

## **Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du Grand Paris**

Pierre-Philippe Combes

Laurent Gobillon

Miren Lafourcade

Avril 2015

# Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du Grand Paris<sup>1</sup>

Pierre-Philippe Combes<sup>2</sup>, Laurent Gobillon<sup>3</sup> et Miren Lafourcade<sup>4</sup>

**Résumé:** L'objectif de ce rapport est d'estimer ce qui, dans la prime salariale dont bénéficient les travailleurs des grandes métropoles françaises, provient de leurs compétences propres et des économies d'agglomération engendrées par la taille des villes. Les estimations sont réalisées sur un panel composé de 3 133 573 observations, et couvrant environ 500 000 salariés des secteurs privé et semi-public sur les six années constituant ce panel (1993, 1996, 1999, 2002, 2005 et 2008). Cet échantillon constitue la plus grande source de variabilité individuelle, temporelle, géographique et sectorielle utilisée à ce jour pour estimer l'impact des économies d'agglomération.

Les disparités de salaires entre villes sont notables, même pour des travailleurs disposant de compétences similaires. Par exemple, un actif occupé travaillant à Paris en 2008 disposait d'un salaire supérieur de 26 % à celui d'un individu identique, mais travaillant dans le Morvan, la zone d'emploi la moins dense du panel en 2008. D'une manière générale, à compétences données, les primes dont bénéficient les salariés du Grand Paris sont assez remarquables : de 15 % à Coulommiers ou Provins à 23 % à Saclay ou Orly. À compétences données, les primes associées aux autres grandes métropoles françaises sont plus faibles, de 10% à Bordeaux à 15 % à Lyon. Les zones d'emploi rurales et peu denses du centre et du littoral breton disposent des primes les plus faibles de toute la France.

Ces disparités, qui ne sont pas imputables au tri spatial des qualifications et aux compétences propres des travailleurs, permettent d'estimer la part de la prime salariale imputable à la taille des villes. L'élasticité de la prime salariale à la densité d'emploi (l'intensité des économies d'agglomération statiques) est estimée à 0,02. Cette valeur, même si elle est faible, induit des disparités géographiques de salaire assez substantielles, car il existe de grands écarts de densité entre villes françaises. Par exemple, si on compare les deux extrêmes de la distribution des densités, Paris et le Morvan, le différentiel de productivité imputable à l'écart de densité entre ces deux villes est de 17% environ.

L'impact de la densité est ramenée à 0,013 si l'on contrôle la mesure de l'élasticité de la prime salariale par la superficie des villes et leur potentiel marchand, qui capte la possibilité qu'elles ont de recourir aux marchés externes les plus proches en sus de leur marché local. Doubler la densité d'une zone permet donc d'y augmenter le salaire moyen de 1%, à superficie et potentiel marchand donnés. L'accès aux marchés externes a un effet sur la prime salariale trois fois supérieur à celui de la densité. Le fait de doubler le potentiel marchand d'une ville y accroît la rémunération moyenne de ses salariés de 1%, toutes choses égales par ailleurs. Si on compare les deux extrêmes de la distribution des potentiels marchands, Orly et Bayonne, le différentiel de productivité imputable à l'écart de potentiel marchand entre ces deux villes est de 10,5% environ.

Les économies d'agglomération dynamiques mesurent la relation entre la densité d'emploi et la progression salariale au cours de la carrière (prime salariale à l'expérience). Ces gains dynamiques

---

1 Cette recherche a été financée par la Société du Grand Paris dans le cadre du Contrat de Recherche CRC n°2013-05. Elle a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'avenir portant la référence ANR-10-EQPX-17 (Centre d'accès sécurisé aux données - CASD).

2 Aix-Marseille School of Economics, 2 Rue de la Charité, 13002 Marseille. Email : ppcombes@gmail.com.

3 INED et Paris School of Economics (PSE), 133 Boulevard Davout, 75980 Paris. Email : laurent.gobillon@ined.fr.

4 Université Paris-Sud (RITM) et PSE, 48 Boulevard Jourdan, 75014 Paris. Email : miren.lafourcade@ens.fr.

sont plus également répartis sur le territoire que les gains statiques. Sur les 20 villes offrant la plus forte valorisation de l'expérience figurent seulement quatre zones d'emploi du Grand Paris. La prime salariale à l'expérience dont bénéficie un travailleur à Paris, relativement à un salarié disposant des mêmes compétences dans le Morvan n'excède pas 1,5 %. Certaines zones d'emploi comme Pauillac, Carcassonne, Salon-de-Provence ou Istres-Martigues, disposent d'une prime à l'expérience similaire, voire supérieure aux zones d'emploi du Grand Paris, bien qu'elles soient bien moins denses.

La taille des villes a un impact positif très significatif sur la prime salariale à l'expérience : une hausse de 1 % de la densité d'une ville permet à ses salariés d'accroître leur rémunération de 0.0007 % par année s'ils restent dans cette ville, et de 0.0011 % par année s'ils la quittent pour aller valoriser leur expérience ailleurs. Il semble donc que la mobilité au départ des zones denses soit payante, et permette aux salariés ayant acquis de l'expérience dans une grande métropole de la valoriser dans n'importe quelle autre ville, et d'en tirer une prime salariale annuelle plus élevée que dans leur ville d'origine.

**JEL Classification:** J31, R12 et R23

**Keywords :** salaires, densité, économies d'agglomération, effets d'apprentissage, zones d'emploi.

Introduction .....	4
1. Productivité, salaire nominal et économies d’agglomération .....	9
1.1 Les mécanismes théoriques sous-jacents.....	9
1.2 Economies d’agglomération statiques : l’approche agrégée .....	12
1.3 Economies d’agglomération statiques : l’approche individuelle.....	14
1.4 Economies d’agglomération statiques : l’approche sectorielle.....	19
1.5 Economies d’agglomération dynamiques.....	20
2. Présentation des données et statistiques descriptives.....	24
2.1 Données individuelles issues du panel DADS .....	24
2.2 Des unités spatiales pertinentes : les zones d’emploi 2010 .....	26
2.3 Statistiques descriptives sur les salaires et la taille des zones d’emploi .....	28
2.4 Statistiques descriptives sur la qualification de la main d’œuvre .....	33
3. Résultats de l’estimation des économies d’agglomération .....	39
3.1 Des économies d’agglomération statiques assez substantielles.....	40
3.2 Des économies d’agglomération dynamiques transférables .....	48
3.3 Problèmes d’endogénéité et robustesse des estimations.....	55
3.4 Des économies d’agglomération similaires pour toutes les villes.....	61
Conclusion.....	65
Références bibliographiques.....	67
Annexes.....	69
Annexe 1 : Distance et temps de transport au sein du Grand Paris.....	69
Annexe 2 : Nomenclature des secteurs d’activité .....	72
Annexe 3 : Salaire net fiscal et taille des zones (moyenne 6 années) .....	75
Annexe 4 : Liste des Groupes et Catégories Socio-Professionnelles .....	76
Annexe 5 : Expérience des salariés selon la zone (moyenne 6 années) .....	77
Annexe 6 : Périmètre couvert par le Grand Paris Express.....	78
Annexe 7 : Prime salariale « moyenne » associée aux zones d’emploi.....	79
Annexe 8 : Zones d’emploi par quintile de densité moyenne (sur les 6 années)	79

## Introduction

L'avènement des très grandes villes à l'échelle mondiale est un fait marquant du siècle écoulé. Les métropoles sont des lieux d'innovation et de diffusion des idées, des points nodaux dans l'organisation des firmes multinationales et des réseaux de transport internationaux. Tels des aimants, elles attirent les salariés les plus diplômés, les cadres ou les travailleurs qualifiés disposant d'une plus grande productivité. De ce fait, elles constituent un terreau particulièrement fertile pour la création d'entreprises et l'entrepreneuriat, et une matrice pour l'amélioration des compétences des acteurs économiques. Bref, elles engendrent des gains d'efficacité qui sont à la fois statiques et dynamiques, qui en font les foyers des plus forts potentiels de croissance.

A titre d'illustration, la mégalopole de New-York, dont la richesse avoisine les 580 milliards de dollars, produit à elle seule l'équivalent du PIB de la Belgique. Si on déflate cette valeur par le coût de la vie, qui est aussi très élevé dans les grandes métropoles, New York rivalise encore avec l'Espagne, Londres avec la Suède ou la Suisse, et Tokyo avec le Canada. La croissance des villes ne semble par ailleurs pas s'essouffler. Depuis 2009, le taux d'urbanisation dépasse le seuil de 50 % de la population mondiale, contre à peine 10% en 1700, et seulement 30% en 1950 (Véron, 2006). Selon les dernières projections des Nations Unies, ce taux devrait même avoisiner 60 % en 2030, soit une population de plus de 5 milliards de citoyens. Si la tendance se poursuit, il faut s'attendre à ce que l'espace urbain en 2030 soit trois fois plus étendu qu'en 2000, soit un gain de 1,2 million de kilomètres carrés, équivalent à deux fois la superficie de l'Hexagone (Seto, Güneralp et Hutryra, 2012). Chaque jour qui passe, le cumul des espaces gagnés par la ville dans le monde serait ainsi de 110km<sup>2</sup>, la superficie de Paris intra-muros. S'il est assez incroyable d'imaginer que les deux-tiers des zones urbaines de 2030 puissent ne pas encore exister au début du 21<sup>ème</sup> siècle, il est moins surprenant de constater que les projections situent la moitié de cette expansion en Asie, notamment en Chine et en Inde (Seto et al., 2012).

Pourtant la France n'est pas en reste. La superficie de l'espace urbain a progressé de 20 % sur la dernière décennie, passant de 100 000 à 119 000 km<sup>2</sup> pour la métropole (Insee, 2013). Fin 2010, les villes françaises occupaient ainsi 22% du territoire et accueillait 47,9 millions d'habitants, soit 77,5% de la population de l'Hexagone. L'agglomération parisienne, avec 10,3 millions de personnes réparties sur 412 communes, reste de loin la plus importante unité urbaine<sup>5</sup> de France.

L'influence des métropoles ne s'arrête cependant pas à la frontière administrative des unités urbaines. Les villes attirent quotidiennement de très nombreux travailleurs résidant en périphérie des agglomérations. Ces flux domicile-travail, leur distance et leur intensité rendent compte d'une organisation fonctionnelle de l'espace permettant de définir des aires urbaines<sup>6</sup>, illustrées par la Figure 1. Selon ce découpage, 95 % de la population française vit actuellement sous l'influence d'une ville. L'aire urbaine de Paris a par exemple gagné 949 300 habitants sur la période 1999-2008, soit une augmentation de 8,5 % (Labrador, 2011). Cette densification a principalement bénéficié au pôle urbain de Paris, qui a gagné 710 200 habitants, contre seulement 239 100 pour la couronne. Les entreprises, en particulier les firmes multinationales, se sont aussi principalement concentrées dans

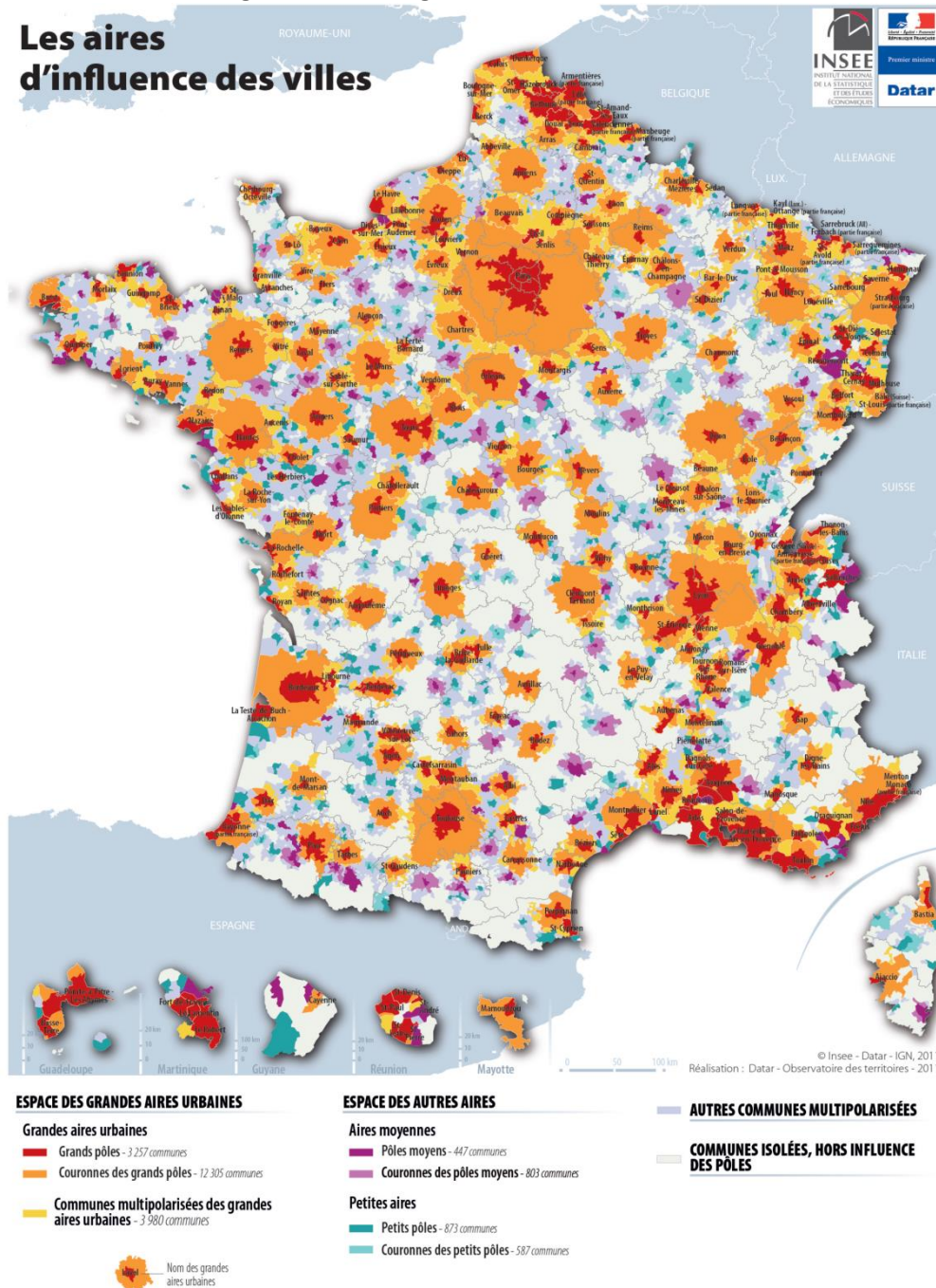
---

<sup>5</sup> L'Insee définit une unité urbaine comme un ensemble d'une ou plusieurs communes présentant une continuité du tissu bâti (pas de coupure de plus de 200 mètres entre deux constructions) et comptant au moins 2 000 habitants. Chaque commune de l'unité urbaine doit posséder plus de la moitié de sa population dans cette zone bâtie. Si la zone bâtie se situe sur une seule commune, on parle de ville isolée. Dans le cas contraire, on a une agglomération multi-communale. Les unités urbaines sont redéfinies à l'occasion des recensements de la population.

<sup>6</sup> L'Insee définit une aire urbaine comme un ensemble de communes, d'un seul tenant et sans enclave, constitué par une unité urbaine de plus de 1 500 emplois (appelée pôle urbain), et par des communes rurales ou unités urbaines (formant la 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 grandes aires urbaines, dont le pôle concentre plus de 10 000 emplois, peuvent s'étendre sur plusieurs régions (c'est le cas de Paris).

le cœur de la région francilienne. Les départements de Paris et des Hauts-de-Seine accueillent désormais la moitié du stock d'établissements contrôlés par des groupes étrangers et, sur la dernière décennie, Paris a concentré 41 % de l'ensemble des créations d'établissements de l'aire urbaine (Le Roux et al., 2013).

Figure 1 : Le zonage en aires urbaines 2010



Source : Insee - Zonage en aires urbaines 2010

Une telle dynamique atteste en soi que les grandes métropoles génèrent des gains qui dépassent les coûts engendrés par une plus forte densité d'acteurs économiques. Il n'est donc pas étonnant que les entreprises y aient une productivité globale des facteurs supérieure à celle dont elles bénéficient ailleurs (Henderson, 2003 ; Combes, Duranton, Gobillon et Roux, 2010). La productivité étant plus élevée dans les grandes villes, les travailleurs y sont aussi bien mieux rémunérés qu'ailleurs. Par exemple, aux Etats-Unis, les salaires versés au sein des aires urbaines de plus d'un million d'habitants sont supérieurs de 30% à ceux versés dans les zones rurales (Glaeser, 2011).

En Europe, les écarts sont également très marqués. Le salaire annuel moyen à Madrid est de 31000€, soit 20 % de plus qu'à Valence, la troisième plus grande ville espagnole, 46 % de plus qu'à Saint-Jacques de Compostelle, la ville médiane, et 52 % de plus qu'en zone rurale (De la Roca et Puga, 2012). En France, la rémunération annuelle d'un salarié francilien à temps plein du secteur privé est supérieure de 24 % à la moyenne du reste de la France (Combes, Duranton et Gobillon, 2015). L'écart entre Paris et Lyon ou Marseille tombe à 15 %, mais il monte à 35 % entre Paris et les villes moyennes, et à 60% entre Paris et les zones à dominante rurale (Combes, Duranton et Gobillon, 2008). Au sein-même de la zone francilienne, les écarts entre d'un côté, Paris et la Petite Couronne (Hauts-de-Seine, Val de Marne et Seine-Saint-Denis) et de l'autre côté, la Grande Couronne (Val d'Oise, Yvelines, Essonne et Seine-et-Marne), sont de l'ordre de 9 %. Plus surprenant, la Petite Couronne semble être l'épicentre du tri spatial des qualifications : les cadres y sont très largement surreprésentés comparé au reste de l'espace francilien, y compris Paris, dont la propension à attirer des travailleurs faiblement qualifiés est plus importante qu'on ne le pense (Combes, Duranton et Gobillon, 2015). Les disparités de salaires sont donc très importantes, aussi bien à l'intérieur de la région francilienne qu'entre l'Île-de-France et le reste de l'Hexagone.

De très nombreux travaux issus de la recherche académique en économie ont analysé les « économies d'agglomération » engendrées par la métropolisation<sup>7</sup>. Duranton et Puga (2004) distinguent trois principales sources de gains, dénommées selon la terminologie anglo-saxonne « Sharing » (Partage), « Matching » (Appariement) et « Learning » (Apprentissage) :

- Tout d'abord, l'agglomération des activités économiques occasionne des économies d'échelle liées au *partage* des biens finaux ou intermédiaires, à la mutualisation du coût des grands équipements et des risques du marché. Les entreprises qui ont une technologie de production à rendements croissants, comme celles du secteur manufacturier, ont une productivité qui augmente avec la taille des unités de production : elles ne peuvent donc fragmenter leur chaîne de production sans accroître considérablement leur coût unitaire de fabrication. Dans ces conditions, elles choisissent de concentrer leur production sur un petit nombre de sites. Les grandes métropoles sont des lieux d'implantation privilégiés car les entreprises peuvent y mettre à profit leurs économies d'échelle.

Les entreprises ont d'autant plus intérêt à s'y concentrer que les coûts d'approvisionnement ou de livraison y sont aussi beaucoup plus faibles, du fait de la proximité d'un grand nombre de clients ou de sous-traitants, mais aussi de coûts unitaires de transport plus faibles, en raison de la présence de grandes infrastructures. En effet, les aéroports, les gares ou les échangeurs autoroutiers sont des équipements collectifs que l'on peut difficilement saupoudrer dans l'espace, du fait de leur indivisibilité. Ils ne peuvent donc être produits que dans les zones ayant atteint une certaine taille critique permettant de mutualiser leur financement sur un plus grand nombre d'utilisateurs : le coût associé à leur indivisibilité est alors compensé par l'utilisation intensive qu'en font les habitants des grandes villes. Les travailleurs ont aussi intérêt à privilégier les grandes métropoles pour bénéficier des opportunités d'emploi offertes par les entreprises, et de déplacements domicile – travail moindres. Il ressort de ces incitations une dynamique d'agglomération simple : les entreprises souhaitent s'implanter à proximité des consommateurs-travailleurs, qui souhaitent également être proches des producteurs-employeurs. L'arrivée de nouvelles entreprises attire de nouveaux ménages, dont la demande favorise l'entrée de nouvelles firmes : par un effet boule de neige, le processus s'auto-entretient et devient dynamique.

Mais les grandes villes permettent aussi aux entreprises de bénéficier d'économies d'échelle externes. Certaines entreprises s'y regroupent en pépinières de manière à former des pôles d'activités spécialisés permettant le partage de facteurs de production spécialisés, que seule une masse critique permet d'obtenir à un coût raisonnable. Ces externalités *intra-sectorielles* sont plus

---

<sup>7</sup> Voir la revue de littérature effectuée par Combes et Lafourcade (2012) pour le compte de la Société du Grand Paris.



communément appelées *externalités de spécialisation*<sup>8</sup>. Elles expliquent pour une large part le succès de certains districts industriels comme la Silicon Valley ou la Route 128<sup>9</sup> aux Etats-Unis (Porter, 1998). Par opposition aux externalités de spécialisation, les *externalités d'urbanisation* sont véhiculées par la diversité du tissu économique des grandes villes et les interactions inter-sectorielles (Jacobs, 1969). Cette diversité permet aux entreprises des grandes métropoles de s'approprier les innovations radicales développées dans les autres secteurs, et d'accroître leur productivité. Elle offre aussi aux entreprises une meilleure résilience aux chocs macro-économiques, qui tendent à contaminer des branches spécifiques, et rend les grandes villes moins tributaires de la spécialisation, en permettant aux emplois détruits de se déverser dans les secteurs non contaminés. Le risque de chômage y est donc plus faible, la pénurie de main d'œuvre et les problèmes d'approvisionnement moins fréquents. Les grandes métropoles jouent ainsi un rôle d'assurance mutuelle pour les entreprises et leurs salariés, en réduisant leurs coûts de transaction.

- Ensuite, l'*appariement* entre l'offre et la demande sur les différents marchés est plus facile dans les grandes métropoles. Depuis Adam Smith (1776), on sait qu'un grand marché permet une division plus fine du travail. Ainsi, plus une zone concentre des entreprises hétérogènes, plus la demande de travail y est importante et diversifiée, ce qui attire des travailleurs disposant de multiples compétences, contribuant à encourager la spécialisation des tâches et à favoriser une meilleure adéquation entre l'offre et la demande de travail. La constitution d'un vaste bassin d'emplois, issu des demandes indépendantes de chaque entreprise, profite ainsi à l'ensemble des entreprises qui vont pouvoir puiser dans ce vivier de travailleurs les qualifications et savoir-faire indispensables au développement de leur activité. Quant aux travailleurs, ils peuvent cibler les offres d'emploi correspondant le mieux à leurs compétences et en tirer de meilleures rémunérations.

- Enfin, les grandes villes favorisent les contacts sociaux, ainsi que les échanges d'information tacites ou codifiées entre des salariés le plus souvent qualifiés, et donc aptes à s'approprier le produit de la R&D et des innovations, à fertiliser les entreprises avec les nouveaux savoir-faire acquis, et à engendrer des retombées dynamiques positives sur leur environnement. Les processus d'*apprentissage* sont donc moins coûteux, plus rapides et plus efficaces au sein des grandes villes, et ces gains dynamiques peuvent être combinés avec les économies d'échelle externes dont y bénéficient les entreprises. Alfred Marshall (1890) est d'ailleurs l'un des tous premiers économistes à avoir souligné les bénéfices statiques et dynamiques engendrés par « l'atmosphère » intangible de districts industriels :

*« Les secrets de l'industrie cessent d'être des secrets ; ils sont pour ainsi dire dans l'air et les enfants apprennent inconsciemment beaucoup d'entre eux. Le travail bien fait est immédiatement reconnu et l'on discute aussitôt des mérites des inventions et des améliorations apportées aux machines, aux procédés et à l'organisation générale de l'industrie : si quelqu'un trouve une idée nouvelle, elle est aussitôt reprise par d'autres et combinée avec des suggestions de leur propre cru ; elle devient ainsi la source d'autres idées nouvelles » (Marshall, 1890, traduction des auteurs).*

En définitive, les économies d'agglomération proviennent de multiples canaux : externalités d'urbanisation liées à la taille globale des villes, externalités de localisation liées à la taille des secteurs présents dans ces villes, externalités de capital humain liées à la présence de travailleurs qualifiés au sein de ces secteurs ...

---

<sup>8</sup> La littérature utilise indifféremment le terme d'externalité de spécialisation, d'externalités de localisation, ou d'externalités Marshalliennes ou MAR – en référence à Marshall (1890), Arrow (1962) et Romer (1990), qui se sont tous trois intéressés à la formation des districts industriels.

<sup>9</sup> La Route 128, qui contourne la ville américaine de Boston sur sa partie ouest, prolonge les activités universitaires de Harvard et du Massachusetts Institute of Technology à travers un tissu dense d'entreprises dédiées aux nouvelles technologies (informatique, nucléaire, fibres optiques, électronique, architecture-urbanisme...). On y trouve des entreprises telles que Digital Equipment Corporation, Data General, Thermo Electron Corporation, Analog Devices, Computervision ou Polaroid.



Bien que les différents canaux par lesquels transitent ces économies d'agglomération fassent toujours l'objet de nombreux débats, leur quantification est primordiale pour analyser l'impact des grands projets d'infrastructure comme celui du Grand Paris Express. L'ampleur de l'investissement – évalué à près de 26 milliards d'euros fin 2014 – et son ciblage prioritaire sur la Petite Couronne, sont en effet de nature à modifier substantiellement la distribution géographique des emplois et des qualifications, aussi bien à l'intérieur de l'aire urbaine de Paris qu'entre Paris et le reste de la France.

Les études ayant estimé les créations d'activités engendrées par le projet du Grand Paris Express prévoient une hausse assez substantielle de l'emploi au sein de la région francilienne, et annoncent des effets positifs qui s'étendent bien au-delà du périmètre des gares du futur réseau (Société du Grand Paris, 2014). Elles prévoient tout d'abord que la seule réalisation du métro contribuera à la création de 15 000 à 20 000 emplois directs non délocalisables liés à la phase du chantier, dans le secteur des travaux publics, des industries ferroviaires et le bâtiment. Les travaux étant réalisés en continu et à plusieurs endroits en même temps<sup>10</sup>, différentes villes situées dans la Petite Couronne seront simultanément affectées. Au-delà des emplois directement liés aux travaux, les études prédisent également que l'arrivée du Grand Paris Express permettra de créer entre 115 000 et 315 000 emplois supplémentaires, par rapport à l'évolution tendancielle de la région. Ces créations proviendront notamment des gains d'accessibilité engendrés par la baisse des temps de transport en commun, et du regain d'attractivité des zones bénéficiaires pour les investissements internationaux.

Il est donc essentiel de bien estimer l'impact que pourraient avoir de telles créations d'emploi si elles étaient avérées. En effet, les gains à attendre du Grand Paris Express sont de trois types. Ils incluent les effets positifs potentiels de la création d'emplois qualifiés dans la région. Or les travailleurs qualifiés ont une productivité intrinsèquement plus élevée que les autres : les salaires devraient donc augmenter dans le bassin parisien, en particulier dans la Petite Couronne. Il ne s'agit cependant que d'un pur transfert depuis les régions d'origine de ces travailleurs qualifiés. Les deux autres gains potentiels sont liés aux économies d'agglomération statiques et dynamiques engendrés par la densification différenciée de l'espace francilien. Les gains de productivité statiques proviennent des externalités d'urbanisation et de spécialisation causées par cette densification à court terme. Quant aux gains de productivité dynamiques, ils sont liés à l'acquisition par les salariés de nouvelles compétences et savoir-faire à plus long terme. Ces gains d'apprentissage devraient encore accroître la productivité des salariés franciliens. La productivité baissera à court terme dans les régions d'origine dont la densité baissera, mais à moyen terme elle pourra y augmenter si l'expérience accumulée dans la métropole du Grand Paris est transférable dans l'espace, et si les travailleurs sont mobiles et changent de ville au cours de leur carrière professionnelle.

Si les deux premiers types de gains ont fait l'objet de nombreux travaux de recherche<sup>11</sup>, entre autres pour la France, ce n'est en revanche pas le cas des économies d'agglomération dynamiques. Il n'existe en effet à ce jour que deux études ayant estimé l'ampleur de ces effets d'apprentissage. La première, réalisée par De la Roca et Puga (2012), montre que les économies d'agglomération dynamiques sont substantielles : 7 années passées dans une grande métropole espagnole augmentent l'efficacité productive d'un salarié d'un montant équivalent à celui engendré par un doublement de la densité (soit 2 % environ). La seconde étude, réalisée par D'costa et Overman (2014), montre qu'il existe aussi des économies d'agglomération dynamiques au Royaume-Uni, mais de moindre ampleur. Les salariés ayant exercé ou exerçant leur activité professionnelle dans une ville britannique bénéficient d'une prime salariale de 1 % environ, en comparaison de ceux n'ayant jamais travaillé en ville. Ce gain n'est cependant pas plus élevé pour ceux ayant travaillé dans une grande métropole – même Londres – que pour les autres : ce qui compte est d'avoir exercé une activité en ville, pas que cette ville ait été particulièrement grande.

---

<sup>10</sup> Les différentes étapes du projet du Grand Paris Express sont illustrées par la carte de l'Annexe 6.

<sup>11</sup> Cf. la revue de littérature détaillée faite par Combes et Lafourcade (2012) pour le compte de la Société du Grand Paris.

L'objectif de ce rapport est de présenter la toute première évaluation des économies d'agglomération dynamiques en France. Les salaires capitalisant nécessairement, à un moment ou à un autre, les gains de productivité engendrés par la concentration des activités économiques, nous revenons dans une première partie sur la relation reliant les salaires et les différents types d'économie d'agglomération, de manière à interpréter correctement les spécifications qui sont estimées dans la suite du rapport. La seconde partie décrit les données utilisées pour réaliser ces estimations, et propose un faisceau de faits stylisés montrant qu'il existe bien une corrélation positive entre les salaires, la taille des villes et la part des travailleurs qualifiés dans la force de travail urbaine. La troisième partie du rapport établit la causalité, et estime ce qui, dans la prime salariale dont bénéficient les salariés des grandes métropoles, est lié à leurs compétences intrinsèques, et aux économies d'agglomération statiques et dynamiques.

## 1. Productivité, salaire nominal et économies d'agglomération

Avant de mesurer les gains de salaire engendrés par les grandes métropoles françaises, en particulier le Grand Paris, il est indispensable de revenir sur les mécanismes économiques reliant la productivité du travail, les salaires et les économies d'agglomération. Un modèle économique simple permet de comprendre les fondements théoriques de cette relation (Combes, Duranton et Gobillon, 2008 ; Combes and Gobillon 2015).

### 1.1 Les mécanismes théoriques sous-jacents

Soit  $Y_{c,t}$  la production d'une entreprise représentative implantée dans la ville  $c$  à la date  $t$ . On peut faire l'hypothèse que cette entreprise utilise deux facteurs de production : du travail, en quantité  $L_{c,t}$ , et un input composite, en quantité  $K_{c,t}$ , englobant tous les autres facteurs nécessaires à sa production (sol, capital, matières premières, biens intermédiaires...).

Le profit réalisé par l'entreprise, noté  $\pi_{c,t}$ , s'écrit :

$$\pi_{c,t} = p_{c,t}Y_{c,t} - \omega_{c,t}L_{c,t} - r_{c,t}K_{c,t},$$

où  $p_{c,t}$  représente le prix du bien produit,  $\omega_{c,t}$ , le taux de salaire nominal sur le marché du travail local, et  $r_{c,t}$  le coût unitaire moyen des autres facteurs de production sur ce même marché.

On fait l'hypothèse que la fonction de production est du type Cobb-Douglas, à savoir :

$$Y_{c,t} = \frac{A_{c,t}}{\alpha^\alpha (1-\alpha)^{1-\alpha}} (s_{c,t}L_{c,t})^\alpha K_{c,t}^{1-\alpha},$$

où  $0 < \alpha < 1$  est la part du travail dans le coût de production de l'entreprise,  $A_{c,t}$ , la productivité totale des facteurs sur le marché du travail local, et  $s_{c,t}$  le niveau des qualifications disponibles sur ce même marché, qui conditionne l'efficacité du travail.

L'entreprise optimise sa fonction de production de manière à maximiser son profit. À l'équilibre concurrentiel, le salaire nominal versé sur le marché du travail local  $c$  est donc égal à la productivité marginale du travail, à savoir :

$$\omega_{c,t} = \left( p_{c,t} \frac{A_{c,t}}{(r_{c,t})^{1-\alpha}} \right)^{\frac{1}{\alpha}} s_{c,t} \equiv B_{c,t} s_{c,t}. \quad (1)$$

La théorie économique permet donc de relier le salaire nominal, égal à la productivité marginale du travail, à la productivité globale des facteurs.

La spécification (1) fait apparaître deux principaux déterminants du niveau de salaire nominal versé sur le marché du travail  $c$  à la date  $t$  :

-  $B_{c,t}$  recouvre les bénéfices potentiels engendrés par une localisation en  $c$  à la date  $t$ , dont l'effet varie avec la taille de la ville. Les grandes métropoles sont tout d'abord à l'origine d'*externalités technologiques*, capturées par  $A_{c,t}$ . Ces externalités peuvent être positives, comme on l'a vu en introduction de ce rapport (présence de grands équipements collectifs, « spillovers » informationnels...), mais elles peuvent être aussi négatives. Une forte densité d'activités peut par exemple occasionner la saturation des réseaux de transport en commun et créer de la congestion. Elle engendre de la pollution locale liée aux activités industrielles, à l'usage résidentiel du sol et à la plus grande fréquence des déplacements. Elle rogne sur les espaces naturels et menace la biodiversité. Ces forces de dispersion viennent atténuer les bénéfices de l'urbanisation ;

Les grandes villes créent aussi des *externalités pécuniaires* qui, contrairement aux externalités technologiques, transitent par la sphère des marchés (et donc par les prix  $p_{c,t}$  et  $r_{c,t}$ ). Ces externalités pécuniaires peuvent faire varier le salaire nominal à la hausse ou à la baisse, comme les externalités technologiques<sup>12</sup>. La concurrence entre les sous-traitants est par exemple plus rude dans les grandes métropoles : le prix des biens intermédiaires (capturé par  $r_{c,t}$ ) y est donc plus faible, ce qui permet aux entreprises de dégager une marge permettant de mieux rémunérer leurs travailleurs. Mais les coûts fonciers (capturés aussi par  $r_{c,t}$ ) augmentent par ailleurs avec la taille des villes, et sont d'autant plus élevés que l'offre de foncier y est peu élastique (comme à Paris). Les entreprises des grandes métropoles compensent ce surcoût par une compression des salaires ;

Des externalités pécuniaires positives et négatives transitent également par le prix de vente des biens finals  $p_{c,t}$ . Une entreprise implantée dans une grande ville dispose d'une recette par unité vendue plus élevée, car elle réalise des économies sur le transport des marchandises. Elle y dispose d'une meilleure compétitivité-prix que dans une petite ville, en raison des économies d'échelle internes dont elle bénéficie. Cependant, sa marge y est aussi plus faible, car elle y est aussi plus exposée à la concurrence des autres entreprises. En définitive, si les externalités positives l'emportent sur les externalités négatives, les entreprises implantées dans les grandes métropoles ont la capacité d'offrir des salaires nominaux plus élevés que dans les petites villes ;

-  $s_{c,t}$  mesure l'état des qualifications disponibles sur le marché du travail local  $c$  à la date  $t$ . Or ces qualifications peuvent ne pas être distribuées aléatoirement entre les villes. En général, les travailleurs qualifiés ont une forte préférence pour les aménités culturelles (musées, théâtres, opéras, universités...), qui peuvent les inciter à sélectionner une grande ville. Si les métropoles attirent une proportion plus importante de cadres supérieurs ou de diplômés du 3<sup>ème</sup> cycle, le salaire nominal moyen y est plus élevé en raison de ce « tri spatial » (les anglo-saxons parlent de « sorting ») des qualifications.

Les grandes métropoles peuvent donc apparaître comme plus productives alors qu'elles ne bénéficient en réalité d'aucune économie d'agglomération : elles accueillent juste une plus grande proportion de travailleurs plus compétents que la moyenne. L'effet positif de la présence de capital humain sur les salaires nominaux ne constitue cependant pas une économie d'agglomération, mais un simple effet de composition locale de la main d'œuvre lié au tri spatial des salariés en fonction de leurs aptitudes individuelles.

---

<sup>12</sup> La distinction entre externalités technologiques et pécuniaires remonte à Scitovsky (1954). Le lecteur intéressé trouvera dans Combes et Lafourcade (2012) une description précise de ces différentes externalités.

Il est important à ce stade d'insister sur deux points fondamentaux. Lorsqu'on cherche à évaluer les effets de la métropolisation sur la productivité, c'est bien le salaire nominal, et non le salaire réel, qui doit constituer la variable dépendante. Il n'y a aucune raison de déflater la variable dépendante du coût de la vie (en particulier des coûts du logement ou du foncier). La relation qu'il faut estimer est celle qui lie le rendement nominal, pas le rendement réel, du travail à la taille des villes. Etudier l'impact de cette taille sur les revenus réels est bien sûr une question intéressante<sup>13</sup>, et importante de surcroît pour les décideurs publics, mais ce n'est pas la même que celle, nécessairement préliminaire, qui consiste à évaluer les économies d'agglomération en termes de productivité. On ne peut étudier les disparités spatiales de salaire réel qu'une fois que les gains de salaire nominal, et la hausse du coût de la vie pour certains acteurs, sont évaluées de façon séparée. La présente analyse ne porte que sur la première composante, les gains de productivité.

De plus, lorsqu'on cherche à quantifier les économies d'agglomération engendrées par les grandes métropoles, c'est l'impact de la variable  $B_{c,t}$  qu'il s'agit d'identifier, et non celui de  $s_{c,t}$ . Cette identification pose cependant la question de l'horizon temporel des gains d'efficacité engendrés par les grandes villes. En effet, si les externalités pécuniaires se répercutent très rapidement dans les prix et sont capitalisées quasi instantanément dans les salaires nominaux, ce n'est généralement pas le cas des externalités technologiques ou des externalités de capital humain, dont l'impact se manifeste à plus long terme. Il existe donc des économies d'agglomération statiques et des économies d'agglomération dynamiques.

Lorsque les économies d'agglomération sont statiques, le terme  $B_{c,t}$  dépend uniquement des caractéristiques de la ville à la date  $t$ , et non des caractéristiques passées de la ville. La taille de la ville en  $t$  n'affecte donc que les salaires nominaux à la date  $t$ . Autrement dit, toute hausse de la taille de la ville à la date  $t+1$  est instantanément capitalisée dans les salaires de l'année  $t+1$ , qui ne dépendent pas de la taille de la ville à la date  $t$ . Si les économies d'agglomération sont statiques, les travailleurs d'une grande ville qui choisissent de migrer vers une petite ville perdent donc instantanément la « prime salariale » dont ils disposaient dans la grande ville.

