

Disparités spatiales de salaire et externalités de capital humain

Paul Charruau
Anne Épaulard



FRANCE STRATÉGIE
ÉVALUER. ANTICIPER. DÉBATTRE. PROPOSER.

Les documents de travail de France Stratégie présentent les travaux de recherche réalisés par ses experts, seuls ou en collaboration avec des experts extérieurs. L'objet de leur diffusion est de susciter le débat et d'appeler commentaires et critiques. Les documents de cette série sont publiés sous la responsabilité éditoriale du commissaire général. Les opinions et recommandations qui y figurent engagent leurs auteurs et n'ont pas vocation à refléter la position du Gouvernement.

Sommaire

Introduction	3
1. Faits stylisés	4
1.1. Des différences de salaires moyens marquées entre les zones d'emploi malgré une décomposition par catégories socioprofessionnelles	4
1.2. Des répartitions spatiales de l'emploi et de la population très qualifiée également très inégales et plutôt en faveur des grandes villes	6
2. Revue de la littérature	13
2.1. La densité des activités économiques augmente la productivité et les salaires, mais peut conduire à des phénomènes de congestion	13
2.2. Des effets de densité qui proviennent en partie d'effets de composition et du « tri spatial » des travailleurs qualifiés	15
2.3. Des externalités positives de capital humain qui améliorent les salaires au niveau local	16
3. Décomposer les effets d'agglomération : quelle influence des externalités de capital humain ?	18
3.1. Externalités de capital humain et effet de substitution imparfaite entre travailleurs qualifiés et non qualifiés : cadre théorique	19
3.2. Équation de détermination des salaires moyens par CSP au niveau des zones d'emploi	21
3.3. Deux principales sources d'endogénéité	22
3.4. Composition des qualifications au sein des CSP	26
4. Résultats	27
4.1. Des effets importants de la concentration de diplômés révélant des externalités de capital humain	27
4.2. Robustesse	29
Conclusion	30
Références bibliographiques	31
Annexes	
Annexe 1 – Zonage d'étude : les zones d'emploi 2010	33
Annexe 2 – Données	34
Annexe 3 – Tableaux de résultats d'estimations	37
Annexe 4 – Tableau de résultats d'estimations – modèle préféré (complet)	41
Annexe 5 – Robustesse : économétrie spatiale	42

Résumé

Cette étude réexamine la question de savoir si les disparités géographiques de salaire trouvent majoritairement leur explication dans les différences de densité de l'emploi ou dans l'existence d'externalités de capital humain. L'analyse repose sur un modèle d'estimation des déterminants des salaires moyens par catégories socioprofessionnelles (CSP) des 304 zones d'emploi de France métropolitaine en 2012. Une procédure d'estimations en deux temps, avec en deuxième étape une stratégie à variables instrumentales, permet d'isoler les effets de composition et de corriger des biais de causalité inverse. Les résultats confirment la présence de gains d'agglomération issus de la concentration des diplômés du supérieur long. Les primes de productivité et de salaires semblent moins le résultat de la seule densification de l'emploi que d'une part importante des personnes très qualifiées dans la population locale. Selon le modèle préféré, une augmentation d'un point du ratio de diplômés du supérieur long engendre des gains de salaires moyens de 0,6 % pour les CSP cadres, 0,5 % pour les CSP professions intermédiaires, et 0,3 % pour les CSP employés. L'écart interdécile de ce ratio de diplômés, c'est-à-dire la différence entre le dixième des zones les plus concentrées en diplômés et le dixième des zones les moins concentrées, est de 8 points.

Mots clés : salaire, territoire, zone d'emploi, effet d'agglomération, effet de densité, externalité, capital humain

Disparités spatiales de salaire et externalités de capital humain

Paul Charruau¹ et Anne Épaulard²

Introduction³

Les dernières décennies ont été marquées par une tendance croissante à la concentration des activités économiques et de l'emploi au sein de grandes aires urbaines. Cette dynamique s'est globalement accompagnée d'un accroissement des inégalités territoriales, notamment en matière de salaires et de productivité. En 2012 par exemple, le salaire moyen de la zone d'emploi de Paris était supérieur de plus de 80 % à celui de la zone d'emploi de Saint-Flour (Auvergne). Un écart semble se creuser entre les zones très denses et le reste du pays.

De nombreux travaux de recherche se sont donc intéressés à la nature et aux causes de ces inégalités géographiques. S'il existe un certain consensus autour d'une relation positive reliant la densité des activités et des personnes à la productivité et aux salaires, les canaux de transmissions de ces gains d'agglomération font toujours l'objet de débats. Leurs origines relèvent à la fois i) des effets de densité liés à la taille des marchés, avec des économies d'échelle favorables, davantage de débouchés, une meilleure optimisation de la chaîne de sous-traitants et un meilleur appariement sur le marché du travail ; ii) des externalités de capital humain liées au niveau de qualifications de la main-d'œuvre locale et aux effets de diffusion des connaissances et des innovations ; iii) des effets de spécialisation inter et intra-sectorielle ; iv) des différences dans les dotations factorielles locales (caractéristiques naturelles, biens publics locaux, gouvernance et institutions). Quelques contributions montrent également des effets de congestion liés à la densité, qui peuvent annuler les gains d'agglomération.

Un pan de la littérature sur les effets d'agglomération s'est concentré sur le capital humain : les travailleurs d'une zone relativement plus « qualifiée » deviennent eux-mêmes plus productifs, en partie grâce aux externalités de capital humain (rendements externes de l'éducation ; Acemoglu et Angrist, 2000). Le niveau de capital humain d'une zone peut donc, en théorie, améliorer la productivité et les salaires au-delà de l'effet direct de l'éducation sur la productivité individuelle des travailleurs. L'identification des externalités que les diplômés sont susceptibles d'exercer sur l'ensemble des travailleurs pose néanmoins de sérieux problèmes, en raison, d'une part, des

¹ Chargé de mission, département Économie, France Stratégie.

² Conseillère scientifique, France Stratégie, et professeure, université Paris-Dauphine.

³ Ce travail a bénéficié des commentaires de Vincent Aussilloux, Pierre-Philippe Combes, Élise Coudin, Boris Le Hir, Fabrice Lengart, Fanny Mikol, Michaël Orand et Sébastien Roux. Les auteurs remercient aussi Claude Motte pour leur avoir fourni les données de population par communes du recensement de 1806.

effets de composition de la population locale, et d'autre part, des risques avérés de causalité inverse entre salaires et densité d'emploi et de population qualifiée.

L'enjeu est de savoir d'où viennent les inégalités géographiques de productivité et de salaire en France et de mesurer l'importance des externalités de capital humain dans ce cadre. Les études empiriques qui mettent en évidence les gains d'agglomération ne cherchent pas à identifier précisément ce qui ressort d'effets « purs » de la densité d'éventuelles externalités de capital humain. Les tentatives de quantification empirique de ces externalités, en particulier dans le contexte français, sont rares et portent généralement sur l'impact sur le salaire toutes CSP confondues, sans prendre en compte d'éventuels effets de composition des qualifications. Il s'agit donc ici d'identifier, le plus clairement possible, l'ampleur des externalités de capital humain dans les marchés du travail locaux en France pour apprécier leur importance parmi les sources potentielles de gains d'agglomération, et notamment vis-à-vis de la densité. L'analyse économique des territoires, en particulier des effets d'urbanisation, peut s'effectuer selon diverses grilles de lecture. Le choix s'est porté ici sur le découpage Insee des zones d'emploi (voir Annexe 1), échelle adaptée à l'étude des marchés du travail locaux, et donc des différences de productivité et de salaires.

La présente étude fait d'abord le constat des disparités territoriales de salaires, faisant le parallèle avec les différences géographiques de densité d'emploi et de concentration de diplômés du supérieur long. Une stratégie économétrique en deux étapes, avec un modèle d'estimations à variables instrumentales, est ensuite mise en œuvre pour mesurer l'influence de la densité et des externalités de capital humain sur les primes salariales entre zones d'emploi, à catégorie socioprofessionnelle (CSP) donnée.

La première section présente certains faits stylisés concernant les niveaux de salaires moyens par zone d'emploi en 2012. Les statistiques descriptives ne permettent pas vraiment de conclure sur l'origine des écarts de salaires, même si des forces d'agglomération semblent bien exister. La deuxième section est une revue de la littérature sur les effets d'agglomération et leur quantification en France et à l'étranger. Globalement, doubler la taille ou la densité d'une zone entraînerait une augmentation de la productivité et des salaires de l'ordre de 3-8 % (Rosenthal et Strange, 2004 ; Combes et Lafourcade, 2012). Cependant, la prise en compte des effets de capital humain dans les modèles a tendance à diminuer l'ampleur des élasticités de la productivité et des salaires à la densité. Une troisième section détaille les problèmes statistiques inhérents à l'identification des externalités de capital humain et explicite la stratégie économétrique employée. La dernière section rassemble les principaux résultats de l'étude.

1. Faits stylisés

1.1. Des différences de salaires moyens marquées entre les zones d'emploi malgré une décomposition par catégories socioprofessionnelles

En France métropolitaine, les salaires varient beaucoup d'un territoire à un autre. En 2012, le salaire horaire net moyen se situe à 13,9 euros, soit 1,9 fois le Smic. Mais il varie quasiment du simple au double entre la zone où il est le moins élevé (10,4 euros dans la zone d'emploi de Saint-Flour, dans le Cantal) et celle où il est le plus élevé (19,5 euros dans la zone d'emploi de Saint-Quentin-en-Yvelines) (voir Tableau 1 page suivante).

Tableau 1. Niveau et dispersion des salaires en France en 2012 selon les zones d'emploi, le genre et la catégorie socioprofessionnelle

Variables	Salaire moyen France métropolitaine	Moyenne	Écart-type	Médiane	Ratio interdécile	Minimum	Zone d'emploi du minimum	Maximum	Zone d'emploi du maximum
Salaire net horaire moyen (SNHM) en 2012 (€)	13,9	12,3	1,3	12,0	1,2	10,4	Saint-Flour	19,5	Saint-Quentin-en-Yvelines
<i>Femmes</i>	12,4	11,1	1,0	10,8	1,2	9,7	<i>Saint-Flour</i>	17,0	<i>Paris</i>
<i>Hommes</i>	14,9	13,2	1,5	12,8	1,2	10,9	<i>Saint-Flour</i>	22	Saint-Quentin-en-Yvelines
SNHM « Supérieur »	25,0	22,4	1,6	22,2	1,2	19,5	Saint-Flour	29,3	Saint-Quentin-en-Yvelines
<i>Femmes</i>	21,7	19,2	1,3	19,0	1,1	16,8	<i>Mauriac</i>	25,0	<i>Paris</i>
<i>Hommes</i>	26,6	23,9	1,8	23,6	1,2	20,3	<i>Issoudun</i>	32,1	<i>Paris</i>
SNHM « Intermédiaire »	14,4	14,0	0,6	13,8	1,1	12,8	Villefranche-de-Rouergue	16,1	Saint-Quentin-en-Yvelines
<i>Femmes</i>	13,3	12,8	0,7	12,6	1,1	11,6	<i>Les Herbiers</i>	15,0	<i>Paris</i>
<i>Hommes</i>	15,1	14,7	0,8	14,5	1,1	13,3	<i>Villefranche-de-Rouergue</i>	18,0	<i>Chablais</i>
SNHM « Employé »	10,3	10,0	0,4	9,8	1,1	9,2	Le Blanc	11,9	Saint-Quentin-en-Yvelines
<i>Femmes</i>	10,1	9,8	0,5	9,6	1,1	9,0	<i>Brioude</i>	11,8	Saint-Quentin-en-Yvelines
<i>Hommes</i>	10,9	10,6	0,4	10,5	1,1	9,6	<i>Sarlat-la-Canéda</i>	12,6	<i>Rambouillet</i>
SNHM « Ouvrier »	10,9	10,6	0,5	10,6	1,1	9,6	Saint-Flour	12,9	Saint-Quentin-en-Yvelines
<i>Femmes</i>	9,6	9,4	0,4	9,3	1,1	8,5	<i>Sartène - Propriano</i>	10,8	Saint-Quentin-en-Yvelines
<i>Hommes</i>	11,1	10,9	0,6	10,8	1,1	9,7	<i>Saint-Flour</i>	13,2	Saint-Quentin-en-Yvelines

Note : le salaire net horaire moyen en 2012 est en moyenne de 12,3 euros pour les 304 zones d'emploi de France métropolitaine ; le salaire net horaire moyen de France métropolitaine est de 13,9 euros (moyenne pondérée).

On peut diviser le territoire métropolitain en 304 zones d'emplois, définies par l'Insee comme « des espaces géographiques à l'intérieur desquels la plupart des actifs résident et travaillent, et dans lesquels les établissements peuvent trouver l'essentiel de la main-d'œuvre nécessaire pour occuper les emplois offerts » (voir Annexe 1).

En moyenne dans les 304 zones d'emploi de France métropolitaine, le salaire net horaire moyen s'élève à 12,3 euros. La distribution de ces salaires est relativement inégale : les zones d'emploi du 9^e décile présentent des salaires moyens au moins 20 % supérieurs aux zones d'emploi du premier décile (le rapport interdécile est de 1,2). Bien qu'inégale, cette distribution est relativement symétrique, comme le révèle une moyenne proche de la médiane.

En réalité, le salaire moyen de chaque zone recouvre des salaires très hétérogènes selon les activités économiques et les catégories d'emplois concernées. Et le premier facteur expliquant les diversités de niveau de salaire moyen sur le territoire français est un effet de composition, lié aux poids respectifs de ces activités économiques ainsi qu'à la proportion de cadres, de professions intermédiaires, d'employés et d'ouvriers que compte chaque zone d'emploi.

La zone d'emploi de Saint-Quentin-en-Yvelines compte une plus forte proportion de cadres supérieurs que la zone d'emploi de Saint-Flour.

Cependant, des écarts substantiels s'observent encore lorsque l'on compare sur le territoire les salaires d'une catégorie socioprofessionnelle (CSP)⁴ donnée. Par exemple, pour les professions intermédiaires, le salaire net moyen parisien (15,8 euros/heure) est supérieur de 23 % à celui de la zone de Saint-Flour (12,8 euros/heure). La dispersion des salaires est la plus forte pour les salaires des hommes cadres, pour lesquels le salaire moyen est inférieur à 23,5 euros nets par heure dans la moitié des zones d'emploi, tout en dépassant 32 euros nets à Paris. Les hommes ou les femmes cadres sont les mieux rémunérés, en moyenne, dans la zone de Paris, de même que les femmes de la catégorie professions intermédiaires. Les ouvriers (hommes et femmes) sont les mieux rémunérés, en moyenne, dans la zone d'emploi de Saint-Quentin-en-Yvelines, de même que les hommes et les femmes employés.

Les zones où les salaires sont les plus élevés en moyenne correspondent pour beaucoup aux zones d'emploi des grandes métropoles et d'autres grandes villes, telles que Clermont-Ferrand, Orléans, Pau ou Tours – quelle que soit la CSP (voir cartes 1 à 4). On observe en particulier une agglomération de zones d'emploi favorisées sur le plan des salaires autour de Paris et de l'axe Lyon-Grenoble. Ces tendances suggèrent l'existence d'effets d'agglomération.

1.2. Des répartitions spatiales de l'emploi et de la population très qualifiée également très inégales et plutôt en faveur des grandes villes

En 2012, le nombre total d'emplois⁵ est d'environ 25,8 millions dans les 304 zones d'emploi (voir Tableau 2). Leur répartition spatiale est particulièrement inégale : 50 % de l'emploi total est concentré dans 32 zones d'emploi correspondant aux grandes capitales régionales et/ou aux grands pôles d'attractivité économique, plus quatre zones d'Île-de-France⁶ – l'écart entre le nombre moyen d'emplois par zone (85 000) et le nombre médian d'emplois (38 000) montre bien cette concentration des emplois dans un petit nombre de zones.

Les zones d'emploi appartenant au 10 pour cent les moins dotées en emplois comptent plus de 12 fois moins d'emplois que les zones appartenant au dernier décile. Paris est la seule zone à dépasser le million d'emplois avec plus 3,5 millions d'emplois au lieu de travail en 2012.

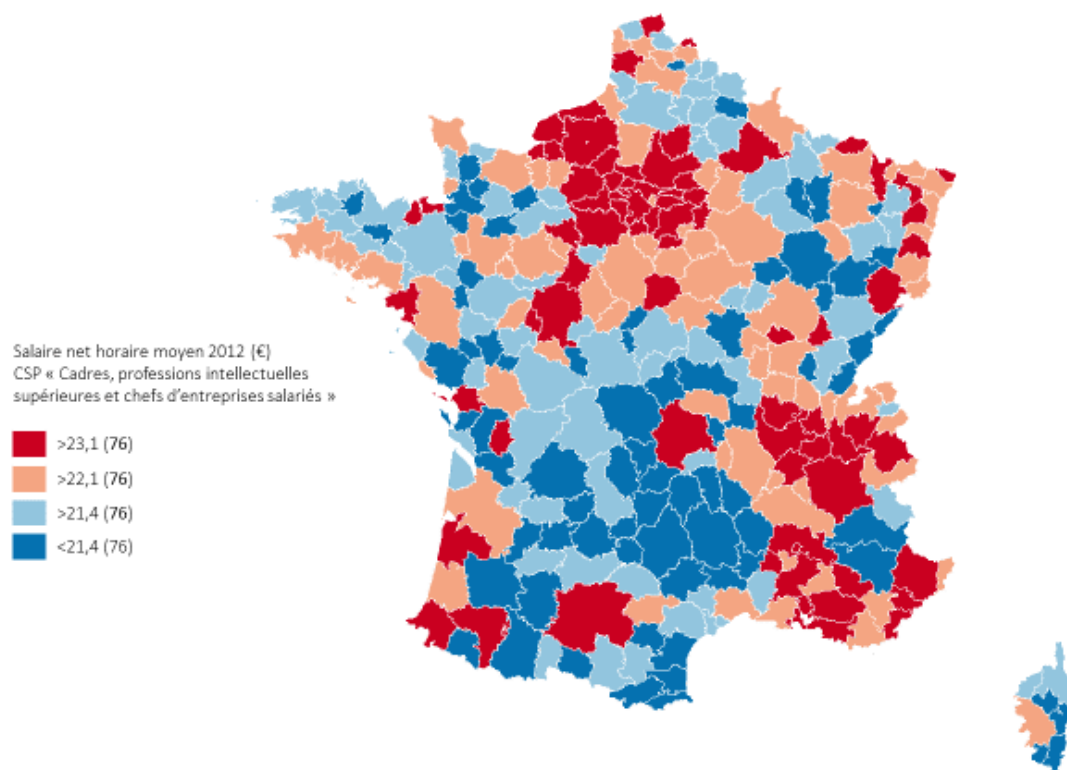
Rapporté à la superficie de chaque zone (en km²), le nombre d'emplois au lieu de travail donne la densité de l'emploi, généralement admise comme une mesure de l'agglomération des activités économiques. Il existe deux principaux avantages à utiliser la densité de l'emploi pour évaluer les effets d'agglomération (voir Melo, Graham et Noland, 2009) : i) c'est un meilleur indicateur des gains de productivité liés à la concentration des activités économiques, alors que les statistiques de population sont très liées aux aménagements urbains et aux phénomènes de congestion ; ii) les mesures de densité sont robustes aux différences de superficie entre les zones.

⁴ Nomenclature des professions et catégories socioprofessionnelles (PCS) de l'Insee – niveau agrégé à un chiffre.

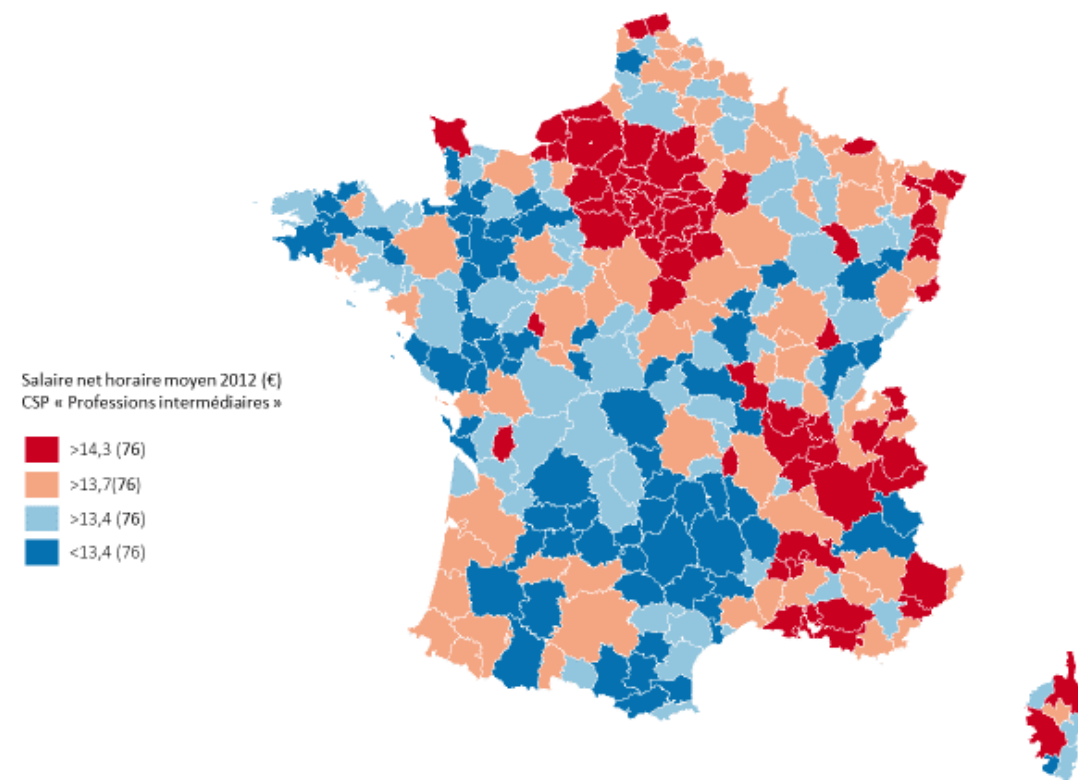
⁵ Ici la notion d'emploi correspond aux emplois au lieu de travail (salariés ou non) déclarés dans le recensement de la population de l'Insee (voir Annexe 2 Données).

