



**HAL**  
open science

# Human capital externalities in local labor markets and evaluation of a university decentralization policy.

Paul Charruau

► **To cite this version:**

Paul Charruau. Human capital externalities in local labor markets and evaluation of a university decentralization policy.. Economics and Finance. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLD011 . tel-03828160

**HAL Id: tel-03828160**

**<https://theses.hal.science/tel-03828160>**

Submitted on 25 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT**

**DE L'UNIVERSITÉ PSL**

Préparée à l'Université Paris-Dauphine

**Externalités de capital humain dans les marchés du travail locaux et évaluation d'une politique de développement universitaire**

Soutenue par

**Paul Charruau**

Le 29 juin 2022

Ecole doctorale n° ED 543

**Ecole doctorale SDOSE**

Spécialité

**Sciences Economiques**

**Composition du jury :**

EI Mouhoub, Mouhoud Professeur des Universités, Université Paris Dauphine – PSL	<i>Président</i>
Clément, Bosquet Professeur des Universités, CES, Université Paris 1 Panthéon-Sorbonne	<i>Rapporteur</i>
Miren, Lafourcade Professeure des Universités, Université Paris-Saclay	<i>Rapporteuse</i>
Clément, Dherbécourt France Stratégie	<i>Examineur</i>
Gabrielle, Fack Professeure des Universités, Université Paris Dauphine – PSL	<i>Examinatrice</i>
Anne, Epaulard Professeure des Universités, Université Paris Dauphine - PSL	<i>Directrice de thèse</i>

*L'Université Paris-Dauphine n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses; ces opinions doivent être considérées comme propres à leurs auteurs.*

# Remerciements

Ce travail n'aurait pu aboutir sans l'aide précieuse de ma directrice de thèse, Anne Epaulard. Je la remercie chaleureusement de son accompagnement et de ses encouragements constants. Depuis les premiers travaux à France Stratégie, jusqu'aux dernières relectures de cette thèse, sa disponibilité, sa curiosité et son esprit d'analyse, sa franchise et sa confiance, m'ont guidé et ont participé au façonnement de mon goût pour la recherche.

Je souhaite également adresser mes remerciements aux chercheurs et chercheuses qui ont contribué à l'amélioration de mes travaux par leurs conseils et remarques. Je pense en premier lieu à El-Mouhoub Mouhoud et Gabrielle Fack, membres de mon comité de suivi de thèse et de mon jury de soutenance. Je remercie aussi Clément Bosquet et Miren Lafourcade d'avoir accepté d'être les rapporteur · e · s de cette thèse et de leurs précieuses recommandations lors de ma pré-soutenance. J'exprime également ma gratitude à Clément Dherbécourt pour avoir accepté de participer au jury. Je remercie enfin, pour leurs commentaires et discussions, les membres du LEDa et les participant · e · s aux diverses conférences auxquelles j'ai eu l'occasion de présenter mes articles de recherche.

J'adresse aussi toute ma reconnaissance à celles et ceux qui organisent et facilitent notre quotidien à l'Université. Je pense particulièrement à Eve Caroli et Elise Huillery pour leur appui en tant que responsables du programme doctoral d'économie, Bertrand Villeneuve et Audrey Desbonnet au niveau du SDFi, ainsi qu'aux gestionnaires administratives, Nathalie Ferragu, Chantal Charlier et Karène Alarich-Val.

J'aimerais remercier affectueusement Marine Malacain et Adrien Williatte pour leur relecture du premier chapitre. Mes sincères remerciements également à David Hans pour son travail de correction des chapitres deux et trois.

Enfin, merci infiniment à tou · te · s celles et ceux qui m'entourent, de près ou de loin, collègues, ami · e · s et proches. Les noms et surnoms s'entremêlent, la liste est longue... Je sais que chacune et chacun se reconnaîtra en songeant aux moments partagés... Mes dernières pensées vont naturellement à ma famille, grand-parents, parents, frères et sœurs.

# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Introduction Générale</b>	<b>9</b>
<b>Summary</b>	<b>19</b>
<b>1 Spatial wage disparities and human capital externalities in France</b>	<b>21</b>
1.1 Introduction . . . . .	22
1.2 Related Literature . . . . .	25
1.3 Descriptive statistics and Data sources . . . . .	29
1.3.1 Geographical wages disparities in France: relation with density and skills	29
1.3.2 Data . . . . .	31
1.4 Empirical Strategy . . . . .	36
1.4.1 OLS two-steps model. . . . .	36
1.4.2 Instrumental variables in second step. . . . .	38
1.5 Results . . . . .	40
1.5.1 Main Results . . . . .	40
1.5.2 Robustness . . . . .	43
1.6 "Learning by working in well-educated labor markets" . . . . .	48
1.6.1 Accumulated experience in highly-educated areas . . . . .	49
1.6.2 Medium-term human capital externalities . . . . .	51
1.7 Local human capital externalities: separating private and public-sector higher- educated workers . . . . .	53