En revanche, si les économies d'agglomération sont dynamiques, les économies d'agglomération à la date  $t$  dépendent des caractéristiques passées. Si les effets d'apprentissage sont plus forts dans les grandes métropoles, comme l'a souligné par exemple Glaeser (1999), ils permettent aux travailleurs de ces villes d'accumuler davantage de capital humain (autrement dit  $B_{c,t}$  ou  $s_{c,t}$  sont influencés par les valeurs passées de la taille de la ville). Comme les travailleurs qualifiés sont plus aptes à s'approprier et/ou à diffuser le produit des externalités technologiques (Lucas, 1988), les effets d'apprentissage dynamiques peuvent être encore amplifiés par la sélection spatiale des salariés.

Lorsque les économies d'agglomération sont dynamiques, les travailleurs d'une grande ville qui choisissent de migrer vers une petite ville sont aussi susceptibles de conserver une partie - au mieux la totalité - de la prime salariale dont ils bénéficiaient initialement. Si l'expérience acquise dans une grande ville est transférable dans le temps et l'espace, il existe un effet permanent des économies d'agglomération sur le capital humain permettant aux salariés travaillant dans une grande ville d'acquérir de l'expérience susceptible d'être valorisée ailleurs. Dans le cadre de l'évaluation empirique, il convient donc de distinguer ce qui, dans la prime salariale acquise dans les grandes métropoles, provient du tri spatial des travailleurs qualifiés, des économies d'agglomération statiques et des économies d'agglomération dynamiques.

---

<sup>13</sup> Cette question théorique a été largement étudiée depuis Roback (1982). La littérature empirique sur le sujet est bourgeonnante. Voir par exemple Albouy (2008 et 2009), Albouy, Leibovici, Warman, (2013) ou Moretti (2013).

## 1.2 Economies d'agglomération statiques : l'approche agrégée

Les économies d'agglomération statiques ont été les plus étudiées. La plupart des études empiriques ont estimé une spécification du type<sup>14</sup>:

$$\log \omega_{c,t} = Z_{c,t}\gamma + u_t + \eta_{c,t}, \quad (2)$$

où  $\omega_{c,t}$  représente le salaire nominal moyen dans la ville  $c$  à la date  $t$ ,  $Z_{c,t}$  est un vecteur de caractéristiques locales véhiculant les économies d'agglomération et les externalités de capital humain,  $u_t$  un effet fixe temporel capturant les chocs macroéconomiques affectant uniformément les villes, et  $\eta_{c,t}$  un terme résiduel d'erreur incluant toutes les autres variables susceptibles d'influencer le salaire de la ville.

Depuis l'étude séminale de Ciccone et Hall (1996), la littérature a montré que la densité locale des emplois, mesurée par les effectifs salariés par kilomètre carré de surface ( $\text{dens}_{c,t} = L_{c,t}/\text{area}_c$ ) véhiculait une grande partie des externalités technologiques et pécuniaires. Si cette variable est la seule incluse dans le vecteur  $Z_{c,t}$ , le coefficient  $\gamma$ , qui mesure l'élasticité du salaire nominal à la densité des emplois dans la ville, capture l'effet « total net » des économies d'agglomération. Total, parce qu'il reflète l'ensemble des gains et coûts liés à une plus forte densité, qu'ils transitent par  $A_{c,t}$ ,  $p_{c,t}$  ou  $r_{c,t}$  dans le modèle, net parce que le total des coûts est déduit du total des gains. C'est bien sûr un peu frustrant pour le décideur politique de ne pouvoir disposer que de l'effet total net de la densité. On aimerait savoir si les économies d'agglomération transitent par le prix des biens vendus ou par les externalités technologiques, et connaître la part des gains absorbés par la capitalisation foncière de ces effets.

Quantifier la résultante de tous ces phénomènes est cependant très utile, et c'est en tout état de cause le premier pas nécessaire à une meilleure connaissance des effets de la métropolisation sur la productivité du travail et les salaires. Si  $\gamma$  est positif, l'ensemble des externalités positives l'emportent sur les externalités négatives et la densification d'une ville engendre une prime salariale pour ses salariés : une hausse de 1 % de la densité des emplois dans la ville y occasionne une hausse de la rémunération nominale moyenne de ses travailleurs de  $\gamma$  %. On peut ainsi évaluer la part des différentiels spatiaux de salaire imputable aux écarts de densités. Si une ville est deux fois plus dense qu'une autre, ses travailleurs y bénéficient d'une prime salariale de  $(2^\gamma - 1) \times 100$  %<sup>15</sup>. D'une manière plus prospective, il est aussi possible de prédire l'impact d'une densification ou d'un desserrement des emplois de la ville sur la rémunération moyenne de ses salariés, ce qui est très utile du point de vue de la politique publique.

La densité de l'emploi ne capture cependant qu'une partie de l'effet « taille » des villes. En effet, une grande ville peut être très étendue, sans nécessairement être très dense. On a donc généralement coutume de considérer la superficie,  $\text{surf}_c$ , en complément de la densité dans les régressions. Même si la densité et la superficie sont deux mesures interconnectées de la taille du marché local, il est important de quantifier séparément leur influence, car les politiques publiques susceptibles d'affecter ces deux déterminants de la productivité ne sont pas les mêmes. Afin de faire jouer l'effet taille, une collectivité territoriale peut décider de modifier la superficie occupée par les activités urbaines – par exemple via un nouveau Plan Local d'Urbanisme (PLU). Elle peut aussi obtenir la désignation de certains quartiers en Zones Franches Urbaines, dans l'espoir d'y attirer de nouvelles

<sup>14</sup> Voir Duranton et Puga (2004) pour une recension des mécanismes théoriques à l'œuvre, ainsi que Rosenthal et Strange (2004), Puga (2010) et Combes et Lafourcade (2012) pour une revue des principaux résultats empiriques.

<sup>15</sup> Si deux villes  $c$  et  $c'$  sont telles que  $\text{dens}_{c'} = 2 \times \text{dens}_c$ , on a  $\log(\omega_{c'}/\omega_c) = \gamma \log(\text{dens}_{c'}/\text{dens}_c) = \gamma \log 2$ . On a donc  $\omega_{c'}/\omega_c = 2^\gamma$ , qui est le coefficient multiplicateur du salaire. Un doublement de la densité engendre ainsi une hausse du salaire moyen dans la zone de  $(2^\gamma - 1) \times 100$  %.

entreprises et de créer des emplois, sans pour autant modifier le plan d'occupation des sols de l'unité urbaine. Tandis que le PLU joue sur la taille physique de la ville, les ZFU visent à en modifier la densité. Ces deux politiques peuvent évidemment être combinées, mais il est intéressant pour le décideur public d'en évaluer séparément les effets.

De plus, les économies d'agglomération ne proviennent pas uniquement du marché local, mais aussi des marchés externes. Si les entreprises implantées à Bordeaux profitent évidemment des avantages offerts par cette grande métropole, les entreprises implantées à Lille exploitent, en sus des économies d'agglomération de la métropole lilloise, celles des agglomérations parisiennes, bruxelloises ou londoniennes, dont elles sont physiquement plus proches que Bordeaux. Les marchés externes ont aussi un effet sur la productivité locale, du fait de la mobilité pendulaire des salariés et du commerce inter-régional des biens. L'impact des marchés externes sur la productivité ou le salaire d'une ville est cependant partiellement atténué par la distance qui les sépare de la ville. C'est donc la facilité avec laquelle une entreprise de la ville peut accéder aux marchés externes, combinée à la taille de ces marchés, qui conditionne sa productivité.

Afin de prendre en compte ces effets, les études incluent généralement dans le vecteur  $Z_{c,t}$  une variable de « Potentiel marchand », calculé comme la somme des densités des marchés externes, escomptées de la distance séparant ces marchés de la ville, comme le suggérait le géographe Harris (1954) :  $MP_{c,t} = \sum_{c' \neq c} (\text{dens}_{c',t} / \text{dist}_{c'c})$ . Cette variable illustre l'idée que la portée spatiale des économies d'agglomération excède généralement la limite administrative des unités urbaines considérées. La notion d'aire urbaine présentée plus haut illustre bien ces effets de débordement. Les marchés externes ont des retombées d'autant plus positives sur la productivité d'une ville qu'ils en sont moins éloignés. La distance capture assez simplement cet effet, tout en ayant l'avantage de constituer une très bonne approximation des coûts et des temps de transport<sup>16</sup>.

Les variables de densité, de surface et de potentiel marchand capturent chacune un aspect particulier de la taille des villes. Il faut donc les intégrer simultanément dans le vecteur  $Z_{c,t}$  pour pouvoir identifier l'impact des économies d'agglomération statiques sur le salaire nominal de la zone. La spécification estimée est alors la suivante :

$$\log \omega_{c,t} = \gamma_1 \log \text{dens}_{c,t} + \gamma_2 \log \text{surf}_c + \gamma_3 \log MP_{c,t} + u_t + \eta_{c,t}.$$

Afin de dissocier l'impact de ces économies d'agglomération statiques des effets de composition de la main d'œuvre en termes de qualifications, de nombreuses études incluent des variables explicatives additionnelles mesurant le capital humain de la ville, comme la part des individus diplômés ou la part des différentes Catégories Socio-Professionnelles (CSP) dans la force de travail locale. Ces variables permettent de comparer les écarts de salaire entre des villes ayant en moyenne la même composition de la main d'œuvre : l'impact des économies d'agglomération est alors évalué à composition moyenne donnée des qualifications.

Cette stratégie présente cependant deux écueils. Tout d'abord, elle ne permet pas d'identifier l'effet des externalités de capital humain. En effet, une partie de l'effet positif du capital humain sur le salaire moyen dans la ville provient du fait que les travailleurs qualifiés, en interagissant avec leurs collègues de travail présents dans la ville, les rendent plus efficaces quel que soit leur niveau de qualification initial. Mais une autre partie provient du fait que les travailleurs qualifiés sont eux-

<sup>16</sup> À l'échelle des zones d'emploi françaises, Combes et Lafourcade (2005) trouvent une corrélation de 0,95 entre d'un côté la distance géodésique, et de l'autre le coût de transport routier généralisé des marchandises ou le temps de trajet des camions par la route. La corrélation entre le temps de transport en commun et la distance géodésique est également très élevée, comme le montre l'exemple du Grand Paris, traité dans l'Annexe 1 de ce rapport.

mêmes plus productifs, le seraient tout autant s'ils habitaient une petite ville, mais ont choisi d'habiter une grande métropole, par un effet de sélection.

Ensuite, l'éducation ou la CSP ne mesurent généralement que très approximativement les compétences réelles des salariés, qui dépendent d'une multitude d'autres caractéristiques individuelles, dont certaines sont observables, comme l'âge, le genre ou l'expérience, mais d'autres beaucoup moins, comme l'origine sociale, la profession des parents, la composition familiale, l'héritage génétique ou le talent. Si l'impact de ces inobservables n'est pas correctement neutralisé, les variables mesurant la taille de la ville capturent simultanément les économies d'agglomération et l'effet de ces caractéristiques individuelles, même si la spécification inclut par ailleurs des variables de contrôle du niveau de capital humain agrégé de la zone.

### 1.3 Economies d'agglomération statiques : l'approche individuelle

L'utilisation de données de panel permet heureusement de pallier en partie cette limite. Certaines bases de données annuelles, comme les Déclarations Annuelles de Données Sociales en France, enregistrent les rémunérations versées par chaque entreprise à chacun de ses salariés, mais aussi l'âge du salarié, son sexe, sa CSP ou le nombre d'années d'activité passées dans le secteur privé. Le fait de disposer de multiples observations dans le temps pour chaque salarié, permet d'expliquer le salaire de l'individu à la date  $t$ ,  $\omega_{i,t}$ , à la fois par les caractéristiques économiques de la ville où il travaille,  $Z_{c,t}$ , et par ses caractéristiques individuelles, comme l'a fait l'étude séminale de Glaeser et Maré (2001).

Afin de comprendre en quoi l'utilisation de données de panel permet une analyse plus fine des économies d'agglomération, reprenons le modèle théorique évoqué plus haut, mais dans lequel on considère cette fois des travailleurs individuels hétérogènes. L'emploi efficace de l'entreprise s'écrit maintenant :

$$s_{c,t}L_{c,t} = \sum_{i \in (c,t)} s_{i,t}l_{i,t},$$

où  $l_{i,t}$  est le nombre d'unités de travail effectuées par le salarié  $i$  à la date  $t$ , et  $s_{i,t}$  son efficacité individuelle à cette même date.

Le profit de l'entreprise, qui utilise désormais des travailleurs n'ayant pas tous les mêmes aptitudes, et donc la même rémunération, devient :

$$\pi_{c,t} = p_{c,t}Y_{c,t} - \sum_{i \in (c,t)} \omega_{i,t}l_{i,t} - r_{c,t}K_{c,t}.$$

La maximisation de ce profit permet d'obtenir le salaire nominal de chaque individu, en fonction des économies d'agglomération de la ville  $c$  où il travaille à la date  $t$ ,  $c(i,t)$ , et de son efficacité individuelle  $s_{i,t}$  :

$$\omega_{i,t} = B_{c(i,t),t} s_{i,t}.$$

L'efficacité de chaque salarié dépend ainsi maintenant de ses caractéristiques individuelles. Certaines, comme l'âge<sup>17</sup> ou l'expérience constituent un vecteur d'observables  $X_{i,t}$ , qui varient au cours du temps. D'autres caractéristiques observables, comme le genre, ne varient pas au cours du temps, tout comme de nombreuses caractéristiques inobservables (c'est le cas, par exemple, de l'origine sociale ou de certains traits de la personnalité). Ces caractéristiques peuvent être capturées par un effet fixe individuel,  $u_i$ , puisqu'on dispose pour un même individu de plusieurs

<sup>17</sup> L'impact de l'âge sur le salaire nominal étant non linéaire du fait des entrées et sorties des individus sur le marché du travail, on peut capturer cet effet en considérant, parallèlement à l'âge, l'âge au carré.



observations au cours du temps. Enfin, certaines aptitudes (ou inaptitudes, comme la maladie) constituent un choc idiosyncratique,  $\epsilon_{i,t}$ , que les données ne permettent d'appréhender que de façon résiduelle.

On peut alors spécifier le salaire nominal versé à chaque individu :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t}\gamma + u_t + \eta_{c(i,t),t} + \epsilon_{i,t}. \quad (3)$$

Cette spécification individuelle à la Mincer (1958) présente deux principaux avantages par rapport à la spécification agrégée (2). Tout d'abord, la mesure des économies d'agglomération est purgée de l'effet des compétences propres du salarié. On peut ainsi évaluer l'écart de salaire qui résulte de la localisation d'un même salarié dans une grande et une petite ville, alors que l'estimation sur données agrégées permet seulement de comparer les écarts de productivité entre des villes ayant en moyenne la même composition de la main d'œuvre, mesurée par des indicateurs agrégés très imparfaits. In fine, le coefficient  $\gamma$  ne reflète plus que les économies d'agglomération véhiculées par la taille de la métropole, et pas le gain de productivité engendré par le tri spatial des salariés dans l'espace.

Ensuite, l'approche individuelle permet aussi d'identifier ce qui, dans la prime salariale offerte par les grandes métropoles, revient aux externalités de capital humain dont les travailleurs qualifiés font bénéficier les autres travailleurs de la ville (que l'on peut capturer par des variables d'éducation ou de CSP incluses dans  $Z_{c(i,t),t}$ ), ou aux aptitudes individuelles du salarié  $s_{i,t} = \exp(u_i + X_{i,t}\theta + \epsilon_{i,t})$ .

D'un point de vue économétrique, la spécification (3) peut être estimée en premières différences, de manière à faire disparaître l'effet fixe individuel  $u_i$ . Un exemple simple permet de comprendre comment se fait l'identification des économies d'agglomération dans ce contexte. Supposons, pour simplifier, que l'efficacité individuelle des travailleurs ne dépende que de caractéristiques individuelles invariantes au cours du temps,  $u_i$  (on néglige pour simplifier l'effet des observables  $X_{i,t}$ ). On a donc :  $\log \omega_{i,t} = u_i + Z_{c(i,t),t}\gamma + u_t + \eta_{c(i,t),t} + \epsilon_{i,t}$ .

Pour les individus qui travaillent dans la même ville pendant deux années consécutives,  $t-1$  et  $t$ , le taux de croissance des salaires nominaux entre ces deux dates s'écrit :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = (Z_{c,t} - Z_{c,t-1})\gamma + \eta_{c,t} - \eta_{c,t-1} + u_t - u_{t-1}.$$

La variation temporelle des caractéristiques de la ville  $c$ ,  $Z_{c,t} - Z_{c,t-1}$ , permet donc d'identifier  $\gamma$ . En revanche, pour les travailleurs qui changent de ville entre les deux dates, la première différence s'écrit :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = (Z_{c',t} - Z_{c,t-1})\gamma + \eta_{c',t} - \eta_{c,t-1} + u_t - u_{t-1}.$$

C'est donc ici la variation temporelle et spatiale des caractéristiques de la ville,  $Z_{c',t} - Z_{c,t-1}$ , qui permet d'identifier  $\gamma$ .

L'estimation des économies d'agglomération par une méthode économétrique simple est donc possible même s'il n'existe aucun travailleur mobile dans le panel, puisqu'il n'y a qu'un seul coefficient à estimer dans les deux cas. L'introduction d'un effet fixe individuel  $u_i$  impose, lorsque le nombre d'observations est très élevé, de recourir à une méthode d'estimation « within », qui produit un estimateur convergent du même paramètre  $\gamma$  que celui des MCO obtenu à partir de la spécification (3) transformée, dans laquelle les variables expliquée et explicatives sont centrées sur leur moyenne individuelle. Autrement dit, l'estimation « within » consiste à prendre comme variable dépendante la différence entre le salaire de l'individu  $i$  à la date  $t$  et son salaire moyen sur

l'ensemble des années disponibles dans le panel,  $\log \omega_{i,t} - \log \omega_{i,\cdot}$ , et à expliquer cet écart par  $Z_{c,t} - Z_{c,\cdot}$ , où  $Z_{c,\cdot}$  est le vecteur des variables explicatives centrées, ce qui permet de faire disparaître par différence  $u_i$ . Toutes les caractéristiques individuelles qui influencent à l'identique le salaire annuel et le salaire moyen de l'individu (comme l'origine sociale), n'ont alors plus d'impact sur le salaire en différence. L'intérêt de l'estimation « within » est donc de neutraliser l'effet de toutes les caractéristiques individuelles inobservées ne variant pas au cours du temps.

L'estimation « within » de la spécification (3) pose cependant plusieurs problèmes. Le premier est lié au fait que l'effet des variables observées invariantes au cours du temps, individuelles ou locales, disparaît aussi par différence. L'impact du genre, par exemple, ne peut pas être étudié, et les économies d'agglomération liées à la superficie de la ville ne peuvent pas non plus être évaluées.

Une deuxième limite, également présente pour l'estimation MCO en niveau, résulte d'un biais d'endogénéité dû à des variables omises. En effet, le vecteur des caractéristiques locales observées  $Z_{c,t}$  ne capte pas tous les facteurs locaux déterminant la productivité, qui sont capturés également par des caractéristiques inobservées,  $\eta_{c,t}$ . Par exemple, certains biens publics, comme les routes, les gares, les hôpitaux ou les centres de recherche, sont construits en priorité dans les grandes métropoles, et ils améliorent la productivité des salariés. A contrario, certaines aménités de consommation, comme les biens culturels ou les restaurants, sont à la fois surreprésentées dans les grandes métropoles et valorisées par les salariés. Ceux qui migrent vers les grandes villes pour bénéficier de ces aménités consomment également plus de logement, ce qui augmente la rente foncière payée par les entreprises, qui sont alors contraintes de réduire les salaires nominaux pour préserver leurs marges.

Ce *biais de variables omises* peut donc conduire à surestimer, autant qu'à sous-estimer, les économies d'agglomération. On peut tenter de corriger ce biais en incluant des variables de contrôle additionnelles, construites par exemple à partir des données produites par l'Institut Géographique National (qui répertorie de nombreux équipements publics<sup>18</sup>). Mais les variables potentiellement omises sont très nombreuses et on peut douter de la capacité à toutes les identifier.

Notons qu'il est crucial pour les politiques publiques d'estimer l'impact non biaisé de la taille des villes. En effet, si ce sont les biens publics locaux, et non la taille, qui engendrent des gains de productivité, alors les collectivités territoriales doivent accroître leurs dépenses d'investissement plutôt que de chercher à attirer des entreprises par le biais d'une défiscalisation ou d'un subside à l'implantation. Si seule la taille influence la productivité, alors les investissements publics seront sans effet sur les salaires. Il est aussi possible que les biens publics locaux attirent les entreprises<sup>19</sup> et les rendent plus productives, ce qui n'est pas une économie d'agglomération, mais que la densification induite par l'arrivée de ces entreprises crée ensuite des économies d'agglomération.

Enfin une dernière limite de l'estimation « within » de la spécification (3) réside dans le fait que, si le panel contient des travailleurs mobiles, le taux de croissance des salaires inclut un terme d'erreur de structure complexe, qui dépend notamment de la différence entre les chocs inobservés ( $\eta_{c,t}$  et

---

<sup>18</sup> Les équipements publics antérieurs à la première année d'observation du panel (par exemple, un aéroport construit en 1970, si le panel commence en 1993) ne posent pas de problème. En effet, ces équipements augmentent la productivité des individus à toutes les dates, mais ils augmentent de la même manière leur productivité moyenne sur la période. Lorsqu'on fait la différence entre ces deux variables par la méthode d'estimation « within », les deux effets s'annulent. L'impact de la taille de la ville sur cette différence n'est donc pas biaisé par l'omission de ces équipements.

<sup>19</sup> Holl (2004) estime par exemple que le taux d'implantation des firmes dans les municipalités espagnoles situées à moins de 10km des autoroutes a été supérieur de 14 % à celui des autres municipalités espagnoles sur la période 1980-1994. L'étude de Strauss-Kahn et Vives (2009) illustre le même genre de résultats pour les sièges sociaux : sur la période 1996-2001, ces derniers se sont 2 (1,4) fois plus implantés à proximité des métropoles américaines disposant d'un gros (petit) hub aéroportuaire, que dans les villes américaines sans aéroport.

$\eta_{c,t-1}$ ) émanant de deux villes ( $c$  et  $c'$ ) qui ne sont pas identiques pour tous les migrants, même ceux travaillant initialement en  $c$ , puisqu'ils ont des destinations  $c'$  potentiellement différentes. Il n'est donc pas possible de trier simplement les individus de manière à corriger l'hétéroscédasticité des résidus et à rendre l'inférence robuste (par la méthode des « clusters » par exemple). On peut bien évidemment ignorer ce problème, mais les écarts-types obtenus par la méthode « within » peuvent alors être sérieusement biaisés, comme l'a montré Moulton (1990). On ne sait donc pas si les économies d'agglomération sont réellement significatives ou pas.

Il est heureusement possible de pallier partiellement ces trois limites simultanément, à l'aide d'une procédure d'estimation à deux étapes proposée par Combes, Duranton et Gobillon (2008) :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + \epsilon_{i,t}, \quad (4)$$

$$\beta_{c,t} = Z_{c,t}\gamma + u_t + \eta_{c,t}.$$

La première étape consiste à identifier séparément la part de la prime salariale liée aux compétences individuelles et celle liée à toutes les caractéristiques observables et inobservables de la ville, capturées par l'inclusion d'un effet fixe ville-temps,  $\beta_{c(i,t),t}$ <sup>20</sup>. Les erreurs  $\epsilon_{i,t}$  ne dépendent alors plus de chocs agrégés, et elles peuvent être supposées identiquement et indépendamment distribuées. La méthode des MCO est alors adaptée, et les écarts-types n'ont pas besoin d'être corrigés. Cette première étape permet de purger la productivité du salarié des effets liés à ses compétences propres, de manière à ne conserver que la productivité « résiduelle » estimée émanant de l'ensemble des caractéristiques locales, observées ou pas,  $\hat{\beta}_{c,t}$ .

Afin d'identifier le rôle des déterminants des économies d'agglomération, et notamment de la densité et du potentiel de marché évoqués plus haut, la seconde étape consiste à expliquer cette productivité « résiduelle » par le vecteur de variables  $Z_{c,t}$ . Cependant, comme cette productivité est issue d'une première estimation, elle contient de l'erreur de mesure (d'autant plus faible qu'elle est précisément estimée). Il faut donc corriger les écarts-types de seconde étape, en utilisant une méthode d'estimation autorisant une certaine hétéroscédasticité des résidus, comme les Moindres Carrés Quasi-Généralisés ou pondérés. De plus, il existe toujours un biais de variables omises dans la seconde étape d'estimation, qu'il convient de corriger par des méthodes économétriques appropriées comme la technique des variables instrumentales, sur laquelle nous nous reviendrons dans la troisième partie de ce rapport.

L'identification des économies d'agglomération dans le cadre du modèle (4) est cependant plus contrainte, car l'estimation de  $\gamma$  en seconde étape requiert la présence de travailleurs mobiles dans le panel. Pour bien comprendre cette contrainte, reprenons l'exemple simplifié déjà évoqué plus haut, dans lequel l'efficacité des travailleurs dépend uniquement de leurs caractéristiques individuelles invariantes au cours du temps,  $u_i$ . La première étape consiste donc à estimer  $\ln \omega_{i,t} = u_i + \beta_{c(i,t),t} + \epsilon_{i,t}$ . Pour les travailleurs qui ne bougent pas, le taux de croissance des salaires nominaux entre deux dates s'écrit :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = \beta_{c,t} - \beta_{c,t-1} + \epsilon_{i,t} - \epsilon_{i,t-1}.$$

<sup>20</sup> Notons que cet effet fixe capture l'influence de toutes les caractéristiques locales autres que la taille de la ville (aménités, biens publics, infrastructures de transport, ...), ce qui permet de neutraliser le biais de variables omises évoqué plus haut. Mais il capte aussi simultanément l'effet des caractéristiques individuelles croissant linéairement au cours du temps, comme l'âge. Son inclusion ne permet donc plus d'identifier l'influence de cette variable (l'effet de l'âge au carré, lui, reste identifiable).

Pour ceux qui changent de zone, on a en revanche :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = \beta_{c,t} - \beta_{c,t-1} + \epsilon_{i,t} - \epsilon_{i,t-1}.$$

Contrairement à l'équation (3), dans laquelle il n'y avait qu'un seul paramètre à identifier ( $\gamma$ ), il faut maintenant estimer tous les effets fixes  $\beta_{c,t}$ , soit autant de paramètres que de villes et d'années disponibles dans l'échantillon.

Si le panel ne contient aucun travailleur mobile, on peut identifier la variation temporelle des effets fixes locaux, mais pas leur variation spatiale. S'il existe aussi des travailleurs mobiles, on peut se rendre compte que le modèle est identifié en écrivant des doubles différences : différence entre le taux de croissance du salaire d'un travailleur  $i'$  ayant migré de  $c$  vers  $c'$ , et le taux de croissance du salaire d'un travailleur  $i$  n'ayant pas bougé de  $c$ , soit :

$$(\log \omega_{i',t} - \log \omega_{i',t-1}) - (\log \omega_{i,t} - \log \omega_{i,t-1}) = \beta_{c',t} - \beta_{c,t}.$$

Pour tout couple de localisation  $(c, c')$ , la différence des taux de croissance du salaire des travailleurs mobiles et immobiles permet d'identifier la différence de productivité entre les deux localisations. Le taux de croissance du salaire des travailleurs immobiles permet quant à lui d'identifier la variation temporelle de la productivité résiduelle. Les effets fixes  $\beta_{c,t}$  sont donc tous identifiables<sup>21</sup> dès lors qu'il existe suffisamment de mobilité des travailleurs entre les villes du panel.

Ce point méthodologique, qui paraît technique de prime abord, est bien plus important qu'on ne le pense. L'identification précise des économies d'agglomération statiques n'est possible que si on dispose d'un panel de travailleurs suffisamment mobiles au cours du temps, et dont la mobilité présente beaucoup de variabilité spatiale. Cette limite explique pourquoi la mesure des économies d'agglomération à l'échelle d'une métropole particulière, aussi grande soit-elle (comme le Grand Paris), n'est tout simplement pas possible.

On peut toutefois mesurer des effets d'agglomération spécifiques à différentes classes de villes définies en fonction de la taille (par les quintiles de densité par exemple). Pour conserver toute l'information nécessaire à la réalisation de l'inférence statistique, on regroupe les villes par classe, et on interagit l'effet de la taille de la ville avec une indicatrice d'appartenance à chaque classe. La spécification estimée est dans ce cadre :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t),t} + \epsilon_{i,t}, \\ \beta_{c,t} &= \sum_{q=1}^5 (Z_{c,t} \times q_c) \gamma_q + u_t + \eta_{c,t}, \end{aligned} \tag{5}$$

où  $q_c$  est une variable binomiale indiquant la classe à laquelle appartient la ville  $c$ , définie par le quintile de densité moyenne (calculée sur l'ensemble des années disponibles dans le panel) auquel appartient la ville  $c$ . C'est l'approche qui sera adoptée dans la troisième partie de ce rapport pour mettre en évidence quelques faits saillants spécifiques aux différentes classes, en particulier à celles incorporant les zones à très forte densité localisées dans la métropole du Grand Paris.

---

<sup>21</sup> Sauf un, que l'on peut normaliser à 0, vu qu'on ne peut identifier que les premières différences.

## 1.4 Economies d'agglomération statiques : l'approche sectorielle

Jusqu'à présent, nous avons considéré que les économies d'agglomération, qu'elles soient véhiculées par des externalités technologiques ou des externalités pécuniaires, ont un impact sur la productivité qui ne dépend pas du secteur d'appartenance des travailleurs. Or, elles peuvent bien évidemment différer d'une industrie à l'autre. La densité ou la proximité des marchés compte bien davantage pour des secteurs comme celui des nouvelles technologies de l'information ou de la communication, qui nécessitent de très nombreuses interactions en face-à-face. Dans le même esprit, les externalités pécuniaires sont vraisemblablement plus fortes pour les secteurs dans lesquels le prix des biens et des inputs dépendent très largement des coûts de transport.

D'un point de vue méthodologique, la prise en compte de cette hétérogénéité sectorielle peut s'effectuer de plusieurs façons. La première consiste à scinder l'échantillon des travailleurs selon le secteur d'appartenance, et à estimer la spécification (3) pour chaque industrie présente dans le panel. La spécification estimée est alors :

$$\log \omega_{i,t} = u_i + X_{i,t} \theta_{k(i,t)} + Z_{c(i,t),t} \gamma_{k(i,t)} + \eta_{c(i,t),k(i,t),t} + \epsilon_{i,t}, \quad (6)$$

où  $k(i,t)$  est le secteur d'appartenance de l'individu  $i$  à la date  $t$ , et  $\eta_{c(i,t),k(i,t),t}$  un effet aléatoire capturant les chocs sectoriels locaux à cette même date. On estime ainsi un coefficient différent pour chaque secteur,  $\gamma_k$ , ce qui permet de comparer l'intensité des économies d'agglomération statiques entre secteurs. Pour pouvoir utiliser cette méthode d'estimation, il faut cependant disposer d'un nombre suffisant d'individus pour chaque secteur. On ne peut donc le faire que sur la base d'une classification sectorielle relativement agrégée, comme par exemple la NES36. Une manière de pallier le manque d'observations éventuel consiste à empiler les industries, et à introduire un terme d'interaction permettant de mesurer le surcroît ou le déficit d'économies d'agglomération caractérisant le secteur  $k$  par rapport à la moyenne des secteurs.

Ces deux méthodes d'estimation présentent néanmoins les mêmes écueils que ceux évoqués pour la spécification (3) : les écart-types ne sont pas précisément estimés du fait de l'hétéroscédasticité des résidus. Il est donc préférable d'adapter l'approche en deux étapes décrite dans la spécification (4), et de rendre la productivité résiduelle estimée lors de la première étape propre au secteur :

$$\ln \omega_{i,t} = u_i + X_{i,t} \theta + \beta_{c(i,t),k(i,t),t} + \epsilon_{i,t}, \quad (7)$$

$$\beta_{c,k,t} = Z_{c,t} \gamma_k + u_t + \eta_{c,k,t}.$$

Les économies d'agglomération mesurées lors de la seconde étape dépendent alors du secteur considéré.

Cette spécificité a cependant une contrepartie : les caractéristiques locales susceptibles d'influencer la productivité résiduelle sectorielle,  $\beta_{c,k,t}$ , sont bien plus nombreuses qu'avant. En effet, jusqu'à présent, nous n'avons considéré que les externalités d'urbanisation liées à la taille globale des villes, et qui ne dépendent pas de leur composition sectorielle. Or, comme nous l'avons vu en introduction, il existe aussi des économies d'agglomération intra-sectorielles. Certaines entreprises, comme celles appartenant aux industries de haute-technologie, produisent des externalités dont profitent essentiellement leurs concurrentes. Elles se regroupent alors en pépinières de manière à tirer parti de l'accumulation localisée d'informations, de connaissances et d'innovations émanant des autres entreprises du secteur. Ce n'est alors plus seulement la taille de l'économie locale qui induit des gains de productivité, mais plus spécifiquement celle du secteur considéré.

Cela implique d'estimer en seconde étape une spécification plus sophistiquée du type :

$$\beta_{c,k,t} = Z_{c,t} \gamma_k + W_{c,k,t} \phi_k + u_t + \eta_{c,k,t},$$

où  $W_{c,k,t}$  est un vecteur de caractéristiques locales propres au secteur  $k$ . Le paramètre  $\phi_k$  mesure l'effet des externalités de localisation propres au secteur  $k$  sur la productivité. Il est d'usage d'inclure dans le vecteur  $W_{c,k,t}$  la spécialisation, mesurée par la part du secteur dans l'emploi de la ville,  $\text{spe}_{c,k,t} = L_{c,k,t} / \sum_k L_{c,k,t}$  (Glaeser, Kallal, Scheinkman et Schleifer 1992 ; Combes, Duranton et Gobillon, 2008). Les qualifications au sein du secteur jouent également un grand rôle dans la sédimentation des bénéfices de la spécialisation. La capacité des travailleurs à s'approprier les retombées positives de ces « spillovers » intra-sectoriels dépend de leur qualification. Les plus qualifiés sont généralement plus aptes à fertiliser leur entreprise avec les savoir-faire acquis dans les autres firmes de la pépinière. Afin de mesurer cet effet, on peut aussi inclure dans  $W_{c,k,t}$  les parts de l'emploi sectoriel local relevant des différentes Catégories Socio-Professionnelles  $j$  présentes dans le panel,  $\text{csp}_{j,c,k,t} = L_{j,c,k,t} / L_{c,k,t}$  (Combes, Duranton et Gobillon, 2008). L'identification d'un tel modèle est cependant très compliquée car elle requiert en première étape l'estimation de très nombreux effets fixes (nombre de villes multiplié par le nombre de secteur et par le nombre d'années disponibles dans le panel), et la présence d'un grand nombre de travailleurs mobiles et immobiliers entre zones et secteurs.

Une alternative, proposée par Combes, Duranton et Gobillon (2008) consiste à estimer :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t} \theta + \beta_{c(i,t),t} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \\ \beta_{c,t} &= Z_{c,t} \gamma + u_t + \eta_{c,t}, \end{aligned} \quad (9)$$

où  $\vartheta_{k(i,t),t}$  est un effet fixe capturant les chocs sectoriels non spécifiques au marché local du travail du salarié  $i$  à la date  $t$ .

Ce modèle est cependant moins général que les modèles (7) et (8), car l'estimation de première étape ne permet pas de prendre en compte l'effet de toutes les variables inobservées ville-secteur-temps susceptibles d'influencer la productivité. De plus, il fait l'hypothèse que les externalités d'urbanisation mesurées en seconde étape ne sont pas propres au secteur (puisque le coefficient  $\gamma$  ne dépend pas de  $k$ ). En contrepartie, il a l'avantage de permettre de dissocier l'impact des externalités d'urbanisation et de localisation, ce qui est intéressant du point de vue de la politique publique. Il est en effet crucial pour un élu local de bien identifier les secteurs - et au sein de ces secteurs, les qualifications - les plus aptes à engranger des économies d'agglomération. Les collectivités territoriales peuvent ainsi actionner des leviers spécifiques favorisant l'arrivée ou le développement de certains secteurs.

C'est l'approche adoptée dans la troisième partie de ce rapport<sup>22</sup>, dans laquelle nous estimerons :

$$\log \omega_{i,t} = u_i + X_{i,t} \theta + \beta_{c(i,t),t} + \varphi_k \log \text{spe}_{c(i,t),k(i,t),t} + \sum_j \phi_{k(i,t)} \text{csp}_{j,c(i,t),k(i,t),t} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \quad (10)$$

$$\beta_{c,t} = \gamma_1 \log \text{dens}_{c,t} + \gamma_2 \log \text{surf}_c + \gamma_3 \log \text{MP}_{c,t} + u_t + \eta_{c,t}.$$

### 1.5 Economies d'agglomération dynamiques

Les économies d'agglomération peuvent cependant aussi avoir un impact permanent, et pas seulement transitoire, sur les salaires. Les canaux par lesquels transitent ces effets dynamiques sont potentiellement nombreux. Tout d'abord, les externalités technologiques favorisent généralement la croissance de la productivité globale des facteurs,  $A_{c,t}$ . De plus, l'accumulation du capital humain est

<sup>22</sup> L'hétérogénéité sectorielle des économies d'agglomération sera analysée dans la phase 2 de notre étude.

souvent plus facile ou plus rapide dans les grandes métropoles, comme l'a suggéré Lucas (1988) : l'efficacité individuelle des travailleurs,  $s_{i,t}$ , y augmente donc plus rapidement que dans les petites villes. Dans ces conditions, le temps passé dans une ville peut compter tout autant que la taille de la ville à une date donnée. Les salariés ayant travaillé dans une grande métropole, même s'ils finissent par la quitter, transfèrent avec eux une partie, voire la totalité, des gains de productivité acquis dans cette ville. Ils sont donc ex post, toutes choses égales par ailleurs, plus productifs que les salariés n'ayant jamais travaillé dans une grande ville.

Mais il existe d'autres canaux de transmission des économies d'agglomération dynamiques. Si la diffusion de l'information sur la qualité des biens et des facteurs de production est meilleure dans les grandes métropoles, cela affecte en retour le taux de croissance du prix des biens finaux,  $p_{c,t}$  ou celui des facteurs de production  $r_{c,t}$  lorsque, par exemple, les entreprises peuvent adapter leurs marges en situation de concurrence imparfaite. Ainsi, même s'il est plus vraisemblable que les économies d'agglomération dynamiques recouvrent des externalités technologiques ou des externalités de capital humain, il n'est pas exclu qu'elles transitent par des mécanismes de marché et recouvrent aussi des externalités pécuniaires. Il est difficile d'isoler précisément le canal par lequel opèrent les économies d'agglomération dynamiques. Tout comme pour les économies d'agglomération statiques, on peut néanmoins quantifier leur effet total net sur le salaire nominal.