⁶ En ordre croissant : Aix-en-Provence, Le Mans, Cannes-Antibes, Créteil, Brest, Dijon, Avignon, Angers, Caen, Metz, Orléans, Nancy, Orly, Tours, Toulon, Clermont-Ferrand, Nice, Saint-Étienne, Montpellier, Strasbourg, Rouen, Grenoble, Rennes, Lille, Nantes, Saclay, Marseille-Aubagne, Bordeaux, Roissy-Sud-Picardie, Toulouse, Lyon, Paris.

Carte 1. Salaires moyens des cadres par zone d'emploi en 2012

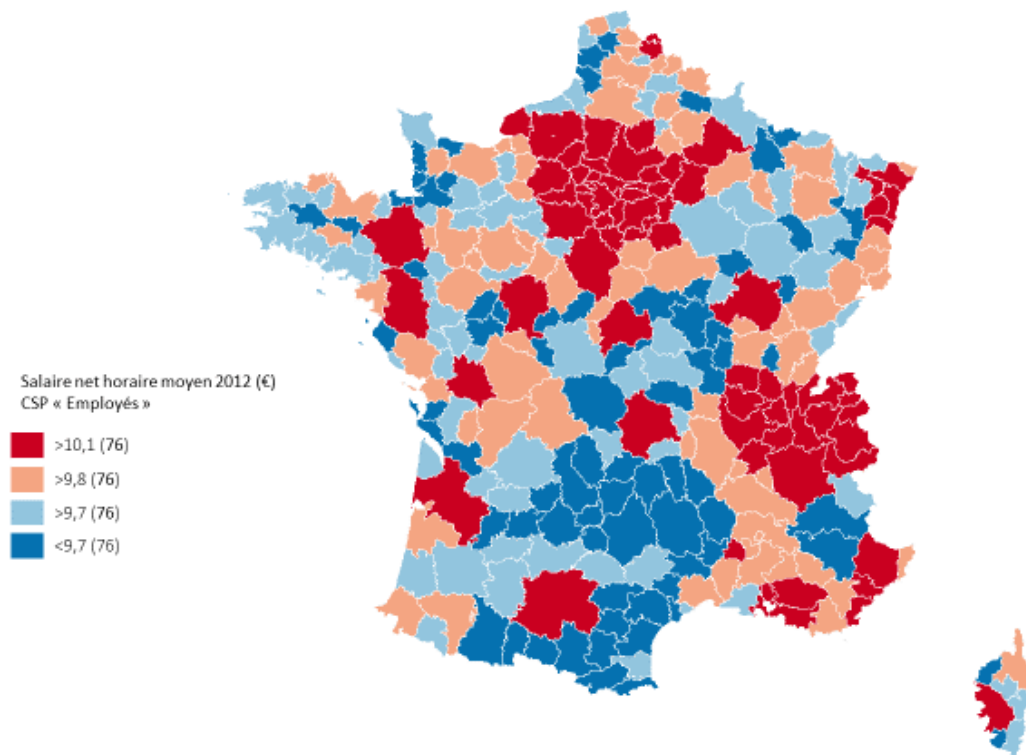


Carte 2. Salaires moyens des professions intermédiaires par zone d'emploi en 2012

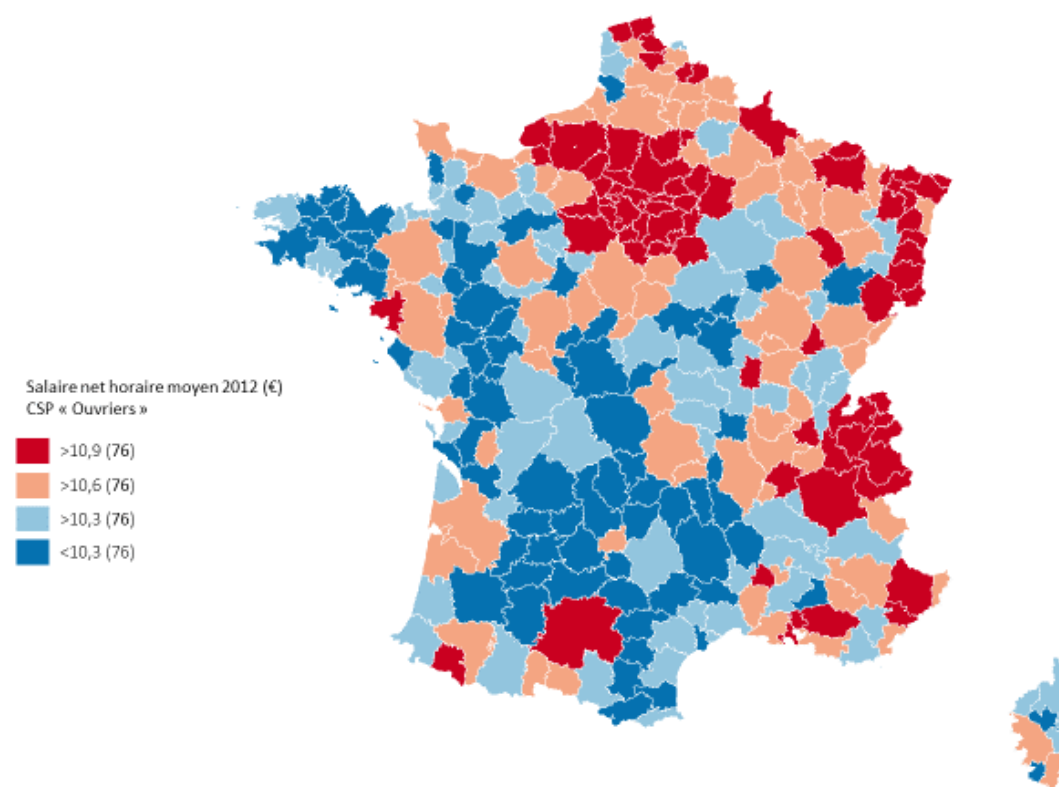


Note : en couleur, quatre tranches de salaire net horaire moyen 2012 en euros ; entre parenthèses, le nombre de zones d'emploi de la tranche. Données Insee DADS 2012.

Carte 3. Salaires moyens des employés par zone d'emploi en 2012



Carte 4. Salaires moyens ouvriers par zone d'emploi en 2012



Note : en couleur, quatre tranches de salaire net horaire moyen 2012 en euros ; entre parenthèses, le nombre de zones d'emploi de la tranche. Données Insee DADS 2012.

L'indicateur de densité de l'emploi révèle aussi l'hétérogénéité des territoires. En moyenne, les zones d'emploi comptent environ 73 emplois par km², mais l'écart-type atteint 329 et le ratio interdécile 7,7 (voir Tableau 2). La moyenne est tirée par les valeurs très hautes des zones d'emploi de faible superficie mais très denses comme Paris ou Lille. Lorsqu'on représente sur une carte de France métropolitaine (voir carte 5) la densité des emplois, on remarque une ressemblance avec les cartes des inégalités de salaires. Celles-ci ne sont cependant pas identiques.

La répartition géographique des plus diplômés est également très dispersée. Le nombre moyen de personnes (non scolarisées) de 15 ans ou plus titulaires d'un diplôme de l'enseignement supérieur long (appelées « diplômés » ici) par zone d'emploi est de 20 300 en 2011 ; ce chiffre est tiré par un petit nombre de zones très peuplées autour des grandes métropoles (voir Tableau 2 et carte 6).

Tableau 2. Statistiques sur la main-d'œuvre locale

Variables	Moyenne	Écart-type	Médiane	Ratio interdécile	Minimum	Zone d'emploi du minimum	Maximum	Zone d'emploi du maximum
Nombre d'emplois (2012)	84 775	221 648	37 896	12,2	4 224	<i>Sartène-Propriano</i>	3 514 354	<i>Paris</i>
Densité de l'emploi (/km ²)	73	329	29	8,1	4	<i>Morvan</i>	5 378	<i>Paris</i>
Nombre de diplômés (2011)	20 300	81 490	6 414	22,4	848	<i>Chatillon</i>	1 347 954	<i>Paris</i>
Part des diplômés	9 %	6 %	8 %	8 pp ⁷	4 %	<i>Thiérache</i>	32 %	<i>Paris</i>

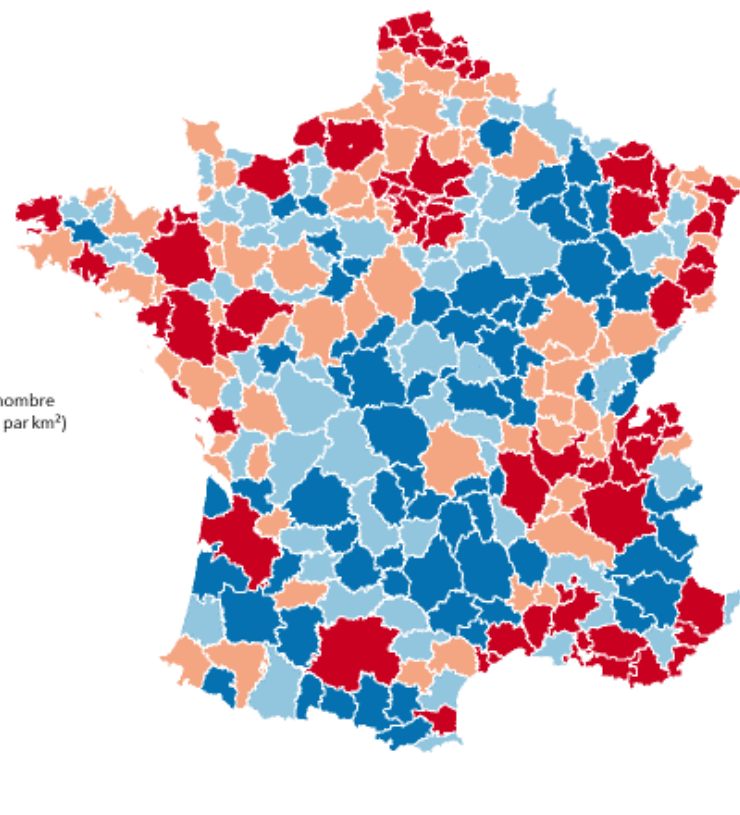
Lorsque ce total est rapporté à la taille de la population locale de 15 ans ou plus (non scolarisée), le ratio est de 9 % en moyenne en 2011 et la médiane de 8 %. On définit cet indicateur comme le ratio de diplômés d'un territoire, à savoir la part des personnes titulaires d'un diplôme bac + 3 ou plus parmi celles qui sont sorties du système scolaire. Les inégalités territoriales sont ici aussi relativement importantes.

Lorsqu'on observe sur une carte de France les zones d'emploi où le ratio de diplômés est particulièrement élevé (supérieur à 10 % dans un quart des zones d'emploi) ou particulièrement faible (inférieur à 7 % dans un quart des zones d'emploi), on constate que cette carte diffère quelque peu de la carte représentant la densité des zones d'emploi. Plus précisément, la carte des zones d'emploi à ratio de diplômés élevé recouvre à peu près celle des zones d'emploi les plus denses, mais celle des zones d'emploi à faible ratio de diplômés se distingue nettement de celle des zones d'emploi les moins denses. Alors que les zones les moins denses se trouvaient pour l'essentiel autour d'une diagonale sud-ouest/nord-est et s'agrégeaient en grappes (voir carte 2), les zones où il y a peu de diplômés du supérieur long se situent plutôt dans la partie nord du pays et sont davantage dispersées sur le territoire.

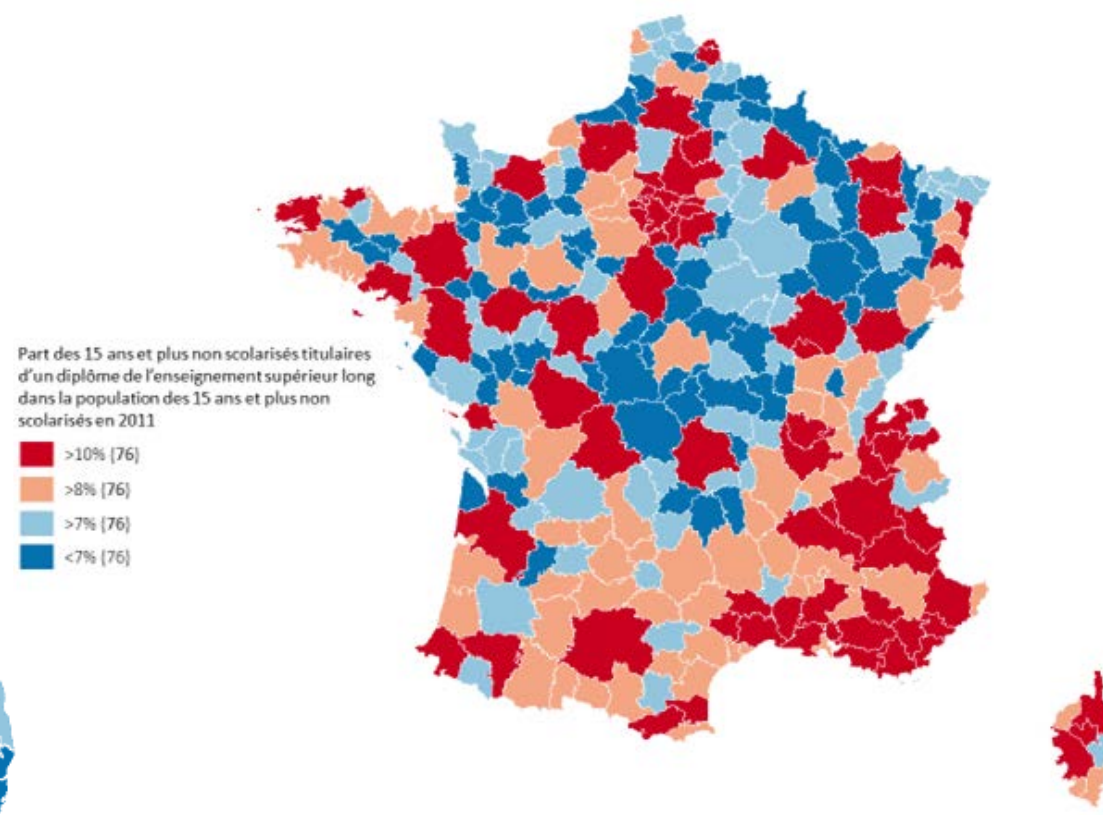
Les graphiques 1 à 4 représentent les courbes de salaires par CSP, affichant sur un même plan le niveau du salaire moyen, la densité de l'emploi et le ratio de diplômés du supérieur long. Aucun schéma dominant ne peut être clairement identifié. Si la densité des zones d'emploi et le ratio de diplômés semblent positivement liés (plus les bulles sont grandes, plus la couleur est bleue), il est difficile d'identifier une relation graphique claire entre ces deux indicateurs et le niveau de salaire moyen – cela quelle que soit la CSP. Par exemple, en ce qui concerne les salaires moyens des CSP cadres, la zone de Rennes est classée 175^e, alors qu'elle se situe au 54^e rang selon la densité d'emplois et au 21^e rang en matière de concentration de diplômés. À l'inverse, la zone de Chinon présente le 5^e salaire moyen le plus élevé, mais n'est classée que 205^e selon la concentration de diplômés et 162^e selon la densité d'emplois. Il semblerait toutefois que les zones avec un ratio de diplômés inférieur à 6 % (bulles orange) restent en bas de la distribution des salaires, tandis que les zones denses en diplômés (bulles bleu foncé) se situent plutôt en haut de la distribution, à CSP données.

⁷ Écart interdécile de 8 points de pourcentage ici (D9-D1).

Carte 5. Densité de l'emploi en 2012 par zone d'emploi

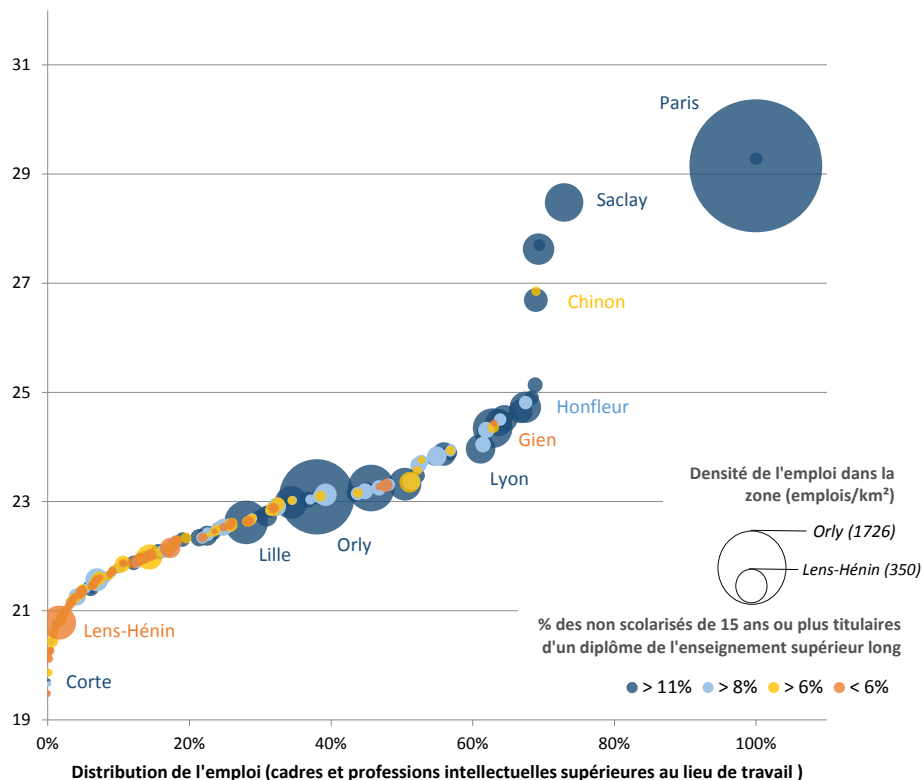


Carte 6. Ratio de diplômés en 2011 par zone d'emploi

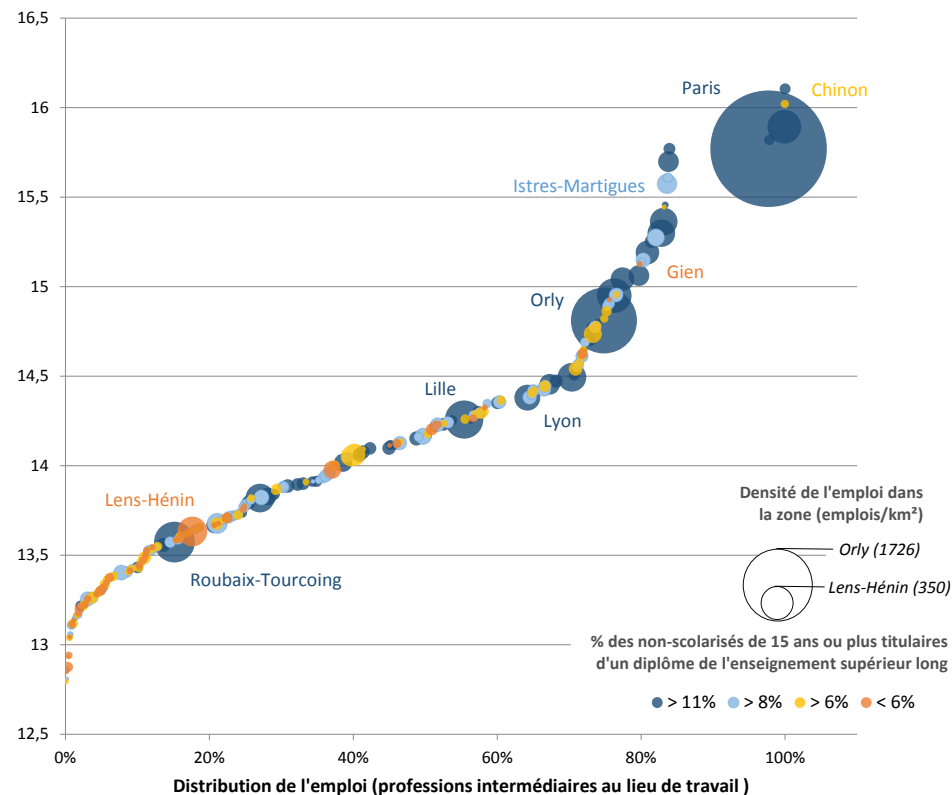


Note : en couleur, quatre tranches de la densité de l'emploi ou de la part des diplômés dans la population ; entre parenthèses, le nombre de zones d'emploi de la tranche. Données Insee, recensements de la population 2011 et 2012.