1.8	Conclusion . . . . .	57
<b>Appendices</b>		<b>59</b>
.1	First step estimation results . . . . .	59
.2	Instrumental variables. . . . .	60
.3	Zones ranking depending on local density and human capital . . . . .	64
<b>2</b>	<b>The Geography of Social Mobility During Working Life</b>	<b>65</b>
2.1	Introduction . . . . .	66
2.2	Related Literature . . . . .	69
2.3	Social mobility over the life cycle: definition, data, and stylized facts . . . . .	71
2.3.1	Social mobility . . . . .	71
2.3.2	Data . . . . .	73
2.3.3	Stylized facts on social mobility during working life . . . . .	75
2.4	Empirical strategy . . . . .	78
2.5	Main results . . . . .	81
2.6	Robustness . . . . .	91
2.7	Plausible mechanisms . . . . .	94
2.7.1	External vs. internal promotion . . . . .	95
2.7.2	The role of work experience in top areas . . . . .	98
2.8	Conclusion . . . . .	102
<b>Appendices</b>		<b>105</b>
.1	Additional Tables . . . . .	105
<b>3</b>	<b>The Local Impact of University Decentralization in France</b>	<b>113</b>
3.1	Introduction . . . . .	114
3.2	Historical Background . . . . .	122
3.2.1	The History of French Universities . . . . .	122
3.2.2	The University 2000 Plan . . . . .	124
3.2.3	The "New Universities" . . . . .	126

---

3.3	Empirical Analysis . . . . .	129
3.3.1	The synthetic control method . . . . .	129
3.3.2	Inference . . . . .	132
3.3.3	Data . . . . .	134
3.4	Results . . . . .	137
3.4.1	Computing Synthetic Controls . . . . .	137
3.4.2	Impacts of "new university" creation . . . . .	137
3.4.3	Placebo Tests . . . . .	139
3.5	Robustness . . . . .	141
3.6	Exploring the Employment Implications of the Creation of "New Universities" . .	148
3.7	Conclusion . . . . .	152
<b>Appendices</b>		<b>155</b>
.1	Trends in Higher-educated Population Share 1968-2014 . . . . .	155
.2	Fit & Balance . . . . .	157
.3	Estimated Effects by Zone with <i>p-values</i> . . . . .	158
.4	In-Space Placebo Tests by Zone . . . . .	159
.5	In-Time Placebo Tests . . . . .	160
.6	Synthetic Weights by Zone . . . . .	161
.7	Diff-in-Diff Event Study Design . . . . .	166
.8	Employment Outcomes Effects by Zone . . . . .	167
.9	Employment Outcomes In-Time Placebo Tests . . . . .	168



# Introduction Générale

En 2018, un an après le début de ce travail de thèse, le mouvement des Gilets jaunes remettait les inégalités géographiques et les « territoires » au centre du débat public en France. De nouveau au premier plan de l’agenda politique, les questions de justice sociale et justice spatiale s’entrechoquaient. D’emblée, et au-delà d’un effet de surprise, les débats médiatiques se sont singularisés par des difficultés à comprendre et analyser les fondements de ce mouvement, ainsi que les ressentiments dont il témoignait – notamment en raison de la sociologie relativement inédite des manifestations et la multiplicité des revendications (Delpirou, 2018). Les éclairages sociologiques, historiques, géographiques et économiques ne manquent pourtant pas sur le sujet des inégalités géographiques, en France et ailleurs.

L’économie géographique s’est développée autour de la volonté de remettre l’« espace », le « territoire », au cœur de l’analyse économique. Le territoire est initialement appréhendé comme le lieu, le cadre, dans lequel s’incarnent les inégalités socioéconomiques. Cette conception du territoire comme « contexte » évolue progressivement pour lui ajouter une dimension « active » : le territoire devient également un facteur d’organisation, il est moteur du dynamisme économique (Davezies, 2012; Courlet and Pecqueur, 2013). C’est notamment l’apport de la « Nouvelle Economie Géographique<sup>1</sup> » initiée par les travaux de Paul Krugman (voir en particulier Krugman (1992, 1991)). En reprenant les intuitions fondatrices d’Alfred Marshall, Krugman remet en évidence les possibilités d’externalités locales à l’origine du processus d’agglomération des activités et des agents économiques, menant au développement inégal des régions.

Cette thèse s’inscrit dans le champ de l’économie géographique. Elle appréhende les inégalités géographiques de salaire et de promotion sociale entre les différentes zones d’emploi françaises.

---

<sup>1</sup>Cette dénomination est sujette à controverse du point de vue de son caractère « novateur » ou de sa dimension « géographique » (voir Martin (1999); Duranton and Rodríguez-Pose (2005); Walther (2011)).

Elle met l'accent sur la concentration géographique des diplômés du supérieur comme source majeure d'économies d'agglomération dans les marchés du travail locaux et étudie la capacité d'une politique spécifique, la création des Universités nouvelles au début des années 1990, à attirer (et maintenir) des personnes diplômées sur un territoire.

**Les économies d'agglomération.** Le concept d'économies d'agglomération repose sur l'idée d'un ensemble de gains d'efficacité engendrés par la proximité des activités et des agents économiques. « *Everybody loves density. Economists like to model and quantify the many benefits of urban density*<sup>2</sup> » (Duranton and Puga, 2020). Cette notion d'avantages liés à la densité géographique recouvre en réalité des phénomènes socioéconomiques multiples. Duranton and Puga (2004) les regroupent en trois grandes catégories selon qu'ils sont liés à des mécanismes de partage (*sharing*), d'appariement (*matching*) ou d'apprentissage (*learning*). En premier lieu, la mise en commun d'un vaste marché de biens et de facteurs de production, le partage d'infrastructures locales, la mutualisation des coûts et des risques, engendrent des gains d'efficacité et de productivité pour les entreprises et les travailleurs des grandes agglomérations – en particulier par la réalisation d'économies d'échelle. C'est le « *sharing* ». En second lieu, l'agglomération des activités et des personnes permet un meilleur appariement entre l'offre et la demande sur les différents marchés, et en particulier sur le marché du travail. C'est le « *matching* ». Enfin, en facilitant les interactions, les territoires agglomérés favorisent la transmission, le développement et l'accumulation des savoirs et des technologies. C'est le « *learning* ».