Pour bien comprendre comment se fait cette identification, reprenons l'exemple simple dans lequel l'efficacité des travailleurs dépend uniquement des caractéristiques individuelles invariantes au cours du temps,  $u_i$ . Le taux de croissance du salaire nominal du salarié  $i$  entre les date  $t-1$  et  $t$  est la première différence des économies d'agglomération statiques :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \varepsilon_{i,t},$$

où  $\varepsilon_{i,t} = \epsilon_{i,t} - \epsilon_{i,t-1}$  est le terme d'erreur égal à la différence des chocs idiosyncratiques.

La manière la plus simple de modéliser les économies d'agglomération dynamiques consiste à introduire un effet fixe additionnel,  $\mu_{c(i,t-1),t-1}$ , mesurant l'effet pour le salarié  $i$  d'avoir travaillé dans la ville  $c$  à la date  $t-1$  :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \mu_{c(i,t-1),t-1} + \varepsilon_{i,t}.$$

On peut alors déduire par récursivité l'impact des économies d'agglomération dynamiques sur le taux de salaire nominal du salarié à la date  $t$  :

$$\log \omega_{i,t} = \log \omega_{i,0} + \beta_{c(i,t),t} + \sum_{\tau=1}^t \mu_{c(i,t-\tau),t-\tau} + \zeta_{i,t}, \quad (11)$$

où  $\zeta_{i,t}$  est un bruit blanc.

Le taux de salaire nominal dépend donc des compétences accumulées par le passé dans les villes où le salarié a travaillé antérieurement,  $\mu_{c(i,t-\tau),t-\tau} \forall \tau < t$ . On voit ici que, même si les économies d'agglomération dynamiques n'affectent que le taux de croissance de l'année en cours des caractéristiques locales ( $A_{c,t}, p_{c,t}, r_{c,t}, s_{c,t}$  et  $s_{i,t}$ ), celles-ci ont un effet permanent sur le niveau absolu du salaire nominal.

La spécification (11) repose cependant sur des hypothèses implicites restrictives. Tout d'abord, elle suppose que les économies d'agglomération dynamiques sont parfaitement transférables au cours du temps. Or il n'est pas impossible qu'après un certain laps de temps, les gains de productivité accumulés par le passé connaissent une certaine forme de dépréciation, auquel cas le terme  $\sum_{\tau=1}^t \mu_{c(i,t-\tau),t-\tau}$  doit être déflaté pour tenir compte de l'obsolescence des connaissances.



De plus, la spécification (11) est certainement plus adaptée aux salariés qui n'ont jamais changé de ville, qu'à ceux qui ont fait le choix de migrer, à un moment ou à un autre de leur carrière. On peut en effet penser qu'un salarié ne valorise pas nécessairement de la même manière les connaissances accumulées par le passé lorsqu'il migre vers une plus grande ville, vers une plus petite ville, ou lorsqu'il reste dans sa ville d'origine, qu'elle soit grande ou petite. En d'autres termes, les économies d'agglomération dynamiques ne sont pas nécessairement parfaitement transférables d'un lieu à l'autre, et leur degré de portabilité dépend à la fois de la ville d'origine et de destination.

Il est donc plus pertinent de considérer que les économies d'agglomération dynamiques dépendent à la fois de l'origine et de la destination des salariés, auquel cas :

$$\log \omega_{i,t} = \log \omega_{i,0} + \beta_{c(i,t),t} + \sum_{\tau=1}^t \mu_{c'(i,t-\tau),c(i,t),t-\tau} + \zeta_{i,t},$$

où  $\mu_{c'(i,t-\tau),c(i,t),t-\tau}$  capture l'effet pour le salarié  $i$  travaillant dans la ville  $c$  à la date  $t$  d'avoir travaillé dans la ville  $c'$  à la date  $\tau < t$ , effet lui-même variable dans le temps

L'estimation de cette spécification pose cependant deux problèmes majeurs. Pour comprendre le premier problème, reprenons l'exemple des salariés qui ne changent pas de ville entre les dates  $t-1$  et  $t$ . Le taux de croissance de ces travailleurs immobiles s'écrit :

$$\log \omega_{i,t} - \log \omega_{i,t-1} = \beta_{c(i,t),t} - \beta_{c(i,t-1),t-1} + \mu_{c(i,t-1),c(i,t),t-1} + \varepsilon_{i,t}.$$

On ne peut pas identifier séparément le gain de productivité émanant des économies d'agglomération dynamiques ( $\mu_{c(i,t-1),c(i,t),t-1}$ ), et celui engendré par une évolution des économies d'agglomération statiques ( $\beta_{c(i,t),t} - \beta_{c(i,t-1),t-1}$ ), liée par exemple à un changement de la densité de la ville. Pour résoudre ce problème, De la Roca et Puga (2012) font l'hypothèse que les économies d'agglomération statiques et dynamiques ne varient pas au cours du temps, ce qui revient à supposer que  $\beta_{c,t} = \beta_c \forall t$  et  $\mu_{c',c,t-\tau} = \mu_{c',c} \forall \tau \leq t$ . Le coefficient  $\mu_{c',c}$ , qui mesure l'effet des compétences accumulées par le salarié dans une seule et même ville, capture donc à la fois l'effet des économies d'agglomération dynamiques et l'impact de l'évolution des économies d'agglomération statiques.

Sous ces hypothèses, il est possible de reformuler la spécification initiale du taux de salaire nominal de première étape, en fonction de l'expérience acquise par les individus dans chaque zone :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t)} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \sum_{c'} \mu_{c',c(i,t)} (e_{i,c',t} - e_{i,t}/C) + \varepsilon_{i,t}, \quad (12)$$

où  $e_{i,t}$  représente l'expérience totale de l'individu  $i$ , c'est-à-dire le nombre total d'années d'activité, qui est une des variables entrant dans le vecteur de caractéristiques individuelles  $X_{i,t}$ , et  $e_{i,c',t}$  le nombre d'années d'activité passées dans la ville  $c'$  jusqu'à la date  $t$ ,  $C$  étant le nombre total de villes. Le terme  $(e_{i,c',t} - e_{i,t}/C)$  mesure donc le surcroît ou le déficit d'expérience accumulée en  $c'$  par le passé par rapport à une distribution spatiale de référence de l'expérience globale entre villes, la référence choisie étant ici une distribution uniforme. Le paramètre  $\mu_{c',c}$  mesure donc la valorisation supplémentaire dans la ville  $c$  de l'expérience passée dans la ville  $c'$ , par rapport à cette référence.

La spécification (12) présente deux principaux avantages. Tout d'abord, elle permet de comparer la valeur des coefficients  $\beta_{c'}$ , la prime salariale engendrée par le fait de travailler aujourd'hui en  $c$ , et  $\mu_{c',c}$ , la prime salariale engendrée par le nombre d'années antérieures passées à travailler en  $c$ . Cette comparaison est intéressante, même si elle ne permet pas d'évaluer parfaitement le poids respectif des économies d'agglomération statiques et des économies d'agglomération dynamiques, pour les raisons décrites plus haut.

Ensuite, elle permet aussi d'évaluer l'ampleur respective des coefficients  $\mu_{c',c}$ . On peut ainsi comparer les gains de productivité dynamiques tirés de l'expérience passée dans une grande ou une petite ville  $c'$ , selon qu'elle est ensuite valorisée dans grande ou une petite ville  $c$ . Cela permet d'évaluer le degré de transférabilité d'une expérience passée selon les villes d'origine et de destination du salarié, et d'étudier le rôle joué plus spécifiquement par la taille des villes d'origine et de destination dans cette portabilité.

L'étape suivante consiste donc assez naturellement à estimer en seconde étape une spécification du type :

$$\mu_{c',c} = Z_{c',\cdot} \psi + Z_{c',\cdot} (Z_{c,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{c',c}, \quad (13)$$

où  $Z_{c,\cdot}$  est le vecteur incluant la moyenne, sur toutes les années disponibles dans le panel, des caractéristiques locales mesurant la taille de la zone  $c$ ,  $Z_{\cdot,\cdot}$  le vecteur incluant la moyenne de ces moyennes, et  $\zeta_{c',c}$  un terme aléatoire. Lorsque l'individu travaille dans une ville ayant les caractéristiques locales moyennes de l'échantillon, le second terme de la spécification (13) disparaît. Le paramètre  $\psi$  mesure donc la valorisation pour le salarié travaillant dans cette ville « moyenne » de l'expérience passée dans la ville  $c'$ . Le paramètre  $\nu$  mesure quant à lui la valorisation additionnelle tirée du fait que la ville  $c$  dans laquelle travaille actuellement le salarié est plus ou moins dense que la ville « moyenne ».

Le principal problème des spécifications (12) et (13) est que le nombre des paramètres à estimer est potentiellement très important (égal au carré du nombre de villes). Si le panel contient de nombreuses villes, comme c'est souvent le cas, on peut ne pas disposer de suffisamment de degrés de liberté pour estimer l'ensemble des  $\mu_{c',c}$ .

Pour pallier cette limite, il est possible d'imposer des restrictions sur certains de ces paramètres<sup>23</sup> et/ou de restreindre la dimension spatiale, en définissant des classes de villes selon leur appartenance à tel ou tel quantile de densité moyenne (calculée sur l'ensemble des années disponibles dans le panel), comme le font De la Roca et Puga (2012)<sup>24</sup>. On estime alors un système d'équation du type :

$$\log \omega_{i,t} = u_i + X_{i,t} \theta + \beta_{c(i,t)} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \sum_{q'} \mu_{q',q(i,t)} \left( e_{i,q',t} - \frac{e_{i,t}}{C} \right) + \epsilon_{i,t}$$

$$\beta_c = Z_{c,\cdot} \gamma + \eta_c,$$

$$\mu_{q',q} = Z_{q',\cdot} \psi + Z_{q',\cdot} (Z_{q,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{q',q}, \quad (14)$$

où  $Z_{q,\cdot}$  est le vecteur des caractéristiques locales moyennes dans la classe  $q$  définie par les quantiles lui servant de bornes,  $Z_{\cdot,\cdot}$  est le vecteur des moyennes de ces moyennes, et  $\zeta_{q',q}$  est un bruit blanc. Le paramètre  $\mu_{q',q}$  correspond à la valorisation dans une ville de la classe  $q$  de l'expérience acquise dans une ville de la classe  $q'$ . Ce gain peut être décomposé en deux éléments. Le paramètre  $\psi$  mesure la valorisation de l'expérience passée dans les villes de la classe  $q'$  pour un salarié travaillant dans la ville « moyenne », et  $\nu$  la valorisation additionnelle tirée du fait que la classe  $q$  correspondant à la ville dans laquelle le salarié travaille actuellement inclut des villes de densité supérieure à cette moyenne.

<sup>23</sup> Par exemple, on peut faire l'hypothèse que l'expérience passée dans les autres villes que celle où travaille actuellement le salarié n'est pas transférable, auquel cas :  $\mu_{c',c} = 0 \forall c' \neq c$ .

<sup>24</sup> De la Roca et Puga (2014) considèrent seulement trois classes d'aires urbaines en Espagne : la première contient les deux aires les plus peuplées (Madrid et Barcelone), la seconde, les trois aires suivantes (Valence, Séville et Saragosse), et la dernière toutes les autres aires.

## 2. Présentation des données et statistiques descriptives

L'estimation des économies d'agglomération nécessite des données de panel ayant une dimension individu-temps. Comme nous l'avons vu dans la première partie de ce rapport, cette double dimension est indispensable pour distinguer ce qui, dans la prime salariale dont bénéficient les travailleurs des grandes métropoles, revient à leurs compétences propres, aux économies d'agglomération statiques ou aux économies d'agglomération dynamiques. La seule source de données sur les salaires ayant cette double dimension en France est le panel des Déclarations Annuelles des Données Sociales (DADS) diffusé par l'INSEE.

### 2.1 Données individuelles issues du panel DADS

La déclaration annuelle de données sociales (DADS) est une formalité déclarative que doit accomplir toute entreprise employant des salariés, en application de l'article R243-14 du code de la Sécurité sociale (Décret du 24 mars 1972) et des articles 87.240 et 241 de la loi 51-711 du 7 juin 1951 du code général des Impôts. D'après le code général des Impôts, est astreinte à cette déclaration toute personne physique ou personne morale domiciliée ou établie en France qui verse des traitements ou des salaires, et ce, même si elle est exonérée de la taxe sur les salaires en totalité ou en partie.

Dans ce document commun aux administrations sociales et fiscales, les employeurs (y compris les entreprises nationales, les administrations publiques et les collectivités locales) sont tenus, annuellement et pour chaque établissement, de communiquer aux organismes de Sécurité Sociale d'une part, à l'administration fiscale d'autre part, la masse totale des traitements versés, les effectifs employés au 31 décembre, et une liste nominative de salariés indiquant, pour chacun d'entre eux, le montant annuel des rémunérations perçues (avantages et indemnités pour frais professionnel inclus), ainsi que des mentions particulières relatives à l'identité du salarié (nom, genre, adresse), au type de poste occupé (CDI, CDD, autre...), aux conditions de l'emploi (temps complet ou partiel, interim), au type d'emploi (apprenti, stagiaire, emploi aidé ou ordinaire), à la Catégorie Socio-professionnelle (chefs d'entreprise, cadres et professions intellectuelle supérieures, professions intermédiaires, employés, ouvriers)<sup>25</sup>, et la durée du travail (périodes d'emploi, nombre d'heures rémunérées en équivalent temps plein).

L'exploitation statistique des DADS se distingue d'une enquête en ce sens qu'elle ne résulte pas d'un questionnaire mais d'un formulaire administratif multipartenaires. L'INSEE est destinataire officiel de la DADS, qui lui est transmise dans le but d'élaborer des statistiques sur l'emploi et les salaires, et ce depuis 1950. Bien que toutes les DADS soient chargées dans sa base de production, qui inclut un volume de données considérable (2 millions d'employeurs, plus de 55 millions de lignes relatives aux salariés), l'INSEE exclut du champ de la diffusion la fonction publique d'Etat, les services domestiques (car les particuliers-employeurs peuvent être dispensés de taxes sur les salaires versées à leur personnel domestique), les activités extraterritoriales, et les établissements implantés à l'étranger employant des salariés qui relèvent de la sécurité sociale française (et donc présents dans les DADS), mais exerçant leur activité hors de la France. Le champ de diffusion final des DADS couvre donc l'essentiel des secteurs privé et semi-public depuis 1950.

À partir de ces DADS, l'INSEE a constitué un panel permettant de suivre, depuis 1976, tous les salariés nés en octobre des années paires (soit environ 1/25<sup>ème</sup> des effectifs salariés du secteur privé et semi-public) et, depuis 2002, tous les salariés nés en octobre quelle que soit l'année d'observation (soit désormais 1/12<sup>ème</sup> des salariés du privé et du semi-public).

---

<sup>25</sup> Le niveau d'éducation ou les diplômes ne sont en revanche, et malheureusement, pas connus.

Ce panel est constitué à partir de fichiers DADS annuels de niveau « poste » (i.e. individu x siret), qui sont ensuite agrégés au niveau de l'entreprise de manière à obtenir des données individu x siren x an. Si un même individu a travaillé dans au moins deux établissements différents au cours d'une année donnée, les postes sont agrégés et ne sont conservés que les caractéristiques de l'établissement pour lequel la durée de paie est la plus longue et, en cas d'ex-æquo, le salaire le plus élevé. Les salaires, les heures et les durées de paie (en jours) sont additionnés<sup>26</sup>.

Bien que le panel DADS existe depuis 1976, nous avons été contraints de limiter la période d'étude. En effet, deux changements drastiques dans la chaîne de production des DADS ont eu lieu en 1993 (passage de la NAP73 à la NAF) et en 2009 (passage de la NAF rev1 à la NAF rev2), et ils ont engendré d'importantes ruptures de séries dans la mesure des salaires individuels à l'échelle des différents secteurs. Nous avons donc restreint l'analyse des économies d'agglomération aux années comprises entre ces deux dates. Par souci de cohérence temporelle, nous nous sommes concentrés sur les secteurs appartenant au champ de diffusion des DADS sur l'ensemble de cette période. L'agriculture, la sylviculture et la pêche, qui ne sont rentrées dans le champ d'exploitation des DADS qu'en 2002, n'ont donc pas été étudiés. La frontière étant parfois floue entre fonction publique hospitalière et territoriale, nous n'avons pas non plus considéré l'administration publique (Poste et Télécommunications incluses)<sup>27</sup>, les collectivités territoriales et les activités associatives. Nous avons également exclu certaines industries, comme l'extraction de minerais, dont la localisation est fortement conditionnée par l'accès aux matières premières<sup>28</sup>, ou comme le transport spatial, trop spécifique. Enfin, bien qu'ils représentent une part importante de l'emploi des grandes métropoles, nous avons été contraints d'exclure la finance et l'assurance, car les DADS comptabilisent les effectifs de ces secteurs au lieu d'implantation du siège social, et non au lieu d'emploi des salariés. En définitive, le nombre de secteurs étudiés s'élève tout de même à 99 sur les 114 définis par le niveau 2 de la Nomenclature Economique de Synthèse européenne. La liste de ces secteurs est présentée dans l'Annexe 2 de ce rapport.

Après exclusion de ces secteurs, le nombre d'observations était encore trop important pour pouvoir construire la base de données et effectuer des traitements statistiques dans un laps de temps raisonnable<sup>29</sup>. Nous avons donc restreint l'échantillon aux salariés à temps complet, et réduit la dimension temporelle en définissant des fenêtres de trois ans, de manière à ne conserver que les années 1993, 1996, 1999, 2003, 2005 et 2008.

L'échantillon restant, composé de 3 133 573 observations et couvrant chaque année environ 500 000 salariés des secteurs privé et semi-public, constitue néanmoins la plus grande source de variabilité individuelle, temporelle, géographique et sectorielle utilisée à ce jour pour estimer l'impact des économies d'agglomération sur les salaires<sup>30</sup>.

---

<sup>26</sup> Ne sont conservées que la plus petite date de début de rémunération et la plus grande date de fin de rémunération. La durée de paie est ensuite mise en cohérence avec ces dates : si elle est supérieure à l'intervalle, elle est ajustée (cela peut se produire, par exemple, dans le cas d'une personne ayant deux postes à temps partiel).

<sup>27</sup> Il faut noter le cas particulier de France Telecom et de La Poste, issus de la privatisation de la direction des Postes et Télécommunications qui n'entraient pas au départ dans le champ des DADS. Les salariés de ces entreprises faisaient auparavant partie de la fonction publique d'État.

<sup>28</sup> Les économies d'agglomération dont ils bénéficient ne sont donc pas liées à la taille des villes.

<sup>29</sup> Soit environ 26h-machine via le serveur du CASD, dont la capacité et la rapidité de traitement sont déjà exceptionnelles.

<sup>30</sup> Le panel utilisé par Combes, Duranton et Gobillon (2008) ne couvrait que des fenêtres de quatre ans sur la période 1976-1996, soit un total de 2 664 474 observations. De la Roca et Puga (2014) disposent de 5 821 846 observations sur la période 2004-2009 mais, du fait de la périodicité mensuelle des données utilisées, la mobilité spatiale inter-temporelle des salariés est bien plus faible dans leur panel. Or, comme nous l'avons vu dans la première partie de ce rapport, c'est cette mobilité qui permet d'estimer précisément les économies d'agglomération.

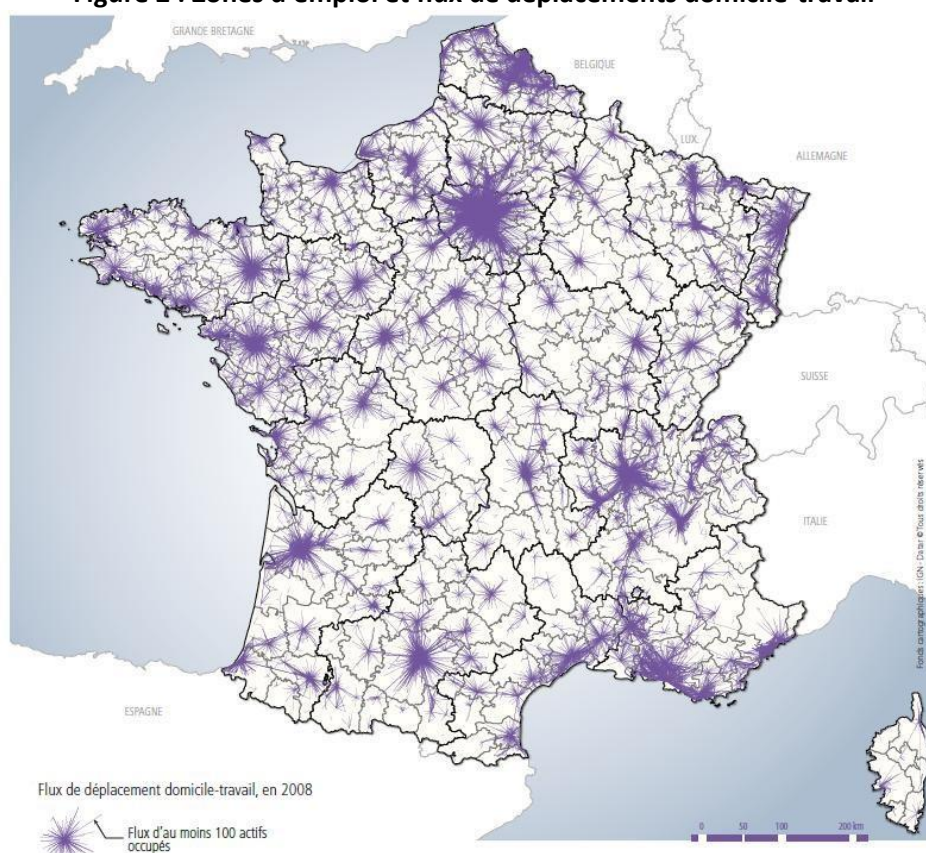
## 2.2 Des unités spatiales pertinentes : les zones d'emploi 2010

Le salaire constitue notre principale variable d'intérêt car, comme nous l'avons vu dans la première partie de ce rapport, les économies d'agglomération sont capitalisées, à plus ou moins long terme, dans cette variable. Il est donc essentiel de choisir des unités spatiales pertinentes pour étudier cette variable, mais aussi pour mesurer les économies d'agglomération, dont on a vu qu'elles pouvaient très largement dépasser la frontière des unités urbaines.

Les zones d'emploi constituent une partition géographique du territoire particulièrement adaptée aux études locales sur le marché du travail, car elles sont définies comme « des espaces à l'intérieur desquels la plupart des actifs résident et travaillent, et dans lequel les établissements peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts » (Mas, 2012). Leur périmètre a été réactualisé en 2010, car les contours initialement définis en 1983-1984 et partiellement révisés en 1993-1994, étaient devenus partiellement obsolètes, en raison de l'évolution des facteurs déterminant les migrations domicile-travail. Les 322 nouvelles zones d'emploi sont entrées en vigueur le 1<sup>er</sup> juillet 2011 : 304 se situent en métropole et 297 sur le continent.

Même si notre période d'étude est antérieure à cette date, nous avons choisi d'utiliser ce découpage pour plusieurs raisons. Tout d'abord, les nouvelles zones d'emploi sont globalement plus homogènes que les précédentes, en raison de leur conception à partir du critère exclusif des déplacements domicile-travail des actifs observés lors du recensement. La figure 2 montre que ces flux sont plus ou moins polarisés selon l'organisation urbaine et le degré de concentration de l'emploi, le réseau routier et les caractéristiques géographiques du territoire français.

**Figure 2 : Zones d'emploi et flux de déplacements domicile-travail**



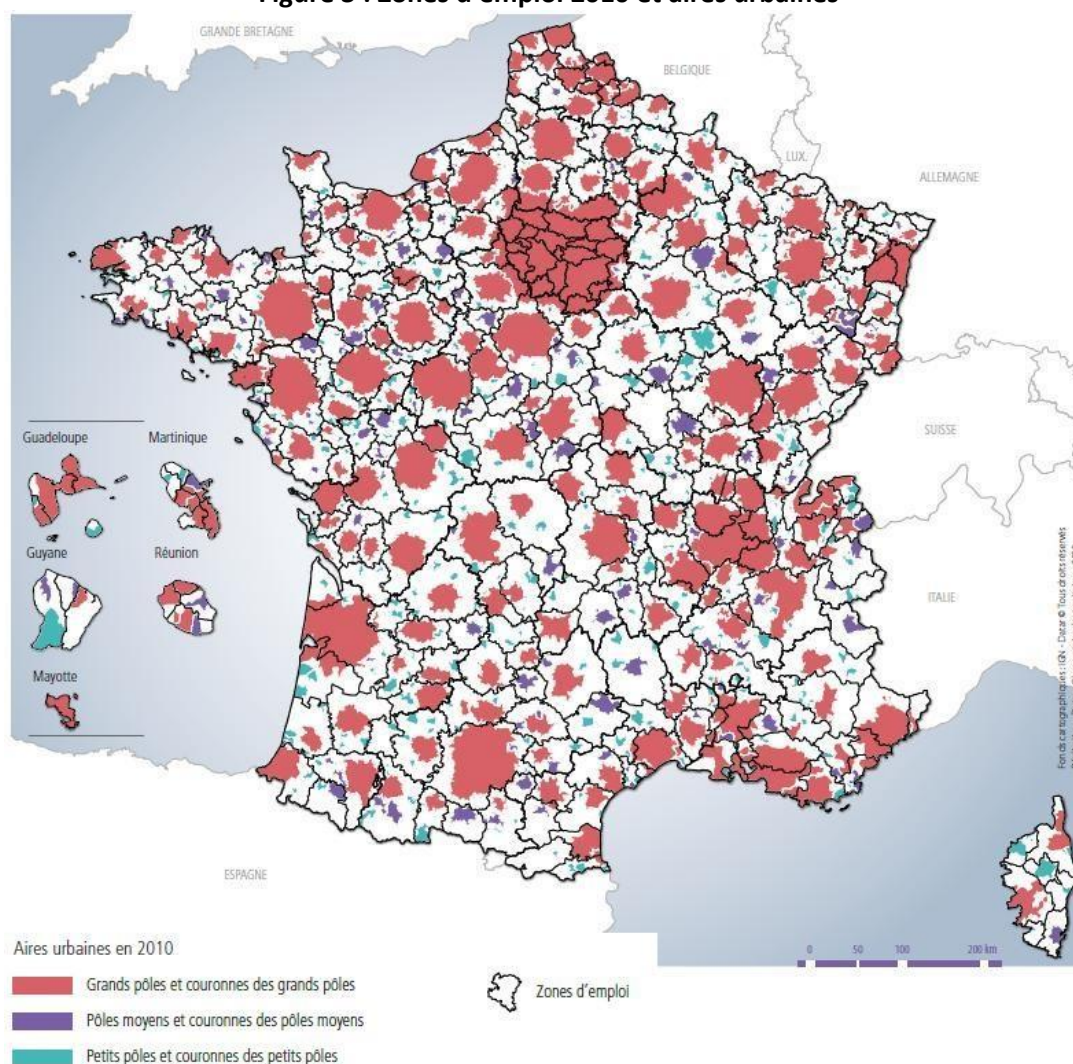
Source : Mas (2012), *Atlas des zones d'emploi 2010*.



L'effet « capitale régionale » est sensible dans toutes les régions, mais à des degrés divers. En Île-de-France, le rayon d'attraction de Paris dépasse largement les limites de la région. En Aquitaine, Midi-Pyrénées ou Bretagne, les flux sont également très polarisés respectivement par Bordeaux, Toulouse et Rennes. À l'inverse les régions Rhône-Alpes et Pays-de-Loire ont un maillage de pôles plus fin, sans toutefois aboutir à des configurations où les flux entre les pôles sont particulièrement intenses, comme en Paca, Nord-Pas-de-Calais ou Alsace. Les zones peu denses, comme le Limousin, l'Auvergne, Champagne-Ardenne, Bourgogne ou la partie alpine de la région PACA ont des flux peu intenses. Les caractéristiques de ces flux expliquent en grande partie les particularités du découpages infrarégional, très morcelé dans les régions PACA, Nord-Pas-de-Calais, Alsace, Rhône-Alpes, Pays-de-Loire, avec des zones d'emploi de taille très variable en Bretagne ou Midi-Pyrénées, et des zones d'emploi très étendues en Limousin ou Champagne-Ardenne.

Ensuite, parmi les outils ayant servi à expertiser le nouveau découpage, le zonage en aires urbaines a été utilisé pour repérer les pôles d'emploi et les limites de leur aire d'influence. Or, comme nous l'avons vu dans l'introduction de ce rapport, les aires urbaines sont aussi construites à partir des flux de déplacements domicile-travail. De façon logique compte tenu de leur construction, les zones d'emploi 2010 sont donc toutes plus ou moins centrées sur une aire urbaine. La superposition des deux découpages, représentée sur la Figure 3, en illustre la grande cohérence.

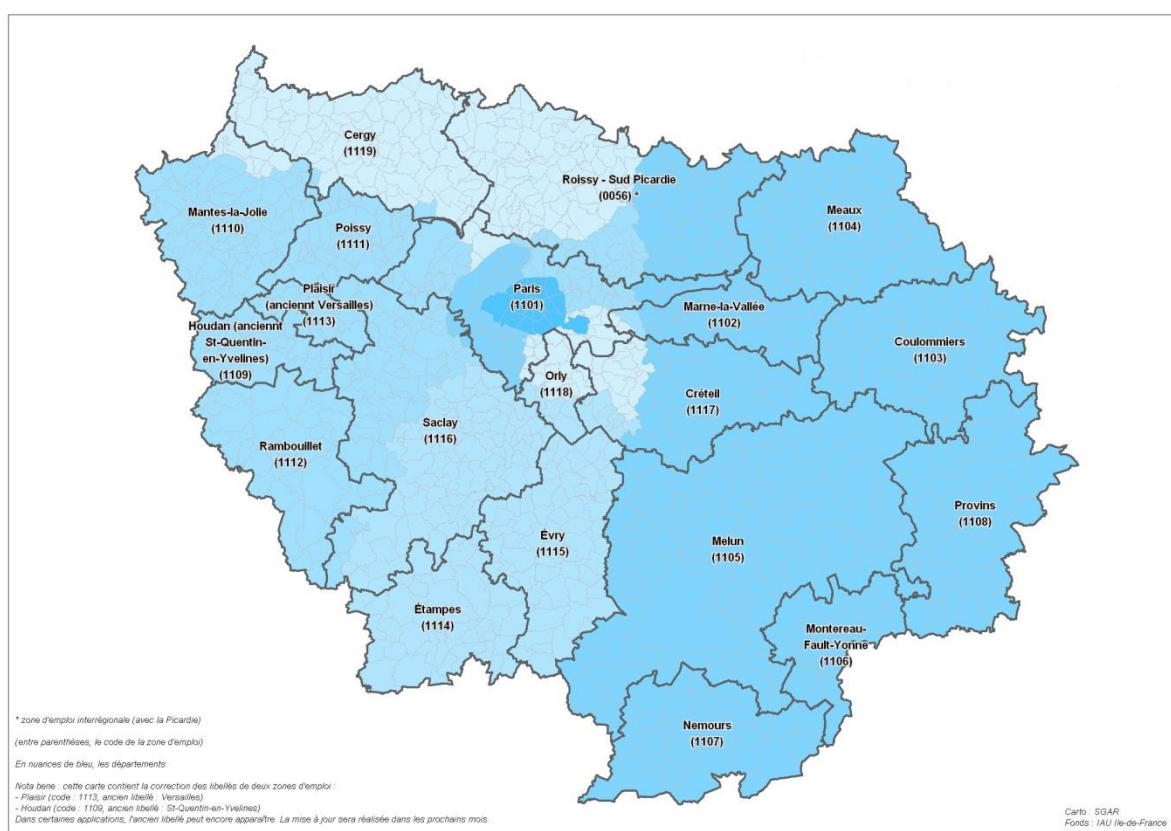
**Figure 3 : Zones d'emploi 2010 et aires urbaines**



Les grandes aires urbaines occupent une grande partie de la surface des zones d'emploi qui les contiennent, voire en dépassent les limites, comme celle de Paris, ou celle de Bordeaux. Dans les territoires moins denses, les zones d'emploi regroupent les aires d'influence de plusieurs pôles, comme celles de Chaumont-Langres et Vitry-le-François/Saint-Dizier en Champagne-Ardenne.

Les zones d'emploi 2010 reflètent donc la réalité actuelle des marchés du travail locaux, mais aussi leurs évolutions récentes : allongement des déplacements domicile-travail, diminution de l'activité des territoires les moins dynamiques et, point déterminant pour notre étude, rayonnement croissant des grandes villes. En contrepartie, elles s'affranchissent dans une plus large mesure qu'avant des limites administratives départementales et régionales, comme l'illustre la Figure 4 pour les 20 zones d'emploi couvrant le périmètre du Grand Paris. Pour autant, le nombre de zones d'emploi interrégionales reste limité. On n'en compte que 11, dont une dans le Grand Paris : la zone de Roissy-Sud-Picardie, qui est à cheval sur l'Île-de-France et la Picardie.

**Figure 4 : Les 20 zones d'emploi du Grand Paris**



### 2.3 Statistiques descriptives sur les salaires et la taille des zones d'emploi

Compte tenu du caractère individuel des informations recensées dans le panel DADS, son utilisation est soumise aux règles du secret statistique, et donc à l'obtention d'une autorisation préalable du Comité du Secret. Toute personne habilitée par ce comité est tenue d'effectuer une déclaration à la CNIL, astreinte au secret professionnel et soumise aux dispositions répressives prévues à l'article L. 226-13 du code pénal. Ces règles de confidentialité imposent, pour traiter les données, de passer via un Centre d'Accès Sécurisé aux Données (CASD), i.e. un équipement spécifiquement conçu par le Genes (Groupe des écoles nationales d'économie et statistique) pour permettre aux chercheurs de travailler à distance sur les données, dans des conditions de sécurité élevées.



Les faits stylisés présentés dans cette section et la suivante ont été produits de manière à respecter les règles du secret statistique, qui imposent la présence d'au minimum cinq individus par cellule. Ils ne sont donc pas aussi détaillés que nous l'aurions souhaité, car nous avons buté sur cette contrainte en raison du caractère très désagrégé de notre analyse (panel au 1/25<sup>ème</sup>, 297 zones, 99 secteurs, 6 Catégories Socio-Professionnelles pour chaque secteur).

Le panel DADS recense deux principales mesures de la rémunération des salariés, notre principale variable d'intérêt. La première mesure est un salaire brut correspondant à l'intégralité des sommes perçues au cours de l'année par le salarié au titre de son contrat de travail, avant toute déduction de cotisations obligatoires. Il est donc calculé à partir de la base CSG, dont l'assiette intègre l'intéressement, la participation et les indemnités de licenciement. La seconde est un salaire net calculé à partir du brut amputé du montant des cotisations sociales obligatoires (sécurité sociale, régimes de retraite et prévoyance complémentaires, assurance chômage, CSG et CRDS).

C'est ce salaire net fiscal que nous avons privilégié pour évaluer l'impact des économies d'agglomération. Dans le cas où un même salarié disposerait de plusieurs rémunérations versées par des employeurs différents, nous avons choisi le salaire net fiscal correspondant à la rémunération annuelle la plus forte, et localisé l'emploi principal dans l'entreprise ayant versé cette rémunération. Afin d'obtenir une mesure du salaire nominal de chaque individu, nous avons divisé son salaire net fiscal annuel par le nombre de jours calendaires. La valeur obtenue est exprimée en euros constants sur toute la période d'observation.

La première carte en haut à gauche de la Figure 5 représente la moyenne des salaires nominaux individuels par zone d'emploi en 2008, exprimée en % de la moyenne des zones (58,60€/jour en 2008, soit le salaire net fiscal moyen à Arras). Les autres cartes représentent trois mesures différentes de la taille des zones en 2008 (emploi, densité d'emploi et potentiel marchand), exprimées en % de la moyenne des zones, et dont il est possible de comparer la variabilité spatiale avec celle du salaire<sup>31</sup>. Sans surprise, les rémunérations les plus élevées concernent les agglomérations denses du bassin parisien, mais aussi des régions PACA, Rhône-Alpes, Alsace et Nord-Pas-de-Calais. Parmi les trente zones d'emploi dont le net fiscal moyen dépasse le plancher du dernier décile, à savoir 63,50€ en 2008, seize sont situées dans le Grand Paris, cinq en Provence (Marseille-Aubagne, Aix-en-Provence, Cannes–Antibes, Bagnols-sur-Sèze et Istres-Martigues), et quatre en région Rhône-Alpes (Maurienne, Vallée de l'Arve, Lyon, Grenoble).

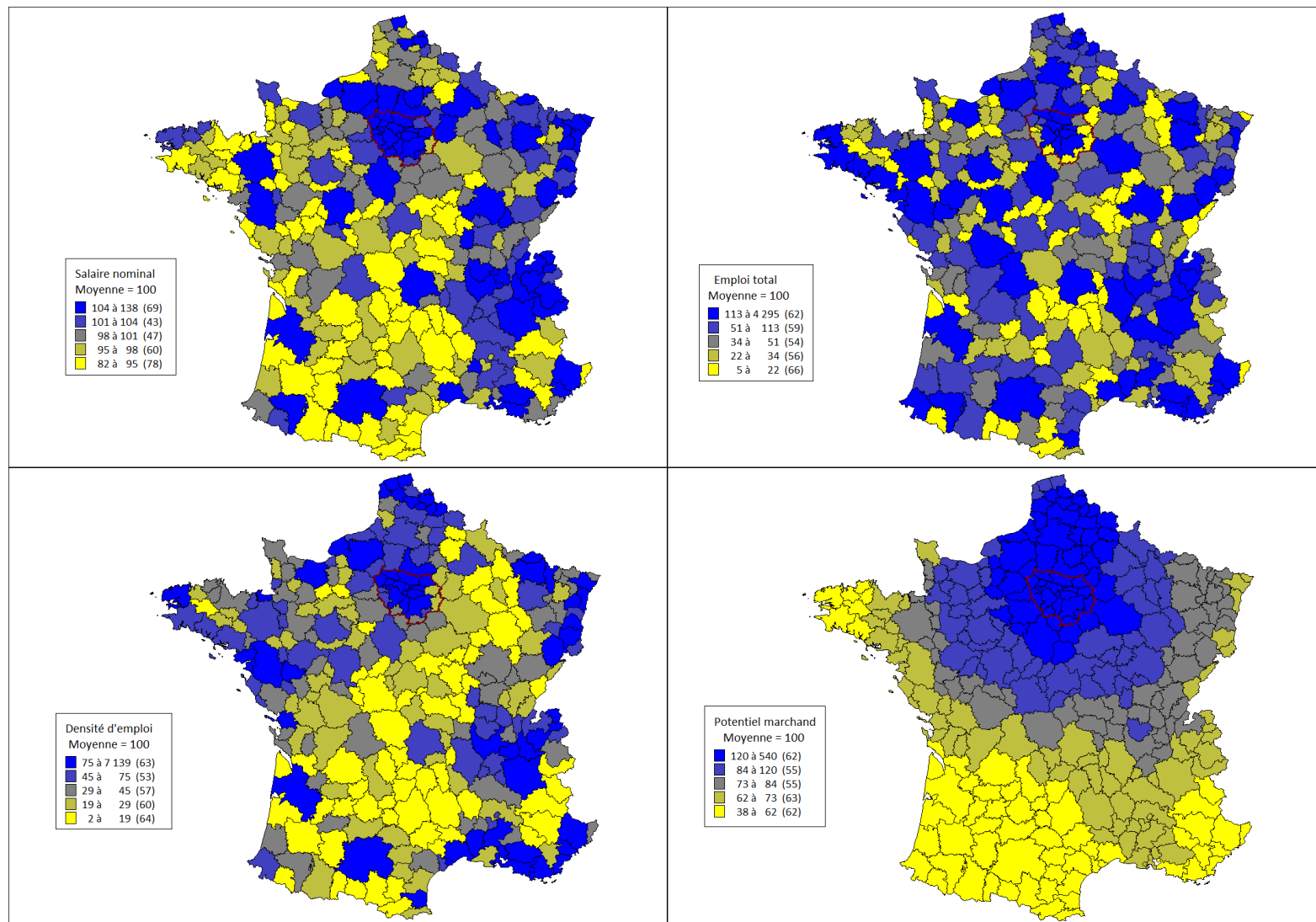
À l'inverse, les zones d'emploi dont les salaires sont les plus faibles se trouvent majoritairement dans les espaces peu denses, comme par exemple le massif du Morvan (dont la zone d'emploi est à la fois la moins dense et la moins rémunératrice de toute la France, avec un salaire nominal moyen de 48,30€/jour), le centre de la Bretagne et les parties les plus rurales du Massif Central et des Pyrénées. Ces espaces sont en fait doublement pénalisés : ils ont une faible densité d'emplois, mais aussi peu d'opportunités de compenser ce manque par le recours aux marchés externes, comme l'illustre la faiblesse de leur potentiel marchand. Les zones d'emploi du Grand Paris bénéficient en revanche d'un double dividende, puisqu'elles sont denses et disposent des potentiels marchands les plus élevés de toute la France.