Graphique 1. Salaires des cadres, densité et concentration de diplômés



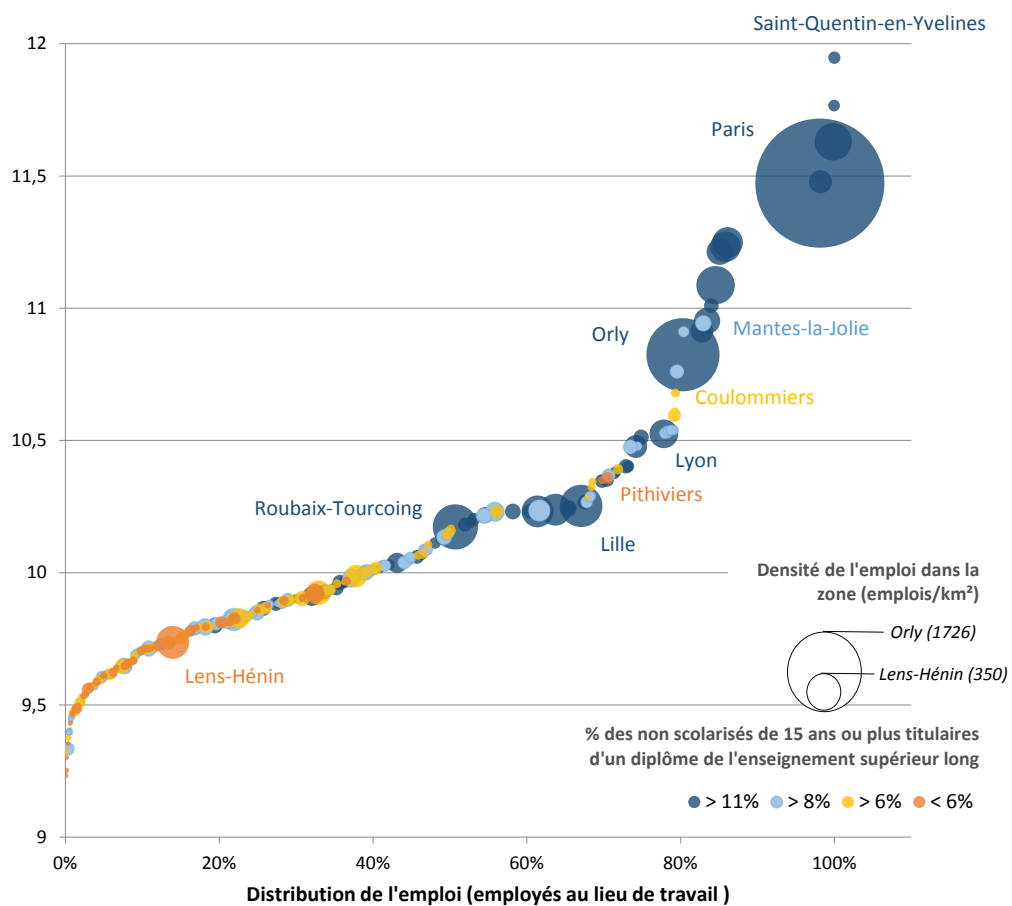
Graphique 2. Salaires des professions intermédiaires, densité et concentration de diplômés



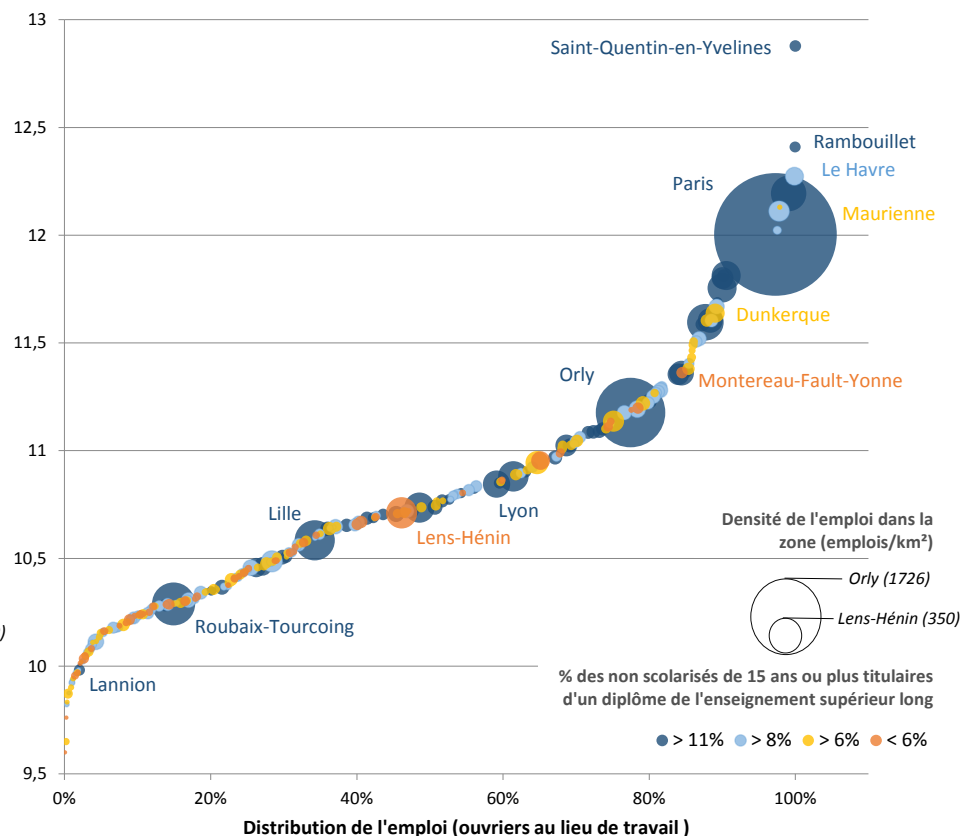
Lecture : la zone d'emploi d'Orly comporte 1 726 emplois par km² et plus de 11 % de sa population (de 15 ans ou plus non scolarisée) est diplômée du supérieur long. Dans cette zone, le salaire net horaire moyen des cadres s'élève à 23 euros ; 38 % de l'emploi total des cadres se situe dans des zones où le salaire moyen des cadres est en dessous de ce niveau. Toujours dans cette zone, le salaire net horaire moyen des professions intermédiaires s'élève à 14,8 euros ; 75 % de l'emploi total des professions intermédiaires se situe dans des zones où le salaire moyen des professions intermédiaires est en dessous de ce niveau.

Données Insee — DADS 2012, recensement de population 2011 et 2012.

Graphique 3. Salaires des employés, densité et concentration de diplômés



Graphique 4. Salaires des ouvriers, densité et concentration de diplômés



Lecture : la zone d'emploi d'Orly comporte 1 726 emplois par km² et plus de 11 % de sa population (de 15 ans ou plus non scolarisée) est diplômée du supérieur long. Dans cette zone, le salaire net horaire moyen des employés s'élève à 10,8 euros ; 80 % de l'emploi total des employés se situe dans des zones où le salaire moyen des employés est en dessous de ce niveau. Toujours dans cette zone, le salaire net horaire moyen des ouvriers s'élève à 11,2 euros ; 75 % de l'emploi total des ouvriers se situe dans des zones où le salaire moyen des professions intermédiaires est en dessous de ce niveau.

Données Insee — DADS 2012, recensement de population 2011 et 2012.

2. Revue de la littérature

Les disparités de salaires évoquées dans la première partie ne sont pas une particularité française. Des phénomènes similaires s'observent à l'étranger, notamment aux États-Unis, en Chine et dans certains pays de l'Union européenne. En raison des implications en matière de politiques publiques, il existe une discussion fournie sur les déterminants de ces disparités.

L'idée centrale du concept d'économies d'agglomération repose sur l'ensemble des gains de productivité engendrés par la proximité des activités et des agents économiques. En particulier, la littérature identifie quatre grands mécanismes sous-jacents (notamment Duranton et Puga, 2004 ; Combes et Lafourcade, 2012) :

- la concentration spatiale des activités économiques peut être source de gains de productivité *via* la diffusion des innovations et des connaissances. Ces bénéfices proviennent des interactions directes entre travailleurs, favorisées par la proximité spatiale des entreprises et des emplois ;
- la taille du marché du travail local, regroupant de nombreuses firmes hétérogènes, encourage une diversification de la demande de travail et contribue à un meilleur appariement avec la main-d'œuvre locale. Cette mise en relation des travailleurs avec les employeurs sur un marché du travail dense améliore la productivité et les salaires en facilitant la division du travail et en limitant à la fois les risques de chômage pour les travailleurs (il est plus facile à chacun de trouver un emploi correspondant à sa qualification et ses souhaits) et le risque d'emploi vacant faute de main-d'œuvre adaptée pour les entreprises.
- l'agglomération des activités économiques peut également accroître l'offre locale de sous-traitants pour les différentes entreprises. Cela diminue les coûts de transaction et facilite l'accès aux intrants nécessaires à la production. Les firmes sont plus productives, ce qui se transmet aux salaires ;
- la taille et la densité du marché a finalement des effets bénéfiques sur les entreprises présentant des technologies à rendements croissants. Les meilleures possibilités de débouchés incitent à l'augmentation de la production pour un coût unitaire plus faible.

D'emblée, il semble difficile d'associer chacun de ces phénomènes à la densité de l'emploi ou à la concentration de diplômés de manière exclusive. Il est clair que la densité comme le capital humain peuvent être à l'origine de gains d'agglomération. Néanmoins, les travaux de recherche ne les considèrent pas toujours simultanément. Lorsque c'est le cas, le capital humain de la main-d'œuvre locale est généralement intégré et interprété dans les analyses en tant que contrôle (par le biais d'effets fixes notamment). Certaines contributions ont tout de même cherché à confronter les deux effets afin d'évaluer leur importance relative.

2.1. La densité des activités économiques augmente la productivité et les salaires, mais peut conduire à des phénomènes de congestion

Il est généralement admis que la densité des activités économiques (souvent mesurée par le nombre d'actifs occupés ou le nombre d'emplois par km²) s'accompagne d'une meilleure efficacité des entreprises qui se transmet, au moins en partie, aux salaires. Globalement, les résultats des études empiriques convergent pour estimer qu'en moyenne doubler la taille d'une zone urbaine entraîne une augmentation de la productivité du travail et des salaires comprise entre 3 % et 8 % (voir en particulier les revues de la littérature proposées par Rosenthal et Strange, 2004, et Combes et Lafourcade, 2012). Les estimations varient selon les contextes et les périodes d'études ainsi que les types de données et les méthodologies utilisées. Melo *et al.* (2009) recensent 34 études des effets d'agglomération : si les 123 estimations qui se concentrent

sur les effets d'agglomération sur les salaires affichent une élasticité moyenne de 0,034⁸, la fourchette est relativement large, allant de -0,096 à 0,143, notamment car elle prend en compte des estimations se limitant à un seul secteur d'activité et faisant appel à des définitions diverses de l'agglomération (densité, superficie, potentiel marchand).

Une des contributions pionnières pour la quantification des effets de densité de l'emploi sur les salaires est proposée par Ciccone et Hall en 1996. En s'intéressant aux différences de productivité du travail entre États américains, ils montrent que celle-ci augmente de 6 % lorsque la densité de l'emploi dans un *county* (comté) est doublée. Toujours dans le contexte des États-Unis, cette fois pour l'année 2010, l'étude de Chauvin, Glaeser, Ma et Tobio (2016) suggère que le doublement de la densité ou de la taille d'une région métropolitaine suscite une augmentation des salaires de l'ordre de 5 %.

Morikawa (2011) estime les effets sectoriels des gains d'agglomération au Japon à l'aide de données individuelles. Il montre un impact important dans les secteurs des services, avec une productivité qui augmente de plus de 10 % lorsque la densité de population de la zone est multipliée par deux.

Au niveau européen, en appliquant sa méthodologie employée dans le contexte américain sur les unités territoriales statistiques (NUTS3) françaises, allemandes, britanniques, italiennes et espagnoles, Ciccone (2002) obtient une élasticité des effets d'agglomération de l'ordre de 0,05 (stable d'un pays européen à un autre). Brülhart et Mathys (2008) prolongent l'analyse de Ciccone suivant la méthode des moments généralisés sur des données de panel, en prenant en compte des dimensions sectorielles. Sur la base de 245 régions NUTS2 comprises dans une vingtaine de pays européens, leur étude confirme l'existence d'effets positifs et significatifs de la densité de l'emploi sur la productivité et les salaires, avec une élasticité de long terme d'environ 0,13. Foster et Stehrer (2009) poursuivent également les travaux de Ciccone avec l'étude d'un panel de 255 régions NUTS2 intégrant six différents secteurs d'activité dans 27 pays européens sur la période 1998-2004. Ils retrouvent des élasticités très proches de celle de Ciccone, autour de 0,05.

Une autre série de contributions s'est focalisée sur l'étude des effets d'agglomération dans le contexte français. Les travaux de Combes *et al.* font référence. Sur une première base de presque 20 millions de données individuelles entre 1976 et 1998, Combes, Duranton et Gobillon (2008) évaluent les effets d'agglomération au sein des zones d'emploi en France. À l'aide d'une simple régression par la méthode des moindres carrés ordinaires (MCO) du salaire moyen local en 1998 sur la densité de l'emploi de la même année, ils estiment une élasticité de 0,05 – en ligne avec la littérature existante. Le développement de leur modèle sur un très large panel de données individuelles, intégrant des effets fixes par zones et par années ainsi que des variables contrôlant des interactions locales entre industries et des dotations factorielles locales, indique ensuite des élasticités relativement plus faibles, de l'ordre de 0,03. Combes *et al.* (2011) proposent une autre étude des déterminants des inégalités spatiale de salaire en France et de leur évolution de long terme (1860-2000). Cette analyse au niveau des départements français révèle un lien positif entre la densité de l'emploi et la productivité du travail de 1860 à 2000, avec des corrélations qui augmentent au fil du temps. Les premières régressions – comprenant uniquement des effets fixes année et secteur – estiment une élasticité de la productivité du travail à la densité de 0,08 ; un doublement de la densité entraîne une hausse d'environ 5,7 %. L'élasticité reste du même ordre (0,09) avec l'introduction d'une série de contrôles. Enfin, la revue statistique proposée par Melo *et al.* (2008) révèle des estimations de l'effet d'agglomération allant de 0,012 à 0,143 dans le cas de la France (54 estimations avec une élasticité moyenne de

⁸ L'élasticité est une grandeur sans unité, elle mesure la réaction d'une variable à une autre. Si l'élasticité du salaire à la densité est de 0,03, cela implique qu'une augmentation de la densité de 1 % engendre une augmentation du salaire de 0,03 %. Cependant, un doublement de la densité (augmentation de 100 %) ne correspond pas à une augmentation du salaire de 3 %. La formule du développement limité à l'ordre 1 n'est plus valable pour des grandes variations (100 %) ; l'augmentation de salaire est donnée par la formule $2^{0,03}-1=2,1$ %.

0,039, une médiane à 0,035 et un écart-type de 0,022). Par ailleurs, les résultats de leur méta-analyse montrent que les élasticités des effets d'agglomération tendent à être relativement plus élevées en France qu'aux États-Unis.

Dans certains cas, la densité peut entraîner une dégradation relative des salaires et de la productivité. Nombres de contributions ont en effet cherché à vérifier certains phénomènes de congestion à partir d'un certain seuil d'urbanisation, et donc la concavité de l'impact de la densité sur la productivité et les salaires. Derrière ces analyses réside l'idée que l'agglomération urbaine a également des coûts, en particulier la congestion routière et des réseaux de transports, la majoration du foncier et de l'immobilier, mais aussi la pollution. Le rendement marginal de l'agglomération diminuerait alors au-delà d'un certain seuil (Combes *et al.*, 2016). Les élasticités estimées correspondent généralement aux gains « nets » d'agglomération. Elles sont généralement positives car les impacts positifs de la densité compensent les aspects négatifs. Si ces derniers ne sont pas compensés, l'élasticité apparaît négative.

2.2. Des effets de densité qui proviennent en partie d'effets de composition et du « tri spatial » des travailleurs qualifiés

La théorie de la croissance endogène insistait déjà sur le rôle du capital humain comme déterminant du développement économique (Lucas, 1988). Dans cette lignée, une partie de la littérature concernant les effets d'agglomération a mis l'accent sur l'influence du capital humain dans la structuration des zones urbaines. En premier lieu dans le but d'isoler cette influence afin d'identifier plus précisément les effets de densité.

L'effet positif de la densité sur la productivité et les salaires peut en effet provenir en partie d'une surreprésentation des travailleurs qualifiés dans les zones plus denses. Combes *et al.* (2012, 2016) parlent de « tri spatial » de la main-d'œuvre, c'est-à-dire une « auto-sélection spatiale des travailleurs » attirés – ou retenus – par certaines préférences spécifiques et des effets de demande de diplômés (voir aussi Ciconne et Hall, 1996 ; Glaeser *et al.*, 2001, 2004 ; Chauvin *et al.*, 2016).

Derrière l'influence du niveau de qualification, il faut bien distinguer ce qui provient des externalités de capital humain, et ce qui est dû aux effets de composition de la main-d'œuvre locale (ce problème d'identification sera rediscuté dans la troisième partie).

Si la répartition géographique des travailleurs qualifiés est inégale et liée à la densité des zones, alors les analyses qui ne prennent pas en compte la composition locale des qualifications surestiment les effets de densité en leur attribuant les gains de productivité provenant du rendement intrinsèquement plus élevé du travail qualifié – plus abondant dans ces zones. Un des premiers facteurs expliquant les disparités de niveau de salaire moyen sur le territoire français est cet effet de composition, lié aux poids respectifs des activités économiques ainsi qu'à la proportion de travailleurs qualifiés. La zone d'emploi de Paris compte une plus forte proportion de cadres supérieurs, relativement plus qualifiés et mieux rémunérés que les autres travailleurs, que la zone d'emploi de Saint-Flour. Le salaire moyen parisien est donc de fait plus élevé.

Un moyen d'isoler l'effet de la densité de ces effets de composition réside dans l'utilisation de données individuelles pour prendre en compte les qualifications et compétences propres à chaque individu. Combes *et al.* (2008) insistent sur l'importance du rôle joué par les qualifications, plutôt que par l'environnement local ou le secteur d'activité, dans la détermination des salaires individuels. En utilisant un grand nombre de données individuelles sur la période 1976-1998, ils montrent un effet des aptitudes individuelles relativement important permettant d'expliquer plus de la moitié des disparités spatiales de salaire. Par ailleurs, un résultat intéressant réside dans la diminution de l'élasticité de la densité aux salaires (de 0,06 à 0,03) après le contrôle pour ces compétences individuelles. Plus récemment, Combes *et al.* (2015a)

estiment les gains d'agglomération dont bénéficient les salariés au niveau des zones d'emploi en France métropolitaine pour six années entre 1993 et 2008. En tenant compte du tri spatial des qualifications et des compétences individuelles, ils montrent, selon leurs estimations les plus conservatrices, une élasticité de la prime salariale à la densité d'emploi de 0,013. Enfin, dans leur méta-analyse, Melo, Graham et Noland (2008) suggèrent que l'omission des contrôles pour la qualité de la main-d'œuvre conduit à des surestimations des effets d'agglomération liés à la densité – et constitue une source importante de variation entre les estimations. Selon eux, la prise en compte des différences de niveau de qualifications pour différentes périodes réduit les estimations des effets d'urbanisation de 5 à 6 points de pourcentage selon les études et les spécifications.

En outre, d'autres travaux traitent les effets de capital humain en soulignant l'hétérogénéité des effets d'agglomération selon les niveaux de diplôme ou les types de professions. Groot et de Groot (2014) remarquent que les gains de salaires dont bénéficient les travailleurs des zones denses aux Pays-Bas diffèrent selon le niveau de diplôme. Sur la période 2000-2010, les travailleurs plus diplômés profitent davantage d'une plus grande densité de l'emploi, même lorsque l'on contrôle des différences individuelles observables et inobservables des travailleurs (panel et variables instrumentales). Combes *et al.* (2016) approfondissent les travaux sur le Grand Paris dans un second rapport et montrent des effets hétérogènes selon les catégories socioprofessionnelles, avec des gains d'agglomération plus importants dans les CSP cadres et professions intellectuelles supérieures. Cependant, ces conclusions ne sont pas toujours vérifiées. À l'opposé, Di Addario *et al.* (2008), dans le contexte italien, considèrent que la prime salariale très faible provenant de l'urbanisation (0,1 % lorsque la population d'un marché du travail local augmente de 100 000 habitants) cache selon eux des gains et pertes différents selon les catégories de travailleurs. Ils montrent aussi que dans les marchés du travail les plus denses, les diplômés du supérieur connaissent des réductions relatives de salaires de 0,4 % à 0,8 %. Ils expliquent ce paradoxe par les préférences spécifiques des populations plus diplômées vis-à-vis de la qualité de vie dans les zones très urbanisées.

2.3. Des externalités positives de capital humain qui améliorent les salaires au niveau local

Comme dit plus haut, le niveau de capital humain peut, en théorie, améliorer la productivité et les salaires au-delà de l'effet direct de l'éducation sur la productivité individuelle des travailleurs qualifiés. Cela passe par les externalités de capital humain. Le concept d'externalités de capital humain met en lumière l'influence des travailleurs qualifiés et de leur répartition géographique sur la productivité et les salaires de l'ensemble des travailleurs de la zone – indépendamment des simples effets de composition.

