij} = \sum_j P_{Fij} = 1
 \end{aligned}$$

De la même façon, on peut montrer que les contributions à la partie inexpliquée des variables catégorielles, mais aussi de chacune de leurs modalités, sont impactées par le choix des contraintes (cf. Maillard S., Boutchenik B. : « Méthodes économétriques de décomposition des inégalités – de la théorie à la pratique », Document de travail, Insee, 2018).

4.3. Comparaison avec un modèle structurel-résiduel économétrique

4.3.1. Dans le cas où on utilise uniquement des variables structurelles catégorielles

Une analyse de la variance sur un ensemble de variables catégorielles donne les mêmes résultats que ceux d'une régression sur ces variables catégorielles dichotomisées (c'est-à-dire transformées en indicatrices d'appartenance à chaque catégorie).

Ainsi si on prend la France métropolitaine comme zone de référence pour tous les départements et si l'on impose la contrainte que $\forall i, \sum_j (P_{Fij} * \hat{\beta}_{Fij}) = 0$, alors les $\hat{\beta}_{Fij}$ de la décomposition de Oaxaca-Blinder sont égaux aux $\hat{\beta}'_{ij}$ du modèle structurel-résiduel économétrique.

$$\text{On a donc } \hat{\beta}_{Fij} = \hat{\beta}'_{ij} = (R_{Fij} - R_{E..})$$

La partie expliquée de la décomposition de Oaxaca-Blinder est égale à :

$$\sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) = \sum_i \sum_j (P_{dij} * \hat{\beta}_{Fij}) - \sum_i \sum_j (P_{Fij} * \hat{\beta}_{Fij})$$

Le second élément $\sum_i \sum_j (P_{Fij} * \hat{\beta}_{Fij})$ peut ainsi s'écrire $\sum_i \sum_j (P_{Fij} * (R_{Fij} - R_{E..}))$

$$= \sum_i \sum_j (P_{Fij} * R_{Fij}) - \sum_i \sum_j (P_{Fij} * R_{E..})$$

$$= \sum_i \sum_j (P_{Fij} * R_{Fij}) - R_{E..} * \sum_i \sum_j P_{Fij}$$

$$= 0$$

car :

$$\sum_i \sum_j (P_{Fij} * R_{Fij}) = \sum_i R_{E..} = K * R_{E..} \quad (K \text{ étant le nombre de variables catégorielles})$$

et

$$R_{E..} * \sum_i \sum_j P_{Fij} = K * R_{E..} \quad \text{car } \sum_j P_{Fij} = 1 \text{ pour tout } i$$

Ainsi, sous la contrainte $\forall i, \sum_j (P_{Fij} * \hat{\beta}_{Fij}) = 0$,

la partie expliquée de la décomposition de Oaxaca-Blinder vaut :

$$\sum_i \sum_j ((P_{dij} - P_{Fij}) * \hat{\beta}_{Fij}) = \sum_i \sum_j (P_{dij} * \hat{\beta}_{Fij}) = \sum_i \sum_j (P_{dij} * \hat{\beta}'_{ij})$$

La partie expliquée de la décomposition de Oaxaca-Blinder est bien égale à l'effet structurel du modèle structurel-résiduel économétrique. Comme elle ne dépend pas des contraintes d'identification imposées aux coefficients, elle lui est égale, quel que soit le jeu de contraintes choisi.

4.3.2. Intérêts comparés de la décomposition de Oaxaca-Blinder et du modèle structurel-résiduel

L'intérêt apparent du modèle structurel-résiduel économétrique, par rapport au modèle de décomposition de Oaxaca-Blinder, est de fournir en une unique analyse de la variance les parties expliquées et inexpliquées de l'ensemble des départements.

Un premier intérêt de la décomposition de Oaxaca-Blinder, par rapport au modèle structurel-résiduel, est de permettre l'utilisation de variables quantitatives : il recourt à des régressions économétriques, et non à une analyse de la variance.

Un autre intérêt de la décomposition de Oaxaca-Blinder est de fournir la contribution de chaque modalité de chaque variable de structure (et donc la contribution de chaque variable de structure) à la partie inexpliquée de l'écart de moyennes. Ce résultat est particulièrement intéressant dans le cadre de notre étude, qui se concentre sur les parties inexpliquées des écarts de revenus d'activité. Il donne en effet des éléments pour comprendre pourquoi la partie inexpliquée est forte pour certains départements, en mettant en évidence les modalités des variables de structure dont l'effet diffère sensiblement entre le département et la France métropolitaine. Ces différences d'effet peuvent résulter de la non-prise en compte de variables observables ou de l'effet de variables inobservables.

La lourdeur apparente de la décomposition de Oaxaca-Blinder (nécessité de réaliser 192 régressions pour obtenir des résultats pour les 96 départements) résulte de la possibilité qu'offre cette méthode de décomposer la partie inexpliquée selon les contributions des variables de structure (figure 15). Cette décomposition dépend cependant du choix des modalités de référence.

En fait la partie inexpliquée totale peut se calculer par différence entre l'écart observé et la partie expliquée.

Or le calcul de cette dernière ne nécessite que les estimations des $\hat{\beta}_{Fi}$, et pas celles des $\hat{\beta}_{di}$.

- Si la condition de « groupes mutuellement exclusifs » est respectée, alors on a seulement besoin de réaliser 96 régressions : une régression par zone de référence « France métropolitaine hors département d ». Ces 96 régressions permettent de calculer les parties expliquées et les parties inexpliquées totales, et de décomposer les parties expliquées.
- Si on pose l'hypothèse que le poids d'un département dans l'ensemble de la France est relativement faible, on peut faire le choix de ne pas exclure de la population de référence les non-salariés du département. Dans ce cas, une seule régression est nécessaire si seule la partie inexpliquée totale nous intéresse. Dans ce dernier cas, on obtient les mêmes résultats que ceux que l'on aurait en mettant en œuvre un modèle structurel-résiduel économétrique.

Figure 15 : Nombre de régressions nécessaires selon les hypothèses posées

Zone de référence	Décomposition de la partie inexpliquée	
	Oui	Non
France métropolitaine hors département d	192	96
France métropolitaine	97	1

5. Tableau de synthèse des différentes méthodes de contrôle des effets de structure

On résume les avantages et inconvénients des différentes méthodes de contrôle des effets de structure testées, dans le cadre de la comparaison d'une sous-zone géographique avec une zone englobante. La sous-zone analysée est donc un élément d'une partition de la zone géographique.

	Modèle structurel-résiduel descriptif	Modèle structurel-résiduel économétrique	Modèle structurel-géographique économétrique	Modèle de décomposition de Oaxaca-Blinder⁸
Nombre de variables utilisables	Une seule (qui peut être le croisement de plusieurs variables)	Plusieurs	Plusieurs	Plusieurs
Type de variables	Catégorielles (qualitatives ou quantitatives discrétisées)	Catégorielles (qualitatives ou quantitatives discrétisées)	Catégorielles (qualitatives ou quantitatives discrétisées)	Catégorielles ou quantitatives
Modélisations économétriques mises en œuvre	X	Analyse de la variance	Analyse de la variance	Régression
Nom donné à la partie de l'écart expliquée par le modèle	Effet structurel	Effet structurel	Effet structurel	partie expliquée
Nom donné à la partie de l'écart non expliquée par le modèle	Effet résiduel	Effet résiduel	Effet géographique	partie inexpliquée
Décomposition de la partie expliquée, selon les contributions des modalités des variables catégorielles	OUI	OUI	OUI	OUI
Décomposition de la partie non expliquée, selon les contributions des modalités des variables catégorielles	OUI	NON	NON	OUI
Nombre de sous-zones traitées par le modèle	Une seule	Autant de sous-zones qu'en comprend la partition de la zone englobante	Autant de sous-zones qu'en comprend la partition de la zone englobante	- Avec décomposition de la partie inexpliquée : une seule sous-zone - Sans décomposition de la partie inexpliquée : autant de sous-zones qu'en comprend la partition de la zone englobante (si l'on utilise la même zone de référence pour toutes les sous-zones)

⁸ Pour la décomposition de Oaxaca-Blinder, on fait l'hypothèse que la partition est suffisamment fine pour que le fait de prendre comme zone de référence la zone englobante ou celle-ci moins la zone analysée ne modifie quasiment pas les résultats. Sinon, la zone de référence change pour la modélisation de chaque sous-zone de la partition.

6. Application des différents modèles aux écarts de revenus d'activité non salariée entre départements

Les différentes méthodes de contrôle des effets de structure (modèle structurel-résiduel économétrique, modèle de décomposition de Oaxaca-Blinder, modèle structurel-géographique) ont été appliquées aux moyennes par département des logarithmes du revenu d'activité des non-salariés.