Au début des années 2000, les décideurs publics s'emparent des enseignements de l'économie de la connaissance et font du capital humain un critère central du développement économique local. Le modèle emblématique est celui de la Silicon Valley. Le regroupement de grandes universités de recherche, de travailleurs hautement qualifiés, d'entreprises très innovantes et d'investisseurs en font un territoire vecteur de développement et de création de richesses. En France, les villes et territoires de Grenoble et Toulouse se sont développés sur des modèles comparables. Le dispositif national des pôles de compétitivité a émergé en 2004 selon cette même logique. On raisonne en réseaux de connaissances et d'innovation, la proximité géographique

---

<sup>2</sup> *Tout le monde aime la densité. Les économistes aiment modéliser et quantifier les nombreux avantages de la densité urbaine.*

favorise les interactions et les processus d'apprentissage et de diffusion des connaissances.

L'agglomération des activités et des personnes peut cependant être aussi à l'origine d'effets négatifs de congestion (infrastructures, transport, services publics, pollution...), comme le souligne une littérature de plus en plus abondante concernant l'arbitrage entre les avantages et les coûts de la densité (voir [Duranton and Puga \(2020\)](#)). Les villes, même dynamiques, sont également des lieux de fixation de la pauvreté ([Brunner and Maurin, 2020](#)). Par ailleurs, [Davezies \(2012\)](#) et [Algan et al. \(2020\)](#) mettent en garde contre le « corollaire » de la métropolisation, à savoir des déséquilibres qui iraient en s'aggravant dans les zones dites « périphériques<sup>3</sup> » ou « en déclin », notamment au regard de l'égalité des chances territoriale et de l'expression d'un mal-être subjectif.

**Les politiques publiques, entre logiques d'efficacité économique et d'égalité des territoires.** La préférence française pour l'égalité ne se limite pas à l'égalité entre individus, mais admet aussi l'égalité entre les territoires. De fait, la Constitution de la Vème République (1958) consacre trois formes d'égalité : entre les citoyens devant la loi, entre hommes et femmes et entre les collectivités territoriales. Plus récemment, la loi constitutionnelle de 2003 évoque explicitement l'objectif de « favoriser l'égalité entre les collectivités locales ».

Les gains (et les coûts) d'agglomération créent cependant une tension entre ces soucis d'égalité entre les territoires et d'efficacité économique. Le récent remplacement du « Commissariat à l'égalité des territoires » en une « Agence de la cohésion des territoires » en 2020 et le passage d'un « Ministère de l'égalité des territoires » à un « Ministère de la cohésion des territoires » en 2018 témoignent de cette tension. Celle-ci se retrouve dans les politiques mises en place. Certaines réorganisations des services publics (hôpitaux et maternités, carte judiciaire...) aboutissent à regrouper des services de l'État dans quelques villes ou zones denses au nom de l'efficacité économique, quand d'autres politiques publiques étendent au contraire à des villes petites et moyennes les logiques d'attractivité et de fixation de populations qualifiées, auparavant réservées aux grandes métropoles. C'est le cas des programmes « Territoires d'industries » (2018), « Action cœur de ville » (2017), « Au cœur des Territoires » (2019), ou des dispositifs « Campus connectés » (2019) et « Petites villes de demain » (2020).

---

<sup>3</sup>Terme popularisé par Christophe Guilluy qui a soulevé de nombreuses controverses parmi les chercheurs.

**Mesurer les effets d'agglomérations, chiffrer les externalités de capital humain et évaluer l'effet des politiques publiques locales.** Savoir où placer le curseur entre logique d'agglomération et logique d'égalité des territoires nécessite de mesurer les effets d'agglomération et d'identifier ce qui leur donne naissance. Les deux premiers chapitres de cette thèse s'inscrivent dans cette logique.

Le premier chapitre<sup>4</sup> reconsidère l'importance respective de la densité d'emploi et du capital humain local dans l'explication des inégalités géographiques de salaires en France sur la période 2009-2015.

A partir des enseignements théoriques concernant les mécanismes à l'origine d'économies d'agglomération, la littérature empirique a démontré de manière relativement consensuelle que la densité des emplois et des personnes (externalités d'urbanisation), la spécialisation sectorielle (externalités de localisation) et la concentration de diplômés (externalités de capital humain) augmentent significativement la productivité et les salaires au niveau local (voir les revues de la littératures proposées par [Combes and Gobillon \(2015\)](#) et [Duranton and Puga \(2020\)](#)). Cela dit, ces caractéristiques locales interfèrent et, même prises isolément, chacune renvoie aux trois types de canaux décrits plus haut (*sharing, matching* et *learning*). Rares sont les travaux qui confrontent ces différents déterminants des effets d'agglomération. En particulier, la question du rôle joué par la seule densité d'emploi par rapport aux effets liés à la composition de ces emplois en matière de capital humain est rarement étudiée. Lorsque c'est le cas, il semblerait que les gains de salaires viennent davantage d'externalités de capital humain plutôt que d'effets purement urbains ([Combes et al., 2011](#); [Rodriguez-Pose and Tselios, 2012](#); [Chauvin et al., 2017](#); [Charruau and Epaulard, 2017](#)).

Cette question est lourde d'implications pour les politiques de développement territorial. Il s'agit de savoir s'il est plus judicieux de concentrer les activités (au moins jusqu'au point où des phénomènes de congestion réduisent les avantages de la densité – voir [Duranton and Puga \(2020\)](#)), indépendamment du niveau de capital humain de la main d'œuvre, ou bien de préférer les politiques tournées vers la formation supérieure et l'attraction et le maintien de diplômés sur un territoire donné.