Pour identifier ces externalités, certaines analyses intègrent directement des indicateurs du niveau de qualification moyen par zone dans les équations de détermination du salaire ; généralement la part des qualifiés dans la population ou le niveau moyen d'éducation de la population.

Rauch (1993) est l'un des tout premiers à explorer les différences de niveau de capital humain entre les villes et à identifier les externalités sur les salaires et la productivité. Il estime une augmentation des salaires de 3 % à 5 % avec une augmentation d'une année du niveau d'éducation moyen dans les régions métropolitaines américaines en 1980.

Moretti (2004b) s'intéresse plus particulièrement aux effets de la proportion de diplômés sur la productivité des usines manufacturières américaines en 1982 et 1992. Selon ses estimations les plus robustes, en moyenne, une augmentation d'un point de pourcentage de la part des diplômés du supérieur dans une ville (en écartant le secteur de la firme à l'étude) est associée à une amélioration de la productivité de l'ordre de 0,5 % à 0,7 %. Ces différentiels de productivité estimés entre les villes avec des niveaux plus ou moins élevés de capital humain sont analogues

aux estimations concernant les disparités de salaires dans le secteur manufacturier. Par ailleurs, Moretti montre que le capital physique extrinsèque (c'est-à-dire le capital physique de la zone autre que celui installé dans l'entreprise) n'a aucun effet sur la productivité de la firme, et donc que les effets estimés ne reflètent pas simplement des économies d'agglomération générés par la densité des activités économiques, mais bel et bien des externalités de capital humain.

Moretti (2004b) s'intéresse également au rendement social de l'éducation (par rapport aux bénéfices individuels) dans les villes américaines à la fin du XX^e siècle. À l'aide d'un modèle à variables instrumentales basé sur des données de recensements en 1980 et 1990, il montre que l'augmentation d'un point de pourcentage de la part des travailleurs diplômés de l'enseignement supérieur dans une zone urbaine entraîne des primes salariales de 1,9 % pour les décrocheurs du secondaire, de 1,6 % pour les diplômés du secondaire, de 1,2 % pour les travailleurs passés par l'université, et de 0,4 % pour les diplômés du supérieur.

Rosenthal et Strange (2008) estiment, dans le cas des États-Unis pour l'année 2000, une élasticité des salaires au nombre de travailleurs dans un rayon de 5 miles de 4,5 % et montrent que cet effet d'agglomération provient en réalité de la présence de travailleurs qualifiés. La proximité entre travailleurs diplômés du supérieur améliore les salaires alors que la proximité avec les moins diplômés a des effets négatifs ; ces impacts sur les salaires apparaissent autant pour les diplômés que pour les non-diplômés.

Abel *et al.* (2011) montrent que le capital humain joue un rôle central dans les gains d'agglomération. En introduisant une variable de stock de capital humain et un terme d'interaction entre cette dernière et la densité de population, ils remarquent que l'élasticité de la productivité du travail – moyenne entre 2001 et 2005 – à la densité augmente avec le niveau de capital humain des aires urbaines américaines. Selon leurs estimations, les effets d'agglomération dépendent du niveau de capital humain des zones métropolitaines. Ces effets sont négatifs dans les zones relativement pauvres en diplômés du supérieur : la densité améliore la productivité uniquement dans les zones avec une forte concentration de personnes très qualifiées. Ils expliquent cela par les externalités d'apprentissage et de connaissances dans les villes, particulièrement dans les secteurs des services aux entreprises, des arts et loisirs, de l'information et de la finance, comme source de gains d'agglomération.

Combes *et al.* (2011) ajoutent des variables capturant les niveaux de qualifications de la main-d'œuvre locale dans leurs spécifications d'estimation des effets d'agglomération. Ils regardent si cette intégration a un effet positif sur la productivité du travail et si elle affecte les autres coefficients estimés. Sur la période 1860-1930, le capital humain affiche un effet positif et significatif sur la productivité du travail, mais ne remet pas en cause les gains d'agglomération liés à la densité – son intégration aux spécifications n'affecte pas les élasticités de la densité et du potentiel de marché. Le même exercice appliqué à l'année 2000 produit des résultats tout à fait différents : après contrôle pour le niveau de capital humain, la densité de l'emploi et le potentiel marchand ne sont plus significatifs, tandis que la variable de capital humain montre un effet important et très significatif. Les auteurs concluent alors sur un effet de la densité des activités économiques largement exagéré en l'absence de contrôle pour le capital humain et une prévalence du niveau de qualification dans l'explication des écarts de productivité du travail entre départements français.

Plus récemment, Chauvin *et al.* (2016) confrontent les effets d'agglomération liés à la densité des régions métropolitaines avec les effets de capital humain. Leur étude des États-Unis révèle une chute des coefficients de la densité avec le contrôle pour le niveau d'éducation. Selon leurs différentes spécifications (MCO ou variables instrumentales), les résultats d'estimation suggèrent une augmentation des salaires de l'ordre de 10 % lorsque la proportion d'adultes titulaires d'un diplôme universitaire (« *Bachelor's degree* ») augmente de 10 points de pourcentage – à niveau d'éducation individuelle constant.

C'est sur ces derniers constats que se construit la présente étude. Il s'agit de mesurer les effets de la densité des activités économiques et de la concentration de diplômés pour évaluer et comparer les forces à l'origine des disparités de salaire moyen entre zones d'emploi en France en 2012.

3. Décomposer les effets d'agglomération : quelle influence des externalités de capital humain ?

Dans le but d'évaluer l'influence de la densité des activités économiques et des externalités de capital humain sur les disparités géographiques de salaire, la stratégie économétrique proposée ici consiste à régresser le logarithme du salaire moyen sur le logarithme de la densité de l'emploi et le pourcentage de diplômés du supérieur long dans la population, ainsi qu'un certain nombre de variables explicatives de contrôle.

Le travail économétrique se fait sur données agrégées (au niveau des 304 zones d'emplois de France métropolitaine) suivant une méthodologie d'estimation de forme réduite en coupe instantanée pour l'année 2012. Ce choix a été motivé par l'absence d'information sur le diplôme dans les données individuelles DADS de l'Insee (d'où l'on tire les salaires moyens), et donc l'impossibilité de contrôler, dans une première étape, des effets de composition en matière de qualification (voir construction du modèle ci-après). Les données agrégées par zone d'emploi permettent d'avoir ces informations à partir des recensements de population de l'Insee (voir Annexe 2 pour les sources de données).

En premier lieu, le salaire moyen en 2012 d'une zone d'emploi ZE peut être déterminé par une équation du type :

$$\ln \text{ Salaire moyen}_{ZE} = \alpha + \beta \ln \text{ Densite}_{ZE} + \gamma \text{ Ratio_Diplomes}_{ZE} + \varepsilon_{ZE}$$

avec $\ln \text{ Salaire moyen}_{ZE}$ le log du salaire net horaire moyen de la zone d'emploi ZE en 2012, $\ln \text{ Densite}_{ZE}$ le log de la densité de l'emploi de la zone d'emploi ZE en 2012, et $\text{ Ratio_Diplomes}_{ZE}$ le ratio de diplômés du supérieur long par rapport à la population de 15 ans et plus non scolarisée dans la zone d'emploi ZE en 2011.

Comme expliqué dans Moretti (2004) et Combes et al. (2012), la variable dépendante de l'équation utilisée pour évaluer les effets d'agglomération doit correspondre au salaire nominal, et non au salaire réel. Pour quantifier les gains d'agglomération liés à la densité des activités et à la concentration de la main d'œuvre qualifiée, il est nécessaire de faire le lien avec la productivité du travail – donc le rendement nominal du travail – sans prendre en compte le coût de la vie au niveau local. En effet, des salaires nominaux (i.e. les salaires que les entreprises versent effectivement) plus élevés dans une zone d'emploi doivent nécessairement refléter une productivité plus grande (Acemoglu & Angrist, 2000). Si les travailleurs ne sont pas plus productifs, alors qu'ils perçoivent des salaires plus élevés, les firmes disparaissent ou se relocalisent dans des zones où les salaires sont plus faibles (Moretti, 2004). Cet argument n'est valide qu'à l'équilibre c'est-à-dire une fois que les éventuelles réallocations spatiales ont eu lieu. Idéalement donc, même en travaillant sur des salaires nominaux, il aurait pu être pertinent de tester l'impact du prix du foncier sur les salaires. Il se peut que dans certaines zones, pour attirer ou retenir les salariés, les entreprises doivent verser des salaires d'autant plus élevés que la vie y est chère, notamment pour se loger. Cela n'a pas été fait, faute de données fines, disponibles et fiables. Pour autant, on peut penser que l'effet éventuel du prix du foncier sur les salaires résulte en fait pour beaucoup des autres facteurs explicatifs déjà intégrés dans la modélisation (densité des activités, spécialisations sectorielles, caractéristiques sociodémographiques de la population résidente).

Tableau 3. Effets de la densité de l'emploi et de la concentration de diplômés

VARIABLES	Ln Salaire moyen en 2012	
<i>In Densité de l'emploi</i>	0.0666*** (0.00545)	0.0342*** (0.00510)
<i>Ratio de diplômés</i>		1.390*** (0.187)
<i>Constante</i>	2.276*** (0.0173)	2.263*** (0.0126)
<i>R² ajusté</i>	0.480	0.675
<i>Observations</i>	304	304

Erreurs-types entre parenthèses.

*** significatif au seuil de 1 %.

Les résultats présentés dans le Tableau 3 montrent bien une surestimation de l'influence de la densité des activités économiques sur les salaires lorsque le niveau de qualification de la population locale n'est pas pris en compte : l'élasticité des salaires à la densité diminue de moitié avec l'ajout du ratio de diplômés dans la spécification.

Les résultats de cette spécification très simple masquent les effets hétérogènes selon les différentes CSP et ne permet pas d'isoler l'impact de l'agglomération d'autres éléments susceptibles d'influencer les salaires moyens, tels que la composition sectorielle ou la composition de la main-d'œuvre locale, en particulier au regard des qualifications.

L'agglomération des activités et des personnes semble bien s'accompagner de primes salariales.

3.1. Externalités de capital humain et effet de substitution imparfaite entre travailleurs qualifiés et non qualifiés : cadre théorique

Si l'estimation d'un coefficient associé à une variable de ratio de diplômés permet de mesurer l'influence de celle-ci sur les disparités de salaire, la mesure ne reflète pas nécessairement les seules externalités de capital humain : le coefficient estimé peut également capter un effet de substitution imparfaite entre travailleurs qualifiés et travailleurs moins qualifiés. Cela entraîne un problème d'identification. Lorsque la part des travailleurs qualifiés augmente, tous les travailleurs bénéficient d'effets d'externalités, mais les moins qualifiés, qui deviennent relativement moins nombreux, profitent d'un effet supplémentaire avec l'amélioration de leur productivité marginale. Cet effet joue dans le sens inverse pour les travailleurs qualifiés. Finalement, les travailleurs peu qualifiés bénéficient de l'augmentation de la part des diplômés dans une zone d'emploi même en l'absence d'externalité, alors que l'effet sur les salaires des qualifiés dépend de l'ampleur relative des externalités de capital humain (Combes et Gobillon, 2015 ; Moretti, 2004c).

Pour illustrer l'existence d'un effet de substitution imparfaite et d'un effet d'externalité sur les salaires de chaque classe de travailleurs, il est possible d'utiliser le modèle simple suivant – également employé dans Combes et Gobillon (2015).

La fonction de production est à rendements d'échelle constants et combine un agrégat travail avec le capital *via* une fonction de production Cobb-Douglas (élasticité de substitution unitaire entre le capital et l'agrégat travail). L'agrégat travail est obtenu en combinant deux types de travailleurs : les travailleurs qualifiés (L_1) et les travailleurs moins qualifiés (L_2), avec une élasticité de substitution constante (CES) entre ces deux formes de travail égale à $\sigma=1/(1-\rho)$, où $\rho < 1$.

$$Y = \left[(A_1 L_1)^\rho + (A_2 L_2)^\rho \right]^{\alpha/\rho} K^{1-\alpha}$$

L'efficacité du travail qualifié (A1) et celle du travail moins qualifié (A2) bénéficient d'externalités de capital humain d'autant plus élevées que le ratio des travailleurs qualifiés par rapport aux travailleurs moins qualifiés ($S = L_1/L_2$) est élevé. On pose $A_1 = S^{\gamma_1}$ et $A_2 = S^{\gamma_2}$, avec γ l'ampleur des externalités de capital humain.

La maximisation du profit conduit à rémunérer les facteurs à leur productivité marginale, d'où (en posant prix=1) :

$$\begin{cases} r = \left[(A_1 L_1)^\rho + (A_2 L_2)^\rho \right]^{\rho/\rho} (1-\alpha) K^{-\alpha} \\ w_1 = \alpha (A_1 L_1)^\rho L_1^{-1} \left[(A_1 L_1)^\rho + (A_2 L_2)^\rho \right]^{\frac{\alpha}{\rho}-1} K^{1-\alpha} \\ w_2 = \alpha (A_2 L_2)^\rho L_2^{-1} \left[(A_1 L_1)^\rho + (A_2 L_2)^\rho \right]^{\frac{\alpha}{\rho}-1} K^{1-\alpha} \end{cases}$$

avec r le taux d'intérêt et w_i le salaire des travailleurs avec la qualification i .

Dont on tire :

$$w_1 = \alpha (S^{\gamma_1})^\rho \left[(S^{\gamma_1})^\rho + (S^{\gamma_2})^\rho S^{-\rho} \right]^{\frac{\alpha}{\rho}-1} \left(\frac{1}{r} (1-\alpha) \right)^{\frac{1-\alpha}{\alpha}}$$

$$w_2 = \alpha (S^{\gamma_2})^\rho S^{1-\rho} \left[(S^{\gamma_1})^\rho + (S^{\gamma_2})^\rho S^{-\rho} \right]^{\frac{\alpha}{\rho}-1} \left(\frac{1}{r} (1-\alpha) \right)^{\frac{1-\alpha}{\alpha}}$$

Les élasticités des salaires au ratio de qualifié S (ε_1 ε_2), peuvent alors s'approximer par :

$$\varepsilon_1 = \gamma_1 - (1-\phi)(1-\rho)(1+\gamma_1-\gamma_2)$$

$$\varepsilon_2 = \gamma_2 + \phi(1-\rho)(1+\gamma_1-\gamma_2)$$

où ϕ représente la part des salaires des qualifiés dans le total des rémunérations.

Il est clair que ces élasticités des salaires au ratio des qualifiés par rapport aux moins qualifiés combinent un effet d'externalité direct (les coefficients γ) et un effet de substitution, positif pour les salaires des moins qualifiés $\phi(1-\rho)(1+\gamma_1-\gamma_2)$ et négatif pour les salaires des qualifiés $-(1-\phi)(1-\rho)(1+\gamma_1-\gamma_2)$, en supposant $(1+\gamma_1) > \gamma_2$.

Il n'est pas tout à fait juste d'interpréter les effets de la structure de l'emploi dans une zone sur les salaires uniquement comme des effets d'externalités. Il est en revanche probable qu'une fois corrigés des effets de substitution, les effets d'externalités soient plus élevés que le coefficient estimé pour les salaires qualifiés et moins élevé pour les autres travailleurs⁹.

⁹ Combes et Gobillon (2015) proposent alors de régresser les salaires nominaux non seulement sur la variable de capital humain (S) mais aussi sur le produit de cette variable par la part des salaires ($S\phi$ et $S(1-\phi)$) afin de séparer les effets d'externalités pures et les effets de substitution.

3.2. Équation de détermination des salaires moyens par CSP au niveau des zones d'emploi

Dans ce cadre, suivant un raisonnement proche de celui développé dans Moretti (2004), la régression du salaire moyen par zone d'emploi est décomposée en quatre spécifications selon les quatre grands groupes socioprofessionnels disponibles dans les données DADS 2012 (Cadres, professions intellectuelles supérieures et chefs d'entreprise salariés ; Professions intermédiaires ; Employés ; et Ouvriers). Les CSP cadres sont généralement plus « qualifiées » que les autres CSP.

Cette décomposition permet de comparer les effets d'agglomération – notamment l'influence de la concentration de diplômés – sur les salaires des différentes CSP, et par extension des travailleurs qualifiés (CSP cadres) par rapport au moins qualifiés (CSP employés et ouvriers en particulier). Elle permet également d'atténuer un autre problème d'identification lié à la difficulté, lorsqu'on analyse le salaire global, d'estimer séparément l'externalité de capital humain et l'effet de prime salariale directe engendré par l'augmentation du nombre de travailleurs qualifiés dans la masse salariale locale (effet de composition). En effet, l'étude du salaire agrégé peut masquer des disparités dans les variations de salaires selon les différentes CSP : un gain de salaire moyen global, dans une zone donnée, pourrait uniquement provenir d'une augmentation de la masse des CSP relativement plus qualifiées, dont la productivité intrinsèque, et donc la rémunération, sont relativement plus élevées. Dans ce contexte, la décomposition des salaires par CSP permet une interprétation plus précise. Même si la productivité agrégée augmente avec la concentration de diplômés, les salaires moyens des CSP ouvriers ou employés – et professions intermédiaires dans une moindre mesure – sont moins directement affectés par la composition de la main-d'œuvre plus concentrée en diplômés. L'effet du ratio de diplômés du supérieur long sur leurs salaires passe davantage par les externalités de capital humain.

Quatre spécifications sont donc déclinées à partir de l'équation de salaire précédente :

$$\ln \text{ Salaire moyen CSP}_{ZE} = \alpha + \beta \ln \text{ Densite}_{ZE} + \gamma \text{ Ratio_Diplomes}_{ZE} + \varepsilon_{ZE}$$

où $\ln \text{ Salaire moyen CSP}_{ZE}$ correspond aux quatre variables suivantes :

- log du salaire net horaire moyen des cadres, professions intellectuelles supérieures et des chefs d'entreprise salariés en 2012 ;
- log du salaire net horaire moyen des professions intermédiaires en 2012 ;
- log du salaire net horaire moyen des employés en 2012 ;
- log du salaire net horaire moyen des ouvriers en 2012.

Les résultats présentés dans le Tableau 4 montrent, comme précédemment, une surestimation de l'effet de la densité de l'emploi sur le salaire moyen lorsque la part des diplômés du supérieur long n'est pas prise en compte, à CSP donnée. Le coefficient associé au ratio de diplômés est non significatif pour la spécification concernant la CSP ouvriers. Notamment parce qu'ils ne prenaient pas en compte les effets de composition, les impacts, précédemment estimés, de la densité et du ratio de diplômés sur le salaire total étaient relativement surestimés (les coefficients étaient respectivement de 0,03 et 1,4, voir Tableau 3) et masquaient l'hétérogénéité selon les différentes CSP (les coefficients sont compris entre 0,015 et 0,02 pour la densité et entre 0,21 et 0,57 pour le ratio de diplômés du supérieur long).

Tableau 4. Effets de la densité de l'emploi et de la concentration de diplômés sur les salaires moyens (en log)

	Cadres, professions intellectuelles supérieures et chefs d'entreprise salariés		Professions intermédiaires		Employés		Ouvriers	
<i>In Densité de l'emploi</i>	0.0387*** (0.00406)	0.0253*** (0.00473)	0.0237*** (0.00242)	0.0168*** (0.00297)	0.0265*** (0.00225)	0.0150*** (0.00246)	0.0260*** (0.00284)	0.0210*** (0.00387)
<i>Ratio de diplômés</i>		0.572*** (0.145)		0.294*** (0.0775)		0.495*** (0.0888)		0.213* (0.112)
<i>Constante</i>	2.968*** (0.0134)	2.963*** (0.0117)	2.550*** (0.00868)	2.547*** (0.00853)	2.203*** (0.00709)	2.198*** (0.00689)	2.274*** (0.00967)	2.272*** (0.00967)
<i>R² ajusté</i>	0.356	0.427	0.239	0.271	0.383	0.508	0.260	0.274

Erreurs-types entre parenthèses.

*** significatif au seuil de 1 %.

** significatif au seuil de 5 %.

* significatif au seuil de 10 %.

Les spécifications présentées ici restent très basiques et de nombreux autres facteurs sont susceptibles d'expliquer les disparités spatiales de salaires – les R² relativement faibles confortent cette possibilité. Elles sont utiles pour cadrer la stratégie économétrique et formuler des premières intuitions, mais leur simplicité pose de sérieux problèmes d'identification.

3.3. Deux principales sources d'endogénéité

Les relations précédentes estimées suivant la méthode des MCO devraient respecter l'hypothèse d'une absence de corrélation entre les variables explicatives et le terme résiduel. Alors qu'elle constitue une condition nécessaire pour obtenir des résultats d'estimation MCO sans biais, l'exogénéité des variables explicatives (que ce soit la densité de l'emploi, la concentration de diplômés ou tout autre déterminant des salaires) n'est que très rarement vérifiée. Certaines zones d'emploi peuvent être intrinsèquement plus productives, encourageant les salaires à la hausse, ce qui attire des travailleurs, et par conséquent engendre une densification des activités. Les variables à l'origine d'effets d'agglomération sont parfois dépendantes des choix de localisation des travailleurs, notamment des qualifiés. Cela entraîne de sérieux problèmes d'identification.

Il existe deux sources principales d'endogénéité potentielle : les variables omises et la causalité inverse.