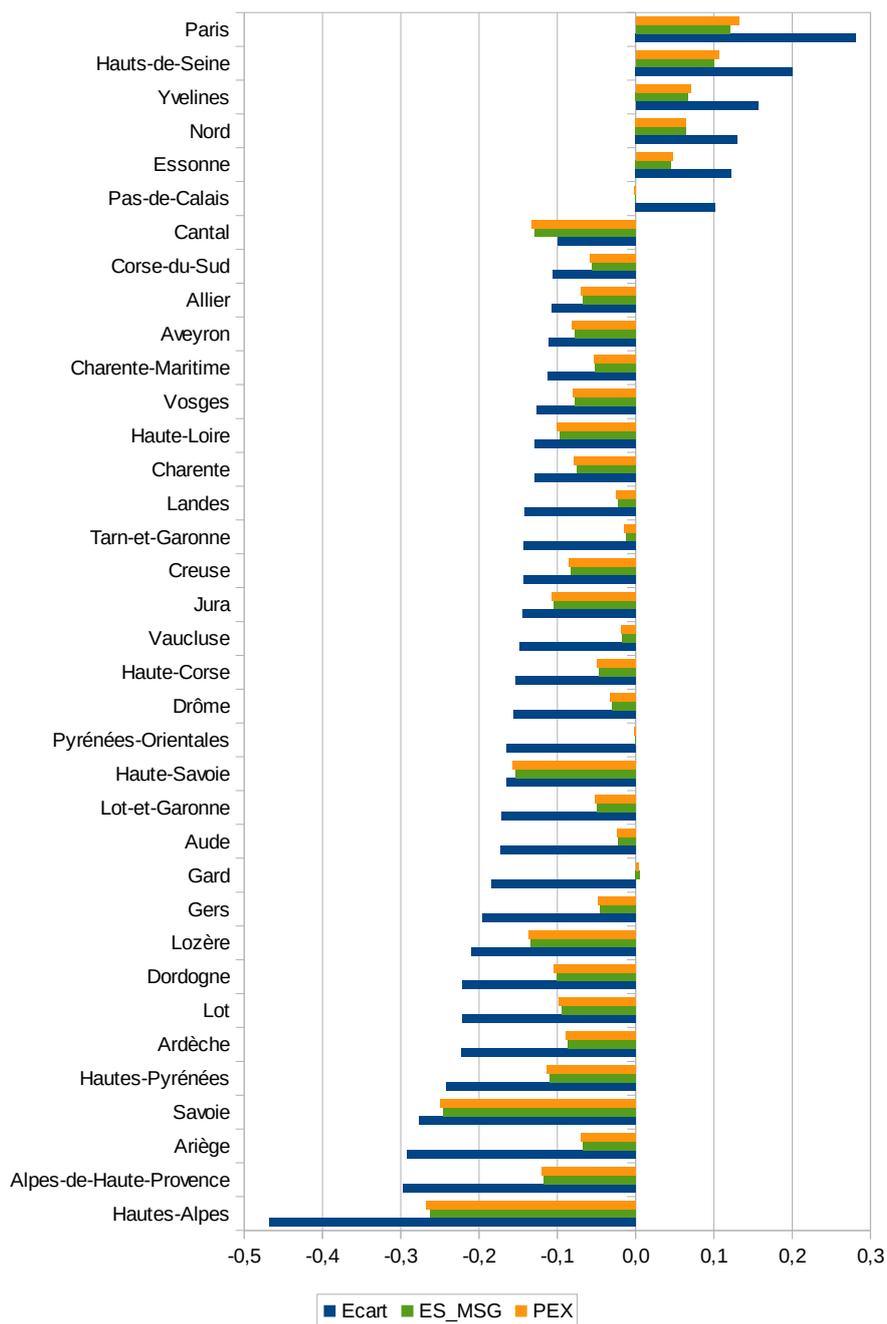
On étudie les effets de la structure par secteur d'activité X la catégorie du non-salarié, ainsi que ceux associés aux autres caractéristiques du non-salarié et de son activité (dont la saisonnalité). L'analyse porte sur les non-salariés classiques de France métropolitaine, actifs fin 2014 et ayant déclaré un revenu positif en 2014.

La mise en œuvre d'un modèle structurel-résiduel économétrique et d'un modèle de décomposition de Oaxaca-Blinder donnent bien des résultats identiques⁹. En revanche, comme attendu, les parties expliquées obtenues à partir de ces deux méthodes diffèrent légèrement de celles issues d'un modèle structurel-géographique (figure 16). Les parties expliquées issues du modèle de décomposition de Oaxaca-Blinder sont légèrement supérieures, en valeur absolue, à celles issues du modèle structurel-géographique.

⁹En prenant France métropolitaine comme référence pour tous les départements dans la décomposition de Oaxaca-Blinder.

Figure 16 : Comparaison des parties expliquées par les variables de structure, selon le modèle mis en œuvre ...

... a) à partir du secteur d'activité X catégorie du non-salarié



Note de lecture : Dans les Hautes-Alpes la moyenne des logarithmes du revenu est inférieure de 0,47 à la moyenne nationale (qui est de 7,61). Les caractéristiques « secteur d'activité X catégorie juridique » contribuent pour 0,27 à cet écart.

On ne représente (en ordonnée) que les départements pour lesquels l'écart initial est supérieur à 0,1 en valeur absolue.

Écart = écart des moyennes des logarithmes du revenu d'activité, entre le département et la France métropolitaine

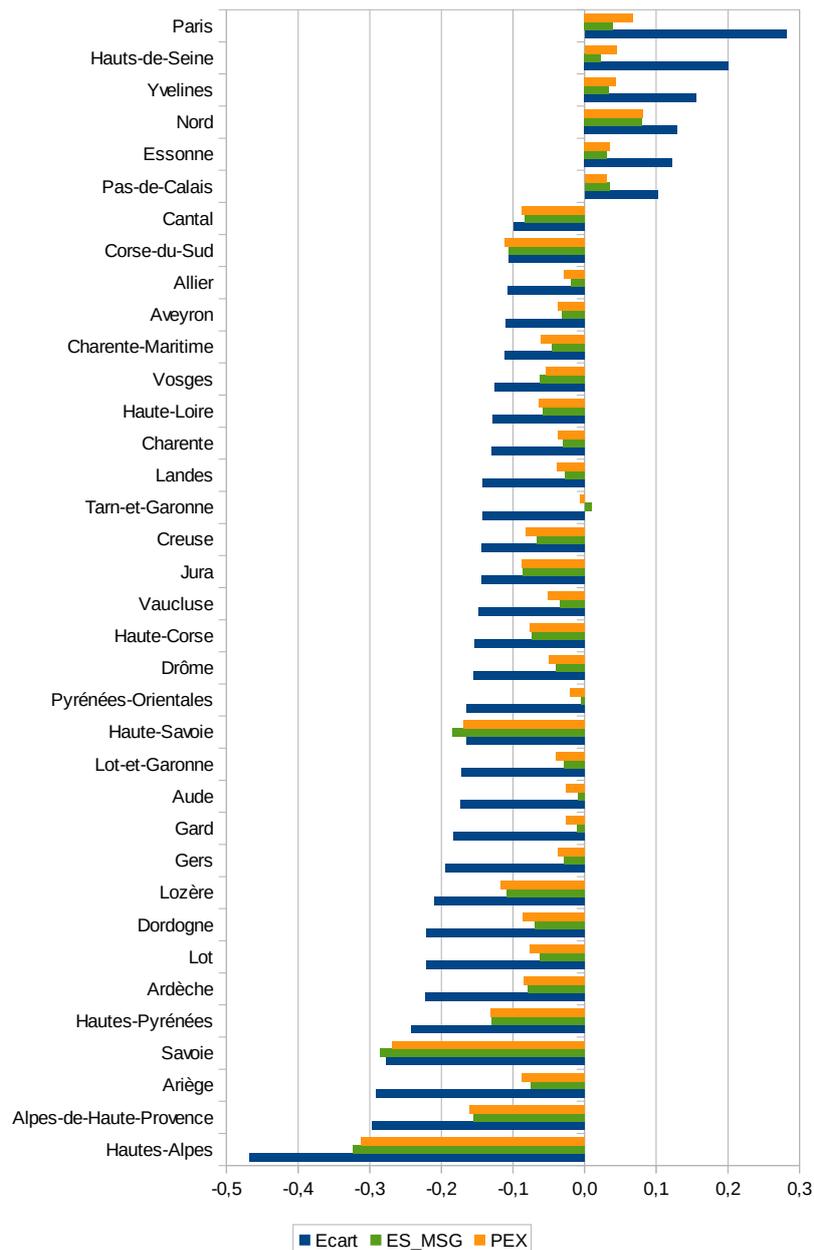
PEX = partie expliquée de la décomposition de Oaxaca-Blinder

ES_MSG = effet structurel du modèle structurel-géographique économétrique.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

... b) à partir du secteur d'activité X catégorie du non-salarié et des autres caractéristiques du non-salarié et de son activité (y compris la saisonnalité)



Note de lecture : Dans les Hautes-Alpes la moyenne des logarithmes du revenu est inférieure de 0,47 à la moyenne nationale (qui est de 7,61). Les caractéristiques « secteur d'activité X catégorie juridique » ainsi que les autres caractéristiques du non-salarié et de son activité contribuent pour 0,31 à cet écart.

On ne représente (en ordonnée) que les départements pour lesquels l'écart initial est supérieur à 0,1 en valeur absolue.

Écart = écart des moyennes des logarithmes du revenu d'activité, entre le département et la France métropolitaine

PEX = partie expliquée de la décomposition de Oaxaca-Blinder

ES_MSG = effet structurel du modèle structurel-géographique économétrique.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxes d'office, hors auto-entrepreneurs).

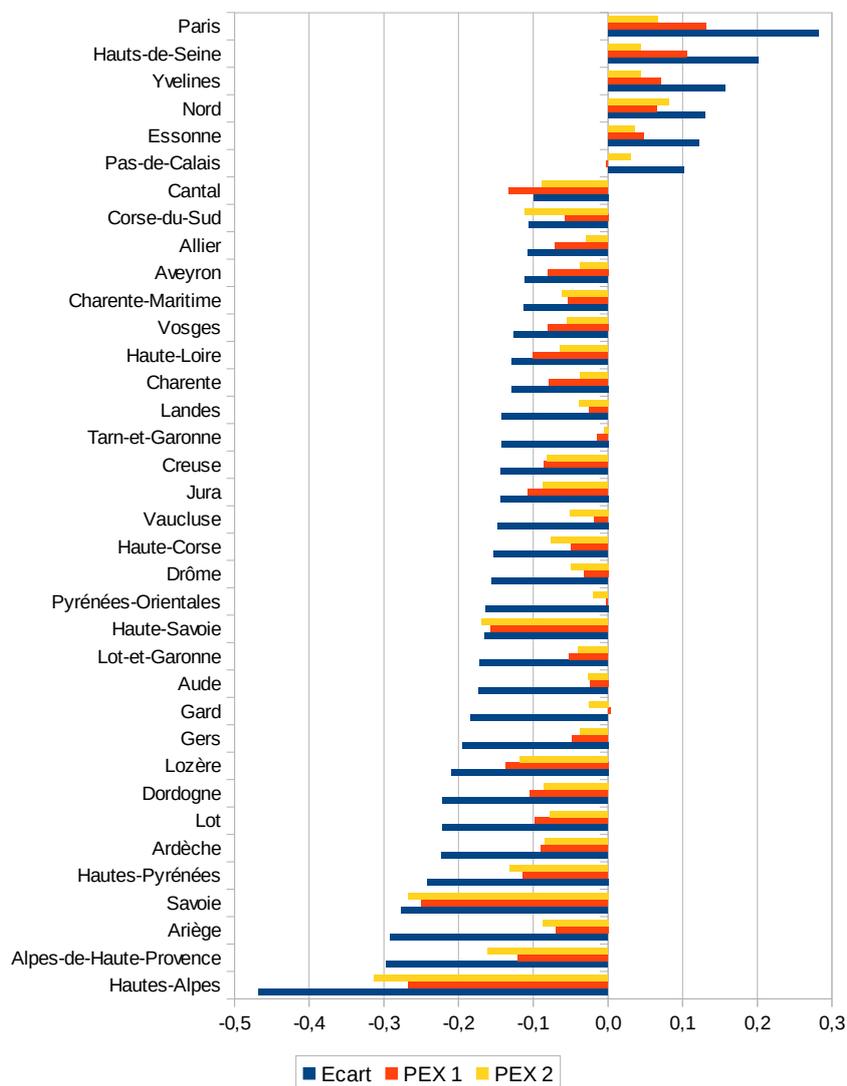
Source : Insee, base non-salariés 2014.