---

<sup>4</sup>L'accès à certaines données utilisées dans le cadre de ce travail a été réalisé au sein d'environnements sécurisés du Centre d'accès sécurisé aux données – CASD (Réf. 10.34724/CASD).

L'exploitation du Panel tous salariés – EDP (Echantillon Démographique Permanent), produit par l'INSEE et le Ministère des Finances (DGFIP), sur la période 2009-2015, fait ressortir des gains d'agglomération reliés davantage aux externalités locales de capital humain, plutôt qu'à des effets de densité pure. Les estimations principales montrent une semi-élasticité du salaire à la concentration de diplômés du supérieur long de 0,13% (significative au seuil de 1%), alors que le coefficient associé à la densité n'est pas significativement différent de zéro. Pour mettre en perspective ce résultat, il convient de comparer des zones d'emplois aux bornes de la distribution du ratio de diplômés : en considérant un écart de 11 points, en 2012, entre une zone appartenant aux 5% les moins dotées (Mauriac par exemple) et une zone des 5% les mieux dotées en capital humain (Toulouse par exemple), on peut s'attendre à une prime salariale d'au moins 1,5% (à Toulouse par rapport à Mauriac) en raison d'une concentration plus importante de diplômés du supérieur long – toutes choses égales par ailleurs.

Une analyse plus approfondie de ces externalités locales de capital humain, basée sur la méthodologie proposée par [De la Roca and Puga \(2017\)](#), montre que l'effet d'externalité du capital humain mesuré précédemment est probablement sous estimé car il omet une partie des effets d'apprentissage liés à ces externalités. L'expérience accumulée par les salariés dans des environnements très concentrés en diplômés du supérieur leur procure une prime salariale substantielle à moyen terme. En effet, en considérant un individu qui reste 6 ans dans sa zone d'emploi, s'il s'agit de Toulouse, on s'attend à une prime salariale de moyen-terme d'au moins 4% – par rapport au salaire s'il avait travaillé à Mauriac durant cette même période – et ce en raison d'une concentration accrue de diplômés du supérieur long. C'est près de trois fois l'effet mesuré précédemment.

Finalement, ce premier chapitre propose une analyse différenciée des externalités locales de capital humain selon qu'elles proviennent de la concentration des diplômés du secteur public, ou des diplômés du secteur privé. Les estimations suggèrent que les inégalités géographiques de salaires proviennent davantage de différences de capital humain dans le privé plutôt que dans le secteur public. S'il existe des externalités locales de capital humain provenant du secteur public, elles ne sont pas à l'origine d'écart géographiques de salaires – le ratio de diplômés du public dans l'emploi local étant très homogène sur le territoire.

Le second chapitre<sup>5</sup> est issu d'un travail commun avec Anne Epaulard. Il porte sur l'influence des marchés du travail locaux sur la mobilité sociale des salariés en cours de carrière.

Les perspectives de mobilité hiérarchique sont des moteurs de satisfaction au travail (cf. [Senik \(2020\)](#)), elle-même gage de satisfaction générale dans la vie sous de nombreux aspects : statut social, épanouissement, valorisation, sentiment d'appartenance et identité sociale ([Harter and Arora, 2010](#); [Abel et al., 2018](#)).

La mobilité sociale ascendante est définie ici comme la progression dans la hiérarchie des catégories socioprofessionnelles<sup>6</sup>. Cette ascension sociale s'accompagne de gains de salaire importants et recouvre d'autres aspects qui ne se limitent pas à la sphère professionnelle. En France, un certain nombre d'écrits à la première personne, dont ceux d'Annie Ernaux ([Ernaux, 2008](#)) et Didier Eribon ([Eribon, 2009](#)), font le récit de la mobilité sociale ascendante, en lui ajoutant souvent une dimension géographique. De fait, les chances de promotion sociale d'un individu au cours de sa vie professionnelle ne sont pas les mêmes sur tout le territoire. Nos données font ressortir que, selon leur zone d'emploi, les chances de promotion des salariés entre 2009 et 2015 varient de 10% à 27%. Cette question de la géographie de la mobilité sociale a des conséquences importantes sur le plan de la politique publique, en particulier au regard des importants efforts financiers européens pour la convergence des régions, mais aussi vis-à-vis du profond sentiment d'injustice sociale et territoriale qu'analyse le sociologue Benoît Coquard ([Coquard, 2019](#)) et dont la crise des Gilets jaunes a témoigné.

L'analyse empirique menée ici participe à une meilleure compréhension des effets d'agglomération à l'origine des inégalités géographiques de promotion sociale en France. Elle est à notre connaissance la première à s'intéresser à la dimension géographique de la promotion sociale au cours de la carrière d'un salarié et vient ainsi compléter les analyses empiriques de la géographie de la mobilité sociale intergénérationnelle (cf. [Chetty et al. \(2014\)](#) pour les Etats-Unis et [Dherbécourt \(2015\)](#) pour la France).

L'exploitation du Panel tous salariés – EDP fait ressortir que les salariés travaillant dans des zones denses, grandes et riches en capital humain en 2009, ont plus de chances d'être promus

---

<sup>5</sup>L'accès à certaines données utilisées dans le cadre de ce travail a été réalisé au sein d'environnements sécurisés du Centre d'accès sécurisé aux données – CASD (Réf. 10.34724/CASD).