Variables omises et hétérogénéité inobservée

Pour fournir une estimation non biaisée, les MCO supposent qu'aucune variable omise ne doit être corrélée avec les variables explicatives intégrées dans la spécification.

Comme son nom l'indique, une variable omise est susceptible d'affecter la variable dépendante (celle que l'on cherche à expliquer), mais n'est pas introduite dans le modèle estimé. Les effets de cette variable seront captés dans les termes d'erreurs. Si une variable omise est corrélée avec une variable explicative observée (la densité par exemple), alors le terme d'erreurs est corrélé avec cette variable explicative et le coefficient associé à cette dernière est donc biaisé.

Par exemple, en cas d'omission d'une variable liée à la composition sectorielle des marchés du travail locaux, la variable de densité capte des effets indépendants de l'agglomération. Dans ce cas, les salaires plus élevés observés pourraient provenir en partie du fait que les zones denses sont en réalité des zones où la spécialisation sectorielle et la création de clusters d'activité sont importantes. Ce serait donc davantage un effet local lié à un secteur spécifique qu'à la concentration de l'emploi total. Si l'on ne prend pas en compte la composition ou la concentration

sectorielle dans la régression, le coefficient de la densité sera sans doute surestimé. En cas de variable omise, l'effet de la densité sur les salaires peut également être sous-estimé, si cette variable omise affecte négativement la productivité et les salaires (phénomènes de congestion par exemple).

En outre, les caractéristiques intrinsèques inobservables des zones d'emploi, de leurs travailleurs et de leurs entreprises, sont fréquemment à l'origine de biais de variables omises. L'hétérogénéité inobservable des firmes, autant du point de vue de la gouvernance d'entreprise (de la politique salariale notamment), qu'en matière de qualité d'ajustement de l'emploi au travers des relations entre employeurs et employés, implique des différences de salaires moyens. Les caractéristiques inobservables des travailleurs, notamment les capacités intellectuelles intrinsèques, le capital social, et certaines aptitudes personnelles (motivation, dynamisme, esprit d'équipe, etc.), affectent également la productivité et les salaires moyens. Ces éléments, non pris en compte dans la spécification, se retrouvent dans le terme d'erreur et peuvent être corrélés autant avec la variable dépendante qu'avec les variables explicatives, ce qui pose nécessairement des biais d'estimation.

Une méthode intuitive de correction de ces biais de variables omises repose sur l'intégration de variables de contrôle. Il s'agit de prendre en compte la composition sectorielle locale, la situation géographique de la zone, la structure démographique de la population, etc. Cependant, la spécification peut devenir très longue sans garantir pour autant la totale absence de variables omises – observables ou inobservables. L'ajout d'un trop grand nombre de contrôles peut également créer des biais de multicolinéarité.

Pour réduire les risques d'endogénéité provenant de variables omises, neuf variables de contrôle sont intégrées à l'équation (2).

- Deux autres indicateurs de la taille des marchés du travail locaux sont ajoutés :
 - En plus de la densité, une seconde variable de superficie de la zone d'emploi (en km²) est ajoutée aux spécifications. Une zone d'emploi peut être très étendue, avec un marché global important, sans forcément être dense. Bien que les variables de densité et de superficie soient relativement interconnectées, l'introduction de la deuxième fournit davantage d'informations et permet de dissocier l'influence de ces deux déterminants.
 - De plus, les effets d'agglomération ne sont pas nécessairement limités à une zone géographique délimitée et son marché local. Généralement, la littérature prend en compte les effets d'accès aux marchés externes à l'aide d'une variable de « potentiel marchand », développée par le géographe Harris (1954). Il s'agit de la somme des densités des activités économiques des marchés externes, divisée par la distance qui les sépare de chaque zone d'emploi – comme approximation des coûts de transport.

$$PM_{ZE} = \sum_{Z \neq ZE} \frac{Dens_emp_z}{Distance_{z,ZE}}$$

Les gains d'agglomération pouvant dépasser la géographie d'une zone d'emploi, l'introduction du potentiel marchand permet de prendre en compte, en plus des effets liés à la taille et la densité du marché local, la part des variations de salaires liée aux interactions avec les autres zones (Combes *et al.*, 2011, 2015).

- Un contrôle pour l'âge moyen de la population locale (et l'âge moyen au carré) est également ajouté¹⁰. Il s'agit de tenir compte de l'expérience moyenne de la main-d'œuvre locale et de compenser le fait que le ratio de diplômés est construit sur l'ensemble de la population de 15 ans ou plus.

¹⁰ La variable d'âge moyen est centrée avant d'être élevée au carré dans le but d'éviter des problèmes de colinéarité.

- Enfin, cinq nouvelles variables de contrôle pour mesurer la part des grands secteurs dans l'emploi local sont intégrées (Agriculture, Industrie, Construction, Tertiaire marchand, Tertiaire non marchand). L'agglomération spatiale de plusieurs entreprises d'un même secteur entraîne nécessairement des externalités intra-sectorielles qui engendrent des gains de productivité. Les effets d'urbanisation peuvent donc être exagérés lorsque l'on omet les effets de spécialisation ou de composition sectorielle locale.

Le modèle MCO final se base sur l'équation de salaire suivante :

$$\ln \text{ Salaire moyen CSP}_{ZE} = \alpha + \beta \ln \text{ Densite}_{ZE} + \gamma \text{ Ratio_Diplomes}_{ZE} + \sigma \text{ Superficie}_{ZE} + \delta \ln \text{ PM}_{ZE} + \zeta_s \ln \text{ Specialisation}_{s,ZE} + \vartheta \text{ AgeMoyen}_{ZE} + \iota \text{ AgeMoyen}_{ZE}^2 + \varepsilon_{ZE} \quad (1)$$

avec Superficie_{ZE} la surface totale de la zone d'emploi ZE ; $\ln \text{PM}_{ZE}$ le log du potentiel marchand de la zone d'emploi ZE qui capte les effets de diffusion de la densité des activités économiques (plus une zone d'emploi est proche d'autres marchés très denses en emploi, plus cet indice est élevé) ; $\ln \text{Specialisation}_{s,ZE}$ le log de la part de l'emploi salarié de la zone ZE dans chacun des cinq secteurs s (Agriculture, Industrie, Construction, Tertiaire marchand, Tertiaire non marchand) ; AgeMoyen_{ZE} l'âge moyen de la population de 15 ans et plus dans la zone ZE .

Les résultats d'estimations sont présentés, par CSP, dans les colonnes 1 et 2 des tableaux 8 à 11 en Annexe 3. L'ajout de variables de contrôle détériore la significativité statistique des effets de la densité de l'emploi ; les primes salariales semblent davantage provenir des effets de capital humain et de potentiel marchand.

Néanmoins, cette spécification reste insuffisante face à une seconde source d'endogénéité provenant d'une potentielle causalité inverse (ou circulaire).

Causalité inverse

Ici, la causalité inverse correspondrait à une sorte de tri spatial des travailleurs (ou des entreprises) en fonction des salaires (ou de la rentabilité). Dès lors qu'elles dépendent des caractéristiques intrinsèques et des décisions de localisation des entreprises et des travailleurs, les trois variables explicatives de concentration de diplômés, de densité de l'emploi et de potentiel marchand présentent de sérieux risques d'endogénéité.

Les travailleurs qualifiés – mobiles – peuvent être attirés par des zones dans lesquelles ils espèrent un salaire et un confort de vie plus élevés (aménagements urbains, activité culturelle, etc.). Alors, ce n'est plus la densité ou la concentration de diplômés qui augmentent la productivité et les salaires, mais la productivité et les salaires qui augmentent la densité et la concentration de diplômés : la causalité est inversée. Ce biais d'endogénéité est évidemment exacerbé lorsque la variable d'intérêt concerne spécifiquement les populations relativement plus qualifiées – mieux rémunérées – et leur répartition géographique. Une récente étude de l'Insee montre par exemple une relation positive entre les évolutions de la part relative des diplômés et de leurs salaires relatifs entre 1991 et 2001 et suggère finalement que la concentration des diplômés dans les marchés du travail denses est en partie liée à l'amélioration de la demande relative de travailleurs qualifiés et à une moindre baisse du salaire relatif (Charnoz et Orand, 2016). Dans ce cas, c'est bien l'évolution de la demande et des salaires qui attire – ou maintient – les diplômés dans les zones denses. Cela ne veut pas dire pour autant que la densité et la concentration de diplômés n'ont plus d'effet sur les salaires. Les travailleurs (qualifiés) attirés vers les zones denses contribuent à une nouvelle augmentation de l'emploi et de la concentration de diplômés dans ces zones, ce qui, s'il y a bien des gains d'agglomération, améliore encore les salaires et donc attire de nouveaux travailleurs : la causalité est circulaire.

Sans corriger totalement ces biais d'endogénéité, les équations de salaires par CSP réduisent en partie le problème de causalité inverse associé au ratio de diplômés, au moins pour les catégories moins qualifiées. S'il est possible que les travailleurs qualifiés soient attirés par les zones présentant des salaires moyens des cadres et professions intellectuelles supérieures plus

élevés (source de causalité circulaire), il paraît moins probable qu'un diplômé préfère une zone d'emploi uniquement parce que la rémunération moyenne des emplois moins intensifs en travail qualifié y est plus importante.

La méthode généralement utilisée pour traiter les problèmes d'endogénéité liés à une causalité inverse (en même temps que les problèmes de variables omises) réside dans l'instrumentation. Il s'agit d'identifier des variables instrumentales (ou instruments) les plus corrélées possible avec les variables explicatives suspectées d'endogénéité, mais les plus exogènes possible du phénomène étudié – les salaires moyens. Ainsi, les instruments n'ont d'influence sur la variable expliquée (le salaire moyen par CSP) uniquement au travers de l'effet qu'ils ont sur les variables suspectées d'endogénéité (le ratio de diplômé ou la densité de l'emploi et le potentiel marchand). Ils permettent alors d'identifier un effet causal.

Une série de cinq instruments est utilisée ici :

- En raison de la tendance séculaire à l'urbanisation, les densités actuelles (d'emploi ou de population) sont liées aux évolutions de populations historiques. La **densité de population en 1806** (Base Cassini, Ehess¹¹) peut donc être utilisée comme instrument, sous l'hypothèse qu'elle n'a pas d'influence sur les salaires 2012 autre que celle qui transite par la détermination des densités actuelles.
- Un second instrument est construit à partir des données sur les Monuments historiques (ministère de la Culture et de la Communication). Selon un raisonnement similaire, le **nombre d'immeubles religieux** (principalement des églises) **protégés au titre des Monuments historiques par km²** reflète la localisation des populations historiques et donc détermine les densités observées en 2012, sans avoir d'autre effet sur les chocs de productivité et de salaires actuels.
- La concentration de diplômés est également corrélée avec les pourcentages de diplômés historiques au niveau des zones d'emploi, notamment en raison de l'inertie géographique relative à l'augmentation des diplômés depuis les années 1960 (en partie due à des inégalités territoriales d'accès à l'enseignement supérieur ; voir Dherbécourt, 2015). La **variation de la part des diplômés du supérieur dans la population de 16 ans et plus entre 1968 et 1982** est donc utilisée comme variable instrumentale.
- Certains travaux utilisent également des **variables reconstruites** pour l'instrumentation, comme développées par E. Moretti (2004c)¹². Ici, un instrument de la part des diplômés en 2011 est reconstruit à partir des structures démographiques des zones d'emploi en 1999. Il s'agit, dans un premier temps, de prédire la structure de la population par tranche d'âge en 2011 sur la base de la part de chaque groupe d'âge dans la population en 1999¹³ – en supposant l'absence de mortalité et de mobilité au sein de chaque zone d'emploi. Dans un second temps, la part des diplômés du supérieur long pour chaque tranche d'âge au niveau national en 2011 est affectée à la structure d'âge précédemment obtenue pour chaque zone d'emploi. La variable obtenue s'apparente au ratio de diplômés que l'on devrait observer en 2011 en l'absence de mortalité et de mobilité depuis 1999 dans chaque zone et si chacun avait le niveau d'éducation moyen national de son groupe d'âge.
- Sur le modèle de la précédente, une seconde variable est reconstruite en variation, entre 2004 et 2011, toujours selon les structures d'âges en 1999. De par sa construction, cette variation de qualifiés est seulement déterminée par la structure d'âge décalée et reste exogène des différences de demande de travail qualifié en 2011.

¹¹ Nous remercions Claude Motte pour la fourniture de ces données.

¹² Moretti adapte la méthode de Katz et Murphy (1992) pour construire un instrument dit « à la Bartik (1991) » qui est le premier à développer cette méthode.

¹³ Une approximation est faite à 2 ans près : les personnes âgées entre 5 et 9 ans en 1999 sont considérées comme appartenant à la tranche 15-19 ans en 2011.

Corrélations des instruments

	Ratio Diplômés	Ln Densité de l'emploi	Ln Potentiel Marchand
Ln Densité de population 1806	0.21	0.64	0.34
Densité Monument Historiques religieux	0.35	0.44	0.47
Variation ratio de diplômés 1968-1982	0.84	0.49	0.29
Ratio de diplômés selon structure démographique 1999	0.53	0.64	0.42
Variation du ratio de diplômé entre 2004 et 2011 selon structure démographique 1999	0.51	0.57	0.30

La méthode employée est celle des moindres carrés en deux étapes (2SLS) : 1) on régresse successivement les variables suspectées d'endogénéité (densité de l'emploi, ratio de diplômés et potentiel marchand) sur les instruments choisis et les autres variables explicatives exogènes ; 2) on régresse la variable expliquée (les salaires moyens) sur les prédictions de premières étapes et les variables explicatives exogènes.

Les résultats d'estimations sont présentés dans les colonnes 3 et 4 des tableaux 8 à 11 en Annexe 3. Les tests de Wald sur les instruments confirment la corrélation avec le ratio de diplômés, le log de la densité de l'emploi et le log du potentiel marchand, avec des *F-statistic* de premières étapes supérieures à 10 – valeur seuil proposée par Staiger and Stock (1997) en dessous laquelle les instruments peuvent être considérés comme faibles. Cela est confirmé par l'examen des statistiques de Cragg-Donald, supérieures aux valeurs critiques proposées par Stock et Yogo (2005). Les p-value de la statistique de Sargan (ou la J-statistique de Hansen) ne permettent pas de rejeter l'hypothèse nulle du test, ce qui aurait impliqué de sérieux doutes quant à la validité des instruments.

Les modèles à variables instrumentales confirment l'importance des effets de capital humain, tandis que les effets de densité semblent davantage provenir de l'accès à d'autres zones très denses plutôt que de la densité de l'emploi locale. La prise en compte des spécialisations sectorielles tend à réduire l'ampleur des effets de capital humain pour les CSP professions intermédiaires, employés et ouvriers. Dans le cas des CSP cadres, notons que les résultats des tests d'endogénéité de Wu-Hausman indiquent que le modèle à variables instrumentales n'est pas vraiment préférable aux MCO – les coefficients sont néanmoins plutôt similaires.

3.4. Composition des qualifications au sein des CSP

Enfin, un dernier écueil provient des différences dans la composition des qualifications d'une même CSP selon les différents territoires. Par exemple, toute chose égales par ailleurs, les employés pourraient être relativement plus qualifiés dans certaines zones d'emploi. Alors, l'impact de la part des diplômés du supérieur long sur les salaires moyens des employés pourrait simplement capter le fait que les travailleurs de cette CSP sont plus qualifiés dans la zone d'emploi, donc en moyenne plus productifs et mieux rémunérés. Il s'agirait donc d'un effet de composition des CSP plutôt que d'externalités de capital humain.

Pour corriger ce nouveau biais d'identification, le modèle est reconstruit selon une procédure en deux étapes. Dans un premier temps, les salaires moyens, pour chaque CSP, sont régressés sur les pourcentages des personnes de la CSP correspondante par niveau de diplôme¹⁴ dans chaque zone d'emploi. Les résidus de ces premières régressions décrivent la part des salaires moyens qui n'est pas expliquée par la structure des qualifications au sein de chaque CSP (appelée *In Salaire moyen CSP_{ZE}'*). La seconde étape consiste à développer l'équation (1) non plus sur les salaires moyens, mais sur les résidus de première étape¹⁵. On mesure alors si ce

¹⁴ Les parts des niveaux de diplôme au sein de chaque CSP sont construites selon quatre catégories sur la base des données de recensement de population de l'Insee 2011 (voir Annexe 2 Données).

¹⁵ Une procédure bootstrap est mise en œuvre ici (500 répliquations).

salaire ajusté des effets de composition intra-CSP connaît des effets d'agglomération. L'équation de détermination devient :

$$\ln \text{ Salaire moyen CSP}_{ZE}' = \alpha + \beta \ln \text{ Densite}_{ZE} + \gamma \text{ Ratio_Diplomes}_{ZE} + \sigma \text{ Superficie}_{ZE} + \delta \ln \text{ PM}_{ZE} + \zeta_s \ln \text{ Specialisation}_{s,ZE} + \vartheta \text{ AgeMoyen}_{ZE} + \iota \text{ AgeMoyen}_{ZE}^2 + \varepsilon_{ZE} \quad (2)$$

Les résultats de ces nouvelles spécifications sont présentés dans les colonnes 5 et 6 des tableaux 8 à 11 en Annexe 3. Comme précédemment, l'examen des statistiques de test confirme la validité du modèle à variables instrumentales.

Comme pressenti, les effets de capital humain ont tendance à diminuer selon ces nouvelles spécifications. La significativité statistique disparaît dans le cas des CSP ouvriers.

Nous retenons la spécification sur résidus avec contrôle de spécialisation sectorielle (colonne 6 des tableaux 8 à 11) comme modèle préféré.

4. Résultats

Les Tableaux 5 (coefficients d'estimations) et 6 (gains escomptés au regard des rapports interdéciles) ci-dessous présentent les résultats d'estimations des modèles à variables instrumentales basés sur les équations (1) et (2), avec en ligne les différentes variables d'intérêt, et en colonne les quatre spécifications par CSP.

4.1. Des effets importants de la concentration de diplômés révélant des externalités de capital humain

D'après le modèle préféré (2), l'augmentation d'un point de pourcentage de la part des diplômés du supérieur long améliore le salaire moyen des CSP cadres de plus de 0,6 %, tandis que les salaires moyens des CSP professions intermédiaires et employés augmentent de respectivement 0,5 % et plus de 0,3 % (voir Tableau 5). Le modèle ne permet pas d'identifier d'externalités de capital humain bénéficiant aux CSP ouvriers¹⁶.

- Alors, on peut s'attendre à des primes salariales d'au moins 5 % pour les cadres, de plus de 4 % pour les professions intermédiaires et d'au moins 3 % pour les employés, lorsque l'on compare une zone des 10 % les plus concentrées en diplômés du supérieur long (zone d'emploi de Dijon par exemple), par rapport à une zone des 10 % les moins concentrées (zone d'emploi de Gien par exemple) (voir Tableau 6).

Ces résultats diffèrent de ceux de Moretti (2004c) qui montraient, aux États-Unis, que la concentration des diplômés du supérieur dans une zone urbaine avait un impact relativement plus important sur les salaires des moins qualifiés que sur les salaires des plus qualifiés. Le résultat inverse ici pourrait trouver une explication dans l'existence d'un salaire minimum. Les effets d'externalités agiraient davantage sur le taux d'emploi et/ou de participation que sur le niveau de salaire, en particulier dans les catégories moins « qualifiées ».

Notons par ailleurs que ces effets relativement plus faibles dans les CSP professions intermédiaires et employés impliquent qu'une augmentation de la concentration de diplômés du supérieur long dans une zone d'emploi accroît quelque peu les inégalités salariales entre CSP dans cette zone, toutes choses égales par ailleurs.

¹⁶ La significativité de ce coefficient lorsqu'il est estimé à partir du modèle (1) semble indiquer que, pour les ouvriers, les primes salariales engendrées par un niveau de capital humain plus élevé proviendraient davantage d'un effet de composition des qualifications à l'intérieur de cette CSP.

Tableau 5. Effets d'agglomération et externalités de capital humain, 2012

	Cadres...	Professions intermédiaires	Employés	Ouvriers
Régression IV Salaires moyens (1)				
<i>Part de la population (15+) diplômée (supérieur long)</i>	1.060*** (0.132)	0.371*** (0.127)	0.655*** (0.0718)	0.355*** (0.115)
<i>Log de la densité de l'emploi</i>	-0.00255 (0.00646)	-0.0172*** (0.00620)	-0.0153*** (0.00352)	-0.0209*** (0.00563)
<i>Log du Potentiel Marchand</i>	0.0631*** (0.0112)	0.0513*** (0.0108)	0.0517*** (0.00610)	0.0324*** (0.00978)
<i>Observations</i>	304	304	304	304
Régression IV Salaires moyens' (2)				
<i>Part de la population (15+) diplômée (supérieur long)</i>	0.623*** (0.151)	0.500*** (0.125)	0.328*** (0.0928)	0.147 (0.118)
<i>Log de la densité de l'emploi</i>	-0.0116 (0.00741)	-0.0139** (0.00612)	-0.0142*** (0.00465)	-0.0203*** (0.00579)
<i>Log du Potentiel Marchand</i>	0.0849*** (0.0129)	0.0349*** (0.0106)	0.0621*** (0.00736)	0.0234** (0.0100)
<i>Observations</i>	304	304	304	304

Erreurs-types entre parenthèses.