Sur les 20 zones d'emploi ayant le plus fort potentiel marchand, 19 sont situées dans le Grand Paris, et l'« outsider », Beauvais, en est limitrophe. Le caractère mono-centrique du potentiel marchand est frappant : les zones d'emploi du Grand Paris, très accessibles et très denses, en constituent le cœur, qui rayonne jusqu'à la frontière avec le Benelux au nord-est et englobe Lyon au sud-est. Le gradient diminue ensuite, à mesure qu'on s'éloigne du centre vers les zones périphériques.

---

<sup>31</sup> La géographie des variables de 2008 est très représentative de celle de la période 1993-2008, comme le montre la figure de l'Annexe 3, qui présente les moyennes calculées sur les six années disponibles dans le panel.

Figure 5 : Salaire net fiscal et taille des zones d'emploi en 2008



Source : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

Concernant les écarts de salaire, la situation est assez contrastée entre, d'une part, les zones du Grand Paris et les grandes métropoles régionales (l'écart de salaire moyen est par exemple de 25% entre Paris ou Saclay et Marseille-Aubagne ou Lille, et de 19% entre Paris ou Saclay et Lyon) et, d'autre part, un nombre élevé de zones d'emploi autour de la moyenne, si bien que le ratio inter-quintile P80/P20 est finalement relativement faible, à 1,11 : l'écart entre les zones situées aux bornes de ces quintiles (à savoir Tergnier et Nevers) est donc de 11%. L'écart entre les zones d'emploi offrant les salaires les plus extrêmes de la distribution (Saclay et Morvan) est en revanche de 67%.

Les écarts de la densité d'emploi sont en revanche beaucoup plus marqués : 2660 entre les deux extrêmes de la distribution (Paris et Morvan), et environ 4 entre les zones délimitant les premier et dernier quintiles de densité (Périgueux et Belfort-Montbéliard-Héricourt). Les inégalités spatiales de potentiel marchand sont fortes, mais pas aussi prononcées que pour les densités : la zone d'emploi la plus accessible (Orly) l'est 14 fois plus que la plus enclavée (Bayonne), et le ratio inter-quintile des potentiels marchands n'est que de 2, soit l'écart existant entre les zones d'emploi de Maubeuge et Agde-Pézenas.

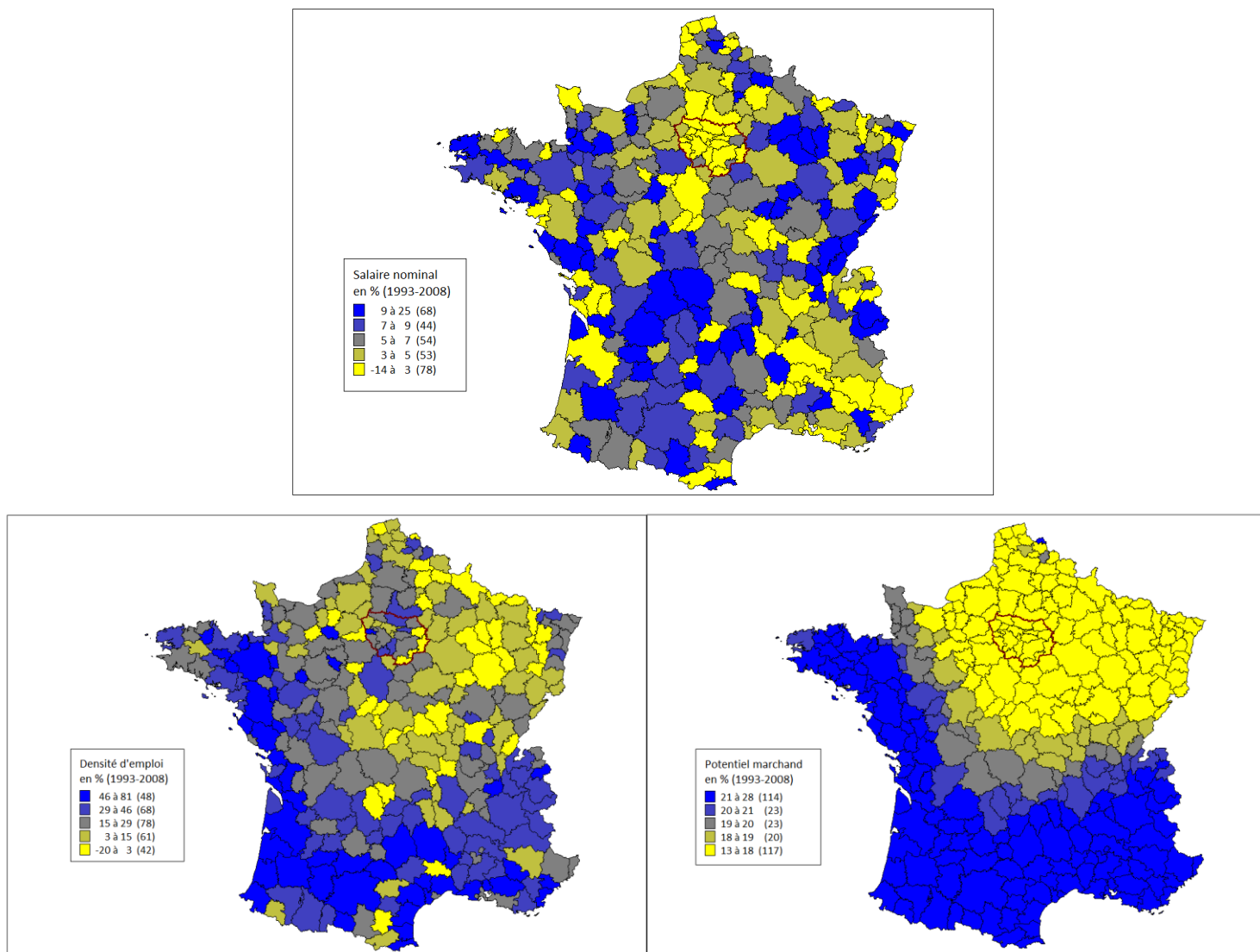
Comme l'illustre la Figure 6, le taux de croissance du salaire nominal sur la période 1993-2008 varie à l'exact opposé des écarts de salaire à la moyenne en 2008. Un processus de convergence est donc clairement à l'œuvre. Bien que les rémunérations aient augmenté en moyenne de 5,9% sur la période, les salaires ont faiblement chuté dans 27 zones d'emploi, dont 10 appartiennent au Grand Paris : la chute est de -6% à Versailles, -3,4% à Évry, -3% à Cergy, -2,5% à Orly ou à Paris, -2% à Melun et -1% à Saclay ou à Créteil. A contrario, les zones d'emploi dont le salaire a le plus augmenté sont situées dans les espaces à faible densité, notamment le centre de la Bretagne, le Massif Central, les Pyrénées et la Savoie.

Ces zones ont en effet connu un rattrapage spectaculaire de leur potentiel marchand. Le processus de densification des emplois a été en effet particulièrement marqué dans le sud-ouest et sur le littoral atlantique. Les villes de Narbonne, Aix-en-Provence et Toulouse ont par exemple vu leur densité d'emploi augmenter respectivement de 75%, 68% et 64% sur la période. Le dynamisme est moindre sur la côte atlantique, mais néanmoins notable. La densité d'emploi a par exemple augmenté de 53% à Nantes, et de 50% environ à Bordeaux et Bayonne. Les zones dont l'accessibilité a le plus augmenté sur la période sont cependant presque toutes situées sur le pourtour méditerranéen : + 27% à Sète, +25% à Marseille-Aubagne ou à Nîmes, +24% à Aix-en-Provence, Orange ou Béziers.

Les zones d'emploi du sud et de l'ouest de la France ont donc bénéficié d'une double dynamique. La densification des emplois a augmenté la taille des marchés locaux, mais elle a aussi accru leur potentiel marchand, par rétroaction. En revanche, les zones d'emploi du nord et de l'est de la France ont perdu sur les deux fronts.

On peut enfin souligner quelques évolutions spécifiques aux zones d'emploi de la Grande couronne parisienne et aux zones limitrophes du Grand Paris, qui semblent avoir bénéficié d'un processus de desserrement des emplois au détriment du cœur de la métropole sur la période 1993-2008. Cette décentralisation a plus particulièrement bénéficié à la zone de Saint-Quentin-en-Yvelines, dont la densité d'emploi a augmenté de 56%, mais aussi dans une moindre mesure à Évry ou Étampes (+34%), et Roissy– Sud-Picardie (+32%).

Figure 6 : Taux de croissance du salaire net fiscal, de la densité d'emploi et du potentiel marchand (1993-2008)



Source : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

## 2.4 Statistiques descriptives sur la qualification de la main d'œuvre

Comme nous l'avons vu dans la première partie de ce rapport, il ne faut pas confondre les gains de productivité liés à la taille des villes, ceux liés aux externalités de capital humain qui s'y produisent et le simple fait que les villes attirent une plus grande proportion de qualifiés, ce qui y accroît mécaniquement les salaires. Les grandes métropoles accueillent davantage de travailleurs qualifiés, qui sont attirés par leurs aménités culturelles ou récréatives, et ce tri spatial des salariés, qui sont plus compétents que la moyenne, a des effets positifs sur la rémunération des salariés.

### *La géographie des différentes Catégories Socio-Professionnelles*

Afin d'illustrer ce tri, la figure 7 représente la part des différentes Catégories Socio-Professionnelles<sup>32</sup> dans l'emploi local, toutes années du panel confondues. En effet, la désagrégation de l'emploi par CSP peut faire apparaître des cas de figure dans lesquels la somme des emplois des différentes catégories est inférieure à 5 salariés. Nous ne sommes donc pas en mesure de produire, comme pour les autres variables, des statistiques annuelles.

Les statistiques agrégées sur plusieurs années parlent cependant déjà beaucoup. Comme on peut le voir sur la figure 7, la répartition spatiale des différentes CSP est loin d'être uniforme. La part des cadres et des professions intellectuelles supérieures dans l'emploi local est bien plus importante dans les zones denses, en particulier celles du bassin parisien. Les cadres représentent ainsi 25% de la force de travail des zones d'emploi de Saclay ou Paris, 18% de celle de Marne-la-Vallée ou Versailles, et 16% de celle d'Évry ou de Cergy. Les grandes villes de Province ne sont pas en reste : l'emploi de Toulouse ou Aix-en-Provence est composé à 16% de cadres, et ce chiffre s'élève à 13% environ à Lyon ou Lille, et 10% environ à Marseille, Nantes, Strasbourg ou Rennes. En revanche, les cadres représentent moins de 2% de l'emploi des zones qui, comme Morvan, Loudéac ou Segré, accueillent bien davantage d'ouvriers. Ces derniers représentent respectivement 50%, 64% et 71% de la force de travail de ces zones, alors que la part des ouvriers n'est « que » de 20% à Paris (le pourcentage le plus faible de toutes les zones d'emplois françaises), et de 25% à Saclay ou à Marne-la-Vallée. Les employés constituent une part non négligeable (plus de 25%) de la main d'œuvre des zones du Grand Paris, mais aussi de celles situées en Aquitaine, sur le pourtour méditerranéen ou à la frontière de la Suisse et de l'Italie. Leur répartition est cependant plus homogène que celle des cadres et des ouvriers, et il en va de même pour les professions intermédiaires.

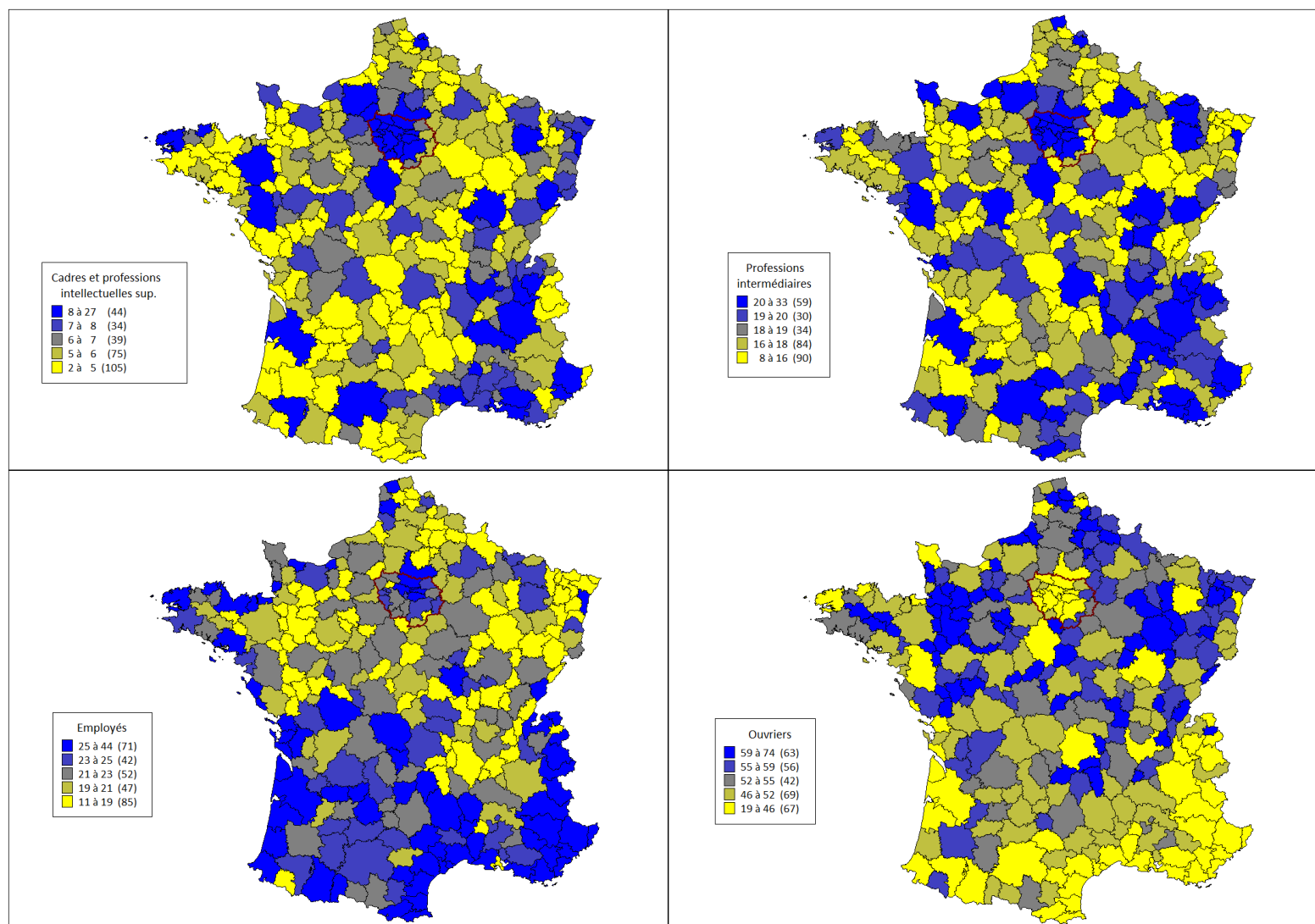
La part des différentes CSP dans l'emploi local, si elle fait apparaître quelques éléments saillants, ne permet cependant pas de connaître le degré de surreprésentation ou de sous-représentation des qualifications dans les différentes zones. Les grandes métropoles, par définition, ont plus d'emplois, et toutes les CSP y sont donc plus largement représentées. C'est d'ailleurs ce qui explique le poids presque aussi important des cadres et des ouvriers dans l'emploi de Paris. Il est donc préférable de construire un indice de spécificité, en rapportant la part des cadres présents dans chaque zone, à la part de la zone dans l'emploi national, soit :

$$\text{Indice}_{j,c} = \frac{csp_{j,c} / \sum_c csp_{j,c}}{L_c / \sum_c L_c} \times 100$$

Un indice supérieur à 100 signifie que la zone attire une proportion de la CSP  $j$  relativement plus importante que le poids de la zone dans l'économie nationale ou, autrement dit que la part de la CSP dans la zone est supérieure à la part de la CSP sur l'ensemble de la France : la CSP  $j$  y est donc surreprésentée. Un indice inférieur à 100 signifie au contraire qu'elle y est sous-représentée.

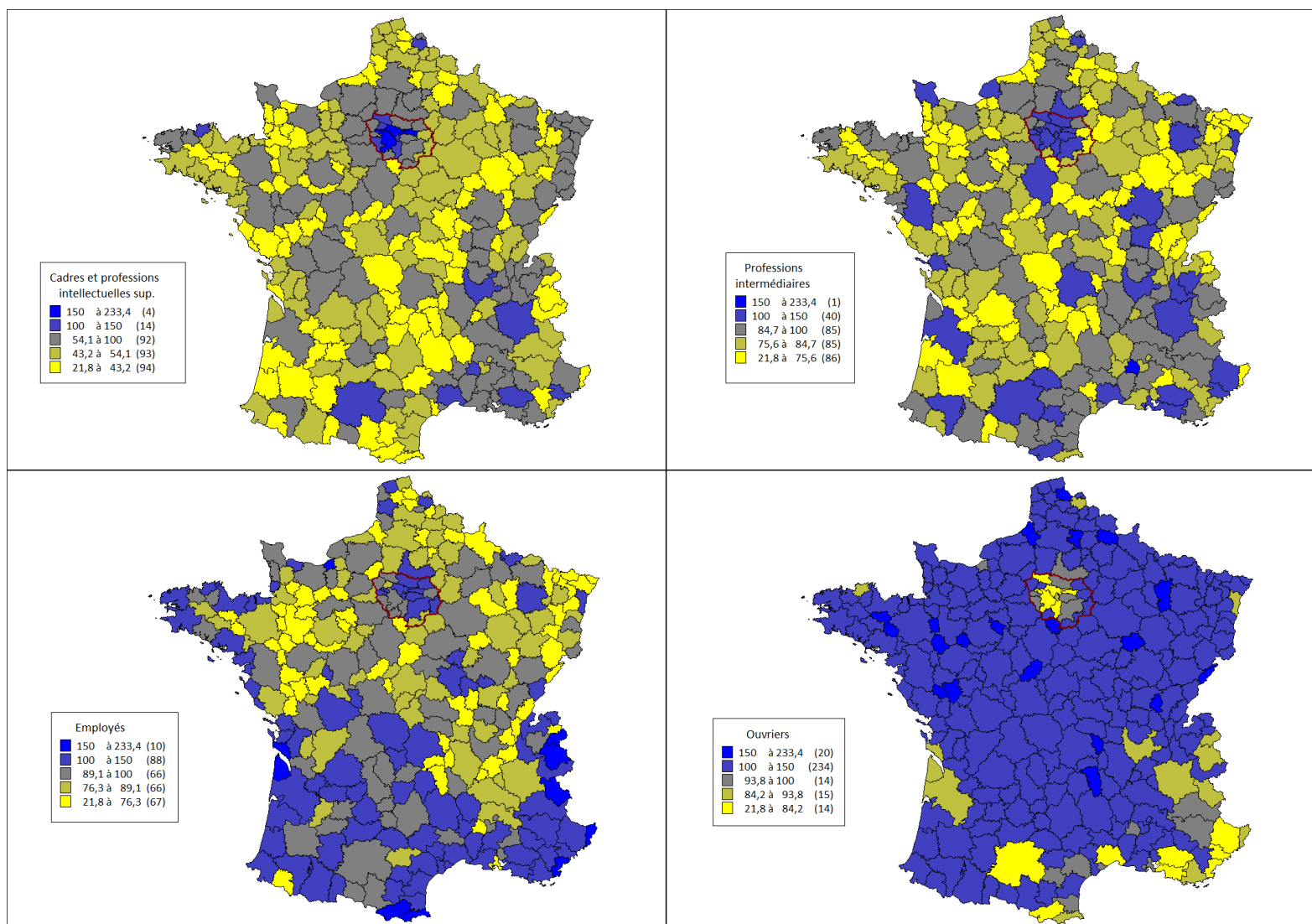
<sup>32</sup> La classification utilisée est présentée dans l'Annexe 4 de ce rapport.

**Figure 7 : Part des différentes Catégories Socio-Professionnelles dans l'emploi local (en %, toutes années confondues)**



Source : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

Figure 8 : Indice de spécificité des différentes Catégories Socio-Professionnelles (base 100, toutes années confondues)



Note: un indice de spécificité supérieur à 100 indique une surreprésentation de la CSP dans la zone. *Source* : INSEE, Panel DADS (secteurs privé et semi-public uniquement).



La figure 8 illustre la variation spatiale de cet indice de spécificité pour les quatre CSP étudiées plus haut. La surreprésentation des cadres dans les zones d'emploi du cœur de l'agglomération parisienne est désormais frappante : leur part y est plus de deux fois supérieure à celle de la zone dans l'emploi national (les indices de spécificité y sont supérieurs à 200). Seules quelques autres zones emploi très denses comme Lyon, Lille ou Toulouse accueillent une proportion relativement plus forte de cadres, mais leur indice de spécificité est plus faible (compris entre 100 et 150). Les cadres sont sous-représentés dans quasiment toutes les autres zones d'emploi.

Il est remarquable de constater que l'indice de spécificité des ouvriers varie à l'exact opposé de celui des cadres. Cette catégorie est très largement sous-représentée dans les zones les plus denses, en particulier dans l'agglomération parisienne, à Toulouse, Marseille, Lille et, dans une moindre mesure, à Lyon. Elle est en revanche surreprésentée dans les zones industrielles de Roubaix-Tourcoing, Pithiviers, Vitry, La Ferté-Bernard, Commercy ou Romorantin-Lanthenay. Les autres zones d'emploi attirent une proportion d'ouvriers peu ou prou similaire à leur part dans l'emploi national. Leur indice de spécificité oscille autour de 100, si bien que la répartition spatiale des ouvriers présente globalement une forte homogénéité spatiale.

La concentration géographique des professions intermédiaires et des employés est moindre, et se reflète davantage dans la part de ces catégories dans l'emploi local. Les employés sont fortement surreprésentés dans les zones d'emploi frontalières de Mont-Blanc, Tarentaise, Briançon, Menton - Vallée de la Roya, Fréjus – Saint Raphaël, Prades et Céret, ainsi que dans les zones côtières de Honfleur, et l'estuaire bordelais (Pauillac et Royan), vraisemblablement du fait de la concentration des activités touristiques. Ils le sont aussi, mais plus faiblement, dans tout le sud de la France, où le tourisme est également dynamique, mais les services à la personne aussi, du fait de la présence de nombreux retraités. Quant à la géographie des professions intermédiaires, elle suit très largement, mais avec un moindre degré de spécificité, la géographie des cadres et des professions intellectuelles supérieures, avec une très forte surreprésentation additionnelle dans la zone de Bagnols-sur-Cèze.

### *La géographie de l'expérience des salariés*

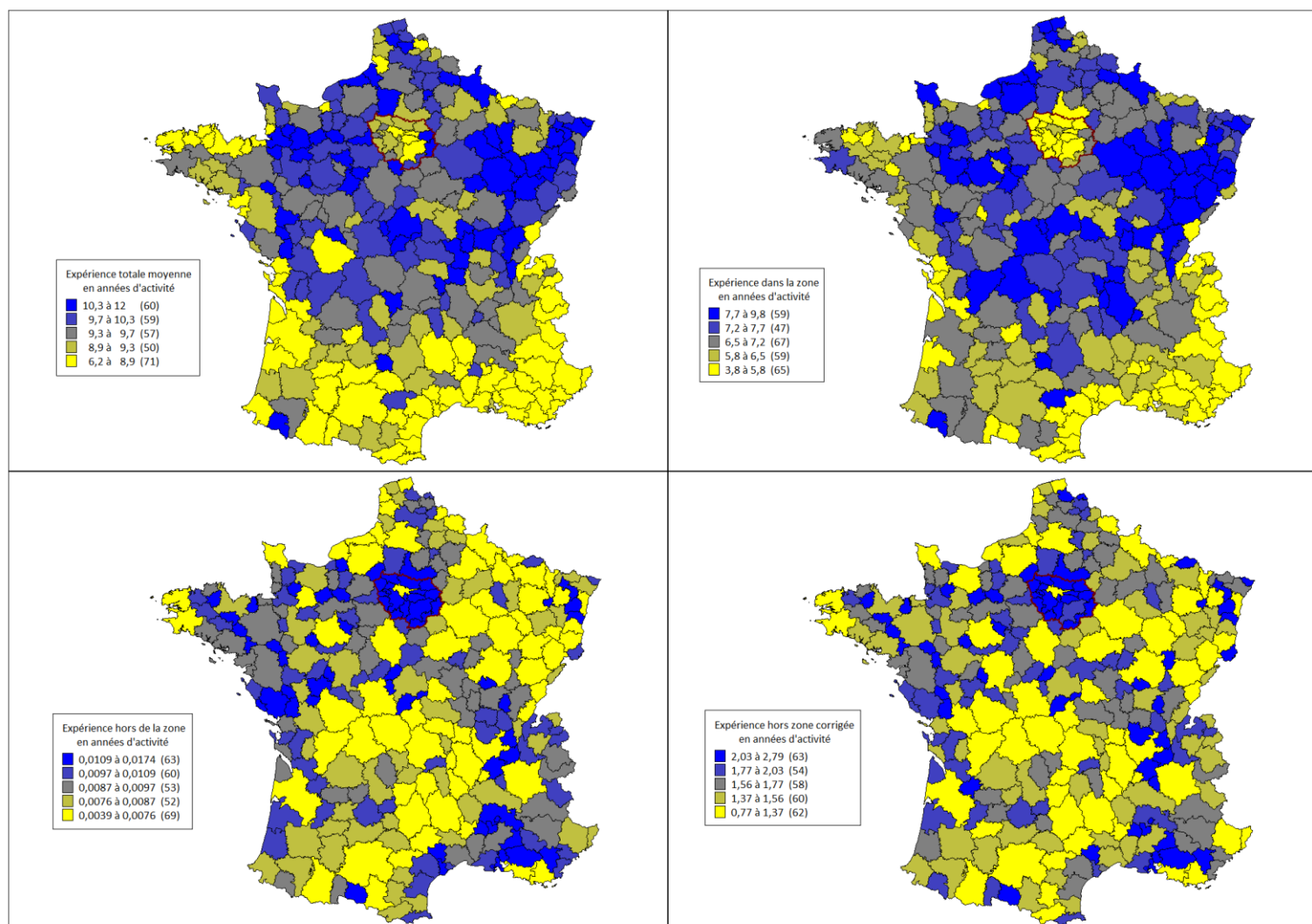
L'appartenance à une CSP particulière, si elle donne de l'information intéressante sur la qualification des travailleurs, n'est cependant qu'une mesure très imparfaite de leurs compétences réelles. D'autres variables complémentaires sont donc intéressantes à étudier. Par exemple, l'expérience des salariés peut être mesurée par le nombre d'années d'activité dans les secteurs privés et semi-public, ce qui revient dans notre cas à compter le nombre d'années passées dans le panel DADS. Le calcul peut être effectué sur l'ensemble de la carrière, indépendamment des zones dans lesquelles les salariés ont exercé une activité, ou bien en prenant en compte la mobilité géographique de chaque salarié, et l'expérience acquise dans les autres zones que celle où il travaille actuellement.

La figure 9 illustre la répartition géographique de quatre variables d'expérience en 2008<sup>33</sup>. La première (en haut à gauche) représente l'expérience totale moyenne des salariés employés dans la zone, mesurée par la moyenne du nombre d'années d'activité passées par chaque individu dans le panel. Pour que ce calcul soit pertinent, il faut cependant éliminer les individus les plus âgés du panel, car leur expérience ne peut être calculée qu'à partir de 1976, et l'erreur de mesure est d'autant plus grande pour ces salariés qu'ils ont en réalité plus d'expérience que celle qui peut être mesurée à partir du panel. L'échantillon a donc été tronqué pour éliminer les salariés ayant plus de 18 ans en 1976, i.e. tous les salariés nés avant 1958 (soit environ 30% des observations du panel).

---

<sup>33</sup> Ici encore, l'année 2008 est représentative de la tendance générale du panel, illustrée dans l'Annexe 5 de ce rapport.

Figure 9 : Expérience moyenne des salariés selon la zone en 2008 (panel tronqué)



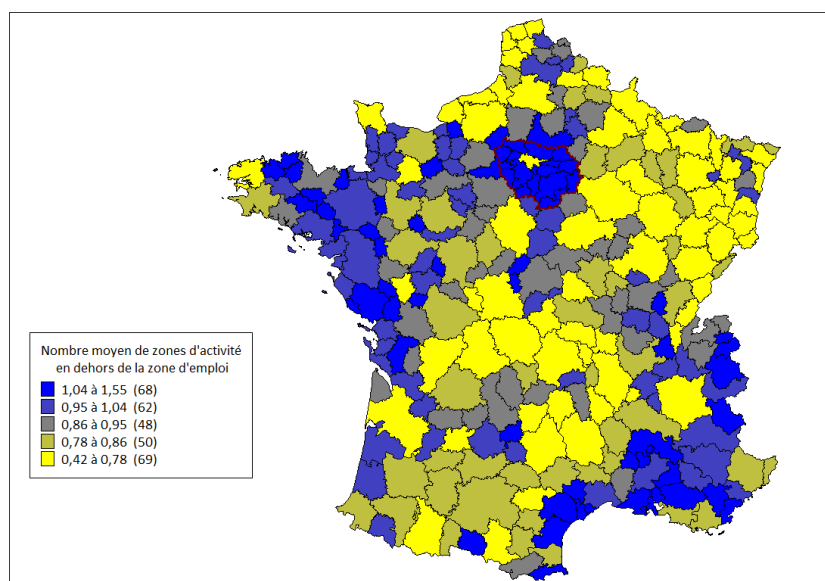
Note : L'expérience hors de la zone est la moyenne du nombre d'années d'activité passées dans toutes les autres zones. L'expérience hors zone corrigée exclut du calcul précédent les zones d'emploi où les salariés de la zone n'ont jamais exercé d'activité par le passé. *Source* : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

Les autres variables d'expérience calculées à partir du panel tronqué sont l'expérience moyenne passée dans la zone (en haut à droite de la figure 9), l'expérience moyenne passée dans toutes les autres zones, y compris celles où les salariés n'ont jamais exercé d'activité (en bas à gauche), et l'expérience moyenne passée dans toutes les autres zones par lesquelles les salariés sont effectivement passés (en bas à droite).

Les zones d'emploi du sud de la France et du Grand Paris se singularisent par une expérience totale moyenne de leurs salariés plus faible que celles de l'arrière-pays : 7,5 ans environ pour les salariés de Nice, Montpellier ou Paris, contre presque 12 pour les salariés de Autun ou Vendôme (la moyenne des zones s'établit à 9,5 en 2008). Les jeunes salariés sont en effet davantage présents dans les grandes métropoles, et notamment celles du bassin parisien : à titre d'exemple, les zones d'emploi de Marne-la-Vallée, Cergy et Évry ont 2 fois plus de jeunes de moins de 20 ans que de personnes de plus de 60 ans (Mas, 2012). La mobilité de ces jeunes est souvent forte, et les transitions vers les autres zones d'emploi très fréquentes, ce qui explique également la faible expérience locale moyenne des salariés de ces zones (à peine 4 ans pour les salariés de Marne-la-Vallée, Saint-Quentin-en-Yvelines, Versailles ou Créteil), et leur forte expérience en dehors de la zone (si on comptabilise uniquement les zones d'emploi effectives : 2,7 ans en moyenne à Orly, Versailles, Saint-Quentin-en-Yvelines et Cergy, 2,6 à Saclay, Provins, Rambouillet ou Poissy et 2,5 à Étampes ou Évry). Dans le sud, en particulier les zones du littoral méditerranéen, les taux de chômage élevés peuvent expliquer le manque d'expérience relatif des salariés. En effet, les individus qui perdent leur emploi et connaissent des épisodes de chômage sortent du panel DADS, puisqu'ils ne sont plus rémunérés. Les sorties du marché du travail, surtout si elles sont fréquentes, réduisent l'expérience locale des salariés des zones ayant de faibles taux d'emploi, mais aussi leur expérience totale si ces salariés ne sont pas mobiles (pour des raisons liées, par exemple, à l'attachement culturel ou aux aménités climatiques), et ne transitent pas par d'autres zones pour trouver un nouvel emploi. La figure 10 illustre le nombre moyen de transitions pour les salariés de chaque zone d'emploi en 2008, qui est effectivement très faible dans toutes les zones d'emploi (un peu moins d'une année).

Même si la variabilité spatiale du nombre de transitions est faible, il est tout de même intéressant de noter que les salariés travaillant en Bretagne, dans le sud-est et dans le bassin parisien (à l'exception notable de Paris), ont davantage bougé, et cette mobilité leur a permis d'acquérir une plus grande expérience en dehors de leur zone d'emploi actuelle.

**Figure 10 : Nombre moyen de zones dans lesquelles les salariés ont exercé une activité en dehors de la zone d'emploi actuelle en 2008 (panel tronqué)**



Source : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

### 3. Résultats de l'estimation des économies d'agglomération

Les études ayant estimé les créations d'activités engendrées par le projet du Grand Paris Express prévoient une hausse assez substantielle de l'emploi au sein de la région francilienne, et annoncent des effets positifs qui s'étendent bien au-delà du périmètre des gares du futur réseau (Société du Grand Paris, 2014). Elles prévoient tout d'abord que la seule réalisation du métro contribuera à la création de 15 000 à 20 000 emplois directs non délocalisables liés à la phase du chantier, dans le secteur des travaux publics, des industries ferroviaires et le bâtiment. Les travaux étant réalisés en continu et à plusieurs endroits en même temps<sup>34</sup>, différentes villes situées dans la Petite Couronne parisienne seront simultanément affectées.

Au-delà des emplois directement liés aux travaux, les études prédisent également que l'arrivée du Grand Paris Express permettra de créer entre 115 000 et 315 000 emplois supplémentaires, par rapport à l'évolution tendancielle de la région. Ces créations proviennent notamment des gains d'accessibilité engendrés par la baisse des temps de transport en commun, et du regain d'attractivité des zones bénéficiaires pour les investissements internationaux. Si ces prédictions sont corroborées, on devrait assister à une modification substantielle de la distribution géographique des emplois et des qualifications, aussi bien à l'intérieur de l'Île-de-France qu'entre la zone francilienne et le reste de la France. Il est donc essentiel de bien estimer l'impact que peuvent avoir les créations d'emploi à l'échelle d'une grande métropole comme Paris, mais aussi d'évaluer leurs effets redistributifs potentiels à l'échelle de toute la France.

Comme nous l'avons vu dans la première partie de ce rapport, les gains à attendre du Grand Paris Express sont de trois types. Ils incluent les effets positifs potentiels de la création d'emplois qualifiés dans la région. Or les travailleurs qualifiés ont une productivité intrinsèquement plus élevée que les autres : les salaires franciliens devraient donc augmenter. Les deux autres gains potentiels sont liés aux économies d'agglomération statiques et dynamiques engendrés par la densification différenciée des zones d'emploi du bassin parisien. Les gains de productivité statiques proviennent des externalités d'urbanisation et de spécialisation causées par cette densification à court terme. Quant aux gains de productivité dynamiques, ils sont liés à l'acquisition par les salariés de nouvelles compétences et savoir-faire à plus long terme. Ces gains d'apprentissage devraient accroître la productivité des salariés franciliens, mais aussi potentiellement celle des autres salariés, si l'expérience accumulée dans la métropole parisienne est transférable dans l'espace, et si les travailleurs sont mobiles et changent de ville tout au long de leur carrière.

Le modèle économique permettant d'identifier séparément ces trois types de gains, tout en ayant l'avantage de minimiser les biais de variables omises potentiels est le suivant :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + \sum_{c'} \mu_{c',c(i,t)}(e_{i,c',t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_{c',c} &= Z_{c',\cdot} \psi + Z_{c',\cdot} (Z_{c,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{c',c}. \end{aligned} \quad (0)$$

où  $e_{i,t}$  est l'expérience totale du salarié  $i$  à la date  $t$  ( $c'$  est-à-dire le nombre total d'années d'activité) et  $e_{i,c',t}$  le nombre d'années d'activité passées dans la ville  $c'$ . Le terme  $(e_{i,c',t} - e_{i,t}/C)$  mesure le surcroît ou le déficit d'expérience accumulée en  $c'$  par rapport à une distribution spatiale uniforme de référence de l'expérience totale entre les villes (au nombre de  $C$  dans le panel).

La première étape consiste à estimer l'impact sur les salaires de plusieurs catégories de variables. Le vecteur  $X_{i,t}$  mesure l'effet des caractéristiques individuelles du salarié  $i$  observables à la date  $t$ ,

<sup>34</sup> Les différentes étapes du projet du Grand Paris Express sont illustrées par la carte de l'Annexe 6.

comme le genre, l'âge, le carré de l'âge (pour tenir compte de la décroissance des rendements marginaux de l'âge) ou l'expérience totale ( $e_{i,t}$ ), mesurée par le nombre d'années d'activité (non colinéaire à l'âge de l'employé du fait des entrées et sorties du marché du travail).