<sup>6</sup>Voir la classification INSEE détaillée dans [Razafindranovona \(2017\)](#).

vers une catégorie socioprofessionnelle supérieure durant les six années suivantes. De nouveau, les estimations suggèrent que les effets d'agglomération en matière de promotion professionnelle proviennent principalement d'externalités locales de capital humain (et de la proximité d'autres marchés denses), plutôt que d'effets exclusivement liés à la taille ou la densité des zones d'emploi. Toutes choses égales par ailleurs, l'augmentation d'un point du ratio de diplômés du supérieur long dans la population locale en 2009 accroît la probabilité de promotion entre 2009 et 2015 de 0.23 points de pourcentage. En reprenant la comparaison entre Toulouse et Mauriac, cela se traduirait par des chances de promotion d'au moins 2,5 points plus élevées pour un salarié travaillant dans la zone de Toulouse – plutôt qu'à Mauriac. Des estimations supplémentaires à l'aide d'un modèle non linéaire montrent que ces effets, qu'ils soient la conséquence de la taille, la densité ou le capital humain local, n'apparaissent que dans les zones d'emploi du haut de la distribution.

Dans un second temps, l'analyse des possibles mécanismes sous-jacents indique que les spécificités des firmes dans les zones denses – généralement plus grandes, plus innovantes et avec davantage de niveaux hiérarchiques (Aghion et al., 2019; Spanos, 2019) – ne suffisent pas à expliquer les écarts en matière de promotion. En effet, si les promotions (qu'elles soient internes ou externes) sont favorisées dans les zones denses et riches en diplômés, elles le sont encore davantage s'agissant des promotions accompagnées d'un changement d'entreprise (promotions externes). Enfin, l'expérience accumulée dans les zones les plus denses, grandes ou riches en capital humain apparaît « transférable » : elle augmente les chances de promotion après une mobilité vers des zones moins denses, plus petites ou moins riches en capital humain.

Les deux premiers chapitres ont donc montré que la concentration de diplômés du supérieur dans une zones d'emploi était à l'origine d'effets d'agglomération augmentant la productivité et les salaires et favorisant la promotion sociale en cours de carrière. Ces résultats donnent ainsi raison aux décideurs locaux qui cherchent à créer, attirer et retenir des populations diplômées sur leur territoire. Le troisième chapitre de cette thèse s'intéresse à la politique la plus intuitive : la création d'une université.

Mis en œuvre au début des années 1990, le plan de modernisation des universités françaises « U 2000 » émerge en réponse à une forte croissance des effectifs étudiants et aux créations tous

azimuts d’antennes universitaires locales, avec pour l’ambition d’opérer certains rééquilibrages territoriaux – quantitatifs et qualitatifs. S’il s’agit tout d’abord de rééquilibrer la carte universitaire dans un contexte de massification de l’enseignement supérieur, l’idée que l’Université peut participer au développement des territoires est très présente chez les promoteurs du plan « U 2000 » (Filâtre and Soldano, 2012). Ce plan a donné lieu à la création de quatre Universités nouvelles – hors Ile-de-France – réparties sur dix sites : l’Université d’Artois sur les sites d’Arras, Béthune, Douai et Lens ; l’Université Bretagne-Sud sur les site de Lorient et Vannes ; l’Université du Littoral Côte d’Opale sur les sites de Boulogne-sur-mer, Calais, Dunkerque (et Saint-Omer qui sera retiré de l’analyse) ; et La Rochelle Université sur le site de La Rochelle. Ces créations universitaires constituent une expérience naturelle idéale pour l’évaluation de l’impact local d’une politique publique de développement universitaire.

A partir des données harmonisées des recensements de la population 1968-2017 de l’INSEE et en utilisant la méthode des contrôles synthétiques, l’analyse économétrique montre que la création des Universités nouvelles a conduit à une augmentation significative de la concentration locale de diplômés du supérieur (non scolarisés) dans leurs zones d’emploi d’implantation. Différents tests de robustesse et l’utilisation complémentaire de la méthode des doubles différences confirment ce résultat. En moyenne, la création d’un site universitaire a augmenté progressivement la part des diplômés pour atteindre un gain de quatre points en 27 ans – ce qui représente une augmentation de 18% par rapport à la situation contrefactuelle. En 2017, la concentration de diplômés du supérieur atteint en moyenne 24% dans les zones étudiées, alors qu’elle n’aurait été que de 20% sans la création des Universités nouvelles. On note cependant des effets hétérogènes selon les cas. L’inférence statistique au cas par cas permet de convaincre d’un effet significativement positif dans les zones de Lens, Béthune, Douai, Calais, La Rochelle et Lorient. Concernant les autres sites (Arras, Boulogne-sur-mer, Dunkerque et Vannes), il semble moins évident que leur situation en matière de diplômés aurait été moins favorable sans leurs créations universitaires.

Finalement, l’examen des potentielles implications en matière d’emploi local suggère que, en moyenne, les gains en capital humain se sont accompagnés d’une augmentation de l’emploi qualifié (cadres et professions intellectuelles supérieures). Les créations universitaires donnent

également lieu à une densification de l'emploi local. Enfin, les décentralisations universitaires n'ont pas eu d'effet significatif sur le taux de chômage.



# Summary

This thesis examines the geographical inequalities in wages and social promotion across French employment zones. It focuses on the concentration of higher-educated people as a major cause of agglomeration economies in local labor markets.

The first chapter reconsiders the matter of density and human capital in the case of French local labor markets over the recent period 2009-2015. It has three main contributions. The first contribution is to provide a new exploration of the role of local human capital in agglomeration gains in France with respect to the role of pure density. Using a panel of individual administrative data and taking into account potential sorting bias, it shows that agglomeration gains in French local labor markets are much more linked to human capital externalities than pure local density effects. The second contribution is to explore the dynamic effects related to local human capital externalities. It finds that there are substantial learning advantages in zones where the human capital is abundant, which leads to a substantial medium-term wage premium. Finally, a third contribution of this paper is to investigate the role of local human capital externalities according to whether they come from private or public higher-educated workers. It finds that geographic wage inequalities appear to stem more from differences in human capital in the private sector than in the public sector. This is probably more a consequence of public policies aimed at distributing public employment equitably across territories, rather than the "proof" that higher educated workers in the public sector do not generate human capital externalities.