*** significatif au seuil de 1 %, ** significatif au seuil de 5 %, * significatif au seuil de 10 %.

Tableau 6. Gains escomptés au regard des rapports interdéciles

	Cadres...	Professions intermédiaires	Employés	Ouvriers
Régression IV Salaires moyens (1)				
<i>Part de la population (15+) diplômée (supérieur long)</i>	9 % ***	3 % ***	5 % ***	3 % ***
<i>Log de la densité de l'emploi</i>	NS	-4 % ***	-3 % ***	-4 % ***
<i>Log du Potentiel Marchand</i>	7 % ***	6 % ***	6 % ***	4 % ***
Régression IV Salaires moyens' (2)				
<i>Part de la population (15+) diplômée (supérieur long)</i>	5 % ***	4 % ***	3 % ***	NS
<i>Log de la densité de l'emploi</i>	NS	-3 % **	-3 % ***	-4 % ***
<i>Log du Potentiel Marchand</i>	10 % ***	4 % ***	7 % ***	3 % **

Lecture : d'après les résultats du modèle préféré (2), une variation du ratio de diplômés correspondant à l'écart interdéciles (8) engendre une prime salariale de 5 % pour les cadres ; de même, une variation de la densité correspondant au ratio interdéciles (8) engendre une dégradation du salaire des professions intermédiaires de -3 % ; enfin, une variation du potentiel marchand correspondant au ratio interdéciles (3) engendre une prime salariale de 10 % pour les cadres.

Notes : gains (ou pertes) de salaires moyens à attendre en comparant une zone d'emploi appartenant au premier décile (de la distribution de la variable correspondante), à une zone d'emploi appartenant au dernier décile ; les ratios interdéciles (D9/D1) de la densité et du potentiel marchand sont de respectivement 8 et 3 ; l'écart interdéciles (D9-D1) du ratio de diplômés est de 8. Par exemple pour le modèle (2), pour le salaire moyen des CSP employés : $3\% = e^{0,08 \times 0,328} - 1$; $-3\% = -8^{0,0142} - 1$; $7\% = 3^{0,0621} - 1$

*** significatif au seuil de 1 %. ** significatif au seuil de 5 %. * significatif au seuil de 10 %. NS : non significatif.

En définitive, lorsque l'on contrôle des effets de composition et des biais de causalité, les externalités de capital humain semblent jouer un rôle majeur dans l'explication des gains de salaires moyens entre zones d'emploi, à CSP données. Celles-ci dominent l'effet de la densité de l'emploi qui est légèrement négatif – l'effet n'est pas statistiquement significatif dans le cas des CSP cadres. Néanmoins, l'influence de la densité semble passer davantage par l'accès à d'autres marchés très denses en emploi, effet repéré par le potentiel marchand qui affiche des élasticités de 0,08 (cadres), 0,03 professions intermédiaires, 0,06 (CSP employés) et 0,02 (ouvriers).

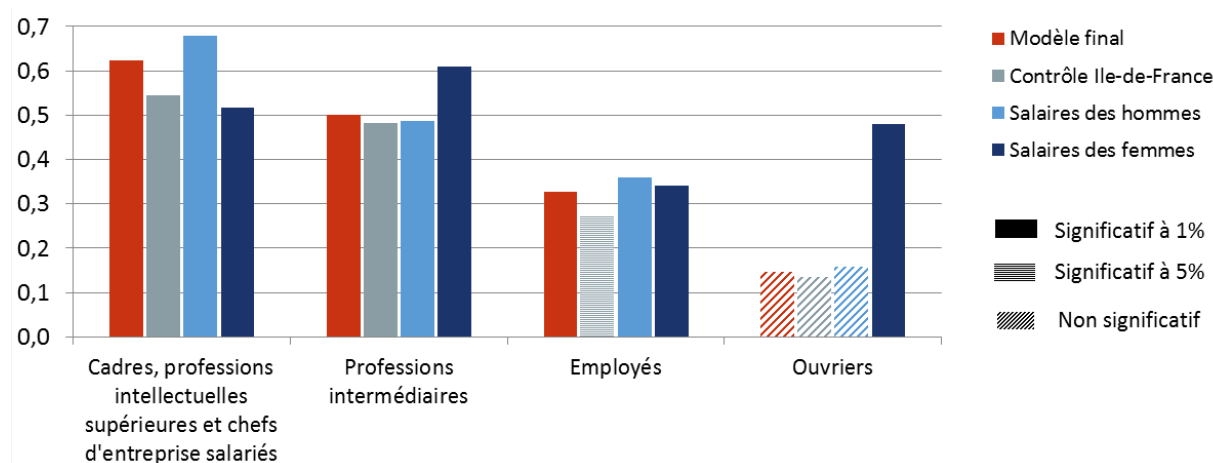
4.2. Robustesse

Pour vérifier la robustesse de ces principaux résultats, une série de spécifications alternatives sont développées à partir du modèle préféré (2).

Dans un premier temps, il s'agit de refaire les estimations en utilisant les salaires moyens des femmes puis les salaires moyens des hommes, toujours en distinguant les quatre CSP. Ces spécifications par genre montrent des effets différents à CSP données. Les semi-élasticités du salaire à la part de diplômés (du supérieur long) sont sensiblement les mêmes pour les hommes et pour les femmes (colonnes 7 et 8 des tableaux 8 à 11 en annexe 3). Néanmoins, pour les CSP ouvriers, les femmes semblent bénéficier d'externalités de capital humain, avec un coefficient de 0,48 statistiquement significatif au seuil de 1 % ; à l'inverse, les salaires des hommes ouvriers ne semblent pas bénéficier d'externalités. Les résultats sont donc robustes à une différenciation des salaires par genre.

Le modèle est également réestimé en intégrant une indicatrice d'appartenance à la région Île-de-France. Il s'agit de regarder si les résultats ne sont pas tirés par les zones d'emploi d'Île-de-France. L'existence d'externalités de capital humain n'est pas remise en cause par l'ajout de cette indicatrice – néanmoins, les coefficients associés aux ratios de diplômés pour les cadres et les employés ne sont significatifs qu'à 5 % et 1 % respectivement. L'ampleur des semi-élasticités diminue légèrement. Notons que la significativité des élasticités de la densité et du potentiel marchand aux salaires est très dégradée par l'introduction de l'indicatrice Île-de-France. Les externalités de capital humain ne sont pas un phénomène uniquement lié à la présence de la région Île-de-France – et des caractéristiques spécifiques des zones d'emploi qui la composent : plus petites, plus denses, plus « qualifiées » notamment. Ces résultats sont tout de même à interpréter avec prudence au regard de la faiblesse des instruments (voir statistiques de Wald et *F-statistic* de première étape dans les tableaux en Annexe 3, colonne 9).

Graphique 1 : Robustesse des estimations de l'impact de la concentration des diplômés du supérieur long sur les salaires des CSP



Finalement, des méthodes d'économétrie spatiale sont mises en œuvre pour tenir compte d'une éventuelle autocorrélation spatiale des données (voir Annexe 5). Les statistiques révèlent des risques d'autocorrélation spatiale pour les variables de salaires et pour les résidus d'estimations, notamment lorsque l'on utilise une définition relativement étroite du « voisinage », à l'aide d'une matrice de contiguïté. Suite à la mise en œuvre de méthodes spécifiques pour capter ces effets de voisinage, la significativité des résultats quant à l'effet du ratio de diplômés sur les salaires n'est pas modifiée. Les effets mesurés sont donc bien robustes au problème d'autocorrélation spatiale et nos résultats finaux ne proviennent pas de phénomènes de débordements géographiques qui ne seraient pas captés par le modèle.

Conclusion

La présente étude propose un modèle d'estimation des déterminants des salaires moyens en 2012 dans les 304 zones d'emploi de France métropolitaine dans le but de mieux comprendre la nature et les causes des inégalités territoriales de productivité et de salaires – à CSP donnée. L'analyse a mis l'accent sur l'existence d'externalités de capital humain qui engendrent des primes salariales dans les zones où la part des diplômés du supérieur long dans la population est importante. Pour obtenir une mesure plus robuste de ces effets, un ensemble de contrôles et de corrections économétriques a été mis en œuvre afin de traiter certains biais d'estimation, et en particulier les problèmes d'identification liés à la présence d'effets de composition, et à de potentielles causalité inverse et variables omises. En définitive, la stratégie économétrique repose sur une procédure en deux étapes afin d'estimer les effets d'agglomération sur des salaires ajustés de la composition des qualifications de chaque CSP à l'aide d'un modèle à variables instrumentales.

L'analyse empirique montre que la densité des activités économiques ne suffit pas à expliquer les écarts de productivité et de salaires. Sans nier l'existence du fait métropolitain, les résultats montrent qu'une bonne partie des écarts de salaire entre zones d'emploi en 2012, à CSP données, viennent en premier lieu de l'existence d'externalités de capital humain. Ce constat va dans le sens des travaux de Combes *et al.* (2011) selon lesquels un niveau d'éducation plus élevé est la principale force des effets d'agglomération dans les départements français en 2000.

D'après le modèle préféré, l'augmentation d'un point de pourcentage de la part des diplômés du supérieur long améliore le salaire moyen des cadres de plus de 0,6 %, tandis que les salaires moyens des professions intermédiaires augmentent de 0,5 % et les salaires des employés de plus de 0,3 %. Mais, l'étude ne permet pas d'identifier précisément ce qui déclenche ces effets d'externalités positifs. Une partie de ces effets est liée aux interactions directes des travailleurs qualifiés entre eux, et avec le reste de la population. Celles-ci favoriseraient la diffusion des connaissances et des innovations ainsi que les mécanismes d'apprentissage « sur le tas ». Ces effets (de type « Silicon Valley ») bénéficient potentiellement à tous, mais ils n'expliquent pas la totalité des gains de salaires et de productivité associés à la présence de diplômés. À cela s'ajoute probablement un effet de « complémentarité » entre les travailleurs plus qualifiés et les moins qualifiés. Ainsi, une forte densité de travailleurs qualifiés peut accroître la demande de services spécifiques (on peut penser à la garde d'enfants, ou à d'autres prestations de services comme les restaurants, les loisirs) et les salaires de ces professions.

Une part importante de diplômés, souvent concomitante à une forte densité des activités économiques, peut également avoir des inconvénients et limiter, *in fine*, les gains d'agglomération. Ainsi, dans les estimations économétriques, une fois pris en compte la part des très qualifiés, la spécialisation sectorielle et la structure démographique des zones d'emploi, la densité des activités économiques semble jouer négativement sur les salaires. Cela peut venir d'effets de congestion et/ou de la difficulté à installer dans ces zones certaines activités économiques (encombrement des réseaux de transport, majoration du foncier et de l'immobilier, pollution, etc.).

Finalement, si les primes salariales liées à la concentration de diplômés sont importantes, cela ne signifie pas qu'il soit nécessairement rentable pour un salarié de changer de zone d'emploi. Aux coûts de déménagement ou déplacements journaliers peut s'ajouter le surcoût de la vie dans les zones les plus denses, en particulier le coût du logement.

Références bibliographiques

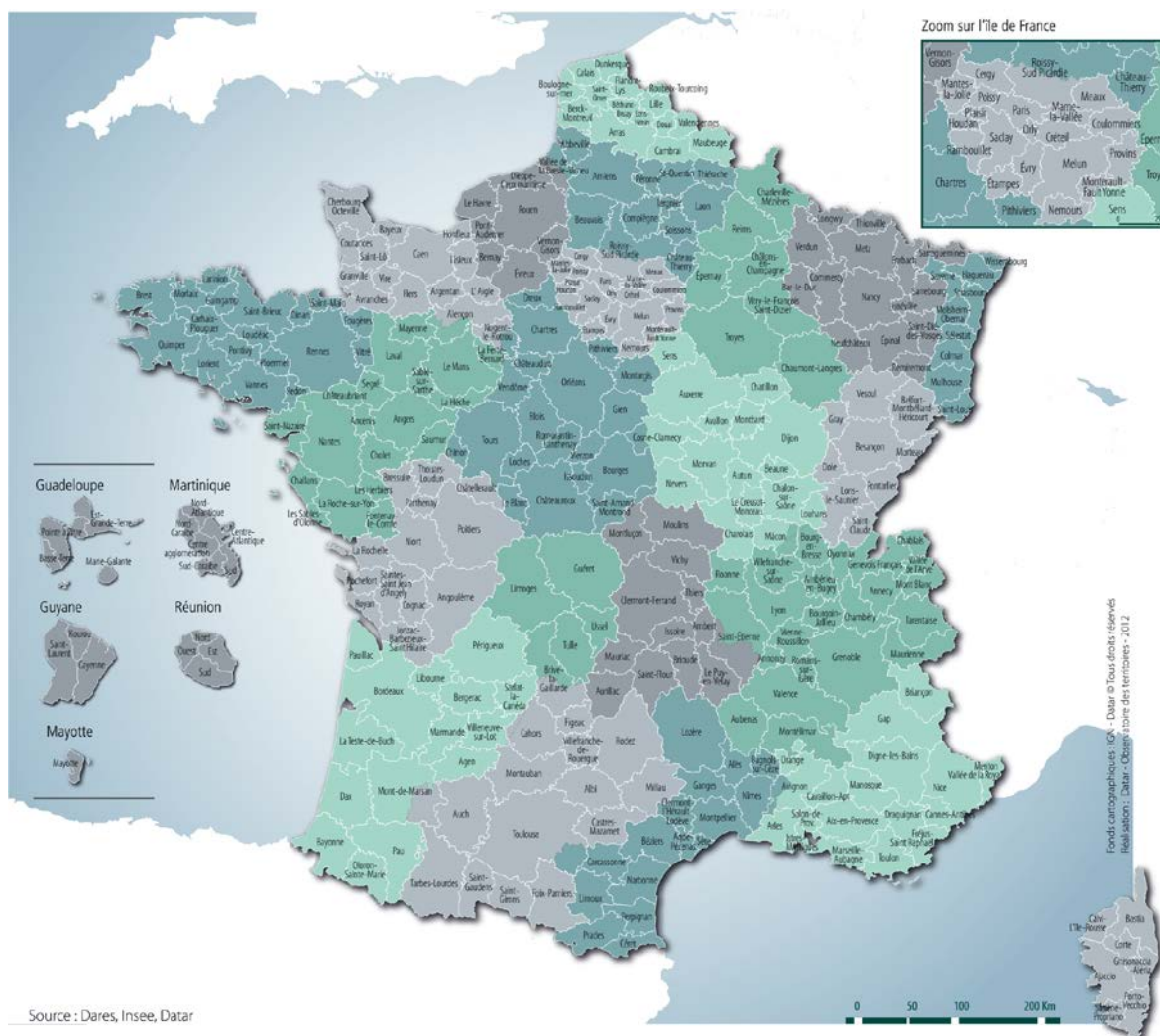
- Abel J. R., Dey I. et Gabe T. M. (2012), « [Productivity and the density of human capital](#) », *Federal Reserve Bank of New York Staff Reports*, n° 440, March 2010 ; revised September 2011.
- Acemoglu D. et Angrist J. (2000), « [How large are human-capital externalities? Evidence from compulsory schooling](#) », *NBER Macroeconomics Annual*, vol. 15, The University of Chicago Press, p. 9-59.
- Brühlhart M. et Mathys N. A. (2008), « Sectoral agglomeration economies in a panel of European regions », *Regional Science and Urban Economics*, vol. 38 (4), p. 348-362.
- Charnoz P. et Orand M. (2016), « [Une concentration spatiale accrue des diplômés du supérieur en France entre 1991 et 2011](#) », *Emploi, chômage, revenus du travail, édition 2016*, Insee Références - Dossier.
- Chauvin J. P., Glaeser E., Ma Y. et Tobio K. (2016), « [What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States](#) ». *NBER Working Paper*, n° 22002, National Bureau of Economic Research, Inc.
- Ciccone A. (2002), « Agglomeration effects in Europe », *European Economic Review*, vol. 46 (2), p. 213-227.
- Ciccone A. et Hall R. E. (1996), « Productivity and the density of economic activity », *American Economic Review*, vol. 86 (1), p. 54-70.
- Combes P.-P., Duranton G. et Gobillon L. (2008), « Spatial wage disparities: Sorting matters! », *Journal of Urban Economics*, vol. 63 (2), Elsevier, mars, p. 723-742.
- Combes P.-P., Duranton G., Gobillon L. et Roux S. (2010), « Estimating agglomeration effects with history, geology, and worker fixed-effects », in Glaeser E. L. (éd.), *Agglomeration Economics*, The University of Chicago Press, p. 15-65.
- Combes P.-P., Lafourcade M., Thisse J.-F. et Toutain J.-C. (2011), « The rise and fall of spatial inequalities in France: A long-run perspective », *Explorations in Economic History*, vol. 48 (2), Elsevier, p. 243-271.
- Combes P.-P. et Lafourcade M. (2012), *Revue de la littérature académique quantifiant les effets d'agglomération sur la productivité et l'emploi*, Rapport final, réalisé pour la Société du Grand Paris.
- Combes P.-P., Gobillon L. et Lafourcade M. (2015a), « [Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du Grand Paris](#) », *CEPREMAP Working Papers (Docweb)*, n° 1504, CEPREMAP.
- Combes P.-P. et Gobillon L. (2015b), « The empirics of agglomeration economies », in Duranton G., Henderson V. et Strange W. (eds.), *Handbook of Regional and Urban Economics*, vol. 5A, Elsevier, Amsterdam, p. 247-348.
- Combes P.-P., Gobillon L. et Lafourcade M. (2016), « [Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du Grand Paris](#) », *CEPREMAP Working Papers (Docweb)*, n° 1602, CEPREMAP.
- Dherbécourt C. (2015), « [La géographie de l'ascension sociale](#) », *La Note d'analyse*, n° 36, France Stratégie, novembre.
- Di Addario S. et Patacchini E. (2008), « Wages and the city: Evidence from Italy », *Labour Economics*, vol. 15 (5), p. 1040-1061.

- Duranton G. et Puga D. (2004), « Micro-foundations of urban agglomeration economies », in Henderson J. V. et Thisse J.-F. (eds.), *Handbook of Regional and Urban Economics*, édition 1, vol. 4, chapitre 48, Elsevier, p. 2063-2117.
- Foster N. et Stehrer R. (2009), « Sectoral productivity, density and agglomeration in the wider Europe », *Spatial Economic Analysis*, vol. 4 (4), p. 427-446.
- Le Gallo J. (2002), « [Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire](#) », *Économie & Prévision*, 2002/4, n° 155, p. 139-157.
- Glaeser E. L. et Saiz A. (2004), « The rise of the skilled city », *Brookings-Wharton Papers on Urban Affairs*, 5 (2004), p. 47-94.
- Groot S. P. T. et de Groot H. L. F. (2014), « [Estimating the skill bias in agglomeration externalities and social returns to education: Evidence from Dutch matched worker-firm micro-data](#) », *Discussion Paper*, n° 2014-088, Tinbergen Institute.
- Katz L. F. et Murphy K. M. (1992), « Changes in relative wages, 1963-1987: Supply and demand factors », *The Quarterly Journal of Economics*, vol. 107 (1), p. 35-78.
- Kelejian H. H. et Prucha I. R. (2010). « Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances », *Journal of Econometrics*, vol. 157 (1), p. 53-67.
- Lucas R. E. (1988), « On the mechanics of economic development », *Journal of Monetary Economics*, vol. 22 (1), p. 3-22.
- Melo P. C., Graham D. J. et Noland R. B. (2009), « A meta-analysis of estimates of urban agglomeration economies », *Regional Science and Urban Economics*, vol. 39 (3), p. 332-342.
- Moretti E. (2004a), « Human capital externalities in cities », in Henderson V. et Thisse J.-F. (eds), *Handbook of Regional and Urban Economics*, vol. 4, NorthHolland, Amsterdam, p. 2243-2291.
- Moretti E. (2004b), « Workers' education, spillovers, and productivity: Evidence from plant-level production functions », *American Economic Review*, vol. 94 (3), p. 656-690.
- Moretti E. (2004c), « Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data », *Journal of Econometrics*, vol. 121 (1-2), Elsevier, p. 175-212.
- Morikawa M. (2011), « Economies of density and productivity in service industries: An analysis of personal service industries based on establishment-level data », *Review of Economics and Statistics*, vol. 93 (1), p. 179-192.
- Rauch J. E. (1993), « Productivity gains from geographic concentration of human capital: Evidence from the cities », *Journal of Urban Economics*, vol. 34 (3), p. 380-400.
- Rosenthal S. S. et Strange W. C. (2004). « Evidence on the nature and sources of agglomeration economies », in Henderson V. et Thisse J.-F. (eds), *Handbook of Regional and Urban Economics*, vol. 4, North-Holland, Amsterdam, p. 2119-2171.
- Rosenthal S. S. et Strange W. C. (2008), « The attenuation of human capital spillovers », *Journal of Urban Economics*, vol. 64 (2), p. 373-389.
- Staiger D. et Stock J. K. (1997), « Instrumental variables regression with weak instruments », *Econometrica*, n° 65, p. 557-586.
- Stock J. H. et Yogo M. (2005), « Testing for weak instruments in linear IV regression », chapitre 5 in Stock J. H. et Andrews D. W. K. (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge University Press, Originally published 2001 as *NBER Technical Working Paper No. 284*.

ANNEXES

A1. Zonage d'étude : les zones d'emploi 2010

L'étude des disparités de salaires se base ici sur l'échelle géographique des zones d'emploi 2010.