Le vecteur  $W_{c,k,t}$  permet d'évaluer l'impact des caractéristiques locales sectorielles susceptibles d'influencer la productivité, comme la spécialisation de la zone ou sa composition en termes de groupes socioprofessionnels. Les effets fixes  $\beta_c$  et  $\vartheta_{k,t}$  capturent respectivement l'impact sur le salaire de toutes les caractéristiques locales omises et inobservées, et des chocs sectoriels globaux (qui frappent les villes de manière uniforme). La variable  $(e_{i,c',t} - e_{i,t}/C)$  mesure le surcroît ou le déficit d'expérience accumulée par le salarié dans la ville  $c'$  (qui peut être celle dans laquelle il travaille à la période courante) par rapport à une distribution spatiale uniforme de référence de l'expérience totale entre les villes (au nombre de  $C$  dans le panel), soit l'expérience acquise par l'individu dans chaque zone s'il les avait toutes fréquentées. Le paramètre  $\mu_{c',c}$  mesure donc la valorisation dans la ville  $c$  du surcroît d'expérience passée dans la ville  $c'$  par rapport à cette moyenne.<sup>35</sup>

La seconde étape consiste à estimer la prime salariale liée au fait de travailler à la période courante dans la ville  $c$ , qui dispose des caractéristiques moyennes  $Z_{c,t}$  (densité d'emplois, superficie et/ou potentiel marchand). Elle permet d'identifier séparément le gain statique moyen, mesuré par le paramètre  $\gamma$ , et les gains dynamiques, mesurés respectivement par  $\psi$  (valorisation moyenne de l'expérience passée dans la ville  $c'$ ), et par  $\nu$  (valorisation additionnelle de l'expérience passée en  $c'$  engendrée par le fait de travailler en  $c$ ).

Comme le modèle (0) est assez complexe à estimer, et que ses résultats ne peuvent être comparés directement à ceux produits par d'autres études empiriques, nous commençons par estimer des spécifications « réduites », n'intégrant que l'effet des économies d'agglomération statiques (sections 3.1 et 3.2). Nous estimons ensuite les gains d'apprentissage liés aux économies d'agglomération dynamiques (section 3.3), et étudions dans quelle mesure ils varient selon la taille des villes.

### 3.1 Des économies d'agglomération statiques assez substantielles

La manière la plus simple d'évaluer l'ampleur des économies d'agglomération statiques, tout en tenant compte des compétences propres des salariés, consiste à estimer équation Mincerienne du type :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t}\gamma + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \quad (1)$$

dans laquelle on mesure simultanément l'impact des caractéristiques locales de la zone ( $Z_{c,t}$ ) et l'effet des compétences propres du salarié, décomposées en un effet fixe individuel  $u_i$  (mesurant l'impact des caractéristiques inobservables invariantes au cours du temps, comme le genre ou l'origine sociale), et un vecteur  $X_{i,t}$  (mesurant l'impact des caractéristiques observables qui ne sont pas déjà capturées par  $u_i$  et  $\vartheta_{k,t}$ , l'effet fixe secteur-temps).

Par souci de comparaison avec les études antérieures, nous nous focalisons tout d'abord sur l'effet de la densité, mesurée par le nombre d'actifs occupés par unité de surface de la zone<sup>36</sup> (on a donc  $Z_{c,t} = \log dens_{c,t}$ ).

<sup>35</sup> Notons que les paramètres  $\mu_{c',c}$  sont identifiés pour tous les couples  $(c',c)$  à l'exception de ceux relatifs à une des zones d'origine  $c'$ , pour lesquels ils sont normalisés arbitrairement à 0. Nous choisissons comme référence les zones situées dans le premier quintile de densité.

<sup>36</sup> Ce nombre correspond aux effectifs comptabilisés dans le panel DADS. Il est donc calculé à partir des salariés des secteurs privés et semi-publics uniquement. Cependant la corrélation spatiale avec l'emploi total et la population est extrêmement forte : utiliser l'une ou l'autre de ces mesures n'affecte pas l'élasticité estimée.

La colonne (1) du Tableau 1 présente le résultat de cette estimation, qui permet tout de même d'expliquer 21 % de la variabilité des salaires de l'échantillon (qui compte, rappelons-le, 2 781 380 observations). L'élasticité du salaire par rapport à la densité est estimée à 0,021. L'interprétation de cette valeur est simple : si on compare deux villes, et si la première est deux fois plus dense que la seconde, un actif occupé y disposera d'un salaire supérieur de  $(2^{0.021} - 1) \times 100 = 1,5\%$  à celui d'un salarié disposant des mêmes compétences, mais travaillant dans la seconde. En supposant pour le moment que l'effet est causal et l'estimation non-biaisée par des problèmes d'endogénéité, doubler la densité d'emplois d'une zone reviendrait donc à y accroître les salaires de 1,5 %. Ce chiffre mesure l'impact des économies d'agglomération statiques, net de l'effet des caractéristiques individuelles observables et inobservables des salariés (en particulier les qualifications).

**Tableau 1 : Economies d'agglomération statiques liées à la densité des villes**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Densité	0.021	0.018	0.021	<b>0.021</b>	0.037	0.028	0.026	0.020	<b>0.020</b>	0.021	<b>0.017</b>	0.020
Spécialisation	-	0.014	-	<b>0.013</b>	-	0.023	0.024	-	<b>0.014</b>	0.014	<b>0.015</b>	Ind.
CSP2	-	0.042	-	<b>0.073</b>	-	0.627	0.621	-	<b>0.062</b>	0.062	<b>0.067</b>	Ind.
CSP3	-	0.153	-	<b>0.132</b>	-	0.834	0.839	-	<b>0.144</b>	0.144	<b>0.168</b>	Ind.
CSP4	-	0.040	-	<b>0.036</b>	-	0.340	0.333	-	<b>0.036</b>	0.036	<b>0.033</b>	Ind.
CSP5	-	-0.027	-	<b>-0.021</b>	-	0.098	0.092	-	<b>-0.024</b>	-	-	Ind. 0.024 <b>0.026</b>
EF individu	oui	oui	oui	<b>oui</b>	âge, sexe	âge, sexe	âge, sexe	oui	<b>oui</b>	oui	<b>oui</b>	oui
R <sup>2</sup> 1 <sup>ère</sup> étape	0.21	0.21	0.23	<b>0.23</b>	0.30	0.33	0.32	0.28	<b>0.29</b>	0.29	<b>0.28</b>	0.29
R <sup>2</sup> 2 <sup>ème</sup> étape	-	-	0.29	<b>0.29</b>	0.36	0.43	0.48	0.25	<b>0.25</b>	0.71	-	0.25
Estimation en 2 étapes	-	-	oui	<b>oui</b>	oui	oui	oui	oui	<b>oui</b>	oui	-	oui
Panel tronqué	-	-	-	-	-	-	-	oui	<b>oui</b>	oui	<b>oui</b>	oui
Salariés avec 1 obs.	-	-	-	-	-	-	oui	-	-	-	-	-
Pondéré en 2 <sup>ème</sup> étape	-	-	-	-	-	-	-	-	-	oui	-	-

Notes : (i) Coefficients significatifs au seuil de 5% ou moins (1% dans l'immense majorité des cas) ; (ii) Toutes les régressions incluent le carré de l'âge du salarié et un effet fixe secteur-temps. L'expérience totale du salarié est aussi prise en compte dans les colonnes (8) à (12) ; (iii) CSP2 : Artisans, commerçants et chefs d'entreprise ; CSP3 : Cadres et professions intellectuelles supérieures ; CSP4 : Professions Intermédiaires ; CSP5 Employés ; CSP6 : Ouvriers (référence). Les autres PCS (cf. Annexe 4) ont été exclues ; (iv) Classification sectorielle utilisée : NAF114 (cf. Annexe 2) ; (v) Les variables de spécialisation et les CSP sont calculées au niveau ZE-NAF 114, mais les coefficients associés ne dépendent pas du secteur, sauf dans la colonne (12) ; (vi) Les colonnes en gras-violet correspondent à nos spécifications favorites.

L'estimation de la spécification (1), si elle neutralise bien l'effet des caractéristiques individuelles, ne tient pas compte des effets de composition sectorielle, qui peuvent rendre les travailleurs plus productifs indépendamment de leurs compétences intrinsèques (s'ils interagissent fréquemment avec des collègues compétents du même secteur, par exemple).

La colonne (2) du Tableau 1 présente les résultats d'une seconde estimation, dans laquelle la mesure des économies d'agglomération est purgée de ces effets de composition sectoriels :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + Z_{c(i,t),t} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \quad (2)$$

Le vecteur  $W_{c,k,t}$  capture l'impact des caractéristiques sectorielles locales susceptibles d'influencer la productivité des salariés du secteur, à savoir la spécialisation,  $spe_{c,k,t}$ , mesurée par le log de la part de l'emploi du secteur dans la zone, et la part des effectifs occupés par les différentes Catégories Socio-Professionnelles ( $csp_{j,c,k,t}$ )<sub>j=1,...,6</sub> dans l'emploi sectoriel de la zone.

L'élasticité du salaire par rapport à la spécialisation est environ deux fois plus faible que celle de la densité totale : les externalités d'urbanisation semblent donc plus fortes que les externalités de localisation. Au sein de chaque secteur, la présence de chefs d'entreprises (artisans ou commerçants), de cadres ou professions intellectuelles supérieures ou de professions intermédiaires dans la zone, y accroît la productivité des salariés du secteur, en comparaison d'une situation de référence dans laquelle il n'y a que des ouvriers. En revanche, les employés ont un impact plus faible que les ouvriers sur la productivité. À l'inverse, conformément aux travaux existants pour la France ou d'autres économies, l'effet le plus fort est celui des cadres et professions intellectuelles supérieures. L'inclusion des variables de spécialisation et de composition dans l'estimation fait très légèrement baisser l'élasticité de la densité (le coefficient estimé passe de 0,021 à 0,018). Dans la spécification (1), malgré la présence de l'effet fixe individuel contrôlant pour le tri spatial des qualifications, la densité capturerait donc partiellement l'effet des externalités de localisation et de capital humain.

D'une manière générale, comme nous l'avons vu dans la première partie de ce rapport, de nombreuses autres variables omises sont susceptibles de biaiser l'estimation de l'effet de la densité, et les écarts-types de cette variable. La seule manière d'éradiquer ce biais consiste à procéder en deux étapes, comme l'ont fait Combes, Duranton et Gobillon (2008). Dans la première étape, la rémunération est déterminée par les compétences propres du salarié et, le cas échéant, par les effets de composition sectorielle de la zone, ainsi que par un effet fixe  $\beta_{c(i,t),t}$  capturant l'impact de toutes les autres caractéristiques locales omises ou inobservées susceptibles d'influencer la productivité. La seconde étape permet ensuite d'estimer ce qui, dans la productivité locale résiduelle estimée en première étape,  $\hat{\beta}_{c,t}$ , est lié à la densité de la zone. Selon que l'on considère ou non les effets de composition sectoriels, on estime :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \quad (3)$$

$$\beta_{c,t} = Z_{c,t}\gamma + u_t + \eta_{c,t},$$

ou bien 
$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \quad (4)$$

$$\beta_{c,t} = Z_{c,t}\gamma + u_t + \eta_{c,t},$$

Les résultats de ces deux estimations sont présentés dans les colonnes (3) et (4) du Tableau 1. A elle-seule, la densité explique près du tiers de la variabilité spatiale des salaires non imputable au tri spatial des travailleurs et/ou aux externalités de localisation (le  $R^2$  de 2<sup>ème</sup> étape est égal à 0.29). Le pouvoir explicatif de l'étape 2 du modèle est un peu plus faible que dans Combes, Duranton et Gobillon (2008). Mais cette étude s'appuyait sur l'ancienne version des zones d'emploi, plus adaptée

à la période 1993-2008 que le zonage 2010<sup>37</sup>. La densité reste néanmoins un très bon prédicteur des écarts de rémunération entre villes. Le coefficient associé à la densité est plus élevé que dans la colonne (2). Des caractéristiques locales omises corrélées négativement à la densité semblent donc biaiser son effet vers le bas. Le biais associé à l'omission des effets de composition sectorielle et socio-professionnelle est quasi nul dans le modèle à deux étapes : l'élasticité estimée pour la densité est de 0,021, que l'on inclut ces effets ou pas.

Jusqu'à présent, nous avons inclus dans toutes les régressions un effet fixe individuel,  $u_i$ , permettant d'identifier séparément la prime salariale liée aux compétences propres de l'individu, et celle liée au fait qu'il travaille dans une zone dense. Afin de saisir l'importance de cet effet fixe et donc implicitement, de la nécessité d'utiliser des données de panel pour estimer les économies d'agglomération, les colonnes (5) et (6) du Tableau 1 présentent les résultats de l'estimation des modèles (3) et (4), dans lesquelles on « remplace » cet effet fixe par des caractéristiques individuelles observables, comme le genre ou l'âge. La précision de l'inférence augmente, puisque le modèle n'est plus estimé en « within » lors de la première étape (seul moyen d'éliminer  $u_i$  lorsqu'on a un échantillon constitué de près de 500 000 individus). En revanche, les coefficients de toutes les variables explicatives explosent. Si on compare les colonnes (3) et (5), la hausse atteint 75% pour la seule densité (passage de 0,021 à 0,037). La hausse est moindre (33%) si on compare les colonnes (4) et (6), qui intègrent en sus les effets de composition sectorielle. Mais les différentes CSP ont un impact presque 10 fois plus élevé qu'avant sur le salaire. Elles capturent en effet maintenant l'influence des variables individuelles inobservées, tout comme le fait aussi la densité. Comme les salariés qualifiés sont surreprésentés dans les zones denses, les coefficients de la densité et des CSP sont très fortement biaisés à la hausse du fait de ce tri spatial, qui n'est pas contrôlé.

La colonne (7) permet d'étudier un autre biais d'estimation éventuel : celui lié au fait que la méthode d'estimation en « within » n'identifie l'effet des économies d'agglomération qu'à l'aide des observations répétées du panel. Les individus qui ne sont présents dans le panel qu'à une seule date sont donc exclus de l'inférence (il y en a 352 193, soit un peu moins de 1/10 de l'échantillon). Or les économies d'agglomération statiques pourraient avoir une amplitude différente pour cette population, qui dispose selon toute vraisemblance de caractéristiques particulières (une faible expérience en l'occurrence). Afin de tester cette conjecture, la colonne (7) du Tableau 1 présente les résultats de l'estimation réalisée sur ces seuls individus, à l'aide de la spécification (2). Les résultats ne sont pas sensiblement différents de ceux mis en évidence dans la colonne (6). Les effets d'agglomération estimés en « within » sont donc bien représentatifs des deux types de populations.

Les spécifications estimées dans les colonnes (8) et (9) sont les mêmes que celles correspondant aux colonnes (3) et (4), mais elles incluent dans le vecteur des caractéristiques individuelles observables  $X_{i,t}$  l'expérience totale du salarié. Comme nous l'avons vu dans la seconde partie de ce rapport, cette expérience peut être mesurée par le nombre d'années passées par l'individu dans le panel DADS. Mais comme ce panel ne commence qu'en 1976, l'approximation est d'autant plus mauvaise que les salariés sont plus âgés, car ils sont en fait occupés depuis bien plus longtemps que ne le suggère le panel. Afin de réduire cette erreur de mesure, il est donc préférable d'exclure les individus les plus âgés de l'échantillon (i.e. ceux nés avant 1958). L'estimation est donc réalisée sur un panel tronqué composé de 1 931 838 individus (observés au moins deux fois dans la période couverte par le panel). La prise en compte de l'expérience dans le cadre du panel tronqué accroît sensiblement la qualité de l'ajustement de première étape, dont les  $R^2$  atteignent maintenant respectivement 0,28 et 0,29 dans les colonnes (8) et (9).

---

<sup>37</sup> Nous privilégions néanmoins le zonage 2010 dans ce rapport car il est mieux adapté à la géographie actuelle des villes, et la phase 3 de notre étude consistera à effectuer des prédictions de l'impact du Grand Paris Express. C'est donc sur le contour actuel des zones d'emploi qu'il faut réaliser les estimations.



Les trois dernières colonnes du Tableau 1 testent la robustesse des résultats précédents. La spécification estimée dans la colonne (10) est la même que celle utilisée dans la colonne (9), mais les estimateurs de seconde étape sont maintenant pondérés par le nombre d'individus observés dans la zone. En effet, dans la mesure où la productivité résiduelle de seconde étape ne dépend plus des individus, une estimation par les MCO non pondérée confère le même poids à toutes les zones, alors qu'elles n'ont pas toutes le même nombre d'actifs occupés. Or si on s'intéresse aux trajectoires individuelles des salariés, qui sont plus nombreux dans les grandes métropoles, il vaut mieux attribuer à ces dernières un poids plus important dans l'estimation. Les résultats présentés dans la colonne (11) révèlent que les estimateurs pondérés ne diffèrent pas significativement des estimateurs non pondérés.

En revanche, le pouvoir explicatif du modèle est bien meilleur lorsque l'estimation est pondérée : la valeur du  $R^2$  de 2<sup>ème</sup> étape passe à 0.72, alors qu'elle n'était que de 0.28 dans la colonne (9). Un second moyen de conserver la perspective individuelle qui nous intéresse est de réaliser l'estimation en une seule étape, comme dans la colonne (2), mais à partir du panel tronqué, de manière à tenir compte de l'expérience des salariés. La colonne (11) présente les résultats de cette inférence. Les coefficients diffèrent sensiblement de ceux présentés dans la colonne (10), mais la qualité de l'ajustement est moindre ( $R^2$  de 0.28) et surtout, les estimateurs sont entachés d'un biais de variable omises du même ordre de grandeur que celui mis en évidence dans la colonne (2) pour le panel non tronqué. La dernière colonne du Tableau 1 présente les résultats d'une ultime estimation à deux étapes, dans laquelle les variables de composition et de spécialisation sont interagies avec des indicatrices précisant le secteur d'appartenance du salarié, et définies au niveau 1 de la Nomenclature Economique de Synthèse (NES36). Le coefficient obtenu pour la densité (0.020) et la qualité de l'ajustement ( $R^2$  de 0.28) sont strictement identiques à ceux de la colonne (9).

En définitive, les estimateurs présentés dans les colonnes (4) et (9) du Tableau 1 ont l'avantage d'être à la fois précis, robustes et faciles à implémenter. Si on se réfère aux effets fixes estimés dans la colonne (4) du Tableau 1, les disparités de salaires entre villes sont notables, même pour des travailleurs disposant de compétences similaires.

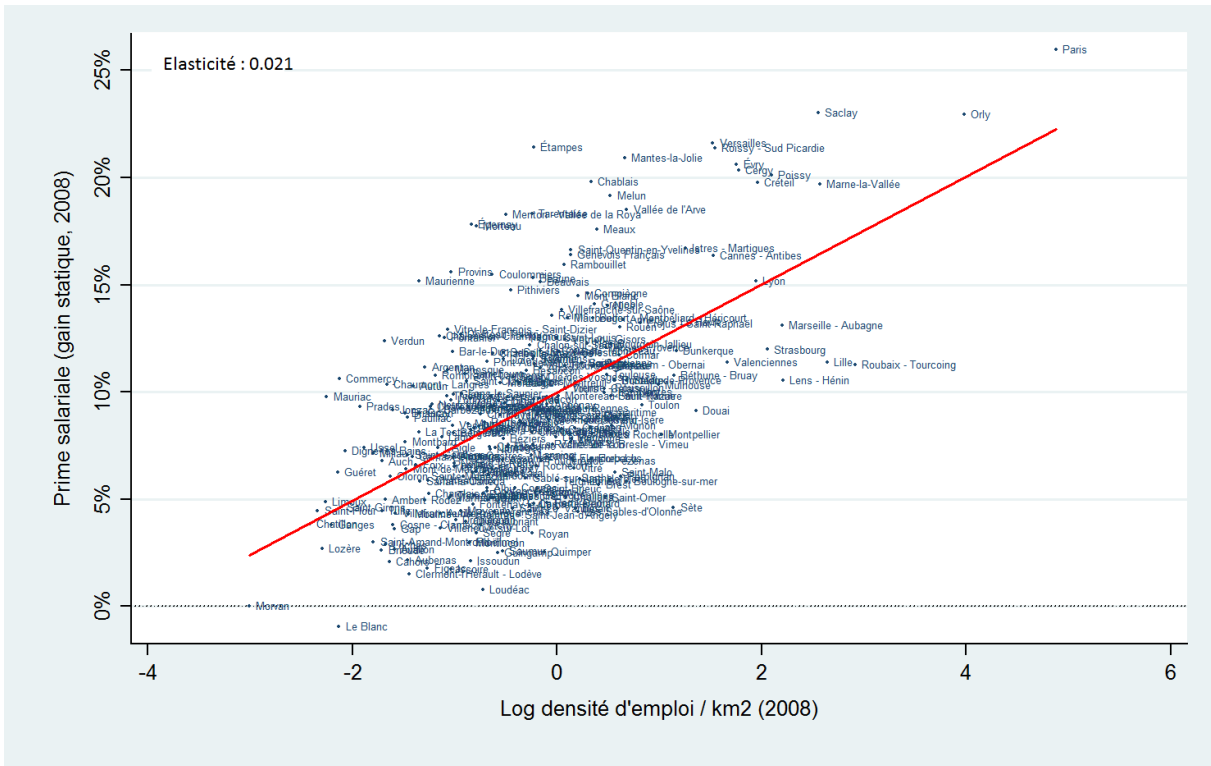
Comme l'illustre la Figure 11 pour l'année 2008, un actif occupé travaillant à Paris dispose d'un salaire supérieur de 26% à celui d'un individu strictement identique<sup>38</sup>, mais travaillant dans le Morvan, la zone d'emploi la moins dense du panel en 2008. D'une manière générale, la prime dont bénéficient les salariés du Grand Paris est assez remarquable : +23% à Saclay ou Orly, +20% à Versailles, Étampes, Roissy-sud-Picardie, Mantes-la-Jolie, Évry, Cergy, Poissy, Créteil, Marne-la-Vallée ou Melun, +17% à Meaux ou Saint-Quentin-en-Yvelines, +15% à Coulommiers ou Provins. À compétences données, les primes associées aux autres grandes métropoles sont plus faibles : par exemple +15% à Lyon, +14% à Nice ou à Grenoble, +13% à Marseille-Aubagne, +12% à Strasbourg ou Lille, ou +10% à Bordeaux. Les zones d'emploi rurales et peu denses du centre (Le Blanc, Montluçon, Issouire, Figeac, Cahors, Issoudun, Lozère) et du littoral breton (Loudéac, Quimper, Ploërmel) disposent des primes les plus faibles de toute la France.

On le voit aussi bien sur la Figure 11 que dans la colonne (4) du Tableau 1, l'élasticité de la prime salariale par rapport à la densité d'emploi est estimée à 0.02, une valeur très proche des estimations existantes (Combes et Lafourcade, 2012). Cette valeur, même si elle est faible, induit des disparités géographiques de salaire assez substantielles, car les écarts de densité entre villes françaises sont très forts. Par exemple, si on compare les deux extrêmes de la distribution, Paris et Morvan, le différentiel de productivité imputable à l'écart de densité entre ces deux villes est de  $(2660^{0.02} - 1) \times 100 = 17\%$  environ. Si on compare les zones d'emploi situées aux bornes des premier et dernier quintiles, à savoir Périgueux et Belfort-Montbéliard-Héricourt, l'écart est en revanche bien plus faible, à  $(4^{0.02} - 1) \times 100 = 3\%$  environ.

---

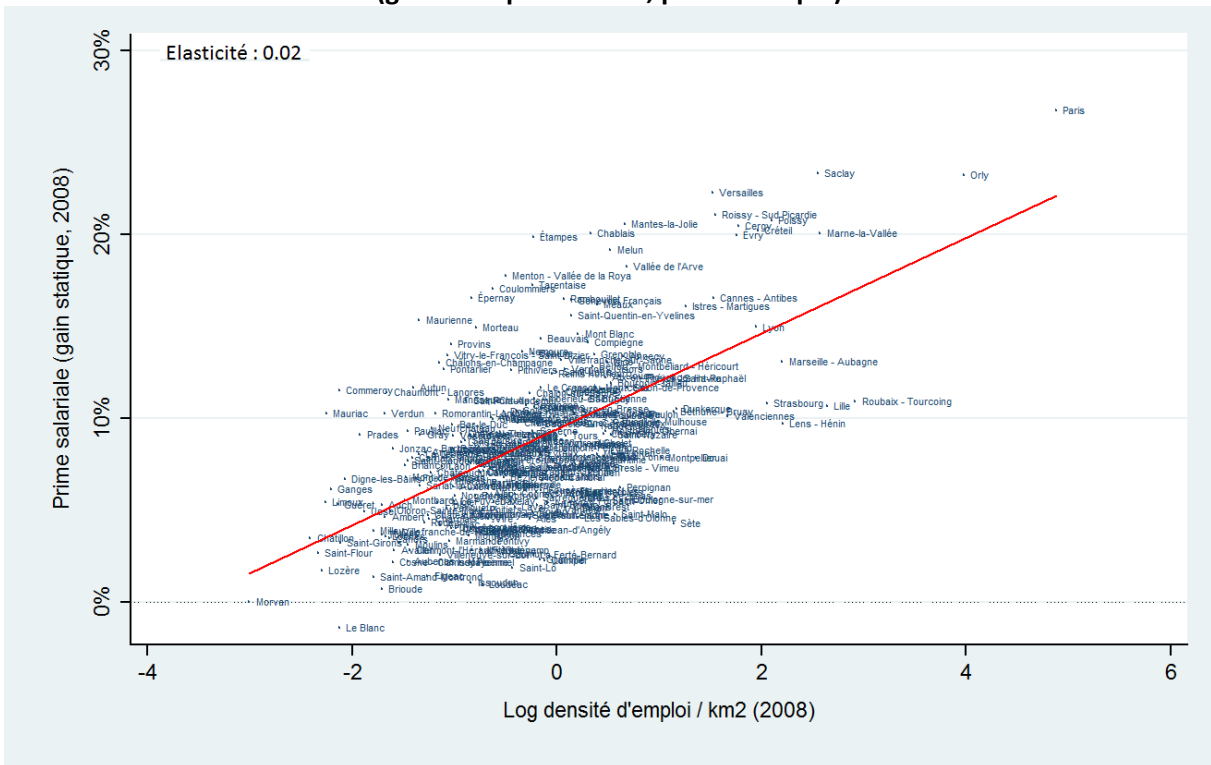
<sup>38</sup> C'est-à-dire disposant des mêmes caractéristiques individuelles observables et inobservables.

**Figure 11 : Prime salariale associée aux différentes zones d'emploi (gain statique en 2008, panel non tronqué)**



Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan en 2008), calculé à partir de l'exponentielle des effets fixes  $\beta_{c,2008}$  estimés en colonne (4) du Tableau 1.

**Figure 12 : Prime salariale associée aux différentes zones d'emploi (gain statique en 2008, panel tronqué)**

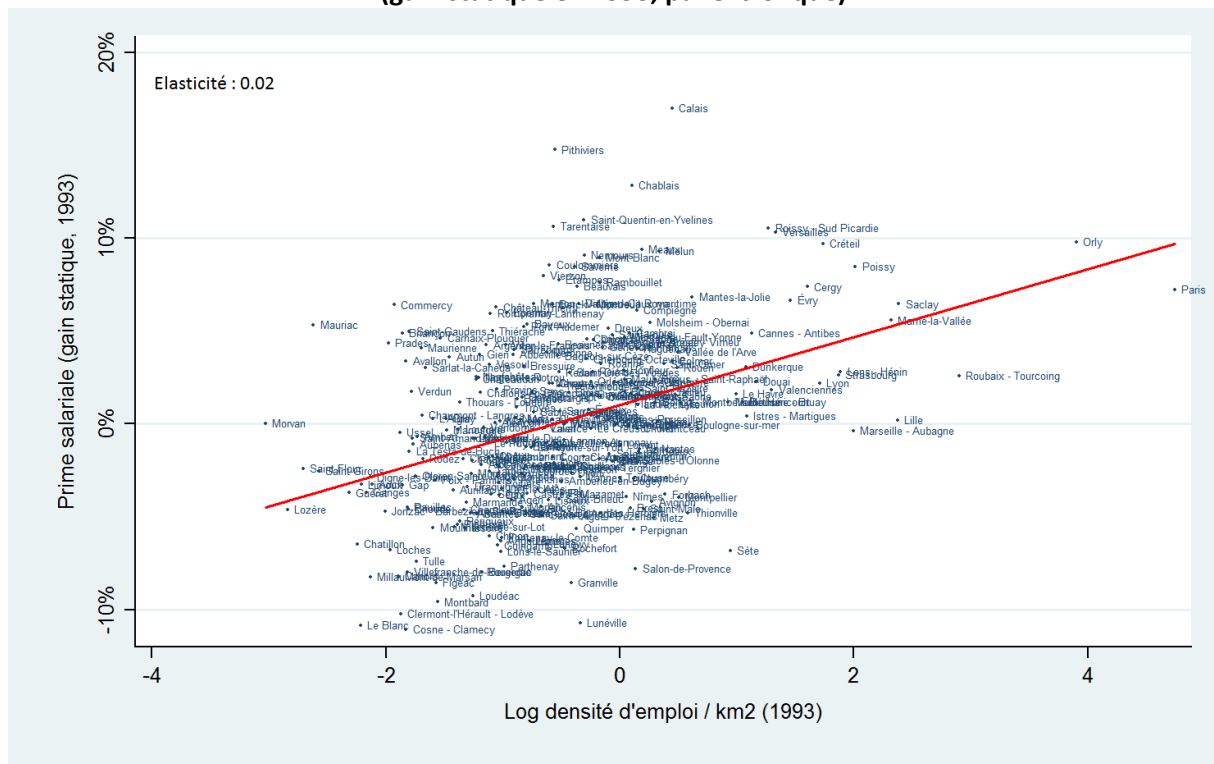


Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan en 2008), calculé à partir de l'exponentielle des effets fixes  $\beta_{c,2008}$  estimés en colonne (9) du Tableau 1.

La hiérarchie des villes est globalement la même si on mesure les disparités géographiques de salaire à partir des effets fixes estimés dans le cadre du panel tronqué (colonne 9 du Tableau 1), comme l'illustre la Figure 12. Les inégalités de salaire correspondant à l'année 2008 sont néanmoins légèrement exacerbées, en comparaison du cas où le panel n'est pas tronqué.

En revanche, la hiérarchie des villes et l'intensité des économies d'agglomération ont évolué au cours du temps. La Figure 13 met en évidence des disparités spatiales de salaires différentes en 1993 : relativement à Morvan, la prime salariale des zones d'emploi du Grand Paris n'excède pas 11 % (valeur maximale constatée à Saint-Quentin-en-Yvelines). Des villes comme Calais, Pithiviers ou Chablais sont caractérisées par des primes salariales élevées en 1993, même si elles ne sont pas parmi les plus denses du panel. La densité, si elle explique une grande part de la variabilité spatiale des salaires, n'explique pas tout. Il est donc important de considérer d'autres caractéristiques locales susceptibles d'expliquer ces disparités.

**Figure 13 : Prime salariale associée aux différentes zones d'emploi (gain statique en 1993, panel tronqué)**



Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan en 1993), calculé à partir de l'exponentielle des effets fixes  $\beta_{c,1993}$  estimés en colonne (9) du Tableau 1.

Comme nous l'avons rappelé dans la première partie de ce rapport, la densité ne capture qu'un aspect particulier de la taille des villes, lié à son « épaisseur ». Or l'étendue/l'étroitesse du marché, ainsi que l'accès aux marchés externes, comptent peut-être tout autant que la densité. Afin d'évaluer le rôle joué par ces autres déterminants, le Tableau 2 présente les résultats de plusieurs estimations, dans lesquelles le vecteur des caractéristiques locales  $Z_{c,t}$  inclut aussi la superficie et le potentiel marchand de la zone.

La prise en compte de la superficie et du potentiel marchand réduit l'élasticité du salaire par rapport à la densité, qui chute de près de 30 % par rapport au Tableau 1, pour s'établir à une valeur très stable de 0.013. Doubler la densité d'une zone permet donc d'y augmenter le salaire moyen de  $(2^{0.013} - 1) \times 100 = 0,9\%$  environ, à superficie et potentiel marchand donnés. Il est assez remarquable de constater que l'accès aux marchés externes a une élasticité environ trois fois supérieure à celle de la densité, alors que l'effet de la superficie est négligeable, même s'il est significatif.

**Tableau 2 : Economies d'agglomération statiques liées à la taille globale des villes**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Densité	0.014	0.014	0.012	0.014	<b>0.013</b>	0.021	0.013	<b>0.012</b>	0.013
Potentiel marchand	0.047	0.040	0.044	0.038	<b>0.037</b>		0.048	<b>0.042</b>	0.038
Superficie	0.007	0.002	0.007	0.004	<b>0.003</b>		0.003	<b>0.006</b>	0.005
R <sup>2</sup> 1 <sup>ère</sup> étape	0.21	0.23	0.22	0.23	<b>0.28</b>	0.28	0.28	<b>0.27</b>	0.26
R <sup>2</sup> 2 <sup>ème</sup> étape	-	0.45	-	0.46	<b>0.40</b>	0.32	0.48	-	0.80
Effet fixe individu	oui	oui	oui	oui	<b>oui</b>	oui	oui	<b>oui</b>	oui
Spécialisation			oui	oui	<b>oui</b>	oui	oui	<b>oui</b>	oui
CSP2-CSP6			oui	oui	<b>oui</b>	oui	oui	<b>oui</b>	oui
Aménités en 2 <sup>ème</sup> étape						oui	oui		
Estimation en 2 étapes		oui		oui	<b>oui</b>	oui	oui		
Panel tronqué					<b>oui</b>	oui	oui	<b>oui</b>	oui
Pondéré en 2 <sup>ème</sup> étape									oui

Notes : (i) Coefficients significatifs au seuil de 5% au pire (de 1% dans l'immense majorité des cas) ; (ii) Toutes les régressions incluent le carré de l'âge du salarié et un effet fixe secteur-temps. L'expérience totale du salarié est aussi prise en compte dans les colonnes (5) à (9) ; (iii) Les aménités sont mesurées par des indicatrices précisant si la ville est côtière, si elle borde un lac, si elle est dans une zone montagneuse, ou si elle dispose de monuments classés au patrimoine architectural ; (iv) Les colonnes en gras-violet correspondent à nos spécifications favorites.

Dans l'estimation la plus basique, l'élasticité du salaire par rapport au potentiel marchand est de 0.047, mais cette valeur diminue lorsqu'on passe à une estimation à deux étapes (passage de 0.047 à 0.040), et qu'on tient compte des effets de spécialisation et de composition (passage de 0.044 à 0.038 ou à 0.037 pour le panel tronqué). Ainsi, le fait de doubler le potentiel marchand d'une zone accroît la rémunération moyenne de ses salariés de  $(2^{0.038} - 1) \times 100 = 2,5\%$ , toutes choses égales par ailleurs. Le différentiel de productivité imputable à l'écart de potentiel marchand entre les deux extrêmes de la distribution des potentiels marchands, Orly et Bayonne, est de  $(14^{0.038} - 1) \times 100 = 10,5\%$  environ. Les économies d'agglomération statiques liées au potentiel marchand semblent donc substantielles, et leur impact similaire à celui de la densité, une fois prise en compte la moindre variabilité spatiale du potentiel marchand par rapport à celle de la densité<sup>39</sup>.

Comme nous l'avons vu dans la seconde partie de ce rapport, les zones d'emploi sont cependant très inégales en termes de potentiel marchand, dont la valeur décroît fortement à mesure qu'on s'éloigne du centre vers la périphérie. Or en France, les zones d'emploi périphériques sont presque toutes situées sur le littoral ou dans des espaces très escarpés ou montagneux. L'élasticité du salaire par rapport au potentiel marchand peut donc être biaisée par des aménités liées à la géographie physique. De plus, le fait de négliger les aménités culturelles qui influencent les choix de localisation des ménages (Roback, 1982) peut également biaiser les coefficients estimés.

<sup>39</sup> Pour conclure définitivement sur l'importance relative du potentiel marchand et de la densité, il faut néanmoins effectuer une analyse très précise de la variance.

Afin de corriger ces biais éventuels, les colonnes (6) et (7) du Tableau 2 incluent dans l'estimation de seconde étape des variables permettant de capturer l'effet de diverses aménités (zone côtière ou montagneuse, bordant un lac ou disposant de monuments classés et d'un patrimoine architectural). Dans la colonne (6), la seule variable de taille considérée est la densité. Le fait d'inclure les aménités ne change pas le coefficient de la densité (le coefficient estimé, à 0.021 est le même que dans la colonne 3 du Tableau 1), mais augmente le pouvoir explicatif du modèle en seconde étape (le  $R^2$  est maintenant de 0.32, alors qu'il était de 0.29 dans la colonne 3 du Tableau 1). En revanche, lorsqu'on considère toutes les variables de taille simultanément, la colonne (7) montre que la prise en compte des aménités accroît considérablement l'impact du potentiel marchand, dont le coefficient passe de 0.037 (colonne 5) à 0.048. Le potentiel marchand appréhendait donc auparavant l'effet des aménités, qui lui sont corrélées négativement.

Tout comme dans le Tableau 1, les colonnes (8) et (9) du Tableau 2 permettent de tester la robustesse des coefficients présentés dans la colonne (5) à l'adoption d'une perspective plus individuelle. Si l'estimation en une seule étape ne change pas drastiquement les coefficients des différentes variables (colonne 8), l'estimation pondérée atténue en revanche l'impact du potentiel marchand. On peut enfin noter que le pouvoir explicatif a considérablement augmenté du fait de l'inclusion de la superficie et du potentiel marchand : les  $R^2$  de seconde étape ont ici doublé par rapport à ceux des estimations similaires reportées dans le Tableau 1. Les trois variables de taille expliquent maintenant au pire 45%, au mieux 80% des disparités spatiales de salaire non imputables au tri spatial des travailleurs et aux externalités de localisation.

### 3.2 Des économies d'agglomération dynamiques transférables

Les spécifications étudiées dans la section précédente ont négligé les gains dynamiques potentiels liés à l'expérience passée des salariés, et au fait qu'ils ont accumulé des connaissances qu'ils peuvent valoriser différemment selon les caractéristiques de la ville dans laquelle ils travaillent désormais. La spécification permettant d'estimer ces économies d'agglomération dynamiques en parallèle des économies d'agglomération statiques correspond au modèle (0) présenté au début de la troisième partie de ce rapport. Cette spécification impose cependant d'estimer un nombre de paramètres  $\mu_{c',c}$  égal au carré du nombre de villes considérées, à savoir  $297 \times 297 = 88\,209$  dans notre échantillon. Or nous ne disposons pas d'un nombre de degrés de liberté suffisamment élevé pour estimer tous ces paramètres. Il est possible de pallier cette limite de deux façons : en imposant des restrictions sur certains paramètres, de manière à n'estimer que les gains dynamiques associés à quelques scénarii particuliers de valorisation de l'expérience, ou en restreignant la dimension spatiale de l'échantillon à quelques grandes classes de villes. Le Tableau 3 synthétise les résultats de ces deux types de stratégies.