The second chapter examines the influence of local labor markets on individuals' social mobility during their working life. In France, over a period of six years, individuals working in big or dense areas and areas with abundant human capital are more likely to be promoted to a higher socioprofessional status – a promotion associated with a substantial wage increase

and/or better working conditions. Applying an empirical strategy to remove spatial sorting bias in a sample of about 350,000 workers in metropolitan France over the period 2009-2015, it finds that density, local human capital, and labor market size significantly increase the likelihood of being promoted, but have no impact on demotions. Again, the results suggest that the agglomeration effects on upward mobility come primarily from human capital externalities (and proximity to other dense markets), rather than pure urbanization and scaled effects. In addition, it shows that local density increases both internal (within the same firm) and external (in another firm) promotions, with the impact of density being even larger on external promotion. Finally, experience accumulated in the densest areas is portable, and increases the chances of promotion after relocation to less dense areas.

The third chapter measures the effects on human-capital accumulation at the local level of the "new universities" created as part of the U2000 Plan implemented in France in the early 1990s. Established in 1990, this national program resulted in the creation of eight universities (spread over 15 sites), including four outside the Paris region (over 10 sites). Using the synthetic control method, it shows that the opening of "new universities" has led to a significant increase in the local share of higher-educated people (not including those in school). On average, the creation of "new universities" increased this share by 4 p.p. within 25 years (about 17% of the counterfactual situation), though the effect differs across cases. Moreover, exploring the employment implications of "new universities" creation, it finds credible evidence that, on average, human-capital gains co-occurred with gains in skilled jobs.

## Chapter 1

# Spatial wage disparities and human capital externalities in France

### Abstract

This chapter reconsiders the matter of density and human capital in the case of French local labor markets over the recent period 2009-2015. Using a panel of individual administrative data and taking into account potential sorting bias, we show that agglomeration gains in French local labor markets are much more linked to human capital externalities than pure local density effects. Exploring the dynamic effects of local human capital, we next find substantial learning advantages in zones where the human capital is abundant, which leads to a substantial medium-term wage premium. Finally, we investigate the role of local human capital externalities according to whether they come from private or public higher-educated workers. We find that geographic wage inequalities appear to stem more from differences in human capital in the private sector than in the public sector.

## 1.1 Introduction

What is the main driver of persistent and significant geographical differences in the wages and productivity of French workers? Is it the density of local labor market (proximity of workers and firms)? Is it the proximity to skilled/highly-educated workers? Or is it the proximity to demand and the largest metropolitan cities?

A glimpse at raw wages data does not give a clear answer. While the density of the commuting-zone of Nantes is higher than Grenoble's one, workers earn more in Grenoble at all levels of qualification. This might be explained by the fact that, although less dense than Nantes, Grenoble has a greater concentration of highly-educated people compared to Nantes. Yet, looking at workers in a third commuting-zone, Compiègne, which is neither very dense nor characterized by a highly-educated workforce, wages are higher than in Nantes, and only slightly lower than in Grenoble. In this case, the proximity to Paris may explain a relatively high average wage.

Clarifying these questions is crucial for the design of efficient local public policies. If the productivity gaps are mainly due to pure density effects, there will be great temptation to concentrate jobs geographically (at least to the point where congestion reduces the advantages of density (Duranton and Puga, 2020)) regardless of the education and skill level of the labor force. In contrast, if the local human capital channel is at work, policies may seek to attract and create activities that employ skilled workers, focus resources on higher education, and strive to attract highly-educated individuals.

In his *Principles of Economics* (1890), Marshall already noted that *"when an industry has thus chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighbourhood to one another"*.<sup>1</sup> Two major insights on the origins of regional development emerge here: spatial proximity and human capital. Since then, regional and urban economics have provided extensive evidence that the density of local labor markets and human capital externalities (along with industrial specialization) are the main determinants of agglomeration gains at the origin of spatial disparities in productivity

---

<sup>1</sup>Marshall A. (2013) *Industrial Organization, Continued. The Concentration of Specialized Industries in Particular Localities*. In: *Principles of Economics*. Palgrave Classics in Economics. Palgrave Macmillan, London. Page 225.

and wages (Moretti, 2011; Combes and Gobillon, 2015).

However, these two local determinants simultaneously influence wages through various agglomeration mechanisms (Duranton and Puga (2004) classify them into sharing, matching and learning effects), which may not be exclusive to density or human capital alone, and the empirical literature still struggle to disentangle the role of one with respect to the other.

Density, as the concentration of economic activities, has often sparked the most research interest, leading to a consensus on a positive elasticity of wages and productivity, with respect to density, of about 0.04-0.06 (Rosenthal and Strange, 2004; Melo et al., 2009; Combes and Gobillon, 2015; Ahlfeldt and Pietrostefani, 2019). With a weaker consensus, human capital externalities are also found to increase wages at the local level (Rauch, 1993; Moretti, 2004b,a; Rosenthal and Strange, 2008). But, as noted in Combes and Gobillon (2015), local density is most often not introduced simultaneously in the regressions that evaluate the impact of local human capital on wages. When it is the case, the human capital influence seems to supersede that of density, which drops significantly (Chauvin et al., 2017; Combes et al., 2011; Rodriguez-Pose and Tselios, 2012).

This paper reconsiders the question of density and human capital in the case of French local labor markets over the recent period 2009-2015. It has three main contributions.