Construit sur la base des déplacements domicile-travail, ce zonage d'étude repose moins sur les divisions administratives que sur la géographie des marchés du travail locaux, leur rayonnement territorial et leurs évolutions.

Délimitées à partir des données des recensements de la population de l'Insee, ces zones d'emplois prennent en compte les structures urbaines, la densité de l'emploi, ou encore les réseaux de transports et la topographie des territoires, pour définir des « espaces à l'intérieur desquels la plupart des actifs résident et travaillent »¹⁷. Chaque zone d'emploi est plus ou moins centrée sur une aire urbaine¹⁸, mais leurs contours ne s'en rapprochent pas nécessairement – les grandes aires urbaines peuvent s'étaler sur plusieurs zones d'emploi (Paris) tandis que les zones d'emploi moins dense peuvent contenir plusieurs pôles plus restreints.

¹⁷ ATLAS des Zones d'Emploi 2010, Dares, Datar, Insee.

¹⁸ Définie selon l'Insee comme « un ensemble de communes, d'un seul tenant et sans enclave, constitué par un pôle urbain (unité urbaine) de plus de 10 000 emplois, et par des communes rurales ou unités urbaines (couronne périurbaine) dont au moins 40 % de la population résidente ayant un emploi travaille dans le pôle ou dans des communes attirées par celui-ci ».

Les zones d'emploi, qui reflètent les marchés du travail locaux et leur fonctionnement, constituent donc un découpage géographique pertinent pour l'analyse des dynamiques territoriales et la mise en œuvre de politiques localisées. Elles sont particulièrement adaptées à l'étude des phénomènes locaux tels que les effets d'agglomération et de concentration du capital humain sur la productivité et les salaires.

Par ailleurs, étant donné que chaque zone d'emploi est formée d'un ensemble de communes entières, qu'une commune ne peut appartenir qu'à une seule zone d'emploi et que les communes d'une zone d'emploi sont contiguës (Atlas des zones d'emploi 2010, Dares, Insee, Datar), il est possible de recalculer certains indicateurs au niveau des zones d'emploi à partir des données correspondant aux communes qui les constituent.

Puisque les caractéristiques intrinsèques des zones d'emplois traduisent les dynamiques des marchés du travail et tissus économiques au niveau local, elles sont nécessairement hétérogènes, et notamment en matière de surface, de population, de densité des activités économiques ainsi que de productivité et de salaires.

A2. Données

L'échantillon à l'étude comporte les 304 zones d'emploi de France métropolitaine.

Les données sur les salaires 2012 proviennent des déclarations annuelles de données sociales (DADS) - Fichier Salariés au lieu de résidence, de l'Insee. Elles sont établies sur la base des informations recueillies sur les entreprises – privées et publiques – localisées en France¹⁹. Les statistiques sur les CSP correspondent à la catégorie socioprofessionnelle du poste principal occupé par le salarié dans l'année.

La concentration de diplômés par zone d'emploi est calculée à partir des données de recensement de la population 2011 de l'Insee. L'indicateur est calculé au niveau des zones d'emploi à partir des données communales.

Les données concernant le nombre d'emplois au lieu de travail proviennent du recensement de la population de l'Insee 2012. Les personnes employées au sens du recensement de la population sont celles ayant déclaré avoir un emploi dans le formulaire du recensement. Notion plus restrictive que celle de l'emploi au sens du BIT qui concerne toutes les personnes ayant travaillé pendant une durée quelconque (ne serait-ce qu'une heure) au cours d'une semaine de référence, l'emploi au sens du recensement peut omettre certaines activités occasionnelles qui ne sont pas déclarées au recensement par les individus. Néanmoins, certaines activités, principalement bénévoles, peuvent être déclarées au recensement de la population (voir Insee définition emploi au recensement de la population). Dans le recensement de la population de l'Insee, « le lieu de travail est la zone géographique où une personne exerce son activité professionnelle ». Par convention, certaines professions spécifiques, amenant les personnes à se déplacer plus ou moins fréquemment pour leur travail (chauffeur-routier, chauffeur de taxi, VRP, commerçant ambulancier ou marin pêcheur), sont comptabilisées dans la commune de résidence.

Les densités sont calculées à partir des superficies de chaque zone d'emploi fournies dans les séries historiques de recensement de population de l'Insee et produites par la Direction Générale des Impôts, Service du cadastre, 2011 – les données au niveau des zones d'emploi sont recalculées à partir des données communales.

Les données relatives à la spécialisation sectorielle des zones d'emploi (part de l'emploi par secteur en 2012) sont issues des estimations localisées d'emploi en France métropolitaine et

¹⁹ Insee : « L'ensemble des employeurs et de leurs salariés sont concernés par les DADS, à l'exception des particuliers employeurs et des activités extraterritoriales (division 99 de la NAF rév. 2). Sont inclus les établissements de la fonction publique à caractère industriel et commercial. Les DADS des établissements de la fonction publique de l'État ne sont cependant pas exploitées dans le cadre de l'opération DADS grand format ».

dans les Dom de l'Insee. Pour les zones d'emploi le niveau de diffusion correspond au niveau A5 de la nomenclature des activités économiques productives NAF rév. 2. L'emploi salarié est donc divisé en cinq grands secteurs : Agriculture, Industrie, Construction, Tertiaire marchand et Tertiaire non marchand.

La variable de structure par âge des zones d'emploi est construite à partir des données de population par âge quinquennal au recensement de 2012 de l'Insee. Sous l'hypothèse que les individus sont distribués de manière homogène au sein des classes d'âge, il est possible de calculer un âge moyen par zone. La population étudiée est celle des 15 ans et plus en 2012.

Le niveau des qualifications au sein de chaque CSP est construit selon quatre catégories sur la base des données de recensement de population de l'Insee 2011 :

- les personnes dont le diplôme le plus élevé ne dépasse pas le Brevet d'études professionnelles (c'est-à-dire les catégories 01 : Pas de scolarité ; 02 : Aucun diplôme mais scolarité jusqu'en école primaire ou au collège ; 03 : Aucun diplôme mais scolarité au-delà du collège ; 11 : Certificat d'études primaires ; 12 : BEPC, brevet élémentaire, brevet des collèges ; 13 : Certificat d'aptitudes professionnelles, brevet de compagnon ; 14 : Brevet d'études professionnelles) ;
- les personnes dont le diplôme le plus élevé est de niveau Bac (c'est-à-dire les catégories 15 : Baccalauréat général, brevet supérieur ; 16 : Bac technologique ou professionnel, brevet professionnel ou de technicien, BEC, BEI, BEH, capacité en droit) ;
- les personnes dont le diplôme le plus élevé est de niveau universitaire de premier cycle (c'est-à-dire la catégorie 17 : Diplôme universitaire de 1^{er} cycle, BTS, DUT, diplôme des professions sociales ou de santé, d'infirmier(ère)) ;
- les personnes dont le diplôme le plus élevé est de niveau universitaire de deuxième ou troisième cycle (c'est-à-dire la catégorie 18 : Diplôme universitaire de 2^e ou 3^e cycle (y compris médecine, pharmacie, dentaire), diplôme d'ingénieur, d'une grande école, doctorat, etc.).

La densité de population en 1806 est calculée au niveau des zones d'emploi à partir de la population communale au recensement de 1806 fournit par Claude Motte du laboratoire LaDéHis de l'EHESS.

L'instrument correspondant au nombre d'Immeubles « religieux » protégés au titre des Monuments Historiques est construit à partir de la base Liste et localisation des Immeubles protégés au titre des Monuments Historiques produite par le Ministère de la Culture et de la Communication. Sont retenus les Immeubles correspondant à un édifice religieux (« basilique », « cathédrale », « église », « mosquée », « synagogue », « temple protestant ») datant d'une période antérieur au XX^e siècle.

La variation des ratios de diplômés entre 1968 et 1982 est calculée à partir des recensements de 1968 et 1982 (exploitations complémentaires par sondage) de l'Insee. Les ratios de diplômés retardés correspondent au nombre de personnes de 16 ans et plus titulaires d'un diplôme universitaire de 1^{er}, 2^e ou 3^e cycle, parmi la population totale des 16 ans et plus. Les données au niveau des zones d'emploi sont reconstruites à partir des séries communales.

Le ratio de diplômés 2011 et la variation du ratio 2004-2011 reconstruits sur la base des structures de population de 1999 sont calculés à partir de deux séries de données :

- les données de population par âge quinquennal au recensement de 1999 de l'Insee ;
- le niveau de diplôme selon l'âge en 2011 issu de l'enquête emploi de l'Insee.

Les effectifs par zones d'emploi sont recalculés à partir des séries communales.

Tableau 7. Statistiques descriptives par zone d'emploi (2012), variables de contrôle

Variables	Moyenne	Ecart -Type	Médiane	Ratio- Interdéciles	Minimum	Zone d'emploi du minimum	Maximum	Zone d'emploi du maximum
Superficie (km ²)	1789	1376	1318	7	119	Versailles	8752	Toulouse
Potentiel Marchand	82	74	60	3	24	Porto-Vecchio	667	Créteil
Part de l'emploi dans l'agriculture (%)	2 %	2 %	1 %	13	0 %	Orly	26 %	Pauillac
Part de l'emploi dans le secteur tertiaire marchand (%)	40 %	7 %	39 %	2	24 %	Segré	66 %	Paris
Part de l'emploi dans le secteur tertiaire non marchand (%)	34 %	7 %	34 %	2	16 %	Les Herbiers	55 %	Corte
Part de l'emploi dans l'industrie (%)	18 %	8 %	17 %	3	4 %	Menton - Vallée de la Roya	44 %	La Ferté-Bernard
Part de l'emploi dans la construction (%)	7 %	2 %	7 %	2	3 %	Paris	19 %	Porto-Vecchio
Âge moyen (population de 15 ans et plus)	49	2	50	1	43	Marne-la-Vallée	56	Morvan

A3. Tableaux de résultats d'estimations

Tableau 8. Salaire net horaire moyen des CSP « Cadres, professions intellectuelles supérieures et chefs d'entreprises salariés »

	MCO				Variables Instrumentales				
	log salaire moyen				log salaire moyen' (non expliqué par le niveau des qualifications)				
	(1)	(2)	(3)	(4)	(5)	(6)	Salaires des Femmes (7)	Salaires des Hommes (8)	Indicatrice Ile de France (9)
Ratio de diplômés	0.864*** (0.104)	1.018*** (0.0970)	0.952*** (0.117)	1.060*** (0.132)	0.696*** (0.139)	0.623*** (0.167)	0.517*** (0.144)	0.678*** (0.190)	0.544** (0.228)
log Densité	0.0138*** (0.00495)	0.00676* (0.00403)	0.0110 (0.00687)	-0.00255 (0.00646)	-0.00762 (0.00905)	-0.0116 (0.00876)	-0.0153** (0.00722)	-0.00982 (0.00983)	-0.00929 (0.0113)
log Potentiel Marchand	0.0530*** (0.00619)	0.0493*** (0.00529)	0.0525*** (0.0131)	0.0631*** (0.0112)	0.0802*** (0.0152)	0.0849*** (0.0138)	0.0908*** (0.0123)	0.0859*** (0.0142)	0.0731 (0.0468)
Observations	304	304	304	304	304	304	304	304	304
R ² ajusté	0.581	0.647	-	-	-	-	-	-	-
Contrôle spécialisations sectorielles	Non	Oui	Non	Oui	Non	Oui	Oui	Oui	Oui
Instruments faibles – Cragg-Donald Wald F-Statistic*			10.921	14.791	10.921	14.791	14.791	14.791	2.074
Test des restrictions suridentifiées (Sargan p-value)**			0.8848	0.5801	0.9185	0.7931	0.8343	0.8633	0.6500
Test d'endogénéité***			0.71356	0.21622	0.07316	0.02045	0.00069	0.07486	0.38657
Régressions de première étape****									
F-stat (Ln Densité) (p-value)			37.48	77.85	94.41	77.85	77.85	77.85	75.26
F-stat (Ratio Diplômés)			94.41	41.84	37.48	41.84	41.84	41.84	41.95
F-stat (Ln PM)			15.82	18.19	15.82	18.19	18.19	18.19	4.64

Notes : modèles à variables instrumentales suivant la méthode des moindres carrés en deux étapes (2SLS) ; Variables du Ratio de diplômés, du log de la Densité de l'emploi et du log du Potentiel Marchand instrumentées ; Instruments exclus : Variation du ratio de diplômés du supérieur entre 1968 et 1982 ; Ratio de diplômés 2011 et variation du Ratio 2004-2011 reconstruits à partir de la structure par âge en 1999 ; Densité de population en 1806 ; Nombre d'immeubles religieux classés monument historique par km². *La F-statistique Cragg-Donald Wald doit être supérieure aux valeurs critiques proposées par Stock et Yogo (2005) – la valeur critique pour un biais d'instrument maximal de 5 % est de 9.53.**Les tests des restrictions suridentifiées testent l'absence de corrélation entre les instruments et les termes d'erreur (HO) ; les p-value des statistiques de Sargan (ou Hansen's J lorsque l'on corrige de l'hétéroscédasticité) sont données.***Un rejet de H0 du test d'endogénéité sur les variables instrumentées indique que ces variables pourraient être considérées comme exogènes.****Généralement, les instruments sont considéré faibles lorsque les statistiques F de première étape sont inférieures à 10 (Staiger and Stock, Econometrica, 1997).

Erreurs-types entre parenthèses. *** significatif au seuil de 1 % ; ** significatif au seuil de 5 % ; * significatif au seuil de 10 %.

Tableau 9. Salaire net horaire moyen des CSP « Professions intermédiaires »

	MCO		Variables Instrumentales						
	log salaire moyen		log salaire moyen' (non expliqué par le niveau des qualifications)						
	(1)	(2)	(3)	(4)	(5)	(6)	Salaires des Femmes	Salaires des Hommes	Indicatrice Ile de France
Ratio de diplômés	0.425*** (0.0786)	0.443*** (0.0832)	0.500*** (0.105)	0.371*** (0.127)	0.392*** (0.112)	0.500*** (0.111)	0.609*** (0.111)	0.486*** (0.137)	0.482*** (0.156)
log Densité	0.000966 (0.00336)	-0.00406 (0.00395)	-0.0134** (0.00617)	-0.0172*** (0.00620)	-0.00232 (0.00797)	-0.0139* (0.00715)	-0.0101 (0.00651)	-0.0158* (0.00818)	-0.0139 (0.00976)
log Potentiel Marchand	0.0324*** (0.00485)	0.0303*** (0.00458)	0.0434*** (0.0118)	0.0513*** (0.0108)	0.0277** (0.0122)	0.0349*** (0.00958)	0.0545*** (0.00913)	0.0249** (0.0110)	0.0358 (0.0366)
Observations	304	304	304	304	304	304	304	304	304
R ² ajusté	0.432	0.456	-	-	-	-	-	-	-
Contrôle spécialisations sectorielles	Non	Oui	Non	Oui	Non	Oui	Oui	Oui	Oui
Instruments faibles – Cragg-Donald Wald F-Statistic*			10.921	14.791	10.921	14.791	14.791	14.791	2.074
Test des restrictions suridentifiées (Sargan p-value)**			0.7193	0.6666	0.3303	0.7450	0.8850	0.6384	0.6726
Test d'endogénéité***			0.02932	0.00565	0.08277	0.01774	0.00006	0.07658	0.10704
Régressions de première étape****			37.48	77.85	94.41	77.85	77.85	77.85	75.26
F-stat (Ln Densité) (p-value)			94.41	41.84	37.48	41.84	41.84	41.84	41.95
F-stat (Ratio Diplômés)			15.82	18.19	15.82	18.19	18.19	18.19	4.64
F-stat (Ln PM)			15.82	18.19	15.82	18.19	18.19	18.19	4.64

Notes : modèles à variables instrumentales suivant la méthode des moindres carrés en deux étapes (2SLS) ; Variables du Ratio de diplômés, du log de la Densité de l'emploi et du log du Potentiel Marchand instrumentées ; Instruments exclus : Variation du ratio de diplômés du supérieur entre 1968 et 1982 ; Ratio de diplômés 2011 et variation du Ratio 2004-2011 reconstruits à partir de la structure par âge en 1999 ; Densité de population en 1806 ; Nombre d'immeubles religieux classés monument historique par km². *La F-statistique Cragg-Donald Wald doit être supérieure aux valeurs critiques proposées par Stock et Yogo (2005) – la valeur critique pour un biais d'instrument maximal de 5 % est de 9.53. **Les tests des restrictions suridentifiées testent l'absence de corrélation entre les instruments et les termes d'erreur (HO) ; les p-value des statistiques de Sargan (ou Hansen's J lorsque l'on corrige de l'hétéroscédasticité) sont données. ***Un rejet de H0 du test d'endogénéité sur les variables instrumentées indique que ces variables pourraient être considérées comme exogènes. ****Généralement, les instruments sont considéré faibles lorsque les statistiques F de première étape sont inférieures à 10 (Staiger and Stock, Econometrica, 1997).

Erreurs-types entre parenthèses. *** significatif au seuil de 1 % ; ** significatif au seuil de 5 % ; * significatif au seuil de 10 %.

Tableau 10. Salaire net horaire moyen des CSP « Employés »

	MCO		Variables Instrumentales						
	log salaire moyen		log salaire moyen' (non expliqué par le niveau des qualifications)						
	(1)	(2)	(3)	(4)	(5)	(6)	Salaires des Femmes	Salaires des Hommes	Indicatrice Ile de France
Ratio de diplômés	0.569*** (0.0708)	0.571*** (0.0823)	0.706*** (0.0620)	0.655*** (0.0718)	0.314*** (0.0818)	0.328*** (0.103)	0.340*** (0.109)	0.360*** (0.118)	0.272* (0.159)
log Densité	-0.00187 (0.00241)	-0.00438* (0.00260)	-0.0145*** (0.00365)	-0.0153*** (0.00352)	-0.0123** (0.00545)	-0.0142*** (0.00489)	-0.0153*** (0.00515)	-0.0133*** (0.00496)	-0.0134* (0.00795)
log Potentiel Marchand	0.0374*** (0.00350)	0.0372*** (0.00356)	0.0456*** (0.00696)	0.0517*** (0.00610)	0.0586*** (0.00889)	0.0621*** (0.00840)	0.0678*** (0.00874)	0.0421*** (0.00864)	0.0598* (0.0306)
Observations	304	304	304	304	304	304	304	304	304
R ² ajusté	0.766	0.790	-	-	-	-	-	-	-
Contrôle spécialisations sectorielles	Non	Oui	Non	Oui	Non	Oui	Oui	Oui	Oui
Instruments faibles – Cragg-Donald Wald F-Statistic*			10.921	14.791	10.604	12.962	12.962	12.962	2.219
Test des restrictions suridentifiées (Sargan p-value)**			9.53	9.53	9.53	9.53	9.53	9.53	9.53
Test d'endogénéité***			0.0248 0.00004	0.2904 0.00002	0.1754 0.0221	0.2532 0.0030	0.3282 0.0012	0.2883 0.1314	0.0821 0.1497
Régressions de première étape****									
F-stat (Ln Densité) (p-value)			37.48	77.85	52.36	32.97	32.97	32.97	42.65
F-stat (Ratio Diplômés)			94.41	41.84	29.48	29.99	29.99	29.99	32.34
F-stat (Ln PM)			15.82	18.19	16.04	15.07	15.07	15.07	4.63

Notes : modèles à variables instrumentales suivant la méthode des moindres carrés en deux étapes (2SLS) ; Variables du Ratio de diplômés, du log de la Densité de l'emploi et du log du Potentiel Marchand instrumentées ; Instruments exclus : Variation du ratio de diplômés du supérieur entre 1968 et 1982 ; Ratio de diplômés 2011 et variation du Ratio 2004-2011 reconstruits à partir de la structure par âge en 1999 ; Densité de population en 1806 ; Nombre d'immeubles religieux classés monument historique par km². *La F-statistique Cragg-Donald Wald doit être supérieure aux valeurs critiques proposées par Stock et Yogo (2005) – la valeur critique pour un biais d'instrument maximal de 5 % est de 9.53.**Les tests des restrictions suridentifiées testent l'absence de corrélation entre les instruments et les termes d'erreur (HO) ; les p-value des statistiques de Sargan (ou Hansen's J lorsque l'on corrige de l'hétéroscédasticité) sont données.***Un rejet de H0 du test d'endogénéité sur les variables instrumentées indique que ces variables pourraient être considérées comme exogènes.****Généralement, les instruments sont considéré faibles lorsque les statistiques F de première étape sont inférieures à 10 (Staiger and Stock, Econometrica, 1997).

Erreurs-types entre parenthèses. *** significatif au seuil de 1 % ; ** significatif au seuil de 5 % ; * significatif au seuil de 10 %.