Dans le modèle de régression du logarithme du revenu, intégrer les autres caractéristiques du non-salarié et de son activité (dont la saisonnalité), en plus du croisement secteur d'activité X catégorie du non-salarié, fait augmenter sensiblement le R² ajusté pour la France métropolitaine : il passe de 0,151 à 0,287. Le R² ajusté progresse d'ailleurs pour l'ensemble des régressions par département, le gain variant de + 0,105 à + 0,179.

Bien que les R^2 ajusté des régressions augmentent fortement, la partie expliquée du modèle de décomposition de Oaxaca-Blinder ne s'améliore pas pour tous les départements (figure 17). Ajouter des variables de structure sur les caractéristiques du non-salarié et de son activité semble plutôt dégrader la partie expliquée pour les départements franciliens, où le revenu moyen est largement supérieur à celui observé sur l'ensemble de la France métropolitaine (écarts de revenus positifs). En revanche la partie expliquée s'améliore pour des départements ayant un écart fortement négatif (c'est-à-dire ceux où le revenu moyen est le plus faible), en particulier pour les départements alpins.

Même si la partie expliquée se dégrade pour certains départements, l'indicateur de gain relatif de la variance des écarts, introduit en partie 1.3, passe de - 46 % à - 57 % en ajoutant les autres caractéristiques des non-salariés et de leur activité à la variable de contrôle initiale (secteur d'activité X catégorie juridique) (cf. Annexe 3, modèles M1 et M3).

Figure 17 : Impact de l'ajout des caractéristiques du non-salarié et de son activité sur la partie expliquée par le modèle de décomposition de Oaxaca-Blinder



Note de lecture : Dans les Hautes-Alpes la moyenne des logarithmes du revenu est inférieure de 0,47 à la moyenne nationale (qui est de 7,61). Les caractéristiques « secteur d'activité X catégorie juridique » contribuent pour 0,27 à cet écart. Rajouter les autres caractéristiques du non-salarié et de son activité fait progresser la partie expliquée de 0,04.

On ne représente que les départements pour lesquels l'écart initial est supérieur à 0,1 en valeur absolue.

Écart = écart des moyennes des logarithmes du revenu d'activité, entre le département et la France métropolitaine

PEX1 = partie expliquée de la décomposition de Oaxaca-Blinder sur le croisement secteur d'activité X catégorie du non-salarié.

PEX2 = partie expliquée de la décomposition de Oaxaca-Blinder sur le croisement secteur d'activité X catégorie du non-salarié et les autres caractéristiques du non-salarié et de son activité (y compris la saisonnalité).

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Enfin, l'un des intérêts d'un modèle de décomposition de Oaxaca-Blinder réside dans le fait de pouvoir utiliser des variables de structure quantitatives. Ainsi certaines variables d'environnement économique qui ont été prises en compte dans la version la plus aboutie du modèle n'ont pas été discrétisées et ont été gardées sous leur forme quantitative (revenu moyen ou part des 65 ans ou plus dans le bassin de vie du non-salarié), ce qui permet de réduire le nombre de coefficients à estimer. L'ajout des variables décrivant l'environnement économique dans lequel s'exerce l'activité permet de réduire encore plus fortement la variance des écarts, le gain relatif en termes de variance passant de $- 57 \%$ à $- 72,5 \%$ (cf. Annexe 3, modèles M3 et M9).

7. La transformation logarithmique des revenus

Pour une meilleure spécification de la forme fonctionnelle, on choisit pour variable d'intérêt le logarithme du revenu d'activité à la place du revenu d'activité.

Le passage au logarithme du revenu nécessite de vérifier les hypothèses suivantes :

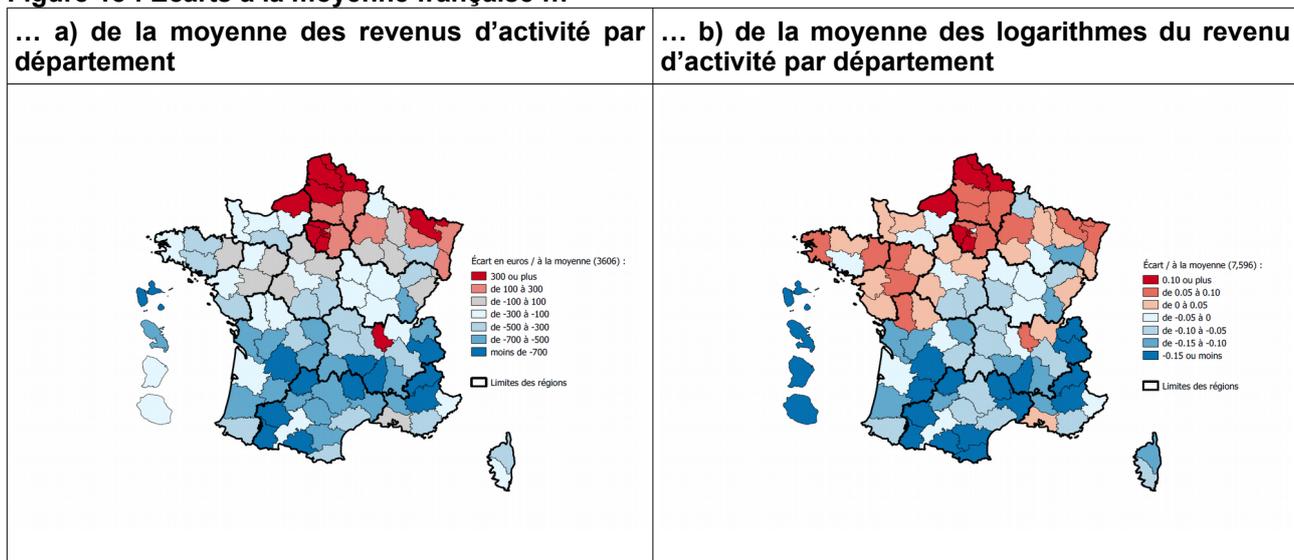
- La corrélation spatiale des moyennes départementales des logarithmes du revenu est proche de celle des moyennes départementales des revenus d'activité.
- Les effets de structure calculés à partir des logarithmes du revenu sont proches de ceux calculés à partir des revenus d'activité.

7.1. Impact du passage au logarithme sur la corrélation spatiale de la variable d'intérêt

La corrélation spatiale des moyennes départementales des logarithmes du revenu d'activité semble être proche de celle des moyennes départementales des revenus d'activité (figure 18). Toutefois, classer les départements en fonction de la moyenne des logarithmes du revenu ne revient pas tout à fait au même que de les classer en fonction de la moyenne des revenus : le passage au logarithme impacte davantage les hauts revenus que les plus faibles. Les départements dont le classement est le plus sensible au passage au logarithme sont globalement les mêmes que ceux qui ont été mis en évidence lors de l'analyse de l'impact du recours au revenu médian plutôt qu'au revenu moyen : la distribution de leurs revenus d'activité diffère le plus de celle observée pour la France (figure 19). Ainsi, le classement de certains départements se dégrade : c'est le cas des DOM, notamment la Guyane (département 9C) et la Réunion (département 9D), mais aussi de la Corse (départements 2A, 2B) et de la Seine-saint-Denis (département 93). À l'inverse, d'autres départements ont un meilleur classement selon la moyenne des logarithmes du revenu que selon la moyenne des revenus d'activité : par exemple, les départements de Bretagne (départements 22, 29, 35 et 56) et une partie de ceux des Pays-de-la-Loire (départements 53, 49 et 85).

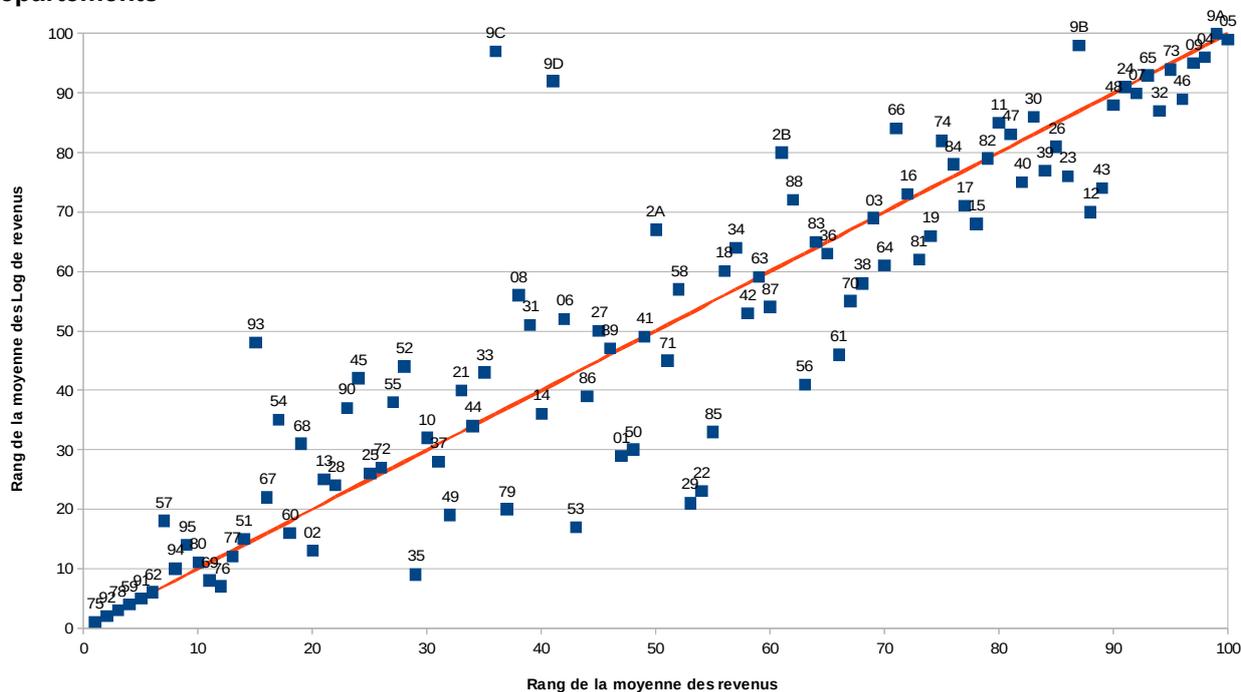
Comme pour le passage au revenu médian, le passage au logarithme impacte donc légèrement la distribution spatiale des revenus. Les changements de classement mis en évidence ci-dessus se retrouvent dans la cartographie des écarts. Notamment, la plupart des départements des régions de Bretagne et des Pays-de-Loire ont des écarts positifs plus élevés, en comparaison avec ceux des autres départements, quand la variable d'intérêt est la moyenne des logarithmes de revenus ; de même, ceux des DOM ou de la Corse, négatifs, sont également accentués par rapport à ceux des autres départements.