Dans les colonnes (1) et (9), nous faisons tout d'abord l'hypothèse que l'expérience acquise par le salarié dans une autre ville que celle où il travaille n'engendre aucun gain dynamique ou, autrement dit, que l'expérience accumulée par le passé n'est pas transférable d'une ville à l'autre, et ne peut être valorisée que dans la ville où elle a été acquise. Le modèle estimé est donc le suivant :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + \mu_{c(i,t)}(e_{i,c(i,t),t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_c &= Z_{c,\cdot} \psi^i + \zeta_c, \end{aligned} \tag{1}$$

La taille de la ville a un impact très significatif et positif sur la prime salariale liée à l'expérience interne à la ville. Lorsqu'on considère uniquement la densité comme mesure de cette taille, le coefficient  $\psi^i$  est estimé à 0.0006. Cela signifie qu'une hausse de 1% de la densité d'une ville permet à ses salariés, s'ils restent dans cette ville, d'y accroître leur salaire annuel de 0.0006 % en moyenne par année supplémentaire passée dans la ville.

**Tableau 3 : Economies d'agglomération statiques et dynamiques liées à la taille globale des villes (panel tronqué)**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Effets dynamiques, ZE actuelle	oui	-	<b>oui</b>	<b>oui</b>	-	-	-	-	oui	-	<b>oui</b>	<b>oui</b>	-	-
Effets dynamiques, toutes ZE confondues	-	oui	-	-	-	-	-	-	-	oui	-	-	-	-
Effets dynamiques, autres ZE que l'actuelle	-	-	<b>oui</b>	<b>oui</b>	-	-	-	-	-	-	<b>oui</b>	<b>oui</b>	-	-
20x20 effets dynamiques, ZE actuelle	-	-	-	-	oui	-	-	-	-	-	-	-	oui	-
20x20 effets dynamiques, toutes ZE confondues	-	-	-	-	-	oui	-	-	-	-	-	-	-	-
20x20 effets dynamiques, autres ZE que l'actuelle	-	-	-	-	oui	-	-	-	-	-	-	-	oui	-
297x297 effets dyn. en 1, statiques en 2	-	-	-	-	-	-	oui	-	-	-	-	-	-	-
297x297 effets dynamiques, 1 seule étape	-	-	-	-	-	-	-	oui	-	-	-	-	-	oui
Pondéré 2 <sup>ème</sup> étape	-	-	-	<b>oui</b>	-	-	-	-	-	-	-	<b>oui</b>	-	-
R <sup>2</sup> 1 <sup>ère</sup> étape	0.28	0.28	<b>0.28</b>	<b>0.28</b>	0.28	0.28	0.28	0.28	0.28	0.28	<b>0.28</b>	<b>0.28</b>	0.28	0.28
R <sup>2</sup> 2 <sup>ème</sup> étape effets statiques	0.27	0.31	<b>0.30</b>	<b>0.71</b>	0.30	0.30	0.33	-	0.48	0.47	<b>0.46</b>	<b>0.82</b>	0.47	-
R <sup>2</sup> 2 <sup>ème</sup> étape effets dyn., ZE actuelle	0.06	-	<b>0.08</b>	<b>0.66</b>	0.55	-	-	-	0.09	-	<b>0.13</b>	<b>0.71</b>	0.56	-
R <sup>2</sup> 2 <sup>ème</sup> étape effets dyn., toutes ZE confondues	-	0.08	-	-	-	0.03	-	-	-	0.16	-	-	-	-
R <sup>2</sup> 2 <sup>ème</sup> étape effets dyn., autres ZE que l'actuelle	-	-	<b>0.06</b>	<b>0.36</b>	0.03	-	-	-	-	-	<b>0.14</b>	<b>0.52</b>	0.05	-
<b>Effets statiques</b>														
Densité	0.019	0.020	<b>0.021</b>	<b>0.022</b>	0.020	0.020	0.020	0.018	0.011	0.013	<b>0.015</b>	<b>0.015</b>	0.013	0.012
Potentiel marchand									0.040	0.035	<b>0.037</b>	<b>0.037</b>	0.037	0.041
Superficie									ns	0.002	<b>0.006</b>	<b>0.009</b>	ns	0.008
<b>Effets dynamiques</b>														
Densité, ZE actuelle	0.0006		<b>0.0007</b>	<b>0.0015</b>	ns		0.0015	0.0015	0.0008		<b>0.0011</b>	<b>0.0019</b>	ns	0.0019
Densité, toutes ZE confondues		0.0008					0.0016			0.0011				
Densité, autres ZE que l'actuelle			<b>0.0011</b>	<b>0.0013</b>	0.0016		0.0013	0.0013			<b>0.0017</b>	<b>0.0021</b>	0.0023	0.0025
Densités ZE orig. et dest. interagies					0.0002	0.0002	ns	ns					0.0002	0.0001
Potentiel marchand, ZE actuelle									-0.0008	-0.0010	<b>-0.0010</b>	ns	ns	-0.0003
Potentiel marchand, autres ZE que l'actuelle											ns	ns	ns	-0.0025
Potentiels marchands ZE orig. et dest. interagies														-0.0005
Superficie, ZE actuelle									ns	0.0007	<b>0.0005</b>	<b>0.0010</b>	ns	0.0011
Superficie, autres ZE que l'actuelle											<b>0.0018</b>	<b>0.0022</b>	ns	0.0006
Superficie ZE orig. et dest. interagies														0.0001

Notes : (i) Coefficients significatifs au seuil de 5% ou moins (1% dans la majorité des cas), sauf si « ns » (non significatif) est mentionné ; (ii) Toutes les régressions incluent les caractéristiques observables du salarié, des effets fixes individuel et secteur-temps, la spécialisation et la part des différentes CSP (cf. Annexe 4) dans l'emploi du secteur ; (iii) Les colonnes en gras-violet correspondent à nos spécifications favorites.

Compte tenu du fait que, dans le panel tronqué, l'expérience passée dans une même ville est en moyenne de 5 ans, l'impact des économies d'agglomération dynamiques liées aux compétences acquises localement est de  $5 \times 0.0006 = 0.003$ , une valeur qui reste néanmoins très inférieure à celle trouvée pour les économies d'agglomération statiques (0.019 dans la colonne 1 du Tableau 3, 0.011 dans la colonne 9 du Tableau 3). La densification d'une zone a donc un double effet : elle accroît de manière instantanée la productivité de ses salariés en causant des externalités technologiques et pécuniaires dans la zone, mais elle y favorise aussi les processus d'apprentissage, ce qui rend les salariés de la zone plus efficaces à plus long terme. Il ne semble exister en revanche aucun gain dynamique associé à la superficie (coefficient non significatif dans la colonne 9). Le potentiel marchand aurait quant à lui un effet légèrement négatif mais on observe par ailleurs que considérer cette variable accroît l'impact de la densité. Il est en fait possible que cette estimation présente un problème de multi-colinéarité (entre densité et potentiel marchand) associé à un pouvoir explicatif de ces variables bien plus faible que pour les effets statiques. Il faut donc considérer avec prudence les résultats des colonnes (9) à (14)<sup>40</sup>.

Contrairement au scénario précédent, on peut aussi considérer que l'expérience est transférable partout de la même façon, et qu'elle peut tout aussi bien être valorisée dans la ville où elle a été acquise que dans les autres villes. Cette hypothèse de portabilité parfaite revient à estimer :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + \sum_{c'} \mu_{c'}(e_{i,c',t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_{c'} &= Z_{c',\cdot} \psi^a + \zeta_{c'}, \quad \forall c'. \end{aligned} \quad (2)$$

Les colonnes (2) et (10) du Tableau 3 présentent le résultat de cette estimation. Le paramètre  $\psi^a$  est estimé à 0.0008 ou 0.0011, selon que l'on inclut ou non la superficie et le potentiel marchand dans la spécification. Une hausse de 1% de la densité d'une ville permet aux salariés de tirer parti de l'expérience qui y a été acquise pour accroître leur salaire annuel de 0.0008 % à 0.0011 % en moyenne par année d'expérience supplémentaire, qu'ils décident de rester ou non dans cette ville. Compte tenu du fait que l'expérience des salariés est en moyenne de 7 ans toutes zones confondues, l'impact des économies d'agglomération dynamiques est au minimum de  $7 \times 0.0008 = 0.0056$  et au maximum de  $7 \times 0.0011 = 0.0077$ .

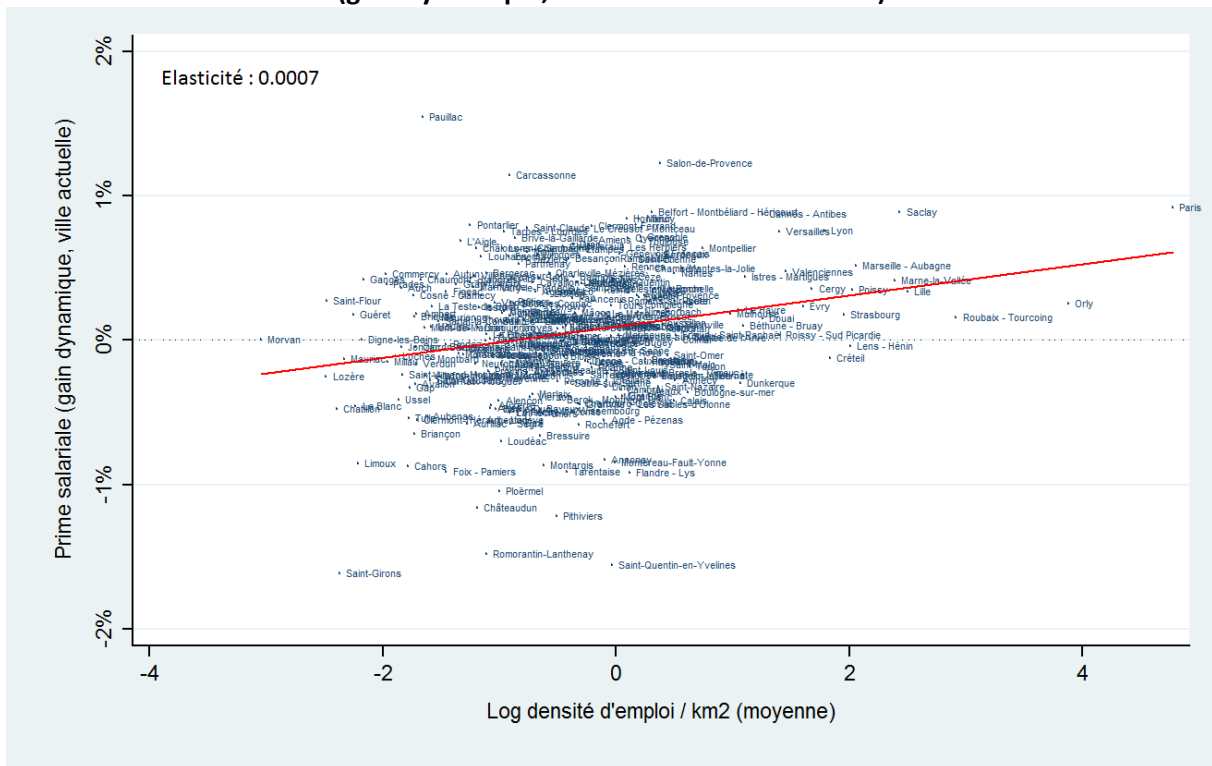
Il est cependant plus vraisemblable que, dans la réalité, il n'y ait pas valorisation à l'identique de l'expérience acquise dans la ville où les salariés travaillent actuellement, et de l'expérience acquise dans les autres villes. On peut donc vouloir estimer une spécification plus flexible que la précédente, et qui permet de distinguer ces deux effets :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} \\ &\quad + \mu_{c(i,t)}(e_{i,c(i,t),t} - e_{i,t}/C) + \sum_{c' \neq c(i,t)} \mu_{c'}(e_{i,c',t} - e_{i,t}/C) + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_c &= Z_{c,\cdot} \psi^i + \zeta_c, \\ \mu_{c'} &= Z_{c',\cdot} \psi^o + \zeta_{c'}, \quad \forall c' \neq c. \end{aligned} \quad (3)$$

Les résultats de ce scénario sont présentés dans les colonnes (3) et (11) du Tableau 3. Les disparités spatiales de salaire imputables aux effets d'apprentissage estimés dans la colonne (3) du Tableau 3 sont illustrées par les Figures 14 et 15.

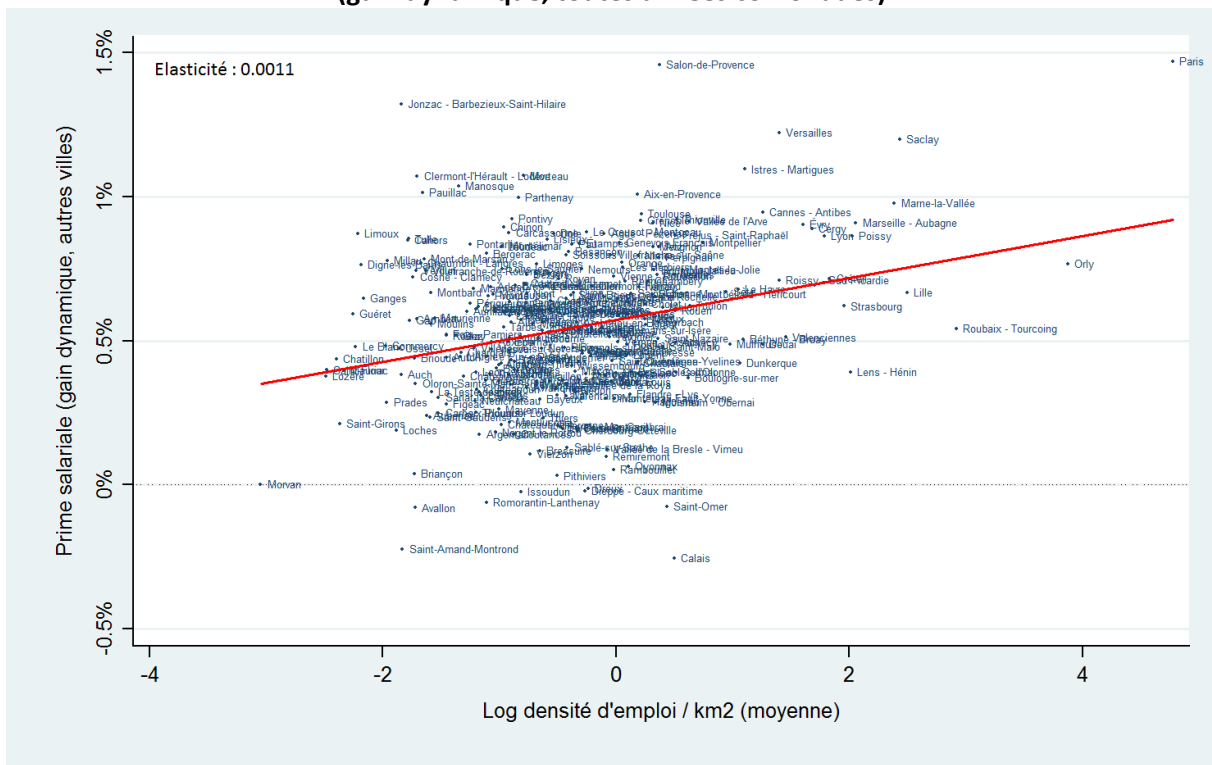
<sup>40</sup> Rappelons que notre étude est la première à estimer ce type de spécification. Les analyses existantes sur les villes espagnoles ou britanniques ne permettent pas d'effectuer des estimations de deuxième étape pour les effets dynamiques.

**Figure 14 : Prime salariale associée à l'expérience passée dans la zone d'emploi actuelle (gain dynamique, toutes années confondues)**



Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan, en moyenne), calculé à partir de l'exponentielle des effets fixes  $\mu_c$  estimés en colonne (3) du Tableau 3.

**Figure 15 : Prime salariale associée à l'expérience passée dans toutes les autres zones d'emploi (gain dynamique, toutes années confondues)**



Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan, en moyenne), calculé à partir de l'exponentielle des effets fixes  $\mu_{c'} \forall c' \neq c$  estimés en colonne (3) du Tableau 3.



Contrairement aux gains statiques, les gains dynamiques ne profitent pas uniquement aux salariés du Grand Paris. Sur les 20 villes offrant la plus forte valorisation de l'expérience (interne ou externe) figurent seulement quatre villes du Grand Paris : Paris, Versailles, Saclay et Marne-la-Vallée. La prime dont bénéficie un salarié à Paris, relativement à un salarié disposant des mêmes compétences dans le Morvan, la ville la moins dense du panel, n'excède pas 1,5 %. Certaines zones comme Pauillac, Carcassonne, Salon de Provence ou Istres-Martigues, disposent d'une prime à l'expérience similaire, voire supérieure aux zones d'emploi du Grand Paris, bien qu'elles soient bien moins denses.

Ici encore, la taille des villes a un impact positif très significatif, aussi bien sur la valorisation de l'expérience interne, que sur celle de l'expérience externe. Dans la colonne (3) par exemple, la valorisation de l'expérience interne est estimée à 0,0007 ( $\psi^i$ ), et celle de l'expérience externe à 0,0011 ( $\psi^o$ ). Une hausse de 1 % de la densité d'une ville permet donc à ses salariés d'accroître leur rémunération annuelle de 0,0007 % s'ils restent dans cette ville, et de 0,0011 % s'ils la quittent pour aller valoriser leur expérience ailleurs. Il semble donc que la mobilité au départ des zones denses soit payante, et permette aux salariés ayant acquis de l'expérience dans une grande métropole de la valoriser dans n'importe quelle autre ville, et d'en tirer une prime salariale annuelle plus élevée que dans leur ville d'origine. Il faut cependant nuancer ce résultat : l'expérience acquise localement est en moyenne bien plus élevée que l'expérience acquise via une mobilité (5 ans contre 1,25 en moyenne, toutes années du panel confondues). Les gains dynamiques tirés de la densification de la zone sont donc de  $0,0007 \times 5 = 0,0035$  en moyenne, si le salarié ne quitte pas sa ville, et de  $0,0011 \times 1,25 = 0,0013$  en moyenne, s'il la quitte.

Il faut également rester prudent dans l'interprétation que l'on fait des gains dynamiques engendrés par la mobilité des salariés, pour deux raisons. Tout d'abord, il se peut que les individus ne décident de migrer que si une entreprise leur offre un emploi mieux rémunéré que celui qu'ils occupent dans leur ville actuelle. Ce n'est donc pas uniquement parce que la ville d'origine est plus dense que l'expérience est mieux valorisée dans la ville de destination, c'est aussi parce que la ville de destination offre des salaires plus élevés aux individus changeant de ville : la causalité peut être inverse, et aller dans les deux sens. Nous reviendrons sur ce biais d'endogénéité dans la section suivante. Ensuite, les colonnes (4) et (12) du Tableau 3 nous amènent également à relativiser l'idée d'une plus grande portabilité de l'expérience externe par année, comparée à celle de l'expérience interne. Lorsqu'on estime les gains dynamiques en pondérant par le nombre d'actifs occupés présents dans chaque zone, pour revenir à la perspective individuelle déjà étudiée dans la colonne (9) du Tableau 1, les deux coefficients ne sont plus statistiquement différents, et le pouvoir explicatif du modèle augmente considérablement. Le  $R^2$  associé à l'estimation de  $\psi^i$  est égal à 0.66 dans la colonne (4) et à 0.71 dans la colonne (12), et il est bien plus élevé que celui associé à l'estimation de  $\psi^o$  (égal à 0.36 dans la colonne 4 et à 0.52 dans la colonne 10).

Une autre manière de pallier le problème associé au fait de devoir estimer un trop grand nombre de paramètres lorsqu'on raisonne à l'échelle de 297 villes, consiste à restreindre la dimension spatiale, de manière à pouvoir estimer tous les paramètres du modèle. C'est ce que nous faisons dans les colonnes (5) à (8) et (13) à (14) du Tableau 3, où l'inférence est réalisée sur 20 classes de villes définies selon leur appartenance aux vingtiles de la distribution des densités moyennes (calculées sur toutes les années disponibles dans le panel). On estime alors un système d'équations du type :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t} \theta + \beta_{c(i,t)} + \sum_{q'} \mu_{q',q(i,t)} (e_{i,q',t} - e_{i,t}/C) + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_{q',q} &= Z_{q',\cdot} \psi + Z_{q',\cdot} (Z_{q,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{q',q}, \end{aligned} \quad (4)$$

où  $Z_{q,\cdot}$  est le vecteur des caractéristiques locales moyennes dans la classe  $q$  définie par les vingtiles lui servant de bornes,  $Z_{\cdot,\cdot}$  est le vecteur des moyennes de ces moyennes, et  $\zeta_{q',q}$  est un bruit blanc.

Le paramètre  $\mu_{q',q}$  correspond à la valorisation dans une ville de classe  $q$  de l'expérience acquise dans une ville de classe  $q'$ . Ce gain peut être décomposé en deux éléments. Le paramètre  $\psi$  mesure la valorisation de l'expérience passée dans les villes de la classe  $q'$  pour un salarié travaillant dans la ville « moyenne », et  $\nu$  la valorisation additionnelle tirée du fait que la classe  $q$  correspondant à la ville dans laquelle le salarié travaille actuellement inclut des villes de densité supérieure à cette moyenne.

Le fait de restreindre la dimension géographique à 20 classes de villes permet d'estimer tous les paramètres  $\mu_{q',q}$ , qui sont au nombre de  $20 \times 20 = 400$ . Afin de comparer les résultats de cette inférence avec ceux mis en évidence dans le cadre de nos scénarii précédents, nous procédons en quatre étapes. Dans les colonnes (5) et (13) du Tableau 3, nous considérons que l'expérience acquise dans une ville de classe  $q$  n'est pas transférable de la même façon partout. L'expérience acquise dans une ville de la classe  $q$  est valorisée différemment, selon que le salarié travaille toujours dans une ville de la classe  $q$ , ou selon qu'il a migré vers une ville d'une autre classe  $q'$ . On a donc :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} \\ &\quad + \mu_{q(i,t)}(e_{i,q(i,t),t} - e_{i,t}/C) + \sum_{q' \neq q(i,t)} \mu_{q',q(i,t)}(e_{i,q',t} - e_{i,t}/C) + \epsilon_{i,t}, \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_q &= Z_{q,\cdot} \psi^i + \zeta_q, \\ \mu_{q',q} &= Z_{q',\cdot} \psi^o + Z_{q',\cdot} (Z_{q,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{q',q}, \quad \forall q' \neq q. \end{aligned} \quad (5)$$

Les résultats ne sont guère probants. L'estimateur du paramètre  $\psi^i$ , qui mesure la valorisation de l'expérience interne à la classe de ville du salarié, n'est jamais significatif. Rappelons tout de même qu'il n'est estimé que sur 20 observations, ce qui peut expliquer cette absence de significativité. En revanche, l'expérience acquise dans les villes des autres classes est transférable. Lorsque seule la densité est prise en compte pour mesurer la taille de la ville, la valeur estimée pour  $\psi^o$  est égale à 0.0016 (colonne 5). Cela signifie qu'une hausse de 1% de la taille des villes de la classe  $q'$  engendre une rémunération additionnelle de 0.0016 % pour les salariés qui y ont acquis de l'expérience, mais travaillent actuellement dans la ville « moyenne » de l'échantillon (c'est-à-dire celle disposant des caractéristiques  $Z_{\cdot,\cdot}$ ). La valorisation additionnelle tirée du fait que la ville actuelle appartient à une classe  $q$  de villes de densité supérieure à cette moyenne est estimée à 0.0002 %. Lorsque le potentiel marchand et la superficie sont pris en compte, la valeur estimée pour le coefficient  $\psi^o$  est plus élevée, à 0.0023. Mais les mêmes limites que celles évoquées plus haut (éventuels problèmes de multi-colinéarité) restent valables (colonne 13). La valeur de  $\nu$  n'est en revanche pas affectée.

La colonne (6) du Tableau 3 présente les élasticités correspondant à une version plus contrainte de la spécification précédente, dans laquelle on fait l'hypothèse qu'il n'existe aucune différence de portabilité entre les deux types d'expérience (locale ou externe), ce qui correspond au modèle :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t}\theta + \beta_{c(i,t)} + \sum_{q'} \mu_{q',q(i,t)}(e_{i,q',t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t} \\ \beta_c &= Z_{c,\cdot} \gamma + \eta_c, \\ \mu_{q',q} &= Z_{q',\cdot} \psi^a + Z_{q',\cdot} (Z_{q,\cdot} - Z_{\cdot,\cdot}) \nu + \zeta_{q',q}, \quad \forall q'. \end{aligned} \quad (6)$$

Les élasticités sont du même ordre de grandeur que les précédentes : la valorisation de l'expérience acquise dans une ville d'une autre classe (possiblement celle dans laquelle se trouve actuellement le salarié) est estimée à 0.0016 pour un salarié travaillant actuellement dans la ville « moyenne ». La valorisation additionnelle tirée du fait que la classe correspondant à la ville dans laquelle le salarié travaille actuellement inclut des villes de densité supérieure à cette moyenne est estimée à 0.002.

Les résultats présentés dans les colonnes (7), (8) et (14) du Tableau 3 permettent de tester la robustesse des estimations précédentes, toutes réalisées en deux étapes, à l'utilisation d'une inférence en une seule étape, qui a l'avantage de pondérer implicitement les observations au niveau zone par le nombre de travailleurs. On dispose alors de deux possibilités pour la mesure des économies d'agglomération statiques. Soit on conserve la logique à deux étapes pour ces gains, et on estime les gains dynamiques lors de la première étape :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t)} + \sum_{c'} \tilde{\mu}_{c',c(i,t)}(e_{i,c',t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \tilde{\epsilon}_{i,t},$$

$$\beta_c = Z_{c,\cdot}\gamma + \eta_c,$$

et on suppose en 1<sup>ère</sup> étape que  $\tilde{\mu}_{c,c} = Z_{c,\cdot}\psi^i$  et  $\tilde{\mu}_{c',c} = Z_{c',\cdot}\psi^o + Z_{c',\cdot}(Z_{c,\cdot} - Z_{\cdot,\cdot})\nu \forall c' \neq c$ . (7)  
Soit on réalise l'intégralité de l'estimation en une seule étape, auquel cas on a :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + Z_{c,\cdot}\gamma + \sum_{c'} \tilde{\mu}_{c',c(i,t)}(e_{i,c',t} - e_{i,t}/C) + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \tilde{\epsilon}_{i,t},$$

avec  $\tilde{\mu}_{c,c} = Z_{c,\cdot}\psi^i$  et  $\tilde{\mu}_{c',c} = Z_{c',\cdot}\psi^o + Z_{c',\cdot}(Z_{c,\cdot} - Z_{\cdot,\cdot})\nu, \forall c' \neq c$ . (8)

Dans les deux cas, la valorisation estimée pour l'expérience est très stable lorsqu'on considère uniquement la densité : la valeur de l'expérience interne ( $\psi^i$ ) est estimée à 0.0015, et celle de l'expérience externe ( $\psi^o$ ) à 0.0013. En revanche, lorsqu'on inclut aussi le potentiel marchand et la superficie, les gains dynamiques augmentent sensiblement : 0.0019 pour la valorisation de l'expérience interne, et 0.0025 pour l'expérience externe modulo les problèmes de co-linéarité évoqués plus haut

En conclusion, il est remarquable de constater que, quel que soit le scénario estimé dans le Tableau 3, les valeurs obtenues pour  $\gamma$ , qui mesure l'impact des économies d'agglomération statiques (toutes années du panel confondues), sont très proches de celles estimées dans la section 3.1. L'élasticité du salaire par rapport à la densité varie dans une fourchette très étroite de 0.018 à 0.022 selon les spécifications, lorsque le vecteur des caractéristiques locales  $Z_{c,\cdot}$  n'intègre que la densité moyenne de la zone. Elle baisse de 30% environ lorsqu'il inclut aussi la superficie et le potentiel marchand moyen (pour s'établir entre 0.011 et 0.015, selon les spécifications). L'élasticité du salaire au potentiel marchand est également très similaire à celle mise en évidence dans le Tableau 2 : elle est comprise dans une fourchette allant de 0.035 à 0.041, selon les spécifications. La superficie n'a pas un impact très stable sur les salaires, tout comme dans les estimations précédentes, mais c'est un résultat assez standard dans la littérature (cf. Combes et Lafourcade, 2012). Le fait de prendre en compte les gains dynamiques ne fait donc pas disparaître les gains statiques, bien au contraire. Leur ampleur est de plus extrêmement stable sur l'ensemble des spécifications estimées<sup>41</sup>.

Quant aux économies d'agglomération dynamiques, elles sont bien moins importantes que les économies d'agglomération statiques, sans pour autant être négligeables. De plus, elles sont transférables d'une ville à l'autre, avec un degré de portabilité différent selon la taille – et surtout la densité – des villes d'origine et de destination. Afin d'illustrer cette portabilité différenciée, on peut réaliser une estimation basée sur une typologie à cinq classes de villes définies selon leur appartenance aux quintiles de la distribution des densités moyennes de l'échantillon (calculées sur toutes les années disponibles dans le panel)<sup>42</sup>. Le Tableau 4 permet de visualiser la matrice des effets d'expérience obtenus dans le cadre de l'estimation de première étape du modèle (6).

<sup>41</sup> L'Annexe 7 de ce rapport présente les disparités salariales engendrées par les économies d'agglomération statiques, toutes années du panel confondues (colonne 3 du Tableau 3). La variabilité spatiale de la densité étant plus faible sur l'ensemble de la période 1993-2008, les écarts salariaux y sont un peu moins exacerbés qu'en 2008 (cf. Figure12).

<sup>42</sup> L'Annexe 8 de ce rapport présente cette typologie.

Comme on peut le voir à l'aide des valeurs indiquées en gras-violet, les 4 coefficients les plus élevés de la matrice se trouvent sur la ligne correspondant au dernier quintile d'origine (« q5 origine »). Si on prend comme référence l'expérience acquise dans les zones rurales du 1<sup>er</sup> quintile, les salariés valorisent bien mieux l'expérience acquise dans les grandes métropoles du dernier quintile et ce, quelle que soit la classe de destination, pour peu qu'elle contienne des villes suffisamment denses. En effet, les zones d'emploi rurales du premier quintile constituent l'exception à cette règle. L'expérience acquise dans les grandes agglomérations du dernier quintile y est moins transférable que celle acquise dans les petites et moyennes villes des 2<sup>ème</sup> et 3<sup>ème</sup> quintiles. On peut aussi observer que, pour un quintile de destination donnée, la valeur des paramètres tend à décroître de bas en haut, avec un effet de moins en moins marqué à mesure qu'on passe du 5<sup>ème</sup> au 1<sup>er</sup> quintile d'origine.

Pour une origine donnée, les résultats sont moins clairs en revanche, et les valeurs des paramètres ne décroissent pas forcément lorsqu'on se déplace de droite à gauche, i.e. des grandes métropoles (5<sup>ème</sup> quintile de destination) vers les petites (1<sup>er</sup> quintile de destination), ce qui corrobore certains effets mis en évidence dans le Tableau 3 : l'expérience acquise dans les classes d'origine incluant des villes disposant peu ou prou de la même densité que celles incluses dans la classe de destination n'implique pas que l'expérience est mieux valorisée dans la classe de destination, alors qu'on aurait pu s'attendre à ce que la similarité des villes d'origine et de destination en termes de densité favorise la transférabilité de l'expérience.

**Tableau 4 : Matrice des effets d'expérience pour 5 classes de densité moyenne**

<b>Destination Origine</b>	<b>q1</b>	<b>q2</b>	<b>q3</b>	<b>q4</b>	<b>q5</b>
<b>q2</b>	0.00178 (0.00065)	-0.00049 (0.00062)	-0.00133 (0.00093)	-0.00175 (0.00089)	<b>0.00119</b> <b>(0.00076)</b>
<b>q3</b>	0.00192 (0.00069)	0.00029 (0.00088)	-0.00151 (0.00065)	-0.00238 (0.00081)	<b>0.00043</b> <b>(0.00070)</b>
<b>q4</b>	0.00109 (0.00066)	0.00114 (0.00084)	-0.00047 (0.00083)	-0.00189 (0.00060)	<b>0.00101</b> <b>(0.00061)</b>
<b>q5</b>	0.00139 (0.00048)	<b>0.00261</b> <b>(0.00075)</b>	<b>0.00244</b> <b>(0.00076)</b>	<b>0.00212</b> <b>(0.00067)</b>	<b>0.00296</b> <b>(0.00052)</b>

Notes : (i) Les valeurs présentées dans ce tableau correspondent aux effets fixes  $\mu_{q',q}$  issus d'une estimation dans laquelle on considère 5 classes de villes ; (ii) q1 est la classe correspondant au premier quintile de densité moyenne, q2, la classe correspondant au 2<sup>ème</sup> quintile, et ainsi de suite ; (iii) Le quintile à l'origine est celui de la zone où l'expérience a été acquise, le quintile à la destination, celui où l'expérience est actuellement valorisée ; (iv) Les zones les moins denses sont prises comme référence, ce qui explique pourquoi il n'y a pas de premier quintile à l'origine ; (v) Le  $R^2$  de première étape est de : 0.28 ; (vi) Le gain statique lié à la densité estimé en seconde étape est de 0.020, et le  $R^2$  de seconde étape de 0.30.

### 3.3 Problèmes d'endogénéité et robustesse des estimations

Si les grandes métropoles permettent aux entreprises d'être plus productives, ces performances sont en partie capitalisées dans les salaires, dont le niveau élevé contribue en retour à rendre ces métropoles plus attractives aux yeux des salariés. Ce n'est donc pas seulement la taille des villes qui influence la productivité, mais aussi la productivité qui influence la taille des villes. Le coefficient  $\gamma$  capture les deux effets, alors qu'on souhaiterait les identifier séparément. Malheureusement, les méthodes économétriques simples comme les MCO ne parviennent pas à distinguer le sens de la causalité. Or cette distinction est cruciale pour les politiques publiques. Si la causalité va de la productivité vers la densité, une densification des emplois suscitée par des exonérations fiscales zonées par exemple n'aura pas d'impact sur la productivité, et elle aura des effets si la causalité va dans l'autre sens. La direction de la causalité est ici fondamentale.

Ce *biais de causalité inverse* peut être neutralisé par l'utilisation de la technique des variables instrumentales. L'objectif est de trouver une ou des variables, "instruments", suffisamment corrélées avec la variable explicative suspectée d'être endogène pour en être un bon « prédicteur », mais pas corrélée avec les variables omises incluses dans le terme d'erreur et n'ayant pas d'influence directe sur la variable dépendante (le salaire). Un exemple simple permet de comprendre la technique des variables instrumentales. Supposons que l'on cherche à estimer l'effet causal de la cigarette sur la santé, une corrélation négative entre l'état de santé et le nombre de cigarettes fumées n'implique pas nécessairement que « fumer nuit gravement à la santé », car d'autres variables potentiellement omises, comme le revenu ou le statut socioprofessionnel, peuvent affecter simultanément la santé et la consommation de cigarettes. La méthode des variables instrumentales consiste à trouver une variable qui n'affecte la santé qu'à travers la consommation de cigarettes, comme les droits d'accise sur le tabac. Ces taxes ne peuvent avoir un effet sur la santé que parce qu'elles réduisent la consommation de cigarettes, toutes choses égales par ailleurs. Une corrélation positive entre le niveau des taxes (qui font baisser la consommation) sur le tabac et l'état de santé signifie donc que fumer nuit effectivement à la santé, et que la relation entre les deux variables est bien causale.

Lorsqu'on évalue l'impact des économies d'agglomération sur le salaire nominal, on procède de la même façon. Au lieu d'utiliser comme variable explicative le vecteur des caractéristiques locales de la ville à la date  $t$ ,  $Z_{c,t}$ , on utilise la prédiction de cette variable par une série d'instruments. Tant que ces instruments n'ont pas d'influence directe sur la variable expliquée, et sont bien corrélés aux caractéristiques locales actuelles, ils captent l'impact direct de ces caractéristiques. On utilise ce pouvoir prédictif pour neutraliser les biais de causalité circulaire et de variables omises. Le problème est bien sûr de trouver de « bons » instruments pour réaliser l'inférence. Ces instruments doivent être corrélés avec la variable endogène,  $Z_{c,t}$ , mais pas avec les variables omises, capturées par le résidu, auquel cas les instruments seraient eux-mêmes endogènes. Mais comme la variable endogène est corrélée avec ces variables omises, une tension apparaît clairement entre les deux objectifs.

En règle générale, les caractéristiques locales retardées de plusieurs décennies, voire de plusieurs siècles, avant la période d'étude sont de bons candidats à l'instrumentation. Les caractéristiques locales passées n'ont en effet pas d'influence directe sur la productivité ou le salaire actuels (à condition que les caractéristiques permanentes des villes, et notamment leur géographie physique, soient considérées comme variables de contrôles), tout en restant bien corrélées avec les caractéristiques locales actuelles. Elles peuvent ainsi capter l'impact direct des caractéristiques actuelles sur les salaires. Dans la suite, nous utilisons les populations urbaines en 1831 et 1881, ainsi que les densités de population pour ces mêmes années, comme instruments de la densité. L'utilisation de ces variables retardées implique cependant d'exclure 24 zones d'emploi qui n'étaient pas urbanisées en 1831 et 1881, et pour lesquelles nous ne disposons pas d'instrument puisque nous n'en connaissons pas les populations urbaines au 19<sup>ème</sup> siècle. L'inférence n'est donc réalisée que sur les 273 zones d'emploi restant dans l'échantillon.

Bien que les variables retardées soient souvent privilégiées, il est cependant possible de trouver d'autres bons instruments. Combes, Duranton, Gobillon et Roux (2010) utilisent par exemple la géologie des sols (en plus des variables retardées) pour instrumenter la densité. La géologie ayant conditionné l'installation des populations il y a plusieurs siècles, elle est corrélée aux densités actuelles, mais moins corrélée avec les variables omises susceptibles d'influencer la variable expliquée que les densités retardées. Dans la suite, nous utilisons la minéralogie des sols, le contenu en eau des sous-sols et de la couche superficielle du sol, comme instruments géologiques. Pour instrumenter le potentiel marchand, nous utilisons une mesure de la périphéralité des zones d'emploi, calculée comme la moyenne de la distance géodésique de la zone à toutes les autres zones.

D'un point de vue technique, l'instrumentation s'effectue en plusieurs temps. Tout d'abord, étant donné que les instruments ne varient pas dans le temps, les variations temporelles des effets locaux ne peuvent plus être utilisées pour identifier l'impact des variables locales. On commence donc par moyenner sur les 6 années considérées dans le panel les effets fixes locaux estimés, ainsi que les caractéristiques locales. Puis on estime par les MCO l'effet d'un vecteur  $W_c$ , incluant les caractéristiques locales « retardées », la géologie des sols et la périphéralité des zones, sur les caractéristiques locales moyennes  $Z_{c,\cdot}$  :

$$Z_{c,\cdot} = W_c \alpha + \eta_c.$$

L'estimateur obtenu pour le coefficient  $\alpha$  permet alors de calculer un prédicteur des caractéristiques locales moyennes, noté  $\widehat{Z}_{c,\cdot}$  :

$$\widehat{Z}_{c,\cdot} = W_c \hat{\alpha}.$$

On estime ensuite par les MCO l'effet de cette prédiction sur la variable expliquée.