The first contribution is to provide a new exploration of the role of local human capital in agglomeration gains in France with respect to the role of pure density. In particular, we ponder whether human capital externalities may be a better explanation for wage disparities than density. Using a panel of individual administrative data and taking into account potential sorting bias, we show that local human capital externalities play a major role in agglomeration gains in France. Our preferred specification indicates a positive and significant semi-elasticity of wage with respect to local human capital<sup>2</sup> of 0.13%, whereas the influence of pure density is not significantly different from zero. Since the most dense areas also tend to be very abundant in human capital, this does not necessarily mean that density is not a source of agglomeration gains, but rather that the majority of these gains involves human capital effects. With a gap of 11 percentage points in 2012 between the top 5% and the bottom 5% of employment zones

---

<sup>2</sup>We define local human capital as the local share of highly-educated people (not including those in school) with a minimum of three years of study after the *baccalauréat*. *Enseignement supérieur long* according to INSEE.

regarding their local share of higher-educated people, we expect a wage premium of 1.5%<sup>3</sup> in a well-educated zone compared to a zone with few graduates.<sup>4</sup>

The second contribution of this paper is to explore the dynamic effects related to local human capital externalities. We find that there are substantial learning advantages in zones where the human capital is abundant. We use the methodology of [De la Roca and Puga \(2017\)](#) and turn it into "Learning by working in well-educated areas". Because they ignore dynamic effects, our previous results underestimate the overall agglomeration gains (static and dynamic) related to local concentration of human capital. To take into account dynamic effects, we disentangle the intrinsic ability of workers and the value of accumulated experience in top areas in terms of local human capital. We find that two thirds of human capital externalities may come from dynamic effects. Considering that workers stay on average 6 years in their employment zone, one percentage point difference in the local share of highly-educated people is associated with a 0.33% increase in their wages. With a gap of 11 percentage points in 2012 between the top 5% and the bottom 5% of employment zones, we expect a wage premium of about 4% in a well-educated zone over a medium-term period.<sup>5</sup> This is three times higher than the wage premium measured without taking into account the role of experience accumulated in highly-educated areas.

A third contribution of this paper is to explore the role of local human capital externalities according to whether they come from private or public highly-educated workers. We draw on a growing literature on public employment in local labor markets, which shows that public employment may be a major force for counterbalancing geographical inequalities ([Jofre-Monseny et al., 2020](#); [Becker et al., 2021](#)). We find that geographic wage inequalities appear to stem more from differences in human capital in the private sector than in the public sector. This is probably more a consequence of public policies aimed at distributing public employment equitably across territories, rather than the "proof" that highly-educated workers in the public sector do not generate human capital externalities.

---

<sup>3</sup> $\exp(0.00134 \times 11)$ .

<sup>4</sup>The wage premium would be 4% between the zone with the smallest ratio of graduates in 2012 (Thiérache with 4% of graduates) and the zone with the largest ratio (Paris with 32% of graduates).

<sup>5</sup>The maximum medium-term wage premium would be of about 10% between the zone of Thiérache and the zone of Paris (see above).

The rest of the paper is organized as follows. Section 1.2 discusses the related literature. Section 1.3.2 introduces a detailed description of data. Section 1.3.1 gives an overview of spatial wage disparities in France. Section 1.4 presents the empirical strategy. Section 1.5 presents the main estimation results and alternative robustness tests, with a focus on the respective role of human externalities and density. Section 1.6 turns to the dynamic effects associated with human capital externalities. Finally, Section 1.7 explores the local human capital externalities from public-sector and private-sector higher-educated workers. Section 1.8 concludes.

## 1.2 Related Literature

**Channels of agglomeration gains.** The broad concept of agglomeration economies refers to all the productivity gains generated by the proximity of economic activities, firms and individuals. Since Duranton and Puga (2004), theoretical research identifies three main theoretical mechanisms at the origin of these agglomeration economies (matching, sharing, and learning), while empirical research has difficulty to precisely quantify each with respect to the others:

- **Matching.** The size and the density of the local labor market, bringing together many heterogeneous firms, encourage a diversification of the labor demand and contribute to better local matching. Linking workers with employers in a thick labor market increases productivity and wages by improving both the frequency and the quality of matches;
- **Sharing (and specialization).** Dense areas facilitate sharing indivisible local public goods and facilities, thereby generate increasing returns by lowering the average production costs. The agglomeration of economic activities can also create gains from variety by increasing the local supply of intermediate inputs and from industrial specialization. This lowers transaction costs and facilitates access to inputs needed for production. Finally, a denser labor market allows for sharing risks by offering a “constant market for skill” (Marshall, 1890);
- **Learning (knowledge spillovers).** The spatial concentration of economic activities can be a source of productivity gains through the diffusion of innovations and knowledge. These benefits come from direct interactions between workers, fostered by the spatial proximity

of firms and jobs.

In light of this theoretical background, most of the empirical literature explores the overall impact of different local characteristics on wages and identifies three main determinants of agglomeration gains: the density or size of the area, the sectoral specialization or diversity, and the human capital externalities (see [Combes and Gobillon \(2015\)](#) and [Duranton and Puga \(2020\)](#) for detailed literature review).