Figure 18 : Écarts à la moyenne française ...



Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors auto-entrepreneurs, hors taxés d'office).
 Source : Insee, base non-salariés 2014.

Figure 19 : Impact du passage aux logarithmes du revenu d'activité, sur le classement des départements



Note de lecture : au-dessus de la droite en rouge, le classement des départements selon la moyenne des logarithmes du revenu d'activité est moins élevé que selon la moyenne des revenus d'activité. Au-dessous, il est meilleur selon la moyenne des logarithmes du revenu que selon la moyenne des revenus.

Par exemple, la Guyane (département 9C) occupe le 36^e rang dans le classement selon la moyenne des revenus et le 97^e rang selon la moyenne des logarithmes du revenu.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

7.2. Impact du passage au logarithme sur l'estimation des effets de structure

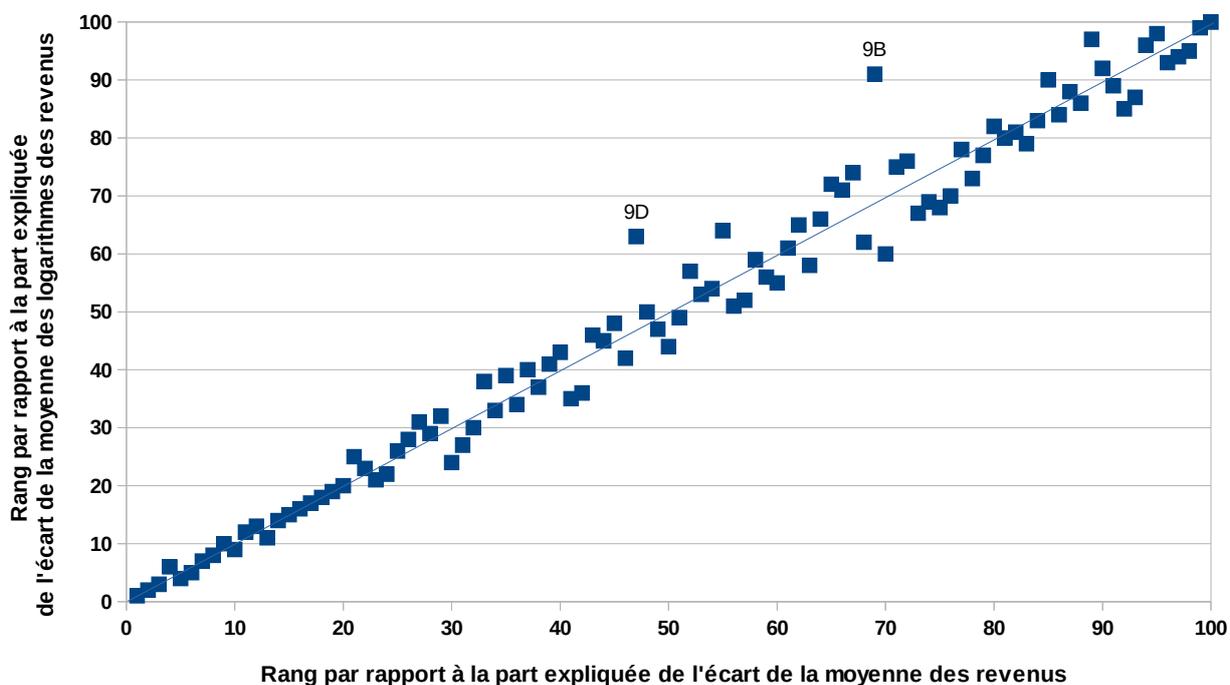
Afin de vérifier la similarité des effets de structure sur la moyenne des revenus d'activité et sur la moyenne des logarithmes du revenu, on compare les résultats de la décomposition de Oaxaca-Blinder appliquée à chacun de ces indicateurs d'intérêt.

Le classement des départements selon la partie expliquée du modèle de décomposition est assez similaire quand on utilise les revenus ou les logarithmes des revenus (figure 20).

En revanche, les classements selon les parties inexpliquées des décompositions sont plus éloignés (figure 21). Certains départements perdent des places (départements 9B, 9C, 9D, 93, 54, 57, 68, 08...) et d'autres en gagnent (départements 22, 29, 35...). On retrouve les départements dont le classement est le plus impacté par le passage au logarithme du revenu d'activité.

Le passage au logarithme se répercute presque complètement sur le classement selon la partie inexpliquée. Or l'indicateur de gain relatif de la variance des écarts utilisé dans l'étude est construit à partir des parties inexpliquées par les effets de structure. Il faut donc regarder avec une certaine prudence les résultats obtenus pour les départements de l'ouest de la France (Bretagne et Pays-de-la-Loire).

Figure 20 : Comparaison des classements des départements selon les parties expliquées des décompositions de Oaxaca-Blinder, suivant la variable d'intérêt utilisée



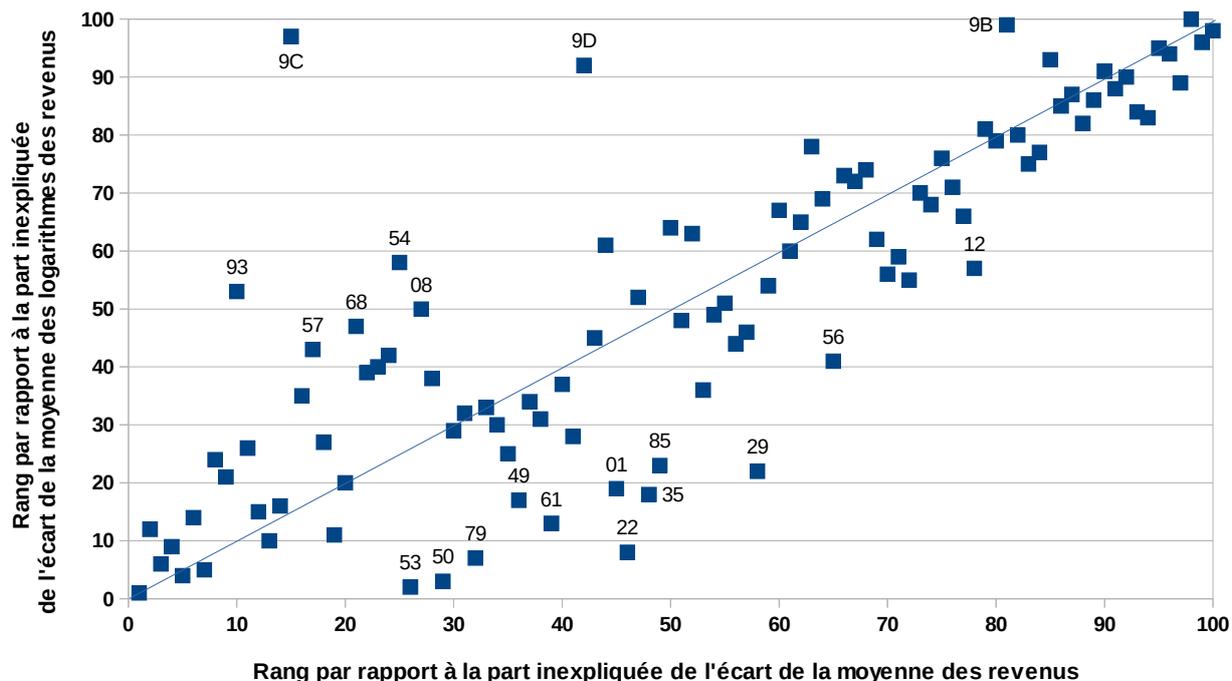
Note de lecture : au-dessus de la droite en bleu, le classement des départements selon la partie expliquée de la décomposition de Oaxaca-Blinder se dégrade si on applique le modèle à la moyenne des logarithmes du revenu plutôt qu'à la moyenne des revenus. Au-dessous, il s'améliore.

Par exemple, la Réunion (département 9D) occupe le 47^e rang dans le classement selon la moyenne des revenus et le 63^e rang selon la moyenne des logarithmes du revenu.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Figure 21 : Comparaison des classements des départements selon les parties inexpliquées des décompositions de Oaxaca-Blinder, suivant la variable d'intérêt utilisée



Note de lecture : au-dessus de la droite en bleu, le classement des départements selon la partie inexpliquée de la décomposition de Oaxaca-Blinder se dégrade si on applique le modèle à la moyenne des logarithmes du revenu plutôt qu'à la moyenne des revenus. Au-dessous, il s'améliore.

Par exemple, la Réunion (département 9D) occupe le 42^e rang dans le classement pour le modèle appliqué à la moyenne des revenus et le 92^e rang pour le modèle appliqué à la moyenne des logarithmes du revenu.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

8. Les améliorations et extensions possibles

8.1. Les limites de l'indice de saisonnalité utilisé

L'indice de saisonnalité national est construit à un niveau fin de la nomenclature d'activités française – NAF rév. 2 (cf. 1.3.4. Une prise en compte partielle du volume de travail). Pour certaines activités, dont les effectifs salariés sont faibles¹⁰, l'indice est donc peu fiable. Toutefois, les activités rares chez les salariés sont généralement aussi rares chez les non-salariés. Leur impact sur l'estimation de la contribution de la saisonnalité aux parties expliquée et inexpliquée du modèle de Oaxaca-Blinder reste donc mineur.