Lorsqu'on veut mesurer uniquement les économies d'agglomération statiques, on estime le modèle 4 présenté dans la section 3.1 :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t} \theta + \beta_{c(i,t),t} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t} \\ \beta_{c,\cdot} &= \widehat{Z}_{c,\cdot} \gamma + \eta_c, \end{aligned}$$

où  $\beta_{c,\cdot}$  est la moyenne des effets fixes estimés dans les colonnes (4) et (9) du Tableau 1, selon que l'on exclut des caractéristiques individuelles en 1<sup>ère</sup> étape l'expérience du salarié (panel non tronqué) ou pas (panel tronqué).

Lorsqu'on veut mesurer simultanément les gains statiques et dynamiques, on estime le modèle 3 présenté dans la section 3.2 :

$$\begin{aligned} \log \omega_{i,t} &= u_i + X_{i,t} \theta + \beta_{c(i,t)} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} \\ &+ \mu_{c(i,t)} (e_{i,c(i,t),t} - e_{i,t}/C) + \sum_{c' \neq c(i,t)} \mu_{c'} (e_{i,c',t} - e_{i,t}/C) + \epsilon_{i,t}, \end{aligned}$$

$$\beta_c = \widehat{Z}_{c,\cdot} \gamma + \eta_c,$$

$$\mu_c = \widehat{Z}_{c,\cdot} \psi^i + \zeta_c,$$

$$\mu_{c'} = \widehat{Z}_{c',\cdot} \psi^o + \zeta_{c'}, \quad \forall c' \neq c,$$

où  $\beta_c$ ,  $\mu_c$  et  $\mu_{c'}$  sont les effets fixes estimés dans les colonnes (3) et (10) du Tableau 3.

Dans les deux cas, les coefficients obtenus pour  $\gamma$ ,  $\psi^i$  et  $\psi^o$  sont les « bons » estimateurs de l'impact de la taille de la ville, car ils sont purgés des biais de variables omises et de causalité inverse via l'instrumentation.

Les Tableaux 5 et 6 présentent les résultats des estimations réalisées avec la technique des variables instrumentales. Dans la colonne (1) du Tableau 5, la seule caractéristique locale considérée est la densité. Lorsque cette dernière est instrumentée par les populations retardées, l'élasticité du salaire par rapport à la densité est estimée à 0.013, soit une baisse de près de 40% par rapport au coefficient estimé dans la colonne 1 du Tableau 1 (0.021). En revanche, le coefficient de la densité explose lorsqu'elle est instrumentée par la géologie (0.036 dans la colonne 2, 0.038 dans la colonne 3). Mais à l'évidence, celle-ci constitue un très mauvais instrument, contrairement aux populations retardées.

**Tableau 5 : Economies d'agglomération statiques liées à la taille globale des villes (instrumentée)**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel tronqué			oui	oui			oui	oui			oui	oui
Densité	0.013	0.036	0.038	0.010	0.010	0.011	0.007	0.008	0.012	0.010	0.011	0.009
Potentiel marchand	-	-	-	-	0.049	0.049	0.050	0.049	0.032	0.040	0.039	0.042
Superficie	-	-	-	-	0.005	0.005	ns	ns	ns	ns	ns	ns
Aménités en 2 <sup>ème</sup> étape					oui	oui	oui	oui	oui	oui	oui	oui
Densité instrumentée	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui
Potentiel marchand instrumenté									oui	oui	oui	oui
Cragg-Donald (F-Stat.)	43.2	5.8	5.8	116.5	34.5	52.1	34.5	52.1	28.5	12.9	28.5	12.9
Test de suridentification (p-value)	0.10	0.17	0.28	0.27	0.36	0.05	0.32	0.05	0.36	0.52	0.24	0.27
<b>Instruments utilisés</b>												
Population urbaine 1831	oui	non	non	oui	non	non	non	non	oui	non	oui	non
Densité de population urbaine 1831	non	non	non	non	oui	oui	oui	oui	non	oui	non	oui
Population urbaine 1881	oui	non	non	oui	non	non	non	non	oui	non	oui	non
Densité de population urbaine 1881	non	non	non	non	non	oui	non	oui	non	non	non	non
Périphéralité	non	non	non	non	non	non	non	non	oui	non	oui	non
Contenu en eau du sous-sol	non	non	non	non	oui	non	oui	non	non	oui	non	oui
Contenu en eau de la couche superficielle	non	oui	oui	non	non	non	non	non	non	non	non	non
Minéralogie du sous-sol	non	non	non	non	non	oui	non	oui	non	non	non	non

Notes : (i) Estimations en deux étapes : la variable expliquée en seconde étape est la moyenne des effets fixes estimés dans les colonnes (4) ou (9) du Tableau 1 (cf. section 3.1, modèle 4) ; (ii) L'estimation est réalisée sur les 273 zones d'emploi urbaines pour lesquelles on dispose des populations au 19<sup>ème</sup> siècle, et qui peuvent donc être instrumentées ; (iii) Les instruments ne sont pas faibles lorsque la statistique de Cragg-Donald est supérieure à 10, et ils passent le test de sur-identification quand la p-value associée est supérieure à 0.05 ; (iv) « ns » signifie que le coefficient estimé n'est pas significativement différent de zéro ; (v) Les résultats détaillés de l'instrumentation sont disponibles sur simple demande.

**Tableau 6 : Economies d'agglomération dynamiques liées à la taille globale des villes (instrumentée)**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Effet statique	oui	oui	oui				
Effet dynamique, ZE actuelle				oui	oui		
Effet dynamique, autres ZE						oui	oui
Densité	0.011	0.009	0.009	0.0011	0.0011	0.0019	0.0016
Potentiel marchand		0.051	0.042	ns	ns	ns	ns
Superficie		0.006	ns	0.0007	0.0006	0.0016	0.0018
Amenité en 2 <sup>ème</sup> étape		oui	oui	oui	oui	oui	oui
Densité instrumentée	oui	oui	oui	oui	oui	oui	oui
Potentiel marchand instrumenté			oui		oui		oui
Cragg-Donald (F-stat.)	116.5	34.5	12.9	169.7	29.3	34.5	29.3
Test de sur-identification (p-value)	0.95	0.25	0.27	0.19	0.27	0.19	0.08
Densité urbaine 1831	non	non	non	oui	non	oui	non
Densité de population urbaine 1831	oui	oui	oui	non	oui	non	non
Population urbaine 1881	oui	non	non	oui	non	oui	oui
Periphéralité	non	non	non	oui	non	oui	oui
Contenu en eau du sous-sol	non	oui	oui	non	oui	non	non

Notes : (i) Estimations en deux étapes : la variable expliquée en seconde étape est la moyenne des effets fixes estimés dans les colonnes (3) ou (10) du Tableau 3 (cf. section 3.2, modèle 3) ; (ii) L'estimation est réalisée sur les 273 zones d'emploi urbaines pour lesquelles on dispose des populations au 19<sup>ème</sup> siècle, et qui peuvent donc être instrumentées ; (iii) Les instruments ne sont pas faibles lorsque la statistique de Cragg-Donald est supérieure à 10, et ils passent le test de sur-identification quand la p-value associée est supérieure à 0.05 ; (iv) « ns » signifie que le coefficient estimé n'est pas significativement différent de zéro ; (v) Les résultats détaillés de l'instrumentation sont disponibles sur simple demande.



En effet, la statistique de Cragg-Donald<sup>43</sup> est ici de 5.8, ce qui est très inférieur au seuil requis pour considérer que l'instrument a un fort pouvoir prédictif ou autrement dit, qu'il n'est pas faible (le minimum requis pour la statistique est de 10).

Lorsqu'on instrumente la densité par les populations urbaines, mais qu'on réalise l'estimation à partir du panel tronqué (colonne 4), le coefficient de la densité baisse à 0.010, soit une chute de 50% par rapport au coefficient de 0.020 mis en évidence dans la colonne (8) du Tableau 1. Il est clair que l'estimation par les MCO surestime l'impact de la densité sur le salaire, comme la littérature l'a déjà montré pour la France (voir Combes, Duranton et Gobillon, 2008 et Combes, Duranton, Gobillon et Roux, 2010).

Deux raisons peuvent expliquer ce biais. La première tient à l'omission des autres caractéristiques locales susceptibles d'influencer le salaire, comme le potentiel marchand, la superficie ou les aménités. La seconde est liée au biais de causalité circulaire évoqué plus haut. L'instrumentation par les variables retardées a l'avantage de purger l'estimateur des deux biais. Il est donc important d'inclure en parallèle de la densité, le potentiel marchand, la superficie et les aménités. Lorsqu'on considère les trois caractéristiques locales et les aménités simultanément, mais qu'on n'instrumente que la densité, le coefficient de la densité baisse significativement, quels que soient les instruments utilisés (populations ou densités de population retardées, géologie). Elle passe de 0.013 à 0.010 ou 0.011 (colonnes 5 et 6) si on utilise le panel non tronqué, et de 0.010 à 0.007 ou 0.008 (colonnes 7 et 8) si on utilise le panel tronqué. Cependant, les instruments sont parfois à la limite de la validité, comme l'illustre la faible p-value du test de sur-identification<sup>44</sup>.

Toutefois, le potentiel marchand est une variable potentiellement aussi endogène que la densité, à partir de laquelle il est d'ailleurs construit (en utilisant la densité des zones voisines). Il est donc aussi préférable de l'instrumenter. C'est ce que nous faisons dans les quatre dernières colonnes du Tableau 5. Quels que soient les instruments (variables historiques ou géologie) et le panel (non tronqué ou tronqué) utilisés, l'élasticité du salaire par rapport à la densité est maintenant très stable, et s'établit autour de 0.010 environ, soit près de 20 % de moins que la valeur reportée dans la colonne (5) du Tableau 2. Le biais d'endogénéité associé à l'estimation des économies d'agglomération statiques est donc du même ordre de grandeur que celui trouvé par Combes, Duranton et Gobillon (2008) et Combes, Duranton, Gobillon et Roux (2010).

Les valeurs reportées dans le Tableau 6 confirment l'existence d'un biais d'endogénéité plus important pour les économies d'agglomération statiques, lorsqu'elles sont évaluées en tenant compte des gains dynamiques : le coefficient estimé pour la densité est de 0.011 (colonne 1) lorsqu'on ne tient pas compte du potentiel marchand et de la superficie, et de 0.009 (colonne 2), lorsqu'on les prend en compte. C'est donc cette fois 50% de moins que la valeur de 0.021 reportée dans la colonne (3) du Tableau 3. En revanche, la prime à l'expérience interne engendrée par une densification de 1% des villes est estimée à 0.0011, que l'on instrumente ou pas le potentiel marchand. Cette valeur est donc la même que celle reportée dans la colonne (10) du Tableau 3. L'élasticité de la prime à l'expérience externe par rapport à la densité n'est pas non plus très affectée par l'instrumentation : le coefficient estimé dans la colonne (7) du Tableau 6 (0.0016) est très similaire de celui reportée dans la colonne (10) du Tableau 3 (0.0017). Il semble donc que les économies d'agglomération dynamiques ne soient entachées d'aucun biais de causalité circulaire, contrairement aux économies d'agglomération statiques.

---

<sup>43</sup> La statistique de Cragg-Donald permet d'effectuer un test sur la nullité du coefficient associé à un instrument dans la régression de première étape (qui consiste à expliquer la variable endogène par les instruments) lorsqu'il est ajouté aux autres instruments, et donc de tester la valeur prédictive additionnelle de la variable endogène par l'instrument choisi.

<sup>44</sup> Le test de sur-identification permet de tester l'hypothèse que le terme d'erreur n'est pas corrélé avec les instruments, auquel cas ces derniers sont valides.

### 3.4 Des économies d'agglomération similaires pour toutes les villes

Comme nous l'avons souligné dans la première partie de ce rapport, pour pouvoir identifier ce qui, dans la prime salariale bénéficiant aux salariés des grandes métropoles, revient aux aptitudes et compétences propres de ces salariés, ou aux économies d'agglomération statiques et dynamiques, il faut que le panel contienne beaucoup de travailleurs mobiles au cours du temps, et que l'origine et la destination de ces migrants présente suffisamment de variabilité spatiale. Une identification des économies d'agglomération sur le seul périmètre du Grand Paris n'est donc malheureusement pas possible, car il n'y a pas suffisamment de zones d'emploi pour estimer ces effets avec précision. On ne peut donc pas savoir si les effets d'agglomération dont bénéficie la métropole du Grand Paris sont plus ou moins grands que ceux dont bénéficient d'autres métropoles. Cela n'est cependant pas fondamentalement problématique pour deux raisons. La première est que les zones d'emploi du Grand Paris ne sont pas toutes homogènes en matière de densité d'emploi. Comme on peut le voir dans l'Annexe 8, treize d'entre-elles appartiennent à la classe définie par les bornes du dernier quintile de densité moyenne, mais sept autres appartiennent à d'autres classes : Provins se situe dans la classe définie par le deuxième quintile, Coulommiers, Étampes et Nemours dans celle définie par le troisième, Saint-Quentin-en-Yvelines, Rambouillet et Montereau-Fault-Yonne dans celle définie par le quatrième. Il n'y a donc pas de raison a priori que les économies d'agglomération ont la même intensité pour toutes les zones d'emploi du Grand Paris.

La seconde raison est qu'on peut tout à fait mesurer des effets d'agglomération spécifiques à des classes de villes définies selon leur appartenance à différents quantiles de densité moyenne (calculés sur l'ensemble des années disponibles dans le panel). Puisque chaque classe contient des villes dont la taille est peu ou prou similaire, l'élasticité du salaire par rapport à la taille calculée sur l'ensemble des villes de la classe reflète bien le comportement de chaque ville au sein de la classe. On peut ainsi tester l'égalité des coefficients estimés entre classes, et voir si l'ampleur des économies d'agglomération varie d'une classe à l'autre sans que cela soit problématique.

C'est ce que nous faisons dans le Tableau 7, qui présente le résultat de plusieurs estimations basées sur la classification par quintile de densité moyenne utilisée plus haut. Dans la colonne (1), nous considérons un modèle basique dans lequel on fait l'hypothèse qu'il n'y a pas d'économies d'agglomération dynamiques, que les économies d'agglomération statiques ne sont véhiculées que par la densité d'emploi, et que leur ampleur est susceptible de varier selon le quintile de densité moyenne auquel appartient chaque ville. On a donc :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \sum_{q=1}^5 (Z_{c(i,t),t} \times q_{c(i,t)}) \gamma_q + \epsilon_{i,t},$$

où  $q_c$  est une indicatrice permettant de savoir à quel quintile de densité moyenne appartient la zone d'emploi  $c$ .

Comme on peut le constater, les coefficients estimés pour chaque quintile ne sont pas significativement différents les uns des autres. De surcroît, ils sont extrêmement proches de l'effet « moyen » estimé dans la colonne (2) du Tableau 1 (0.018). Le fait d'inclure le potentiel marchand et la superficie en parallèle de la densité ne change pas beaucoup les choses. Les résultats présentés dans la colonne (2) montrent en effet que l'élasticité du salaire par rapport à la densité, lorsqu'elle est significativement différente de zéro, est la même pour tous les quintiles, et sa valeur (0.011) très proche de la valeur moyenne estimée dans la colonne (2) du Tableau 2 (à 0.014). L'élasticité du salaire par rapport au potentiel marchand ne varie pas sensiblement pour les quatre premiers quintiles de densité (0.027 environ), mais elle est considérablement plus grande pour le dernier quintile (0.046), ce qui semble indiquer que l'impact du potentiel marchand est plus fort dans la classe des villes les plus denses.

**Tableau 7 : Economies d'agglomération liées à la taille globale des villes par quintile de densité moyenne (sur toutes les années du panel)**

Régression	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimation en deux étapes			oui	oui	oui	oui	oui	oui
Effets dynamiques, ZE actuelle / Autres ZE							oui	oui
Pondéré en 2 <sup>ème</sup> étape					oui	oui		oui
R <sup>2</sup> 1 <sup>ère</sup> étape	0.27	0.26	0.28	0.28	0.28	0.28	0.28	0.28
R <sup>2</sup> 2 <sup>ème</sup> étape – Effets statiques	-	-	0.26	0.43	0.74	0.83	0.31	0.70
R <sup>2</sup> 2 <sup>ème</sup> étape – Effets dynamiques, ZE actuelle	-	-	-	-	-	-	0.13	0.68
R <sup>2</sup> 2 <sup>ème</sup> étape – Effets dynamiques, autres ZE	-	-	-	-	-	-	0.07	0.36
<i>Effets statiques</i>								
Densité, Q1	0.018	0.011	0.023	0.024	0.023	ns	0.022	0.022
Densité, Q2	0.019	0.011	0.025	0.008	0.025	-0.017	0.024	0.023
Densité, Q3	0.017	ns	0.023	0.007	0.024	-0.012	0.022	0.022
Densité, Q4	0.018	0.012	0.024	0.019	0.025	ns	0.023	0.022
Densité, Q5	0.018	0.011	0.026	0.007	0.027	0.015	0.025	0.023
Potentiel marchand, Q1		ns		0.021		ns		
Potentiel marchand, Q2		0.027		0.047		0.032		
Potentiel marchand, Q3		0.025		0.044		0.028		
Potentiel marchand, Q4		0.029		0.031		0.026		
Potentiel marchand, Q5		0.046		0.043		0.048		
Superficie, Q1		-0.004		-0.005		-0.011		
Superficie, Q2		ns		ns		-0.007		
Superficie, Q3		ns		ns		-0.005		
Superficie, Q4		0.003		ns		ns		
Superficie, Q5		0.009		0.005		0.001		

---

*Effets dynamiques*

Densité, ZE actuelle, Q1	0.0009	0.0013
Densité, ZE actuelle, Q2	0.0010	0.0015
Densité, ZE actuelle, Q3	0.0012	0.0016
Densité, ZE actuelle, Q4	0.0013	0.0017
Densité, ZE actuelle, Q5	0.0011	0.0017
Densité, autres ZE, Q1	0.0012	0.0013
Densité, autres ZE, Q2	0.0014	0.0014
Densité, autres ZE, Q3	0.0014	0.0015
Densité, autres ZE, Q4	0.0014	0.0013
Densité, autres ZE, Q5	0.0014	0.0014

---

Notes : (i) Dans les colonnes (3) et (4), les variables expliquées en seconde étape sont les effets fixes estimé dans les colonnes (8) du Tableau 1 et (2) du Tableau 2 (cf. section 3.1, modèle 1) ; dans la colonne (7), ce sont les effets fixes estimés dans la colonne (3) du Tableau 3 (cf. section 3.2, modèle 3) ; (ii) « ns » signifie que le coefficient estimé n'est pas significatif.

Lorsqu'on estime le modèle à deux étapes de manière à gommer le biais de variables omises en première étape :

$$\log \omega_{i,t} = u_i + X_{i,t}\theta + \beta_{c(i,t),t} + W_{c(i,t),k(i,t),t} \phi_{k(i,t)} + \vartheta_{k(i,t),t} + \epsilon_{i,t},$$

$$\beta_{c,t} = \sum_{q=1}^5 (Z_{c,t} \times q_c) \gamma_q + u_t + \eta_{c,t},$$

on retrouve des résultats très similaires aux précédents, mais avec des estimateurs plus précisément estimés. Dans la colonne (3) par exemple, où on considère uniquement la densité, les coefficients estimés pour cette variable ne diffèrent pas significativement selon les quintiles.

Lorsqu'on inclut aussi dans la spécification le potentiel marchand et la superficie (colonne 4), les élasticités du salaire par rapport à la densité et au potentiel marchand diffèrent plus significativement selon les quintiles. Mais ce résultat n'est pas robuste à la pondération des estimations de seconde étape par le nombre d'individus appartenant à chaque classe, comme l'illustrent les colonnes (5) et (6). Enfin, si on estime simultanément les économies d'agglomération statiques et dynamiques véhiculées par la seule densité, comme nous le faisons dans la colonne (7), on n'observe aucune différence significative selon les quintiles, ni pour les économies d'agglomération statiques (l'élasticité du gain statique par rapport à la densité oscille autour de 0.022 pour tous les quintiles), ni pour les économies d'agglomération dynamiques, que celles-ci soient liées à l'expérience interne (dont l'élasticité oscille autour de 0.0011, une valeur plus élevée que l'effet moyen reporté dans la colonne 3 du Tableau 3, qui est de 0.0007), ou à l'expérience externe (dont l'élasticité oscille autour de 0.0014, également plus élevée que celle reportée dans la colonne 3 du Tableau 3, à savoir 0.0011). Le fait de pondérer (colonne 8) ne change pas ce constat, mais permet de retrouver des estimateurs des économies d'agglomération dynamiques plus proches des valeurs moyennes reportées dans la colonne (4) du Tableau 3.

En conclusion, les économies d'agglomération ne semblent pas différentes d'une classe de densité à l'autre, et en particulier pas plus fortes pour la classe des villes les plus denses. Les élasticités « moyennes » estimées dans les sections précédentes sont donc très représentatives du comportement de toutes les villes de notre panel.

## Conclusion

L'objectif de ce rapport était d'estimer ce qui, dans la prime salariale dont bénéficient les travailleurs des grandes métropoles françaises, provient de leurs compétences propres et des économies d'agglomération statiques et dynamiques engendrées par la taille des villes. Les estimations ont été réalisées sur un panel composé de 3 133 573 observations, et couvrant environ 500 000 salariés des secteurs privé et semi-public pour les six années constituant ce panel (1993, 1996, 1999, 2002, 2005 et 2008). Cet échantillon constitue la plus grande source de variabilité individuelle, temporelle, géographique et sectorielle utilisée à ce jour pour estimer l'impact des économies d'agglomération.

Les résultats de notre étude sont riches d'enseignement. Tout d'abord, les disparités de salaires entre villes sont notables, même pour des travailleurs disposant de compétences similaires. Par exemple, un actif occupé travaillant à Paris dispose d'un salaire supérieur de 26 % à celui d'un individu strictement identique<sup>45</sup>, mais travaillant dans le Morvan, la zone d'emploi la moins dense du panel en 2008. D'une manière générale, à compétences données, les primes dont bénéficient les salariés du Grand Paris sont assez remarquables : +23 % à Saclay ou Orly, +20 % à Versailles, Étampes, Roissy-sud-Picardie, Mantes-la-Jolie, Évry, Cergy, Poissy, Créteil, Marne-la-Vallée ou Melun, +17 % à Meaux ou Saint-Quentin-en-Yvelines, +15 % à Coulommiers ou Provins. À compétences données, les primes associées aux autres grandes métropoles françaises sont plus faibles : par exemple +15 % à Lyon, +14% à Nice ou à Grenoble, +13 % à Marseille-Aubagne, +12 % à Strasbourg ou Lille, ou +10 % à Bordeaux. Les zones d'emploi rurales et peu denses du centre (Le Blanc, Montluçon, Issoire, Figeac, Cahors, Issoudun, Lozère) et du littoral breton (Loudéac, Quimper, Ploërmel) disposent des primes les plus faibles de toute la France.

Ces disparités, qui ne sont pas imputables au tri spatial des qualifications et aux compétences propres des travailleurs, permettent d'estimer la part de la prime salariale imputable à la taille des villes. Si on considère comme unique mesure de cette taille la densité d'emploi, l'intensité des économies d'agglomération statiques est estimée à 0.02. Cette valeur, même si elle est faible, induit des disparités géographiques de salaire assez substantielles, car les écarts de densité entre villes françaises sont très forts. Par exemple, si on compare les deux extrêmes de la distribution des densités, Paris et Morvan, le différentiel de productivité imputable à l'écart de densité entre ces deux villes est de  $(2660^{0.02} - 1) \times 100 = 17\%$  environ. Si on compare les zones d'emploi situées aux bornes des premier et dernier quintiles de densité, à savoir Périgueux et Belfort-Montbéliard-Héricourt, l'écart est en revanche bien plus faible, de  $(4^{0.02} - 1) \times 100 = 3\%$  environ.

L'impact de la densité est quelque peu atténué si on considère deux mesures additionnelles de la taille des villes, à savoir leur superficie et surtout, leur potentiel marchand qui capte la possibilité qu'elles ont de recourir aux marchés externes les plus proches en sus de leur marché local. La prise en compte de ces deux variables fait baisser l'élasticité de la prime salariale par rapport à la densité à 0.013. Doubler la densité d'une zone permet donc d'y augmenter le salaire moyen de  $(2^{0.013} - 1) \times 100 = 1\%$ , à superficie et potentiel marchand donnés. Cependant, l'accès aux marchés externes a un impact sur la prime salariale trois fois supérieur à celui de la densité. Le fait de doubler le potentiel marchand d'une ville y accroît la rémunération moyenne de ses salariés de  $(2^{0.038} - 1) \times 100 = 2,5\%$ , toutes choses égales par ailleurs. Si on compare les deux extrêmes de la distribution des potentiels marchands, Orly et Bayonne, le différentiel de productivité imputable à l'écart de potentiel marchand entre ces deux villes est de  $(14^{0.038} - 1) \times 100 = 10,5\%$  environ.

---

<sup>45</sup> C'est-à-dire disposant des mêmes caractéristiques individuelles observables et inobservables.

Contrairement aux gains statiques, les gains dynamiques ne profitent pas uniquement aux salariés du Grand Paris. Sur les 20 villes offrant la plus forte valorisation de l'expérience - interne ou externe à la ville - figurent seulement quatre zones d'emploi du Grand Paris : Paris, Versailles, Saclay et Marne-la-Vallée. La prime salariale à l'expérience dont bénéficie un travailleur à Paris, relativement à un salarié disposant des mêmes compétences dans le Morvan, n'excède pas 1,5 %. Certaines zones d'emploi comme Pauillac, Carcassonne, Salon-de-Provence ou Istres-Martigues, disposent d'une prime à l'expérience similaire, voire supérieure aux zones d'emploi du Grand Paris, bien qu'elles soient bien moins denses.

Ici encore pourtant, la taille des villes a un impact positif très significatif, aussi bien sur la valorisation de l'expérience interne, que sur celle de l'expérience externe. Une hausse de 1 % de la densité d'une ville permet à ses salariés d'accroître leur rémunération de 0.0007 % par année s'ils restent dans cette ville, et de 0.0011 % par année s'ils la quittent pour aller valoriser leur expérience ailleurs. Il semble donc que la mobilité au départ des zones denses soit payante, et permette aux salariés ayant acquis de l'expérience dans une grande métropole de la valoriser dans n'importe quelle autre ville, et d'en tirer une prime salariale annuelle plus élevée que dans leur ville d'origine. Il faut cependant nuancer ce résultat : l'expérience acquise localement est en moyenne bien plus élevée que l'expérience acquise via une mobilité (5 ans contre 1.25 en moyenne, toutes années du panel confondues). Les gains dynamiques tirés de la densification de la zone sont donc de  $0.0007 \times 5 = 0.0035$  en moyenne, si le salarié ne quitte pas sa ville, et de  $0.0011 \times 1.25 = 0.0013$  en moyenne, s'il la quitte.

Les économies d'agglomération dynamiques, si elles sont bien moins importantes que les économies d'agglomération statiques, ne doivent cependant pas pour autant être négligées. En effet, elles sont transférables d'une ville à l'autre, avec un degré de portabilité qui croît avec la taille – et surtout la densité - de la ville d'origine et de destination.



## Références bibliographiques

- Albouy D. (2008), "Are big cities really bad places to live? Improving quality-of-life estimates across cities", NBER Working Paper n°14472. Version révisée 2012 : <http://davidalbouy.net/improvingqol.pdf>.
- Albouy D. (2009), "What are cities worth? Land rents, local productivity, and the capitalization of amenity values". NBER Working Paper n°14981.
- Albouy D., F. Leibovici, et C. Warman, (2013), "Quality of life, firm productivity, and the value of amenities across Canadian cities", *Canadian Journal of Economics* 46(2):379-411.
- Bacolod M., Blum B.S. et Strange W.C. (2009), "Skills in the city", *Journal of Urban Economics* 65(2), 136–153.
- Bai J. (2009), "Panel data models with interactive fixed effects", *Econometrica* 77(4), 1229-1279.
- Combes P.-P., G. Duranton, and L. Gobillon (2008), "Spatial Wage Disparities: Sorting Matters!", *Journal of Urban Economics* 63, 723-742.
- Combes P.-P., G. Duranton, and L. Gobillon (2011), "The identification of agglomeration economies", *Journal of Economic Geography* 11, 253-266.
- Combes P.-P., G. Duranton et L. Gobillon (2015), "Salaires et salariés en Ile-de-France", à paraître dans la *Revue économique*.
- Combes P.-P., G. Duranton, L. Gobillon, D. Puga and S. Roux (2012), "The productivity advantages of large cities: Distinguishing agglomeration from firm selection", *Econometrica* 80 (6), 2543–2594.
- Combes P.-P., G. Duranton, L. Gobillon, and S. Roux (2010), "Estimating Agglomeration Economies with History, Geology, and Worker Effects", In Edward L. Glaeser (ed.) *Agglomeration Economics*, NBER, Cambridge, MA.
- Combes P.-P., G. Duranton, L. Gobillon, and S. Roux (2012), "Sorting and local wage and skill distributions in France", *Regional Science and Urban Economics* 42, 913–930.
- Combes P.-P. et L. Gobillon (2015), "The empirics of agglomeration economies", à paraître In Gilles Duranton, Vernon Henderson and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5, Elsevier.
- Combes P.-P. et M. Lafourcade (2012), « Revue de la littérature académique quantifiant les effets d'agglomération sur la productivité et l'emploi », Rapport réalisé pour la Société du Grand Paris, dans le cadre de la mission d'études des éléments de l'évaluation socioéconomique du réseau de transport du Grand-Paris (Lot 3).
- Combes P.-P., M. Lafourcade, J.-F. Thisse and J.-C. Toutain (2011), "The rise and fall of spatial inequalities in France: A long-run perspective", *Explorations in Economic History* 48, 243-271.
- De La Roca J. et D. Puga (2012), "Learning by working in big cities", CEPR Discussion Paper n°9243. Version révisée 2014 : <http://diegopuga.org/papers/esurban.pdf>.
- D'Costa S. et H.G. Overman (2014), "The urban wage growth premium: Sorting or learning?", *Regional Science and Urban Economics* 48(C), 168-179.
- Duranton G. et D. Puga (2004), "Micro-foundations of urban agglomeration economies", In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, Amsterdam: North-Holland, 2063–2117.
- Henderson J.V. (2003), "Marshall's scale economies", *Journal of Urban Economics* 53(1), 1–28.
- Glaeser E.L. (1999), "Learning in cities", *Journal of Urban Economics* 46(2), 254–277.
- Glaeser E.L. (2011), *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, London: MacMillan.
- Glaeser E.L., Kallal, H., Scheinkman, J. A. et Schleifer, A. (1992), "Growth in cities", *Journal of Political Economy* 100(6), 1126–1152.

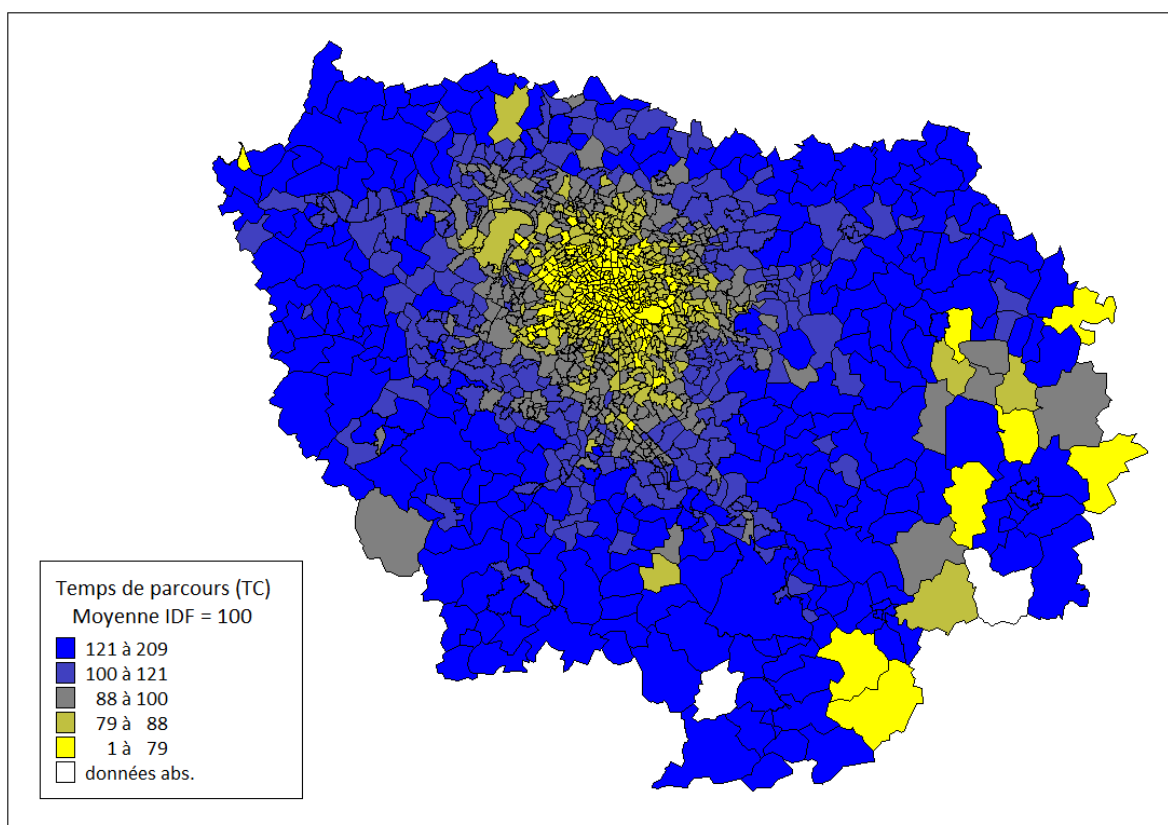
- Glaeser E.L. et D.C. Maré (2001), "Cities and skills", *Journal of Labor Economics* 19(2), 316-342.
- Holl A. (2004), "Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain", *Regional Science and Urban Economics* 34(3), 341-363.
- Insee (2013), *Tableaux de l'économie française* - édition 2013.
- Labrador J. (2011), « Zonage en aires urbaines 2010 : le centre se densifie, le périurbain s'étend », Insee Ile-de-France à la page n°374 - Octobre.
- Krugman P. (1991), "History versus Expectations", *Quarterly Journal of Economics*, 651-677.
- Lucas R.E. (1988), "On the mechanics of economic development", *Journal of Monetary Economics* 22(1), 3-42.
- Mas S. (2012, coord.), *Atlas des zones d'emploi 2010*, Publication de la DARES et de l'INSEE.
- Mincer J. (1958), "Investment in human capital and personal income distribution", *Journal of Political Economy* 66 (4), 281-302.
- Moretti E. (2013), "Real wage inequality", *American Economic Journal: Applied Economics* 5(1), 65-103.
- Moulton B.R. (1990), "An illustration of the pitfall in estimating the effects of aggregate variables on micro units", *Review of Economics and Statistics* 72(2), 334-338.
- Puga D. (2010), "The magnitude and causes of agglomeration economies", *Journal of Regional Science* 50(1), 203-219.
- Roback, J. (1982), "Wages, rents and the quality of life", *Journal of Political Economy*, 90(6), 1257-1278.
- Rosenthal S.S. et W. Strange (2004), "Evidence on the nature and sources of agglomeration economies", In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4, Amsterdam: North-Holland, 2119-2171.
- Seto K.C., Guneralp B. et Hutyra L.R. (2012), "Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools", *Proceedings of the national academy of Sciences of the United States of America* 9(40), 16083-16088.
- Scitovsky T. (1954). Two concepts of external economies, *Journal of Political Economy* 62(2), 143-151.
- Société du Grand Paris (2014), *Rapport d'activité 2013*.
- Strauss-Kahn V. et X. Vives (2009), "Why and where do headquarters move?", *Regional Science and Urban Economics* 39(2), 168-186.
- Véron J. (2006), *L'urbanisation du monde*, La Découverte, Collection Repères n°447.

## Annexes

### Annexe 1 : Distance et temps de transport au sein du Grand Paris

Des temps de trajet par les transports en commun ont été calculés à l'horizon 2013 à partir d'un modèle développé par la Direction Régionale et Interdépartementale de l'Équipement et de l'Aménagement d'Île-de-France (DREI)<sup>46</sup> pour le compte de la Société du Grand Paris. Il s'appuie sur un zonage très fin, composé de 1289 unités infra-communales ou intercommunales (les modus), qui peuvent être ré-agrégées à l'échelle des zones d'emploi franciliennes. Le temps de trajet total entre deux zones, dont la Figure A.1 présente la moyenne pour chaque modus, est la somme des temps de rabattement (vers les connecteurs à l'origine et à la destination), des temps de transport en commun (durée des arrêts intermédiaires comprise), des temps d'attente aux correspondances (au point d'arrêt de correspondance entre l'arrivée et le départ), et du temps de marche à pieds (pour les correspondances avec liaison piétonne entre points d'arrêt, également entre différentes zones d'arrêt d'un arrêt).

Figure A.1 : Temps de transport en commun à l'horizon 2013 au sein du Grand Paris



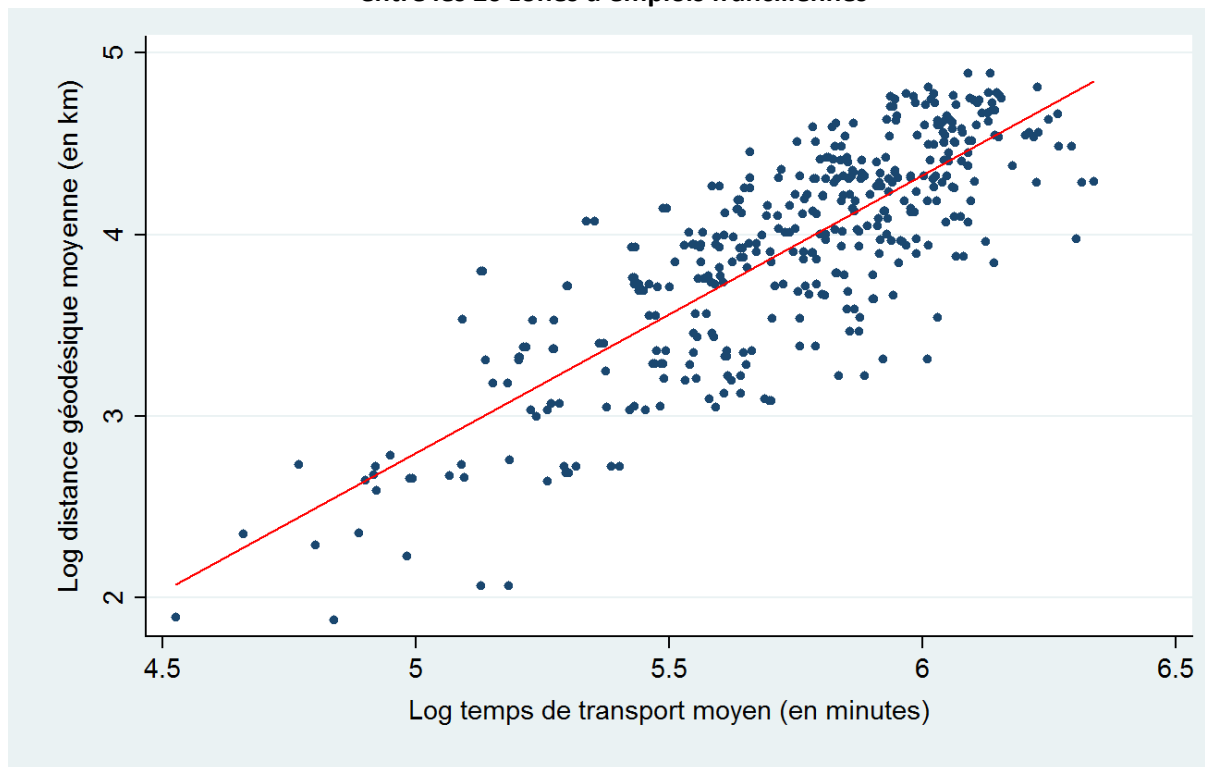
Source : DRIEA/SCEP, Société du Grand Paris et calcul des auteurs.