Tableau 11. Salaire net horaire moyen des CSP « Ouvriers »

	MCO				Variables Instrumentales				
	log salaire moyen				log salaire moyen' (non expliqué par le niveau des qualifications)				
	(1)	(2)	(3)	(4)	(5)	(6)	<i>Salaires des Femmes</i>	<i>Salaires des Hommes</i>	<i>Indicatrice Ile de France</i>
Ratio de diplômés	0.325*** (0.103)	0.309*** (0.0844)	0.461*** (0.102)	0.355*** (0.115)	0.166 (0.131)	0.147 (0.134)	0.480*** (0.119)	0.160 (0.148)	0.136 (0.194)
log Densité	-0.00118 (0.00445)	-0.00911*** (0.00351)	-0.0135** (0.00598)	-0.0209*** (0.00563)	-0.00772 (0.00930)	-0.0203*** (0.00674)	-0.00690 (0.00490)	-0.0233*** (0.00737)	-0.0211** (0.0104)
log Potentiel Marchand	0.0299*** (0.00544)	0.0261*** (0.00460)	0.0219* (0.0114)	0.0324*** (0.00978)	0.0108 (0.0138)	0.0234** (0.00984)	0.0112 (0.00902)	0.0223** (0.00969)	0.0296 (0.0447)
Observations	304	304	304	304	304	304	304	304	304
R ² ajusté	0.512	0.566	-	-	-	-	-	-	-
Contrôle spécialisations sectorielles	Non	Oui	Non	Oui	Non	Oui	Oui	Oui	Oui
Instruments faibles – Cragg-Donald Wald F-Statistic*			10.921	14.791	10.604	14.791	14.791	12.962	2.219
Test des restrictions suridentifiées (Sargan p-value)**			0.2449	0.4245	0.3475	0.2759	0.9754	0.2996	0.1457
Test d'endogénéité***			0.00761	0.06041	0.0192	0.04942	0.18532	0.0752	0.1414
Régressions de première étape****			37.48	77.85	52.36	77.85	77.85	32.97	42.65
F-stat (Ln Densité) (p-value)			94.41	41.84	29.48	41.84	41.84	29.99	32.34
F-stat (Ratio Diplômés)			15.82	18.19	16.04	18.19	18.19	15.07	4.63
F-stat (Ln PM)			15.82	18.19	16.04	15.07	15.07	15.07	4.63

Notes : modèles à variables instrumentales suivant la méthode des moindres carrés en deux étapes (2SLS) ; Variables du Ratio de diplômés, du log de la Densité de l'emploi et du log du Potentiel Marchand instrumentées ; Instruments exclus : Variation du ratio de diplômés du supérieur entre 1968 et 1982 ; Ratio de diplômés 2011 et variation du Ratio 2004-2011 reconstruits à partir de la structure par âge en 1999 ; Densité de population en 1806 ; Nombre d'immeubles religieux classés monument historique par km². *La F-statistique Cragg-Donald Wald doit être supérieure aux valeurs critiques proposées par Stock et Yogo (2005) – la valeur critique pour un biais d'instrument maximal de 5 % est de 9.53.**Les tests des restrictions suridentifiées testent l'absence de corrélation entre les instruments et les termes d'erreur (H0) ; les p-value des statistiques de Sargan (ou Hansen's J lorsque l'on corrige de l'hétéroscédasticité) sont données.***Un rejet de H0 du test d'endogénéité sur les variables instrumentées indique que ces variables pourraient être considérées comme exogènes.****Généralement, les instruments sont considéré faibles lorsque les statistiques F de première étape sont inférieures à 10 (Staiger and Stock, Econometrica, 1997).

Erreurs-types entre parenthèses. *** significatif au seuil de 1 % ; ** significatif au seuil de 5 % ; * significatif au seuil de 10 %.

A4. Tableau de résultats d'estimations – modèle préféré (complet)

Tableau 12. Modèle à variables instrumentales (2SLS) : effets sur les salaires moyens par CSP

Variable dépendante : log du <i>salaire net horaire moyen 2012 non expliquée par les différences de niveaux de qualifications au sein de chaque CSP</i>	Cadres	Professions intermédiaires	Employés	Ouvriers
<i>Ratio de diplômés</i>	0.623*** (0.167)	0.500*** (0.111)	0.328*** (0.103)	0.147 (0.134)
<i>Ln Densité de l'emploi</i>	-0.0116 (0.00876)	-0.0139* (0.00715)	-0.0142*** (0.00489)	-0.0203*** (0.00674)
<i>Ln Potentiel Marchand</i>	0.0849*** (0.0138)	0.0349*** (0.00958)	0.0621*** (0.00840)	0.0234** (0.00984)
<i>Superficie</i>	-0.00705*** (0.00265)	-0.00493** (0.00198)	-0.00226 (0.00152)	-0.00747*** (0.00195)
<i>Ln Agriculture</i>	0.00618 (0.00395)	-0.00357 (0.00321)	-0.000347 (0.00227)	-0.00946*** (0.00307)
<i>Ln Industrie</i>	0.0177 (0.0155)	0.0441*** (0.0103)	0.00733 (0.00929)	0.0394*** (0.0140)
<i>Ln Tertiaire marchand</i>	0.109*** (0.0413)	0.0907** (0.0357)	0.0248 (0.0252)	0.110*** (0.0368)
<i>Ln Tertiaire non marchand</i>	-0.0337 (0.0287)	0.0379* (0.0204)	-0.0207 (0.0167)	0.0231 (0.0255)
<i>Ln Construction</i>	0.0120 (0.0155)	0.0150 (0.0105)	0.00735 (0.00732)	-0.00164 (0.0109)
<i>Âge moyen (centré)</i>	0.00412* (0.00211)	-0.00274* (0.00159)	-0.00173 (0.00137)	-0.00800*** (0.00153)
<i>Âge moyen au carré (centré)</i>	-0.00200*** (0.000428)	-0.00108*** (0.000354)	-0.00117*** (0.000285)	-0.00104*** (0.000357)
Constante	-0.193 (0.133)	0.104 (0.104)	-0.199** (0.0810)	0.132 (0.114)
Observations	304	304	304	304
Instruments faibles – Cragg-Donald Wald F-Statistic*	14.791	14.791	12.962	14.791
Test des restrictions suridentifiées (Sargan p-value)**	0.7931	0.7450	9.53	0.2759
Test d'endogénéité***	0.02045	0.01774	0.2532 0.0030	0.04942
Régressions de première étape****		77.85		77.85
F-stat (Ln Densité) (p-value)	77.85	41.84	32.97	41.84
F-stat (Ratio Diplômés)	41.84	18.19	29.99	18.19
F-stat (Ln PM)			15.07	15.07

Notes : modèles à variables instrumentales suivant la méthode des moindres carrés en deux étapes (2SLS) ; Variables du Ratio de diplômés, du log de la Densité de l'emploi et du log du Potentiel Marchand instrumentées ; Instruments exclus : Variation du ratio de diplômés du supérieur entre 1968 et 1982 ; Ratio de diplômés 2011 et variation du Ratio 2004-2011 reconstruits à partir de la structure par âge en 1999 ; Densité de population en 1806 ; Nombre d'immeubles religieux classés monument historique par km². *La F-statistique Cragg-Donald Wald doit être supérieure aux valeurs critiques proposées par Stock et Yogo (2005) – la valeur critique pour un biais d'instrument maximal de 5 % est de 9.53. **Les tests des restrictions suridentifiées testent l'absence de corrélation entre les instruments et les termes d'erreur (HO) ; les p-value des statistiques de Sargan (ou Hansen's J lorsque l'on corrige de l'hétéroscédasticité) sont données. ***Un rejet de H0 du test d'endogénéité sur les variables instrumentées indique que ces variables pourraient être considérées comme exogènes. ****Généralement, les instruments sont considéré faibles lorsque les statistiques F de première étape sont inférieures à 10 (Staiger and Stock, Econometrica, 1997).

Erreurs-types entre parenthèses.

*** significatif au seuil de 1 %.; ** significatif au seuil de 5 %.;

* significatif au seuil de 10 %.

A5. Robustesse : économétrie spatiale

En raison du caractère localisé des données mobilisées pour cette analyse, le modèle économétrique s'expose à un risque d'autocorrélation spatiale, c'est-à-dire une interdépendance entre les observations géographiquement proches.

L'autocorrélation spatiale est positive lorsque les valeurs (faibles ou fortes) d'un indicateur sont concentrées dans une zone géographique (représentation en grappes). Elle est négative lorsque la valeur d'une variable pour chaque zone d'emploi est très différente des valeurs dans les zones d'emplois qui l'entourent (représentation en damier). Finalement, il y a absence d'autocorrélation spatiale lorsque la distribution géographique est aléatoire.

L'autocorrélation spatiale peut venir i) de processus d'interactions entre zones proches à l'origine d'une organisation spécifique des espaces socio-économiques (par exemple les niveaux de salaires sont similaires entre zones d'emploi à des distances voisines), ii) et/ou d'une autocorrélation des termes d'erreurs en raison de variables omises spatialement interdépendantes, notamment si le phénomène étudié dépasse l'unité géographique choisie (Le Gallo, 2002) (par exemple, les salaires peuvent s'ajuster au sein d'un territoire plus vaste que la zone d'emploi).

Pour repérer l'autocorrélation spatiale, les indices de Moran ou de Geary²⁰ prennent en compte la proximité géographique et la similitude ou la disparité des valeurs d'une variable ou des résidus entre les différentes zones d'emploi.

Ces indices sont calculés sur la base d'une matrice de poids (matrice carrée 304X304 dans notre cas) qui contient les proximités spatiales de chaque zone d'emploi à toutes les autres. La construction d'une matrice de proximité donnée se base sur une conception particulière du « voisinage » (par exemple la distance ou la contiguïté). Les indices d'autocorrélation spatiale dépendent bien évidemment de la conception de voisinage retenue et les résultats des tests d'autocorrélation spatiale diffèrent selon que l'on emploie l'une ou l'autre des matrices.

Les deux types de matrices de proximité généralement utilisées sont les matrices de contiguïté (chaque terme prend la valeur 1 si les deux zones d'emploi ont une frontière commune, 0 sinon) et les matrices de distance inverse (chaque terme correspond à l'inverse de la distance qui sépare les deux zones²¹). Ces matrices sont généralement normalisées pour permettre des comparaisons entre modèles et raisonner à partir de poids relatifs. En effet, cette normalisation consiste généralement à diviser chaque terme de la matrice par la somme des termes de sa ligne (la somme des poids d'une ligne est donc égale à 1), ou bien à diviser chaque terme par le minimum entre la plus grande somme par ligne et la plus grande somme par colonne (cette normalisation à partir d'un scalaire unique est préférable selon Kelejian et Prucha (2010) car elle préserve la symétrie et la spécification de base du modèle).

L'indice de Moran affiche des valeurs comprises entre -1 (autocorrélation spatiale négative ou dispersion parfaite) et 1 (autocorrélation positive parfaite). Une valeur nulle correspond à une absence d'autocorrélation spatiale (la distribution est aléatoire selon les zones d'emploi).

$$I_{Moran} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

²⁰ Indice de Moran généralement préféré à celui de Geary dans la littérature. En réalité, les deux tests conduisent à des résultats proches. L'indice de Moran indique s'il existe une ressemblance significativement plus grande entre points voisins, par rapport à un modèle aléatoire ; tandis que l'indice de Geary indique si les différences entre localisations proches sont significativement plus petites, toujours par rapport à ce que l'on peut attendre d'un modèle aléatoire.

²¹ Ici, la distance entre deux zones d'emploi est celle qui sépare leur principale aire urbaine – donnée généralement par le nom de la zone d'emploi.

Avec i et j deux zones d'emploi ; n le nombre de zone d'emploi ; x_i la valeur des résidus pour la zone i ; \bar{x} la moyenne de x ; w_{ij} les distances de la matrice d'interactions spatiales.

L'indice de Geary prend des valeurs entre 0 et l'infini et vaut 1 en l'absence d'autocorrélation spatiale. Entre 0 et 1 il y a autocorrélation positive (et inversement pour des valeurs de l'indice supérieures à 1).

$$I_{Geary} = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{2W \sum_{i=1}^n (x_i - \bar{x})^2}$$

Pour vérifier la présence ou l'absence d'autocorrélation spatiale des termes d'erreurs, qui pourrait engendrer des biais de significativité des coefficients (et donc impliquerait une analyse basée sur l'économétrie spatiale), on teste la significativité du degré d'autocorrélation spatiale (H_0 : absence d'autocorrélation spatiale) des résidus du modèle à variables instrumentales préféré (2), sur la base d'une matrice de contiguïté et d'une matrice de distance inverse.

Tableau . Indices et tests d'autocorrélation spatiale des résidus

Hypothèse nulle : indépendance spatiale globale

	Indice de Moran	<i>P-value</i>	Indice de Geary	<i>P-value</i>
Matrice de distance inverse (normalisation minmax)				
Cadres...	0.031	0.000	0.966	0.064
Professions intermédiaires	0.014	0.002	0.945	0.011
Employés	0.025	0.000	0.928	0.000
Ouvriers	0.017	0.000	0.953	0.017
Matrice de contiguïté (normalisation minmax)				
Cadres...	0.218	0.000	0.632	0.000
Professions intermédiaires	0.109	0.001	0.771	0.000
Employés	0.267	0.000	0.652	0.000
Ouvriers	0.142	0.000	0.690	0.000

Lorsque l'on utilise une matrice de distance pour juger des interconnexions géographiques entre zones d'emploi – méthode moins restrictive que la matrice de contiguïté qui suppose des interactions uniquement entre zones qui « se touchent » –, les indices de Moran (resp. Geary) sont tous très proches de 0 (resp. 1). Néanmoins, les *p-values* du test d'autocorrélation spatiale conduisent au rejet de l'hypothèse nulle d'absence d'autocorrélation spatiale. Il y aurait donc une très légère autocorrélation spatiale positive des résidus. Ces résultats sont confirmés par les tests utilisant la matrice de contiguïté. Cette matrice reflète une définition plus restrictive du « voisinage ».

Au vu de ce diagnostic, il est utile de prendre en compte cette autocorrélation spatiale et vérifier la robustesse des résultats à l'aide d'un modèle d'auto-régression simultanée avec décalage spatial et erreur spatiale en présence de variables instrumentales.

Le modèle prend la forme suivante :

$$y = Y\pi + X\beta + \lambda W y + u$$

$$u = \rho M u + \epsilon$$

avec

y un vecteur ($nx1$) des valeurs observées de la variable dépendante (ici le salaire horaire net) ;

Y une matrice ($n \times p$) des valeurs des p variables endogènes et π le vecteur ($p \times 1$) des coefficients correspondants ;

X une matrice ($n \times k$) des observations des k variables exogènes et β le vecteur ($p \times 1$) des coefficients correspondants ;

W et M des matrices spatiale de poids ($n \times n$) contenant des zéros sur la diagonale principale (W et M sont identiques ici) ;

$W\gamma$ et Mu des vecteurs ($n \times 1$) correspondant aux valeurs décalées spatialement, avec λ et ρ les paramètres autorégressifs spatiaux correspondants. Chaque ligne de la matrice $W\gamma$ peut s'interpréter comme la somme pondérée des γ dans les zones voisines pour chaque zone d'emploi. Le paramètre λ révèle la dépendance spatiale entre les valeurs de la variable expliquée (les salaires), alors que le paramètre ρ capte la dépendance spatiale des termes d'erreur ;

ϵ un vecteur ($n \times 1$) des termes d'erreur.

Modèle d'autorégression-spatiale (SARAR)				
Matrice de distance inverse (normalisation minmax)				
	Cadres	Professions intermédiaires	Employés	Ouvriers
<i>Ratio de diplômés</i>	0.372*** (0.136)	0.326*** (0.106)	0.239*** (0.0722)	0.0290 (0.102)
<i>Ln Densité de l'emploi</i>	-0.00646 (0.00551)	-0.00507 (0.00452)	-0.00823*** (0.00305)	-0.0130*** (0.00451)
<i>Ln Potentiel Marchand</i>	0.00414 (0.0108)	-0.0107 (0.00761)	0.0320*** (0.00644)	-0.00485 (0.00716)
<i>Superficie</i>	-0.00643*** (0.00216)	-0.00295 (0.00188)	-0.00184 (0.00120)	-0.00494*** (0.00177)
<i>Ln Agriculture</i>	0.00776*** (0.00277)	-0.00133 (0.00234)	0.000336 (0.00154)	-0.00785*** (0.00228)
<i>Ln Industrie</i>	0.0407*** (0.0113)	0.0504*** (0.00927)	0.0147** (0.00625)	0.0383*** (0.00897)
<i>Ln Tertiaire marchand</i>	0.148*** (0.0327)	0.0963*** (0.0271)	0.0334* (0.0183)	0.109*** (0.0260)
<i>Ln Tertiaire non marchand</i>	0.00237 (0.0214)	0.0495*** (0.0178)	-0.00988 (0.0120)	0.0233 (0.0172)
<i>Ln Construction</i>	0.00529 (0.0115)	0.0127 (0.00991)	0.00774 (0.00647)	-0.00273 (0.00941)
<i>Âge moyen (centré)</i>	0.00104 (0.00153)	-0.00263** (0.00123)	-0.00233*** (0.000838)	-0.00647*** (0.00126)
<i>Âge moyen au carré (centré)</i>	- (0.000955***)	-0.000587* (0.000307)	- (0.000200)	- (0.000290)
Constante	0.246** (0.108)	0.303*** (0.0836)	-0.0526 (0.0610)	0.228*** (0.0790)
λ	0.286***	0.306***	0.114***	0.241***
ρ	-0.00751	-0.127	-0.0397	-0.0110
Observations	304	304	304	304

Modèle d'autorégression-spatiale (SARAR)				
<i>Matrice de contiguïté (normalisation minmax)</i>				
	Cadres	Professions intermédiaires	Employés	Ouvriers
<i>Ratio de diplômés</i>	0.278*** (0.107)	0.263*** (0.0913)	0.168*** (0.0626)	0.0206 (0.0848)
<i>Ln Densité de l'emploi</i>	0.00125 (0.00478)	0.000583 (0.00419)	-0.00424 (0.00292)	-0.00591 (0.00414)
<i>Ln Potentiel Marchand</i>	0.0135** (0.00650)	-5.03e-07 (0.00458)	0.0198*** (0.00499)	-0.000367 (0.00455)
<i>Superficie</i>	-0.00220 (0.00193)	-0.000300 (0.00176)	0.000157 (0.00111)	-0.00122 (0.00161)
<i>Ln Agriculture</i>	0.00539** (0.00222)	-0.000804 (0.00191)	9.19e-05 (0.00132)	-0.00631*** (0.00191)
<i>Ln Industrie</i>	0.0254*** (0.00911)	0.0377*** (0.00795)	0.0106** (0.00533)	0.0297*** (0.00765)
<i>Ln Tertiaire marchand</i>	0.0799*** (0.0278)	0.0594** (0.0243)	0.0191 (0.0159)	0.0696*** (0.0233)
<i>Ln Tertiaire non marchand</i>	-0.00358 (0.0180)	0.0387** (0.0157)	-0.00851 (0.0104)	0.0200 (0.0148)
<i>Ln Construction</i>	0.00948 (0.00990)	0.0130 (0.00862)	0.00593 (0.00572)	0.00287 (0.00822)
<i>Âge moyen (centré)</i>	0.00309** (0.00121)	-0.000993 (0.00110)	-0.00109 (0.000741)	-0.00291** (0.00118)
<i>Âge moyen au carré (centré)</i>	-0.000879*** (0.000314)	-0.000525* (0.000273)	-0.000742*** (0.000176)	-0.000698*** (0.000255)
Constante	0.0905 (0.0804)	0.177*** (0.0683)	-0.0361 (0.0478)	0.148** (0.0653)
λ	0.191***	0.176***	0.150***	0.182***
ρ	-0.114***	-0.137***	-0.0566	-0.0659*
Observations	304	304	304	304

Étant donné les normalisations des matrices de poids, les paramètres λ et ρ sont compris entre -1 et 1. Les estimations indiquent des λ significatifs et positifs ce qui suggère une interdépendance spatiale des salaires relativement forte. Les salaires moyens dépendent des salaires moyens des zones d'emploi « voisines ».

Les résultats concernant les coefficients ρ diffèrent selon la matrice de voisinage utilisée. Dans le cas de la matrice de distance inverse, ceux-ci ne sont pas statistiquement significatifs. En revanche, lorsque l'on évalue l'autocorrélation spatiale des résidus à partir d'une matrice de contiguïté, la significativité des paramètres ρ signale la présence d'autocorrélation spatiale des résidus.

Néanmoins, quelle que soit la matrice de poids utilisée, l'influence du ratio de diplômés est robuste à la prise en compte de l'autocorrélation spatiale (les coefficients associés à la variable restent très significatifs – sauf pour les CSP ouvriers). En raison du caractère multidirectionnel du modèle d'économétrie spatiale (les estimations sont déterminées simultanément), la valeur des coefficients ne peut pas s'interpréter comme dans le cas linéaire.

RETROUVEZ
LES DERNIÈRES ACTUALITÉS
DE FRANCE STRATÉGIE SUR :



www.strategie.gouv.fr



[francestrategie](https://www.facebook.com/francestrategie)



[@Strategie_Gouv](https://twitter.com/Strategie_Gouv)



FRANCE STRATÉGIE

France Stratégie, laboratoire d'idées public, a pour mission d'éclairer les choix collectifs. Son action repose sur quatre métiers : évaluer les politiques publiques ; anticiper les mutations à venir dans les domaines économiques, sociétaux ou techniques ; débattre avec les experts et les acteurs français et internationaux ; proposer des recommandations aux pouvoirs publics nationaux, territoriaux et européens. Pour enrichir ses analyses et affiner ses propositions France Stratégie s'attache à dialoguer avec les partenaires sociaux et la société civile. France Stratégie mise sur la transversalité en animant un réseau de sept organismes aux compétences spécialisées.