Une autre limite réside dans l'hypothèse que l'on pose initialement : la saisonnalité d'une activité est identique chez les salariés et chez les non-salariés. Pour certaines activités, cette hypothèse n'est pas vérifiée : c'est le cas en particulier d'activités majoritairement exercées par des non-salariés, notamment dans le domaine de la santé. Par exemple, la sous-classe des « activités des professionnels de la rééducation, de l'appareillage et des pédicures-podologues » (APE = 8690E), regroupe près de 100 000 non-salariés classiques à revenu non nul, ce qui la place au deuxième rang des activités exercées par des non-salariés (après celle des infirmiers et des sages-femmes). À l'inverse, cette sous-classe est très peu représentée chez les salariés : elle apparaît au-delà du 500^e rang des activités à la fois présentes parmi les salariés et parmi les non-salariés. Son indice de saisonnalité, calculé à partir des DADS, la définit comme une « activité de forte saisonnalité ». Mais il est possible que les emplois salariés de cette activité correspondent pour une bonne part aux remplacements nécessités par les départs en congés des non-salariés de cette même activité. Ainsi, alors qu'on pourrait penser que cette activité est modérément saisonnière chez les non-salariés, l'indice associé, construit à partir d'un emploi salarié saisonnier, la caractérise comme une activité de forte saisonnalité.

L'invalidité de l'hypothèse de départ pour certaines activités pourrait expliquer l'importance de la part de non-salariés ayant une activité fortement saisonnière dans le secteur de la santé : les « activités des professionnels de la rééducation, de l'appareillage et des pédicures-podologues » (APE = 8690E) et les « activités des infirmiers et des sages-femmes » (APE = 8690D), regroupant la moitié des effectifs du secteur, relèvent de cette classe de saisonnalité (figure 22).

Figure 22 : Répartition des non-salariés classiques à revenu non nul selon le degré de saisonnalité de leur activité

Secteur \ Saisonnalité	Faible à Moyenne	Assez forte	Forte	Très forte	Total
Industrie (hors artisanat commercial)	79,7	17,1	2,3	0,8	100,0
Construction	90,2	9,5	0,3	0,0	100,0
Commerce et artisanat commercial	35,3	25,3	32,9	6,5	100,0
Transports	72,6	25,6	0,0	1,8	100,0
Services aux entreprises et services mixtes	61,6	27,3	7,9	3,1	100,0
Services aux particuliers (hors santé)	24,8	6,2	12,4	56,6	100,0
Santé humaine et action sociale	42,3	2,4	55,3	0,0	100,0
Secteur indéterminé	100,0	0,0	0,0	0,0	100,0
Ensemble	53,3	13,7	21,8	11,1	100,0

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

¹⁰ Du fait de l'utilisation du fichier postes DADS au 12^e, notamment

8.2. La prise en compte de la pluriactivité

Pour mieux appréhender l'effet du volume de travail sur les disparités de revenus d'activité entre départements, on peut intégrer un indicateur de pluriactivité dans la régression, en complément du degré de saisonnalité de l'activité exercée par le non-salarié. Cet indicateur est corrélé au volume de travail exercé par le non-salarié, dans la mesure où un non-salarié cumulant son activité avec un ou des postes salariés aura moins de temps à consacrer à son activité non salariée ou ne l'exercera qu'une partie de l'année.

Deux indicateurs de pluriactivité sont construits et testés :

- **un indicateur dichotomique**
Cet indicateur vaut 1 si le non-salarié a occupé une activité salariée au cours de l'année, 0 sinon.
- **un indicateur détaillé**
Cet indicateur distingue, parmi les non-salariés ayant occupé une activité salariée au cours de l'année, les principalement non-salariés (revenu d'activité non salariée \geq salaires perçus) des principalement salariés (revenu d'activité non salariée $<$ salaires perçus). Une troisième modalité correspond aux non-salariés n'ayant pas occupé d'activité salariée dans l'année.

L'intégration de l'un ou l'autre de ces indicateurs de pluriactivité dans la régression du logarithme du revenu d'activité sur les caractéristiques du non-salarié et de son activité (y compris le caractère saisonnier) permet de faire progresser le R^2 ajusté : il passe de 0,288 à 0,297 si on ajoute l'indicateur dichotomique, à 0,338 si on ajoute l'indicateur détaillé (cf. Annexe 3, modèles M3, M4 et M5). L'effet de l'indicateur dichotomique sur les écarts de revenu d'activité est significativement négatif, comme attendu. En revanche, si on ajoute l'indicateur détaillé, on trouve un effet négatif de la modalité « pluriactif principalement non-salarié » et un effet légèrement positif de la modalité « principalement salarié ». Ce dernier peut sembler contre-intuitif, car dans ce cas le volume de travail non-salarié est souvent limité et donc le revenu non-salarié plutôt faible.

On choisit d'intégrer l'indicateur détaillé dans le modèle de décomposition de Oaxaca-Blinder¹¹. Alors que la prise en compte de la pluriactivité fait progresser de façon non négligeable le R^2 ajusté, elle n'entraîne pas une réduction plus forte de la variance des écarts : -55 % en ajoutant l'indicateur détaillé, contre -57 % sans prise en compte (cf. Annexe 3, modèles M3 et M5).

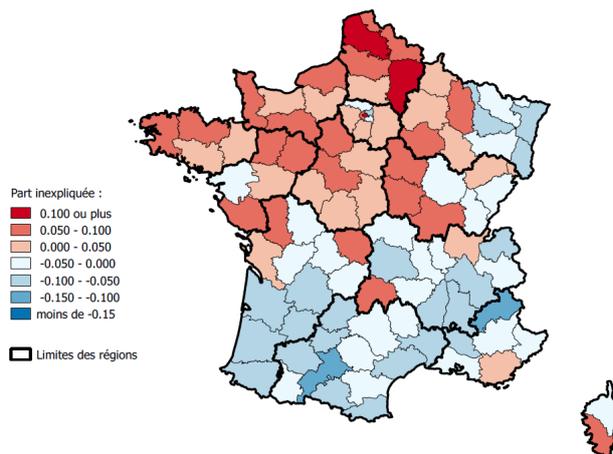
Si l'intégration d'un facteur de saisonnalité a permis d'améliorer la décomposition des écarts de revenus, ce n'est pas le cas pour l'indicateur de pluriactivité en tant que facteur complémentaire de la saisonnalité. Conserver l'indicateur de pluriactivité dans la décomposition des écarts de revenus peut cependant être intéressant pour montrer qu'il contribue peu aux écarts quand il vient en complément de la saisonnalité.

8.3. Effet de la concurrence ou effet de sélection

Même après avoir pris en compte les variables décrivant l'environnement économique dans lequel s'exerce l'activité du non-salarié, la partie des écarts de revenu restant inexpliquée est toujours plus élevée dans les départements du nord que dans ceux du sud (figure 23). Par ailleurs, on peut observer que les départements pour lesquels les écarts avec la moyenne française sont les plus forts sont souvent ceux pour lesquels la proportion de non-salariés dans l'emploi est faible (figure 24).

¹¹ Les variables de structure utilisées sont le croisement secteur d'activité X catégorie du non-salarié et les autres caractéristiques individuelles du non-salarié et de son activité, y compris la saisonnalité

Figure 23 : Écarts des moyennes de logarithmes du revenu non expliqués par la structure par secteur d'activité X catégorie du non-salarié, les autres caractéristiques individuelles du non-salarié et de son activité (y compris la saisonnalité) et l'environnement économique dans lequel s'exerce l'activité

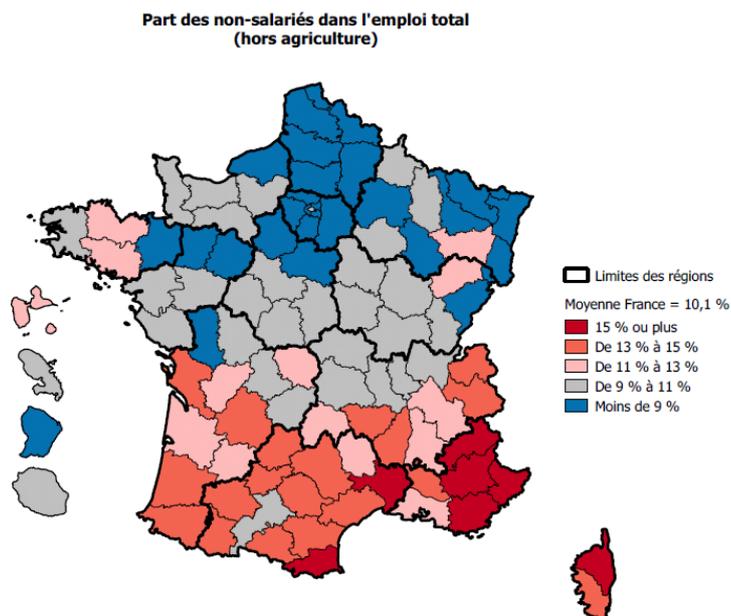


Lecture : Une fois prises en compte les différences de structure, le logarithme du revenu d'activité dans le Pas-de-Calais reste fortement supérieur à la moyenne française.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Figure 24 : Part des non-salariés dans l'emploi en 2014



Lecture : dans le département du Pas-de-Calais, la part des non-salariés dans l'emploi total hors agriculture est inférieure à 9 %.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 et emploi total, hors agriculture.

Source : Insee, base non-salariés 2014 et estimations d'emploi annuelles 2014.

Deux éléments explicatifs peuvent être mis en avant :

- **l'existence d'un effet concurrence**
Une concurrence plus faible, plutôt corrélée à une part de non-salariés dans l'emploi plus faible, peut se traduire par des revenus d'activité moyens plus élevés.
- **l'existence d'un effet de sélection**
On n'a contrôlé que les caractéristiques observées des non-salariés. D'autres caractéristiques inobservées ou non observables peuvent influencer sur le profil des personnes qui optent pour le non-salariat : ce dernier peut notamment différer entre les départements du nord et ceux du sud. Par exemple, la motivation, l'investissement personnel, le réseau du non-salarié, la culture du non salariat, les options alternatives d'emploi, l'attractivité locale, les facilités pour accéder à ce statut... peuvent influencer sur le profil des non-salariés d'un département.