Cette carte fait apparaître un gradient très net des temps de trajet centré autour de Paris et de la Petite Couronne, et qui décroît ensuite de manière assez monotone vers la Grande Couronne et les zones plus périphériques du Grand Paris. Certaines de ces zones, situées à l'est (dans les zones d'emploi de Coulommiers, Provins, Montereau et Nemours), bénéficient néanmoins d'une bonne accessibilité en transport en commun liée aux lignes P et R du Transilien.

<sup>46</sup> Et plus particulièrement le Service de la Connaissance, des Etudes et de la Prospective (SCEP). Nous remercions Adrien Kippelen et Thi-Phuong-Tram Simonet de nous avoir transmis ces données.

Si on considère la moyenne des temps de trajets des modus inclus dans chaque zone d'emploi<sup>47</sup> pour obtenir le temps de trajet bilatéral entre les 20 zones du Grand Paris, on obtient une mesure de l'accessibilité très fortement corrélée à la distance géodésique, comme l'illustre la Figure A.2.

**Figure A.2 : Temps de transport en commun et distances géodésiques entre les 20 zones d'emplois franciliennes**



Source : DRIEA/SCEP, Société du Grand Paris et calculs des auteurs.

La corrélation entre les deux variables (exprimées en log pour faciliter la lecture) est très significative, même si elle est globalement inférieure à celle mise en évidence par Combes et Lafourcade (2005) pour les coûts de transport routiers. Cependant, une partie de cette différence vient du biais engendré par le fait que les temps de transport associés à la zone de Roissy-Sud-Picardie sont uniquement calculés à partir des modus, qui ne couvrent que la partie francilienne-sud de la zone, alors que le calcul des distances tient aussi compte du nord de la zone (qui inclut son centroïde), dont la frontière s'étend bien au-delà du Grand Paris : les distances sont donc mécaniquement plus élevées que les temps pour cette zone, comme on peut le voir sur la Figure A.1. Si on inclut dans la régression par les MCO des distances sur les temps une indicatrice permettant de corriger ce biais, le temps de trajet explique plus de 70% de la variabilité spatiale de la distance (le  $R^2$  de la régression est de 0.72), ce qui est assez remarquable (cf. Tableau 8).

**Tableau 8 : Temps de transport en commun et distances géodésiques : estimation par les MCO**

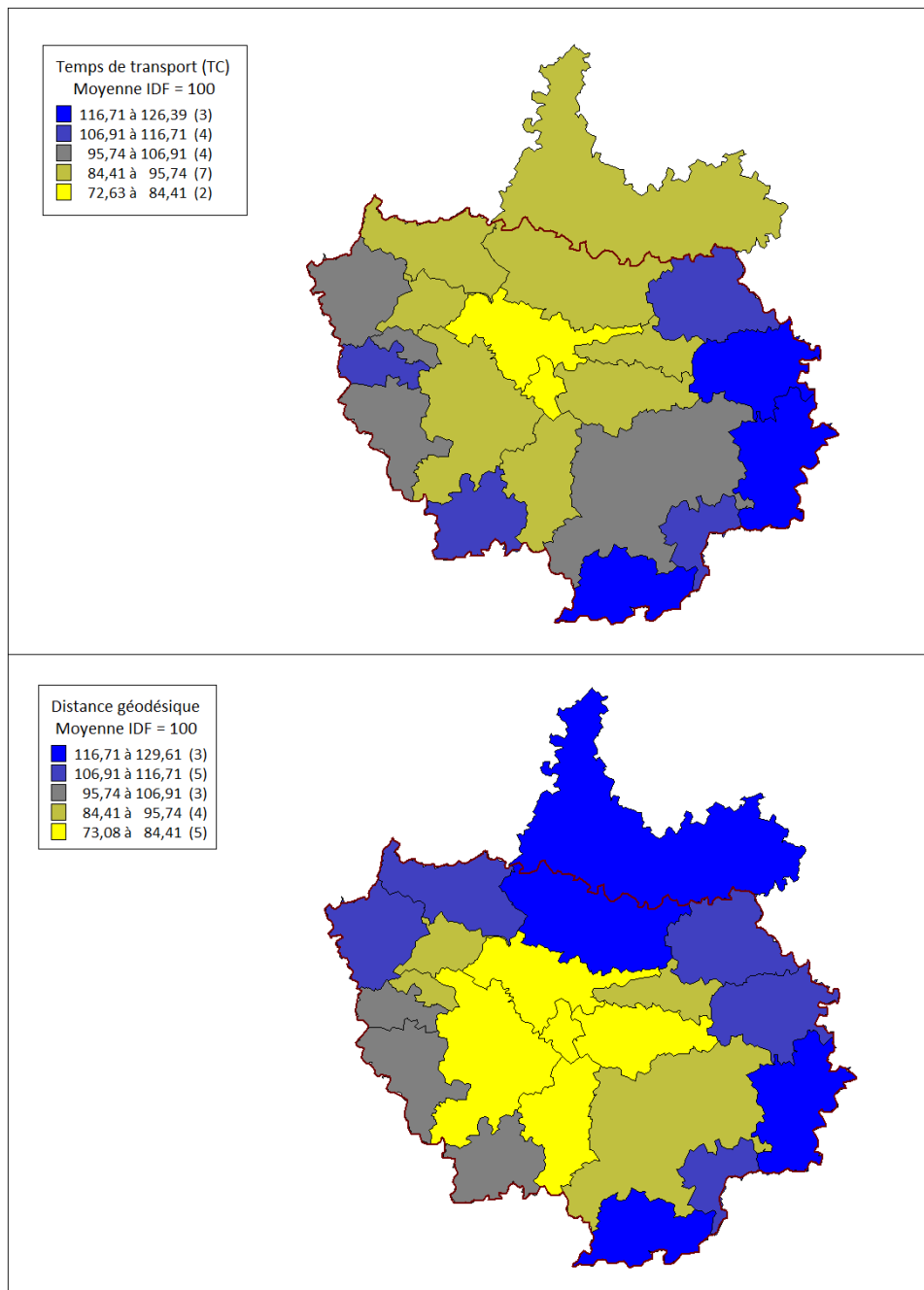
Distance (en log)	Coefficient	Ecart-type	T	P> t	[95% Intervalle confiance]	
Temps (en log)	1.567643	0.0499	31.43	0.000	1.469594	-1.665692
Roissy (indicatrice)	0.4897456	0.0534	9.17	0.000	0.384694	0.594797
Constante	-5.117366	0.2874	-17.81	0.000	-5.682352	-4.552379

<sup>47</sup> Les modus à cheval sur plusieurs zones d'emploi sont affectés à la zone couvrant la plus grande part de leur superficie.

L'élasticité de la distance géodésique par rapport au temps de transport en commun est égale à 1,57, ce qui implique qu'une baisse de 1% du temps de trajet entre deux zones d'emploi équivaut à un rapprochement géographique théorique de 1,57% entre ces zones.

Cette élasticité permet d'estimer le surcroît d'accessibilité et les gains de potentiel marchand (calculés à partir des variations de distance), engendrés par la baisse des temps de transport associée au métro Express. Nous l'utiliserons dans la phase trois de notre étude.

**Figure A.3 : Temps de transport et distance géodésique entre zones d'emploi**



Note : La zone d'emploi de Roissy-Sud-Picardie s'étend au-delà de la frontière administrative de l'Ile-de-France, indiquée en rouge. *Source* : DRIEA/SCEP, Société du Grand Paris et calculs des auteurs.

## Annexe 2 : Nomenclature des secteurs d'activité

Les variables sectorielles utilisées dans la troisième partie de ce rapport ont été construites à partir du niveau 2 de la Nomenclature Economique de Synthèse (NES114), dont la liste suit. Les secteurs indiqués en rouge ont été exclus pour les raisons évoquées dans la section 2.1.

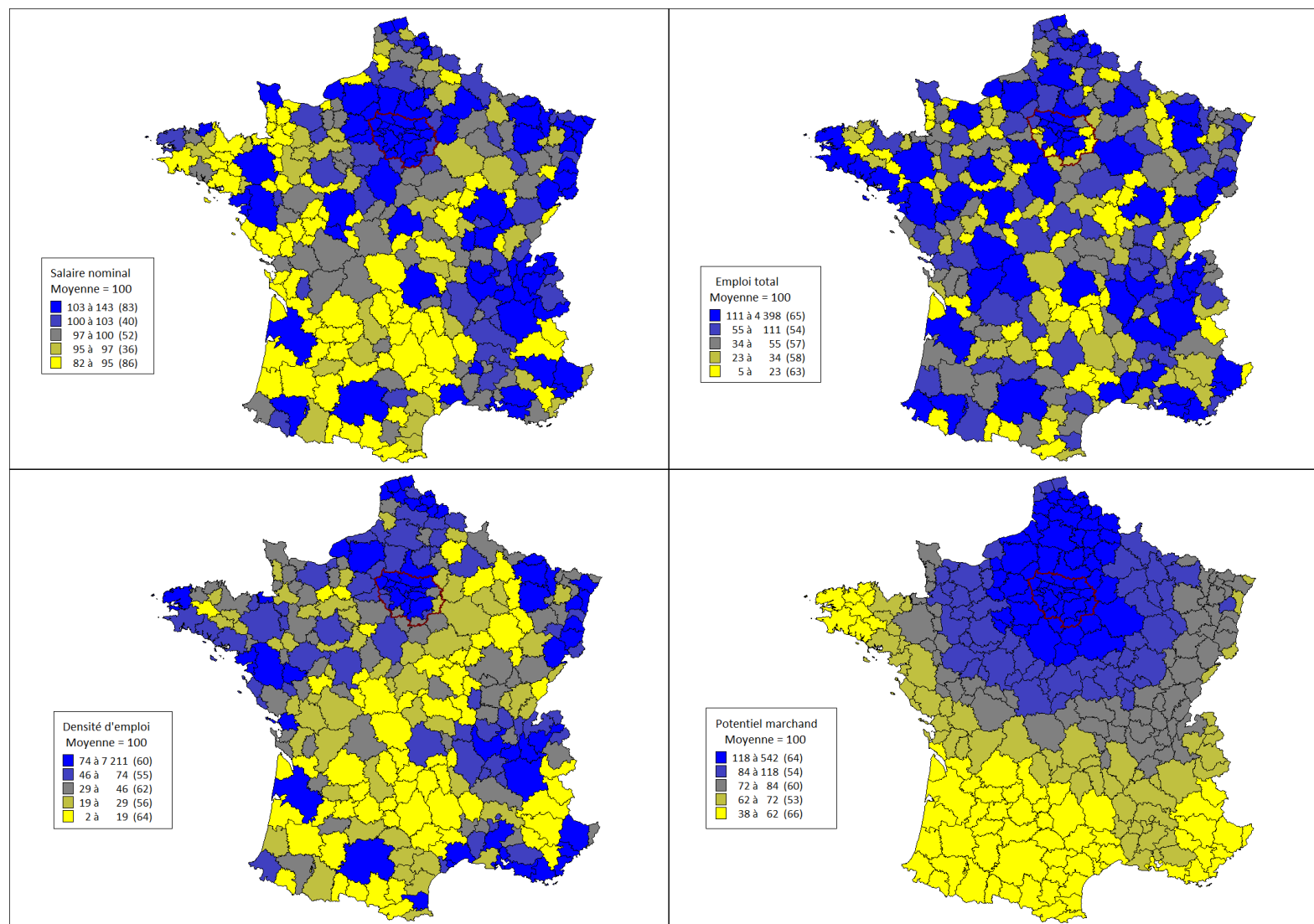
- A01 Agriculture, chasse, services annexes
- A02 Sylviculture, exploitation forestière, services annexes
- A03 Pêche, aquaculture
- B01 Industrie des viandes
- B02 Industrie du lait
- B03 Industrie des boissons
- B04 Travail du grain ; fabrication d'aliments pour animaux
- B05 Industries alimentaires diverses
- B06 Industrie du tabac
- C11 Industrie de l'habillement et des fourrures
- C12 Industrie du cuir et de la chaussure
- C20 Edition, imprimerie, reproduction
- C31 Industrie pharmaceutique
- C32 Fabrication de savons, de parfums et de produits d'entretien
- C41 Fabrication de meubles
- C42 Bijouterie et fabrication d'instruments de musique
- C43 Fabrication d'articles de sport, de jeux et industries diverses
- C44 Fabrication d'appareils domestiques
- C45 Fabrication d'appareils de réception, d'enregistrement et de reproduction (son, image)
- C46 Fabrication de matériel optique et photographique, horlogerie
- D01 Construction automobile
- D02 Fabrication d'équipements automobiles
- E11 Construction navale
- E12 Construction de matériel ferroviaire roulant
- E13 Construction aéronautique et spatiale
- E14 Fabrication de cycles, motocycles, matériel de transport n.c.a.
- E21 Fabrication d'éléments en métal pour la construction
- E22 Chaudronnerie, fabrication de réservoirs métalliques et de chaudières
- E23 Fabrication d'équipements mécaniques
- E24 Fabrication de machines d'usage général
- E25 Fabrication de machines agricoles
- E26 Fabrication de machines-outils
- E27 Fabrication d'autres machines d'usage spécifique
- E28 Fabrication d'armes et de munitions
- E31 Fabrication de machines de bureau et de matériel informatique
- E32 Fabrication de moteurs, génératrices et transformateurs électriques
- E33 Fabrication d'appareils d'émission et de transmission
- E34 Fabrication de matériel médicochirurgical et d'orthopédie
- E35 Fabrication de matériel de mesure et de contrôle

- F11 **Extraction de minerais métalliques**
- F12 Autres industries extractives
- F13 Fabrication de verre et d'articles en verre
- F14 Fabrication de produits céramiques et de matériaux de construction
- F21 Filature et tissage
- F22 Fabrication de produits textiles
- F23 Fabrication d'étoffes et d'articles à maille
- F31 Travail du bois et fabrication d'articles en bois
- F32 Fabrication de pâte à papier, de papier et de carton
- F33 Fabrication d'articles en papier ou en carton
- F41 Industrie chimique minérale
- F42 Industrie chimique organique
- F43 Parachimie
- F44 Fabrication de fibres artificielles ou synthétiques
- F45 Industrie du caoutchouc
- F46 Transformation des matières plastiques
- F51 Sidérurgie et première transformation de l'acier
- F52 Production de métaux non ferreux
- F53 Fonderie
- F54 Services industriels du travail des métaux
- F55 Fabrication de produits métalliques
- F56 Récupération
- F61 Fabrication de matériel électrique
- F62 Fabrication de composants électroniques
- G11 Extraction de houille, de lignite et de tourbe
- G12 Extraction d'hydrocarbures ; services annexes
- G13 **Extraction de minerais d'uranium**
- G14 Cokéfaction et industrie nucléaire
- G15 Raffinage de pétrole
- G21 Production et distribution d'électricité, de gaz et de chaleur
- G22 Captage, traitement et distribution d'eau
- H01 Bâtiment
- H02 Travaux publics
- J10 Commerce et réparation automobile
- J20 Commerce de gros, intermédiaires
- J31 Grandes surfaces à prédominance alimentaire
- J32 Magasins d'alimentation, spécialisés ou non
- J33 Autres commerces de détail, en magasin ou non, réparations
- K01 Transports ferroviaires
- K02 Transport routier de voyageurs
- K03 Transport routier (ou par conduites) de marchandises
- K04 Transports par eau
- K05 Transports aériens
- K06 **Transport spatial**

K07 Manutention, entreposage, gestion d'infrastructures  
K08 Agences de voyage  
K09 Organisation du transport de fret  
L01 Intermédiation financière  
L02 Assurance  
L03 Auxiliaires financiers et d'assurance  
M01 Promotion, gestion immobilière  
M02 Location immobilière  
N11 Activités de poste et de courrier  
N12 Télécommunications  
N21 Activités informatiques  
N22 Services professionnels  
N23 Administration d'entreprises  
N24 Publicité et études de marché  
N25 Architecture, ingénierie, contrôle  
N31 Location sans opérateur  
N32 Sélection et fourniture de personnel  
N33 Sécurité, nettoyage et services divers aux entreprises  
N34 Assainissement, voirie et gestion des déchets  
N40 Recherche et développement  
P10 Hôtels et restaurants  
P21 Activités audiovisuelles  
P22 Autres activités récréatives, culturelles et sportives  
P31 Services personnels  
P32 Services domestiques  
Q10 Education  
Q21 Activités relatives à la santé  
Q22 Action sociale  
R10 Administration publique  
R21 Activités associatives  
R22 Activités extraterritoriales



### Annexe 3 : Salaire net fiscal et taille des zones (moyenne 6 années)



Note : Variables exprimées en % de la moyenne des zones d'emplois. *Source* : INSEE, Panel DADS (secteurs privé et semi-public uniquement).

## Annexe 4 : Liste des Groupes et Catégories Socio-Professionnelles

La nomenclature des Professions et Catégories Socioprofessionnelles (PCS-2003) comporte quatre niveaux d'agrégation emboîtés. Au niveau le plus fin, un poste de la nomenclature PCS correspond à une profession. Au niveau le plus agrégé, celui que nous avons utilisé, on peut identifier 8 groupes socioprofessionnels correspondant au premier chiffre de la PCS, dont la liste suit. Les niveaux d'agrégation intermédiaires correspondent aux Catégories Socio-Professionnelles qui composent chaque groupe. Nous avons exclu les CSP marquées en rouge, soit parce qu'elles n'étaient pas pertinentes dans le cadre de notre analyse (les retraités et les inactifs ne perçoivent pas de salaires), soit parce que nous n'avons pas étudié le secteur correspondant (agriculteurs exploitants).

### **1 - Agriculteurs exploitants**

10 - Agriculteurs salariés de leur exploitation

### **2 – Artisans, commerçants et chefs d'entreprises**

21 – Artisans

22 – Commerçants et assimilés

23 - Chefs d'entreprises de 10 salariés ou plus

### **3 - Cadres et professions intellectuelles supérieures**

31 – Professions libérales et assimilés

32 - Cadres de la fonction publique, professions intellectuelles et artistiques

36 – Cadres d'entreprise

### **4 - Professions intermédiaires**

41 - Professions intermédiaires de l'enseignement, de la santé, de la fonction publique et assimilés

46 - Professions intermédiaires administratives et commerciales des entreprises

47 - Techniciens

48 - Contremaîtres, agents de maîtrise

### **5 - Employés**

51 - Employés de la fonction publique

54 - Employés administratifs d'entreprises

55 - Employés de commerce

56 - Personnels des services directs aux particuliers

### **6 - Ouvriers**

61 - Ouvriers qualifiés

66 - Ouvriers non qualifiés

69 - Ouvriers agricoles

### **7 - Retraités**

71 – Anciens agriculteurs exploitants

72 – Anciens artisans, commerçants, chefs d'entreprise

73 - Anciens cadres et professions intermédiaires

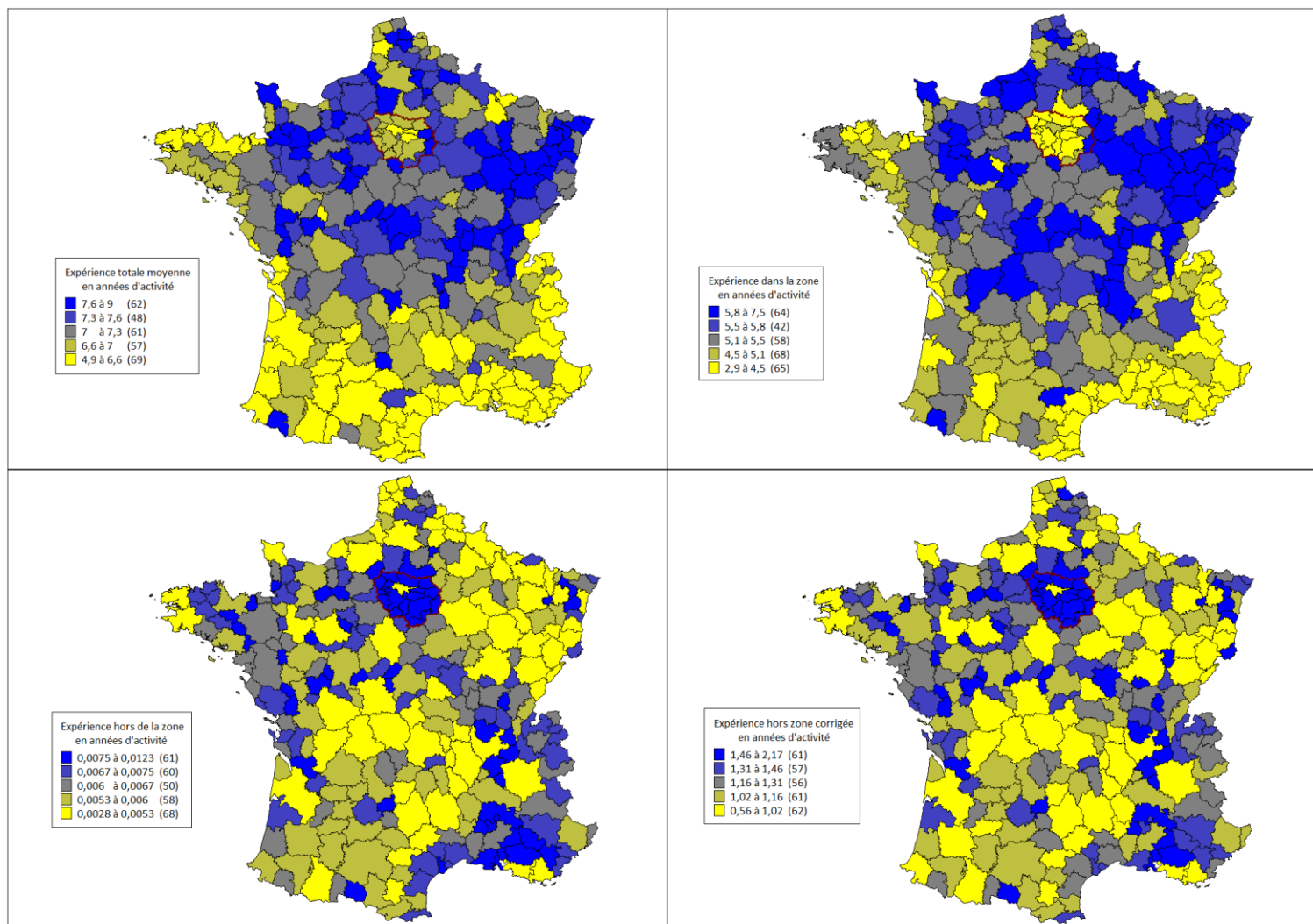
76 - Anciens employés et ouvriers

### **8 - Autres personnes sans activité professionnelle**

81 – Chômeurs n'ayant jamais travaillé

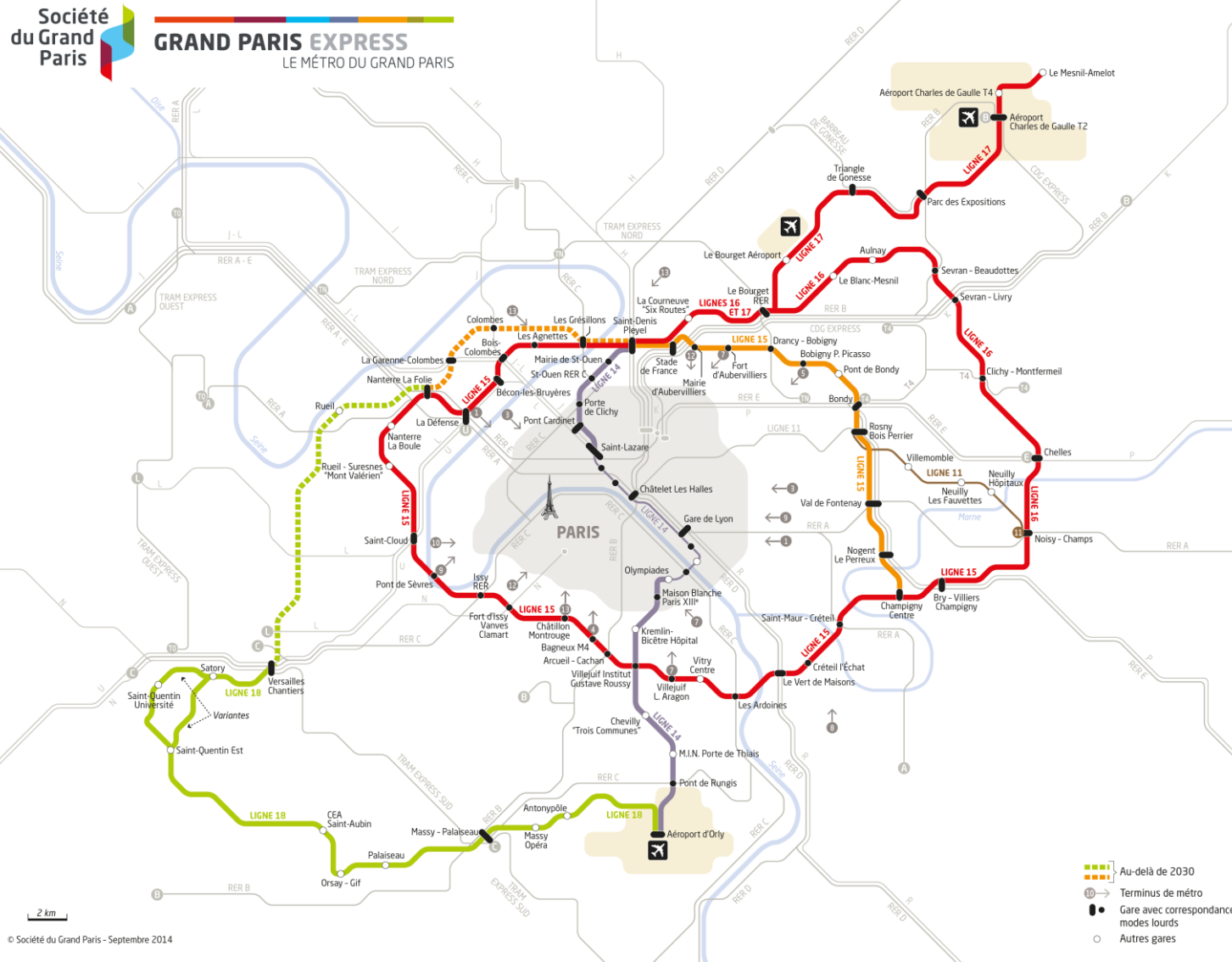
82 – Inactifs divers (autres que retraités)

### Annexe 5 : Expérience des salariés selon la zone (moyenne 6 années)

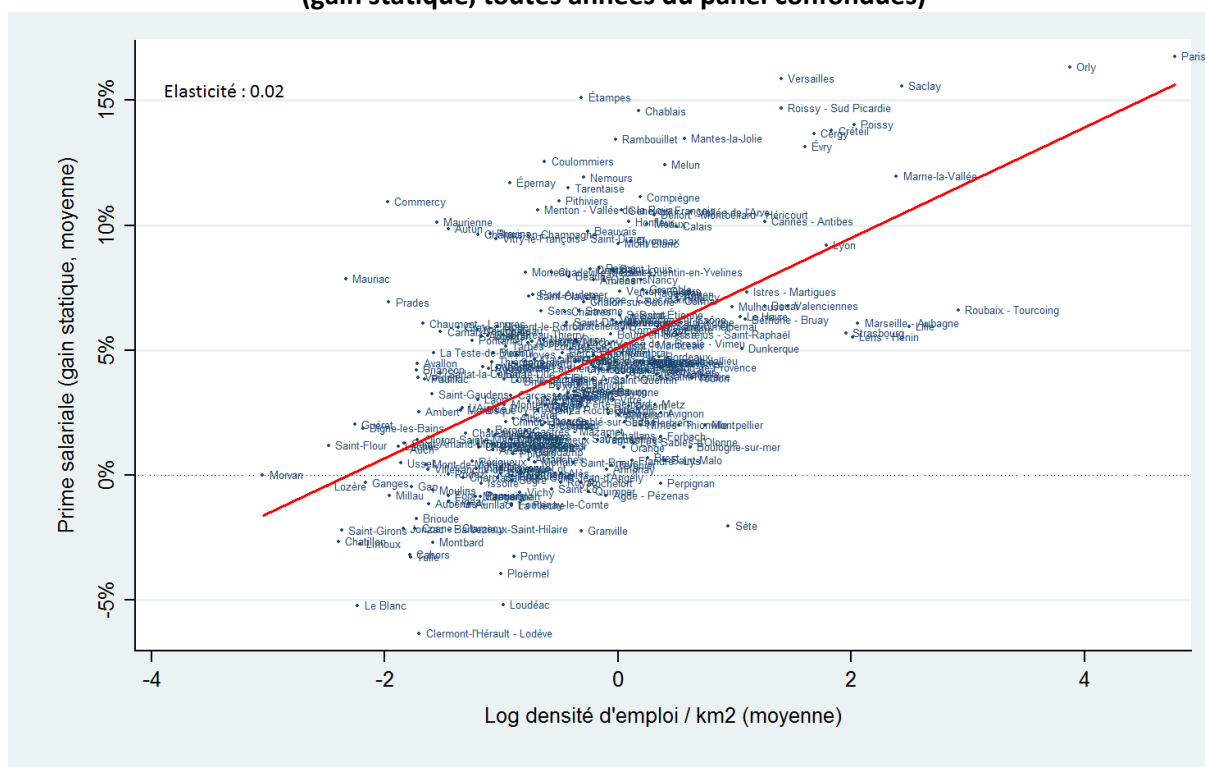


Note : L'expérience hors de la zone est la moyenne du nombre d'années d'activité passées dans toutes les autres zones. L'expérience hors zone corrigée exclut du calcul précédent les zones d'emploi où les salariés n'ont jamais exercé d'activité. *Source* : INSEE, Panel DADS (secteurs privé et semi-public).

## Annexe 6 : Périmètre couvert par le Grand Paris Express



### Annexe 7 : Prime salariale « moyenne » associée aux zones d'emploi (gain statique, toutes années du panel confondues)



Note : La variable en ordonnée correspond au différentiel de salaire entre les villes et la zone d'emploi la moins dense du panel (Morvan, en moyenne), calculé à partir de l'exponentielle des effets fixes  $\beta_c$  estimés en colonne (3) du Tableau 3.

### Annexe 8 : Zones d'emploi par quintile de densité moyenne (sur les 6 années)

q1 – 20% les moins denses (de la moins dense à la plus dense)	q5 – 20% les plus denses (de la plus dense à la moins dense)
2605 Morvan	1101 Paris
9113 Lozère	1118 Orly
8306 Saint-Flour	3110 Roubaix - Tourcoing
2602 Chatillon	3111 Lille
7302 Saint-Girons	1116 Saclay
8305 Mauriac	1102 Marne-la-Vallée
7403 Guéret	9310 Marseille - Aubagne
2407 Le Blanc	1111 Poissy
9102 Limoux	3122 Lens - Hénin
9301 Digne-les-Bains	4205 Strasbourg
9110 Ganges	1117 Créteil
4105 Commercy	8214 Lyon
9116 Prades	1119 Cergy
7303 Millau	1115 Évry
2411 Loches	3115 Valenciennes
7402 Ussel	1113 Versailles
7307 Auch	56 Roissy - Sud Picardie
5404 Jonzac - Barbezieux-Saint-Hilaire	9305 Cannes - Antibes

2402 Saint-Amand-Montrond	3114 Douai
7308 Cahors	9311 Istres - Martigues
7401 Tulle	3123 Béthune - Bruay
9304 Gap	3112 Dunkerque
52 Cosne - Clamecy	2306 Le Havre
8307 Brioude	4208 Mulhouse
9303 Briançon	9112 Sète
2613 Avallon	9111 Montpellier
7213 Oloron-Sainte-Marie	9315 Toulon
4106 Verdun	8220 Vallée de l'Arve
8309 Ambert	3126 Boulogne-sur-mer
9109 Clermont-l'Hérault - Lodève	1110 Mantes-la-Jolie
2106 Chaumont - Langres	4111 Thionville
7206 Pauillac	4207 Colmar
50 Mont-de-Marsan	9314 Fréjus - Saint-Raphaël
7305 Villefranche-de-Rouergue	8218 Annecy
8205 Aubenas	2307 Rouen
7306 Saint-Gaudens	5203 Nantes
2604 Montbard	3125 Calais
8302 Moulins	3124 Saint-Omer
7207 La Teste-de-Buch	1105 Melun
8217 Maurienne	5313 Saint-Malo
5307 Carhaix-Plouguer	9312 Salon-de-Provence
7203 Sarlat-la-Canéda	59 Avignon
7301 Foix - Pamiers	9115 Perpignan
7304 Rodez	4108 Forbach
7309 Figeac	5204 Saint-Nazaire
2607 Autun	7204 Bordeaux
4308 Gray	4202 Molsheim - Obernai
9302 Manosque	5409 La Rochelle
2513 L'Aigle	8209 Bourgoin-Jallieu
2611 Charolais	4107 Metz
2409 Châteauroux	9307 Nice
8304 Aurillac	4302 Belfort - Montbéliard - Héricourt
4304 Pontarlier	8216 Chambéry
7202 Périgueux	5306 Brest
7210 Marmande	4201 Haguenau
7211 Villeneuve-sur-Lot	1104 Meaux
4115 Neufchâteau	5206 Cholet
5401 Thouars - Loudun	61 Toulouse
2204 Laon	8210 Grenoble
2103 Châlons-en-Champagne	

<b>q2 – 20-40% les moins denses (de la moins dense à la plus dense)</b>	<b>q3 – 40-60% les moins denses (de la moins dense à la plus dense)</b>
2405 Châteaudun	4307 Saint-Claude
9313 Draguignan	2403 Vierzon
8311 Issoire	2303 Pont-Audemer
2511 Argentan	5308 Morlaix
2609 Louhans	9316 Cavaillon - Apt
2416 Gien	7209 Agen
8308 Le Puy-en-Velay	9306 Menton - Vallée de la Roya
7201 Bergerac	2505 Vire
2414 Romorantin-Lanthenay	7404 Limoges
1108 Provins	7205 Libourne
2203 Thiérache	2614 Sens
9309 Arles	2501 Bayeux
2612 Auxerre	5412 Bressuire
4309 Vesoul	2606 Nevers
7313 Montauban	5207 Saumur
2107 Vitry-le-François - Saint-Dizier	1103 Coulommiers
2415 Vendôme	8312 Thiers
54 Nogent-le-Rotrou	4101 Longwy
5406 Saintes - Saint-Jean-d'Angély	2417 Montargis
4104 Bar-le-Duc	4112 Épinal
51 Alençon	2504 Lisieux
5210 Mayenne	2101 Charleville-Mézières
5316 Ploërmel	2510 Saint-Lô
5304 Loudéac	5405 Cognac
5202 Châteaubriant	4305 Dole
2201 Château-Thierry	8208 Valence
4306 Lons-le-Saunier	2512 Flers
8206 Montélimar	2419 Pithiviers
2410 Chinon	5209 Laval
7310 Tarbes - Lourdes	2211 Péronne
2104 Épernay	5407 Royan
9103 Narbonne	9104 Alès
9101 Carcassonne	5218 La Roche-sur-Yon
8301 Montluçon	3127 Berck - Montreuil
5212 La Flèche	2404 Chartres
5403 Angoulême	5211 La Ferté-Bernard
5216 Fontenay-le-Comte	5402 Châtelleraut
5208 Segré	2413 Blois
5413 Poitiers	4114 Saint-Dié-des-Vosges
7311 Albi	4102 Lunéville
4109 Sarrebourg	2206 Soissons
5317 Pontivy	8215 Tarentaise
2508 Coutances	2601 Beaune
2401 Bourges	5214 Sablé-sur-Sarthe

2102 Troyes	4301 Besançon
57 Brive-la-Gaillarde	5305 Saint-Brieuc
8303 Vichy	7214 Pau
5411 Parthenay	4206 Wissembourg
7208 Dax	2603 Dijon
2408 Issoudun	9105 Bagnols-sur-Cèze
2301 Bernay	5311 Redon
4303 Morteau	53 Mâcon
9114 Céret	8201 Ambérieu-en-Bugey
7312 Castres - Mazamet	4203 Saverne
5302 Guingamp	5303 Lannion
2209 Abbeville	2507 Cherbourg-Octeville
9108 Béziers	5408 Rochefort
2506 Avranches	2509 Granville
5410 Niort	1114 Étampes
	1107 Nemours

<b>q4 –60-80% les plus denses (de la moins dense à la plus dense)</b>
2608 Chalon-sur-Saône
4110 Sarreguemines
5201 Ancenis
2305 Dieppe - Caux maritime
2207 Beauvais
5309 Quimper
2610 Le Creusot - Montceau
2406 Dreux
2302 Évreux
3121 Arras
2210 Amiens
8212 Roanne
8310 Clermont-Ferrand
5318 Vannes
2105 Reims
5310 Fougères
2418 Orléans
2205 Saint-Quentin
9107 Agde - Pézenas
8204 Annonay
5301 Dinan
5213 Le Mans
4113 Remiremont
5215 Challans
55 Vallée de la Bresle - Vimeu
7212 Bayonne
8202 Bourg-en-Bresse



2202 Tergnier  
8213 Villefranche-sur-Saône  
5314 Vitré  
2412 Tours  
4209 Saint-Louis  
1109 Saint-Quentin-en-Yvelines  
8211 Vienne - Roussillon  
1112 Rambouillet  
1106 Montereau-Fault-Yonne  
8221 Mont Blanc  
5205 Angers  
2304 Vernon - Gisors  
4204 Sélestat  
3117 Maubeuge  
8219 Genevois Français  
3116 Cambrai  
8207 Romans-sur-Isère  
9317 Orange  
5217 Les Herbiers  
5312 Rennes  
2503 Honfleur  
5315 Lorient  
8203 Oyonnax  
3113 Flandre - Lys  
60 Saint-Étienne  
5219 Les Sables-d'Olonne  
8222 Chablais  
9308 Aix-en-Provence  
9106 Nîmes  
2208 Compiègne  
2502 Caen  
4103 Nancy