8.3.1. Essai de prise en compte de l'effet concurrence

On peut approcher la concurrence sur le marché des biens et services produits au sein d'un secteur d'activité par le biais de la densité d'établissements par département (nombre d'établissements d'un secteur rapporté au nombre d'habitants)¹². Une des limites d'un tel indicateur est qu'il ne prend pas en compte la taille des établissements. Par ailleurs, si on construit cet indicateur par secteur agrégé, on regroupe alors des activités aux profils concurrentiels différents.

On peut construire un indicateur amélioré, prenant en compte la taille des établissements et en descendant à un niveau fin d'activité. Pour chaque croisement « bassin de vie X secteur d'activité détaillé », on calcule l'indicateur suivant :

Densité de l'activité dans le bassin de vie / Densité de l'activité en France métropolitaine

Où la densité de l'activité dans une zone =

(Emploi salarié¹³ et non-salarié de l'activité dans la zone / population de la zone)

Le numérateur de la densité est un proxy de l'offre, le dénominateur un proxy de la demande. On compare, pour une activité donnée, la densité de l'emploi de l'activité dans le bassin de vie à celle observée en France.

Un indicateur élevé (donc supérieur à 1) signifie que l'offre mise en regard de la demande est plus importante dans le bassin de vie qu'en France métropolitaine, et que la concurrence y est donc sans doute plus forte.

L'indicateur de concurrence obtenu est discrétisé en cinq classes :

- « inférieur à 0,5 » ;
- « de 0,5 à moins de 0,8 » ;
- « de 0,8 à 1,2 » ;
- « de 1,2 à moins de 1,5 » ;
- « 1,5 ou plus ».

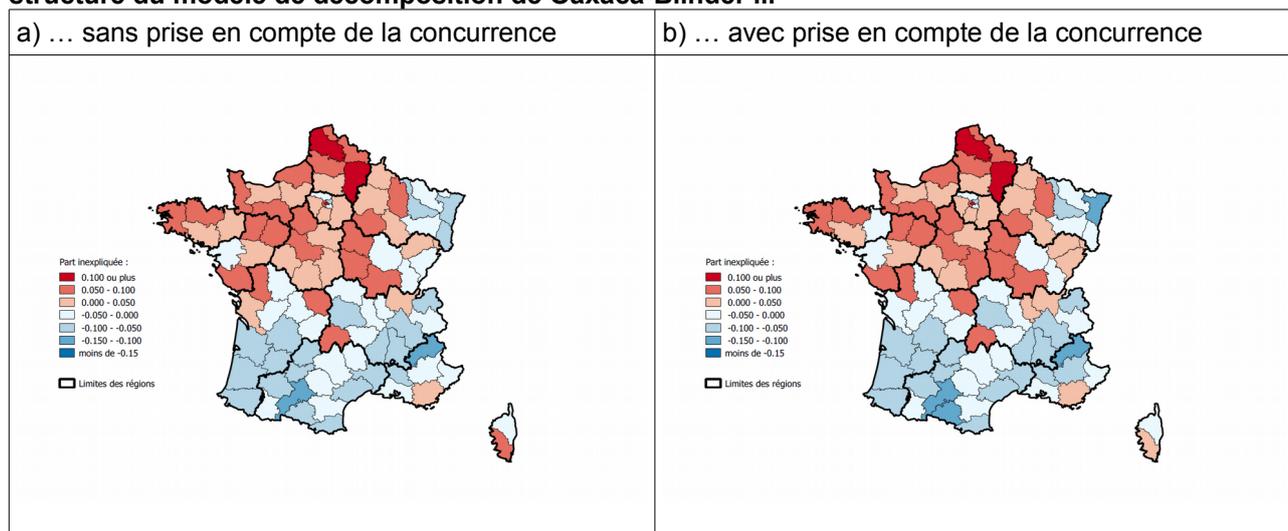
L'intégration de celui-ci dans la régression du logarithme du revenu d'activité sur les caractéristiques du non-salarié et de son activité (y compris le caractère saisonnier) et les variables décrivant l'environnement économique de l'activité ne fait pas progresser le R² ajusté (qui passe de 0,2900 à 0,2905). L'intégration de ce facteur concurrence dans le modèle de décomposition de Oaxaca-Blinder¹⁴ ne permet pas non plus d'améliorer la variance des écarts restant inexpliqués : elle est 72,5 % plus faible que celle des écarts de revenus initiaux, contre 74 % sans ce facteur concurrence. Par ailleurs, les écarts de revenus restant inexpliqués sont toujours plus élevés dans les départements du nord que dans ceux du sud (figure 25).

¹² Dans le cadre de l'Insee Première n° 1185 sur « De l'influence du territoire sur les revenus des non-salariés », cet indicateur a été utilisé dans la régression mise en œuvre pour expliquer le niveau de revenu du non-salarié.

¹³ Source : Insee, CLAP 2014.

¹⁴ Les variables de structure utilisées sont le croisement secteur d'activité X catégorie du non-salarié, les autres caractéristiques individuelles du non-salarié et de son activité, y compris la saisonnalité, les variables décrivant l'environnement économique de l'activité, et l'indicateur de concurrence.

Figure 25 : Écarts des moyennes de logarithmes du revenu non expliqués par les variables de structure du modèle de décomposition de Oaxaca-Blinder ...



Note : Les variables de structure utilisées sont le croisement secteur d'activité X catégorie du non-salarié, les autres caractéristiques individuelles du non-salarié et de son activité, y compris la saisonnalité, les variables décrivant l'environnement économique de l'activité, avec ou sans l'indicateur de concurrence.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

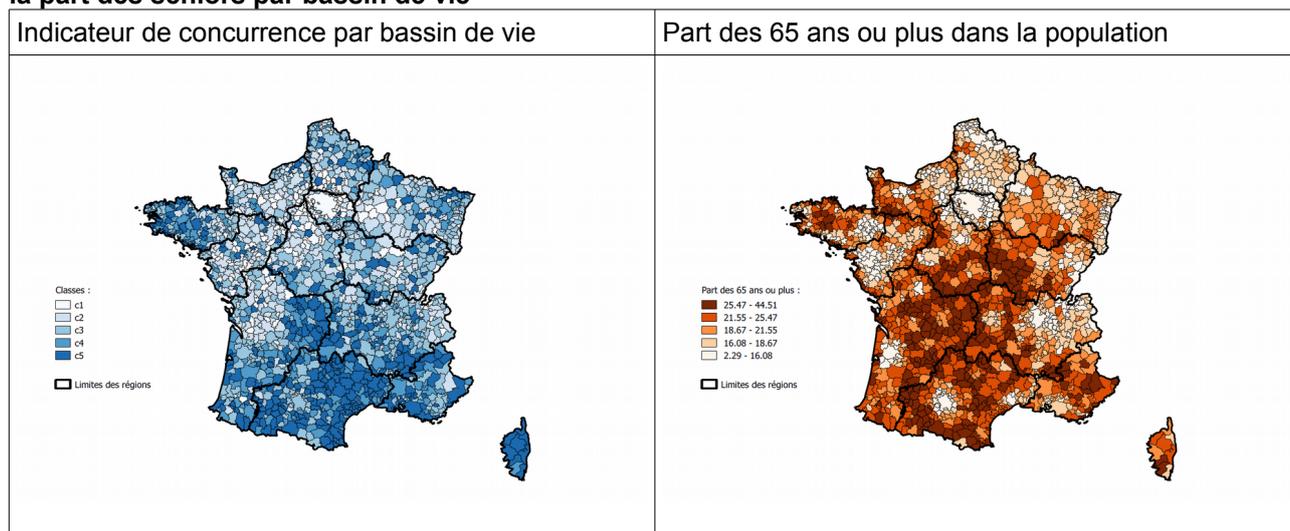
Source : Insee, base non-salariés 2014.

Une importante limite de l'indicateur de concurrence construit réside dans le fait que c'est la population totale qui joue le rôle de proxy de la demande. Pour certaines activités, ce proxy ne convient pas : par exemple, pour les activités des infirmiers et sages-femmes (APE = 8690D), la concurrence peut être très différente dans deux bassins de vie ayant la même densité d'activité, mais dont la pyramide des âges de la population diffère sensiblement (les personnes âgées ayant plus recours à ce type de services que le reste de la population) (figure 26). Pour les activités de restauration traditionnelle (APE = 5610A), la concurrence ne sera pas la même dans deux bassins de vie de même densité d'activité, mais de touristicité différente. En effet la demande locale dépend plus de la population présente que de la population résidente et l'écart entre ces deux populations peut être important dans les zones touristiques (figure 27).

L'indicateur de concurrence construit ici est globalement plus adapté aux activités relevant de la sphère présente (activités mises en œuvre localement pour la production de biens et de services visant la satisfaction des besoins de personnes présentes dans la zone, qu'elles soient résidentes ou touristes) qu'à celles de la sphère productive (dont fait notamment partie l'industrie). Pour ces dernières, ni la demande, ni l'offre concurrentielle ne sont forcément locales. Ainsi, pour certaines activités industrielles, un sous-traitant peut être implanté loin de son donneur d'ordres. De même, la concurrence que subit un non-salarié peut résulter de l'existence de producteurs géographiquement très éloignés (voire à l'étranger). Dans ces cas, cet indicateur, qui s'appuie sur une définition locale de l'offre et de la demande n'est pas très pertinent.

On pourrait améliorer la mesure de la concurrence en recourant à des données comptables, telles que le taux de marge.

Figure 26 : Indicateur de concurrence des activités des infirmiers et sages-femmes, mis en regard de la part des seniors par bassin de vie

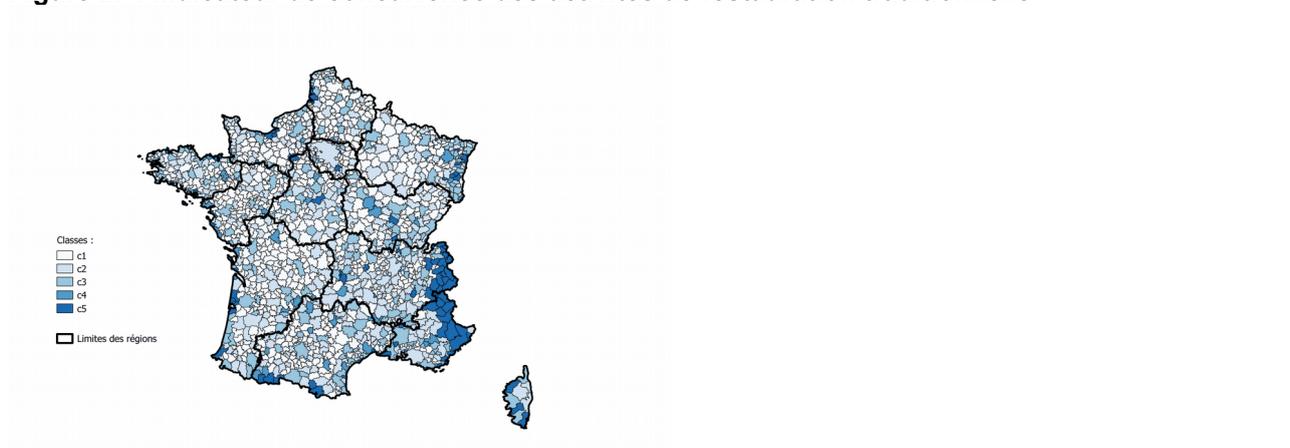


Note : Indicateur de concurrence construit à partir du nombre d'emplois non-salariés et salariés au 31 décembre 2014 , dans les activités des infirmiers et sages-femmes. Pour les non-salariés, personnes exerçant une activité non salariée au 31 décembre 2014, hors agriculture. Les classes de l'indicateur de concurrence vont d'une faible concurrence (c1) à une forte concurrence (c5).

Champ : France métropolitaine.

Sources : Insee, base non-salariée 2014, CLAP 2014, recensement de la population 2014

Figure 27 : Indicateur de concurrence des activités de restauration traditionnelle



Note : Indicateur de concurrence construit à partir du nombre d'emplois non-salariés et salariés au 31 décembre 2014 , dans les activités de restauration traditionnelle. Pour les non-salariés, personnes exerçant une activité non salariée au 31 décembre 2014, hors agriculture. Les classes de l'indicateur de concurrence vont d'une faible concurrence (c1) à une forte concurrence (c5).

Champ : France métropolitaine.

Sources : Insee, base non-salariée 2014, CLAP 2014, recensement de la population 2014.

8.3.2. Mesurer l'effet de sélection

La cartographie des écarts de revenus après prise en compte de la concurrence montre toujours des écarts de revenus restant inexpliqués plus élevés dans les départements du nord que dans ceux du sud (figure 25 b).

Ces écarts peuvent résulter d'effets de sélection. Leur existence pourrait être démontrée, à partir d'autres sources que les bases non-salariés. En effet, il est nécessaire d'utiliser une source regroupant à la fois des non-salariés et des salariés, et pour lesquels on disposerait d'informations identiques. On pourrait mobiliser pour ce faire le panel tous actifs (agrégation du panel tous salariés et du panel non-salariés) quand il sera disponible.

Conclusion

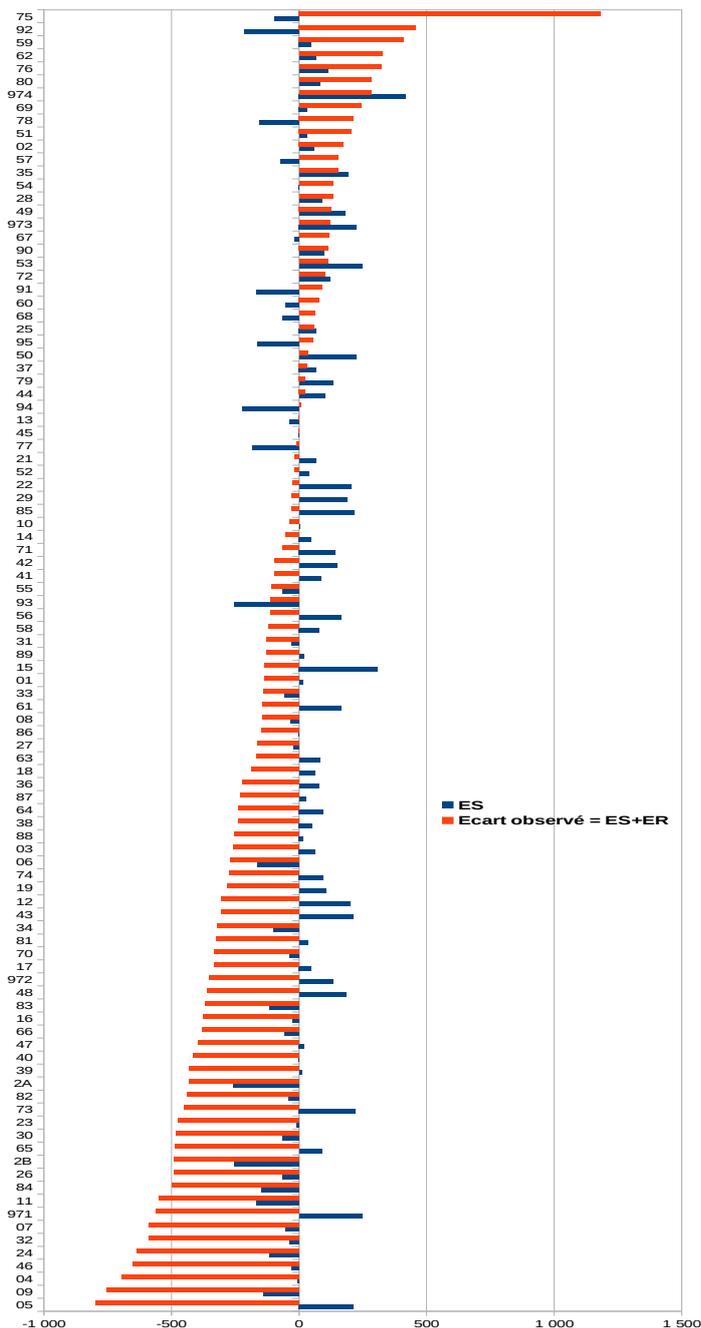
De premiers éléments d'explication aux disparités de revenus d'activité entre départements ont été mis en évidence, à l'aide des modèles de décomposition de Oaxaca-Blinder. Cette analyse pourra cependant être approfondie, en améliorant notamment l'estimation du volume de travail, l'indicateur de saisonnalité et la touristicité des bassins de vie, en utilisant un meilleur proxy de l'effet concurrence, en mesurant le rôle des prix de l'immobilier sur la région d'implantation du non-salarié, ... La principale piste d'approfondissement reste la mise en évidence d'effets de sélection : le profil des personnes optant pour le non-salariat peut en effet différer entre les départements du nord et ceux du sud.

Bibliographie

- Bertran C. : « Le revenu d'activité des non-salariés : plus élevé en moyenne dans les départements du nord que dans ceux du sud », Insee Première n° 1672, novembre 2017.
- Kubrak C. : « Structurel, résiduel, géographique : principe et mise en œuvre des approches comptable et économétrique », Document de travail, Insee, n° H 2018/01, 2018.
- Évain F. : « De l'influence du territoire sur les revenus des non-salariés », Insee Première n° 1185, avril 2008.
- Salembier L., Théron G. : « Revenus d'activité des non-salariés en 2014. Hausse pour les indépendants « classiques », baisse pour les auto-entrepreneurs », Insee Première n° 1627, décembre 2016.
- Fortin N., Lemieux T., Firpo S. : « Decomposition methods in economics », Working Paper 16045, National Bureau of Economic Research, 2010.
- Jann B. : « The Blinder–Oaxaca decomposition for linear regression models », The Stata Journal 8, Number 4, pp. 453–479, 2008.
- Maillard S., Boutchenik B. : « Méthodes économétriques de décomposition des inégalités – de la théorie à la pratique », Document de travail, Insee, 2018

Annexes

Annexe 1 : résultats du modèle structurel-résiduel relatif à l'impact de la structure par grandes catégories de non-salariés sur les écarts de revenus d'activité entre départements



Note : l'écart de revenu d'activité est égal à la différence entre le revenu moyen du département et le revenu moyen en France. Il se décompose en un effet structurel ES (effet de la structure par grandes catégories de non-salariés sur les écarts de revenus observés) et un effet résiduel ER (écart restant inexpliqué par cette structure). Les grandes catégories de non-salariés sont les non-salariés classiques à revenu non nul, les non-salariés classiques à revenu nul et les auto-entrepreneurs.

Variance de l'écart initial = 90 651 ; variance de l'écart résiduel = 98 631. Indicateur de gain relatif de variance = +8,8 %
 Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

Annexe 2 : Indicateurs de revenu par département, pour l'ensemble des non-salariés

Département	Revenu Mensuel moyen (en €)	Revenu Mensuel médian (en €)	Différence Entre revenus mensuels moyens et médians (en €)	Part des Auto-entrepreneurs (en %)	Part des Non-salariés classiques À revenu nul (en %)	Part des Non-salariés classiques À revenu non nul (en %)
75	3 687	1 343	2 344	24,9	11,1	64,0
92	2 962	1 139	1 823	30,2	9,7	60,1
973	2 625	903	1 722	11,5	14,1	74,4
974	2 786	1 208	1 578	6,2	13,4	80,4
59	2 915	1 370	1 546	25,7	6,5	67,8
57	2 658	1 140	1 518	29,6	6,3	64,1
62	2 829	1 333	1 496	26,1	5,6	68,3
93	2 393	904	1 488	33,1	8,3	58,6
95	2 558	1 071	1 486	31,8	6,9	61,3
78	2 717	1 244	1 474	30,2	8,1	61,7
94	2 510	1 046	1 464	31,8	8,5	59,7
91	2 593	1 150	1 443	31,3	7,5	61,3
80	2 787	1 358	1 429	25,1	6,0	68,9
972	2 152	737	1 415	9,1	18,8	72,1
54	2 638	1 224	1 414	27,4	6,3	66,3
68	2 567	1 158	1 408	28,2	7,4	64,4
77	2 496	1 088	1 408	32,1	7,2	60,7
51	2 710	1 304	1 407	26,1	6,6	67,3
76	2 825	1 426	1 399	23,8	6,3	69,9
69	2 747	1 358	1 389	25,9	6,7	67,4
60	2 585	1 201	1 383	28,4	6,9	64,7
67	2 621	1 243	1 378	26,4	7,7	66,0
02	2 679	1 321	1 358	26,1	5,8	68,1
13	2 504	1 165	1 339	27,7	7,0	65,2
08	2 359	1 045	1 314	27,8	6,8	65,4
90	2 619	1 312	1 307	24,2	6,4	69,4
55	2 395	1 101	1 294	30,2	5,5	64,3
971	1 945	653	1 292	9,3	15,5	75,2
45	2 503	1 226	1 277	27,8	5,9	66,3
28	2 636	1 360	1 277	24,8	6,1	69,1
52	2 486	1 232	1 254	27,1	5,5	67,4
2A	2 073	837	1 236	32,1	9,2	58,6
72	2 605	1 375	1 230	24,6	5,4	70,0
10	2 468	1 238	1 230	28,1	5,6	66,4
2B	2 014	798	1 217	33,4	8,0	58,6
06	2 237	1 029	1 207	31,3	7,4	61,3
33	2 364	1 163	1 201	29,1	6,3	64,6
25	2 563	1 363	1 200	25,0	6,5	68,5
37	2 536	1 338	1 198	25,8	5,8	68,4
31	2 379	1 183	1 196	28,4	6,2	65,4
21	2 488	1 298	1 189	24,9	6,6	68,4
58	2 382	1 200	1 182	24,6	6,5	68,9
34	2 184	1 011	1 173	30,4	6,3	63,2
14	2 450	1 287	1 163	25,9	6,2	67,8
88	2 251	1 092	1 159	26,7	6,4	66,9
18	2 317	1 159	1 159	25,2	6,5	68,3
89	2 377	1 219	1 157	27,0	6,1	66,9
66	2 127	977	1 150	29,1	6,3	64,6
27	2 340	1 191	1 149	28,0	6,4	65,6
49	2 631	1 501	1 130	23,3	4,9	71,8

Département	Revenu Mensuel moyen (en €)	Revenu Mensuel médian (en €)	Différence Entre revenus mensuels moyens et médians (en €)	Part des Auto-entrepreneurs (en %)	Part des Non-salariés classiques À revenu nul (en %)	Part des Non-salariés classiques À revenu non nul (en %)
86	2 359	1 229	1 130	28,2	5,5	66,3
41	2 408	1 280	1 129	25,1	5,9	69,0
44	2 528	1 401	1 127	24,9	5,7	69,4
83	2 137	1 014	1 123	30,6	6,6	62,7
36	2 284	1 168	1 116	25,2	6,0	68,7
63	2 338	1 230	1 109	24,9	6,2	68,9
87	2 274	1 166	1 108	26,3	6,5	67,2
42	2 408	1 313	1 095	23,6	5,5	70,9
03	2 245	1 153	1 092	25,1	6,5	68,3
16	2 129	1 038	1 091	28,7	5,8	65,5
79	2 529	1 438	1 091	24,3	5,3	70,4
71	2 438	1 351	1 087	23,9	5,4	70,7
35	2 658	1 585	1 073	21,8	5,9	72,3
53	2 618	1 546	1 071	20,7	5,4	73,9
11	1 955	892	1 063	32,4	6,5	61,1
50	2 538	1 482	1 056	21,9	5,0	73,2
84	2 006	954	1 052	31,7	6,6	61,7
30	2 020	976	1 044	29,5	6,2	64,2
47	2 108	1 067	1 041	26,9	6,1	66,9
01	2 367	1 326	1 041	27,8	5,5	66,7
74	2 230	1 190	1 040	23,0	7,5	69,4
61	2 363	1 329	1 034	23,6	5,2	71,3
70	2 172	1 144	1 028	29,5	5,5	65,1
38	2 266	1 243	1 023	25,7	6,3	67,9
82	2 065	1 045	1 021	29,5	5,6	64,9
64	2 266	1 261	1 005	24,1	6,6	69,4
17	2 172	1 167	1 005	25,7	6,4	67,8
19	2 221	1 217	1 004	24,6	5,8	69,6
81	2 181	1 184	997	27,6	5,2	67,3
48	2 143	1 165	978	22,0	6,0	72,1
56	2 392	1 420	972	22,4	6,1	71,5
26	2 013	1 041	972	29,8	5,9	64,3
15	2 370	1 398	972	18,6	5,5	75,9
23	2 029	1 059	970	28,2	5,8	66,0
85	2 474	1 507	967	21,5	5,5	73,0
40	2 088	1 125	963	27,0	6,6	66,4
24	1 869	912	957	30,9	6,4	62,7
39	2 074	1 119	955	26,8	6,5	66,7
65	2 019	1 072	946	25,1	5,8	69,1
29	2 476	1 540	936	22,5	5,4	72,1
22	2 478	1 546	933	21,9	5,5	72,6
07	1 917	986	931	29,1	6,2	64,7
09	1 751	823	928	33,1	5,1	61,8
43	2 198	1 300	899	21,5	5,7	72,8
04	1 811	917	894	27,4	6,4	66,2
12	2 200	1 306	894	22,2	5,3	72,5
32	1 915	1 026	889	29,2	5,7	65,1
73	2 054	1 166	888	19,8	6,9	73,3
46	1 853	1 000	853	29,0	5,7	65,3
05	1 707	971	736	19,8	7,1	73,1
France	2 504	1 200	1 304	26,6	7,0	66,4

Note : Les parts sont calculées en tenant compte de la durée d'affiliation. Les départements sont triés selon la différence revenu médian – revenu moyen.

Champ : France, personnes exerçant une activité non salariée au 31 décembre 2014 (hors agriculture, hors taxés d'office).

Source : Insee, base non-salariés 2014.

Annexe 3 : Principaux modèles de décomposition de Oaxaca-Blinder mis en œuvre

Facteurs pris en compte	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Modèle 6	Modèle 7	Modèle 8	Modèle 9
"secteur d'activité X catégorie"	X	X	X	X	X	X	X	X	X
Autres caractéristiques du non-salarié et de son activité		X	X	X	X	X	X	X	X
Saisonnalité			X	X	X	X	X	X	X
Pluriactivité dichotomique (oui/non)				X					
Pluriactivité NS, PNS, PS					X				
Variables d'environnement								X	X
Concurrence									X
Effets fixes départementaux						X			
Effets fixes par bassin de vie							X		
R ² ajusté	0,1513	0,2731	0,2876	0,2971	0,3376	0,2933	0,2934	0,2900	0,2905
Indicateur de gain relatif de la variance des écarts	-46 %	-52 %	-57 %	non calculé	-55 %	-----	-----	-74 %	-72,5%

Note : Par souci de comparabilité, tous les modèles ont été appliqués à la même population : celle des non-salariés sans données manquantes pour les variables du modèle 9.

La régression du logarithme du revenu sur le secteur d'activité X la catégorie du non-salarié donne un R² ajusté de 0,1513. Le modèle de décomposition de Oaxaca-Blinder correspondant permet de réduire de 46 % la variance des écarts de revenus entre départements observés.

Champ : France métropolitaine, personnes exerçant une activité non salariée au 31 décembre 2014 et ayant déclaré un revenu positif (hors agriculture, hors taxés d'office, hors auto-entrepreneurs).

Source : Insee, base non-salariés 2014.

Le gain de R² ajusté du modèle de régression du logarithme du revenu et le gain relatif de variance des écarts de revenus entre départements ne sont pas parfaitement corrélés : à une évolution modérée de R² ajusté, peut correspondre une forte diminution de la variance des écarts de revenus (et inversement).

Annexe 4 : codes des régions

Code région	Région
01	Guadeloupe
02	Martinique
03	Guyane
04	La Réunion
11	Île-de-France
24	Centre-Val de Loire
27	Bourgogne-Franche-Comté
28	Normandie
32	Hauts-de-France
44	Grand-Est
52	Pays de la Loire
53	Bretagne
75	Nouvelle Aquitaine
76	Occitanie
84	Auvergne-Rhône-Alpes
93	Provence-Alpes-Côte d'Azur
94	Corse

Annexe 5 : codes des départements

Code département	Département
01	Ain
02	Aisne
03	Allier
04	Alpes-de-Haute-Provence
05	Hautes-Alpes
06	Alpes-Maritimes
07	Ardèche
08	Ardennes
09	Ariège
10	Aube
11	Aude
12	Aveyron
13	Bouches-du-Rhône
14	Calvados
15	Cantal
16	Charente
17	Charente-Maritime
18	Cher
19	Corrèze
21	Côte-d'Or
22	Côtes-d'Armor
23	Creuse
24	Dordogne
25	Doubs
26	Drôme
27	Eure
28	Eure-et-Loir
29	Finistère
2A	Corse-du-Sud
2B	Haute-Corse
30	Gard
31	Haute-Garonne
32	Gers
33	Gironde
34	Hérault
35	Ille-et-Vilaine
36	Indre
37	Indre-et-Loire
38	Isère
39	Jura
40	Landes
41	Loir-et-Cher
42	Loire
43	Haute-Loire
44	Loire-Atlantique
45	Loiret
46	Lot
47	Lot-et-Garonne
48	Lozère
49	Maine-et-Loire

Code département	Département
50	Manche
51	Marne
52	Haute-Marne
53	Mayenne
54	Meurthe-et-Moselle
55	Meuse
56	Morbihan
57	Moselle
58	Nièvre
59	Nord
60	Oise
61	Orne
62	Pas-de-Calais
63	Puy-de-Dôme
64	Pyrénées-Atlantiques
65	Hautes-Pyrénées
66	Pyrénées-Orientales
67	Bas-Rhin
68	Haut-Rhin
69	Rhône
70	Haute-Saône
71	Saône-et-Loire
72	Sarthe
73	Savoie
74	Haute-Savoie
75	Paris
76	Seine-Maritime
77	Seine-et-Marne
78	Yvelines
79	Deux-Sèvres
80	Somme
81	Tarn
82	Tarn-et-Garonne
83	Var
84	Vaucluse
85	Vendée
86	Vienne
87	Haute-Vienne
88	Vosges
89	Yonne
90	Territoire de Belfort
91	Essonne
92	Hauts-de-Seine
93	Seine-Saint-Denis
94	Val-de-Marne
95	Val-d'Oise
9A ou 971	Martinique
9B ou 972	Guadeloupe
9C ou 973	Guyane
9D ou 974	Réunion

Série des documents de travail de la DSDS

- 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