**On the elasticity of wages with respect to density.** There is an extensive empirical literature on agglomeration effects that leads to a consensus on a positive elasticity of wages and productivity with respect to density between 0.04 and 0.06 ([Rosenthal and Strange, 2004](#); [Melo et al., 2009](#); [Combes and Gobillon, 2015](#); [Ahlfeldt and Pietrostefani, 2019](#)). Empirical results are sensitive to the form of specifications and models, especially when considering or not individual characteristics to take into account potential sorting bias. Using French panel data, [Combes et al. \(2008\)](#) show that estimated agglomeration effects on wages are much lower (around 0.03 vs. 0.06) when considering individuals characteristics in a two steps identification strategy. They conclude that up to half of the spatial wage disparities in France at the end of the 20th century are driven by sorting and differences in the skill composition of the workforce. More recent works by these authors have since led to smaller estimated effects of density, with an elasticity of wages with respect to density of 0.01 for the period 1993-2008 ([Combes et al., 2015, 2016](#)). On Spanish data, [De la Roca and Puga \(2017\)](#) estimate an elasticity of wages with respect to city size of 0.02 and confirm the 50% drop in the coefficient when introducing workers fixed effects to address sorting. They continue the analysis of these fixed effects and show that the drop comes mainly from dynamic effects with the accumulation of more valuable experience in big cities. They conclude that the estimations of the static wage premium that use worker fixed effects overestimate the importance of sorting by capturing also the learning advantages of working in dense areas.

**The role of human capital externalities.** In line with the pioneer insights of endogenous growth theories ([Romer, 1986](#); [Lucas, 1988](#)),<sup>6</sup> a part of the literature on spatial wage disparities

---

<sup>6</sup>Taking up the intuitions of [Marshall \(2013\)](#) on the importance of human capital in the development of economic areas, endogenous growth theories emphasize the influence of human capital externalities on productivity and

centered the analysis on human capital effects, seeking to measure whether the geographic concentration of skills and knowledge in the urban areas improves the local productivity and wages (Glaeser et al., 1992; Rauch, 1993; Jaffe et al., 1993). Hence, the focus is on the concentration of highly-educated individuals: the key benefit of agglomeration is the grouping of these workers, especially because this enhances the diffusion of new knowledge and innovations. In contrast with research on density effects, there is no clear consensus on the magnitude of human capital externalities at the local level. Some important papers find evidence of a positive relation between local human capital concentration and wages or productivity, holding individuals' level of education constant (Rauch, 1993; Moretti, 2004c,b; Rosenthal and Strange, 2008), whereas others conclude on no evidence of such external human capital gains for workers (Acemoglu and Angrist, 2001; Ciccone and Peri, 2006). None of these papers introduces controls for competing determinants of agglomeration gains such as the local density (or the size of local labor markets).

#### **Towards disentangling pure density effects from local human capital externalities.**

As human capital externalities are closely related to the theoretical channel of "learning", empirical models that aim to measure them often ignore the other determinants of agglomeration, in particular the local density. However, as Combes and Gobillon (2015) point out, the learning channel, along with the others, may depend on diverse local characteristics, so empirical models focusing on human capital cannot ignore the effects of local labor market's density on wages, since they are known to be substantial and correlated with local human capital. Therefore, if highly-educated areas are also the densest areas, the positive effect of local human capital on wages may result from other density effects. That being said, the reverse is also true. If the estimation of elasticity of wages with respect to density does not take into account the stronger presence of a skilled labor force in dense areas, the positive effect of density may not reflect pure urbanization economies only, but also advantages coming from a higher concentration of human capital.

Some studies actually show that the density premium is greater in better-educated places and corroborate the fact that agglomeration economies are important in dense areas because of knowledge mechanisms (Glaeser and Resseger, 2010; Abel et al., 2012). In addition, when both

---

economic development.

effects (density and local human capital) are included in the same regression, the elasticity of wage with respect to density drops substantively, by about 5 to 6 percentage points according to [Melo et al. \(2009\)](#) meta-analysis. In several cases, the density effect is no longer significant (or even negative) whereas the local human capital shows strong positive effects ([Chauvin et al., 2017](#); [Combes et al., 2011](#); [Rodriguez-Pose and Tselios, 2012](#)).

This paper comes within this literature and places particular emphasis on human capital externalities and their relative importance in French local labor markets, compared to pure density gains.

**Dynamic effects of agglomeration.** Most of the research cited above, whether the authors focus on density, human capital, or both, is primarily focused on static agglomeration effects – i.e. the instantaneous effect on wages, without attention to temporal cumulative gains. Another strand of agglomeration literature has considered specifically dynamic effects of agglomeration, which are generally thought to operate through learning and human capital accumulation mechanisms. One advantage of dense areas is that they enhance the accumulation of human capital and facilitate technological spillovers, experimentation and learning ([Glaeser and Maré, 2001](#); [Wheeler, 2006](#); [D’Costa and Overman, 2014](#); [De la Roca and Puga, 2017](#)).

Section 1.6 of this paper draws on this literature, in particular it follows the methodology of [De la Roca and Puga \(2017\)](#), to explore the dynamic effects on wages related to local human capital externalities.

**Local multipliers of public sectors jobs.** Finally, in the last section of this paper, we go one step further the literature on human capital externalities with a separate analysis of the influence of public-sector and private-sector higher-educated workers. A growing literature has emphasized the role of public employment as a policy tool to develop local labor markets ([Faggio and Overman, 2014](#); [Jofre-Monseny et al., 2020](#); [Becker et al., 2021](#)).

## 1.3 Descriptive statistics and Data sources

### 1.3.1 Geographical wages disparities in France: relation with density and skills

In 2012, French workers earn an average net hourly wage of €13.9<sup>7</sup> (1.9 times the minimum wage). However, this average masks major spatial differences. It is almost doubled between the zone where average wage is the lowest (10.4 euros in the zone of Saint-Flour in south-central France), and the zone where it is the highest (19.1 euros in the zone of Paris).

Figure 1.1: Geography of hourly net wage in 2012

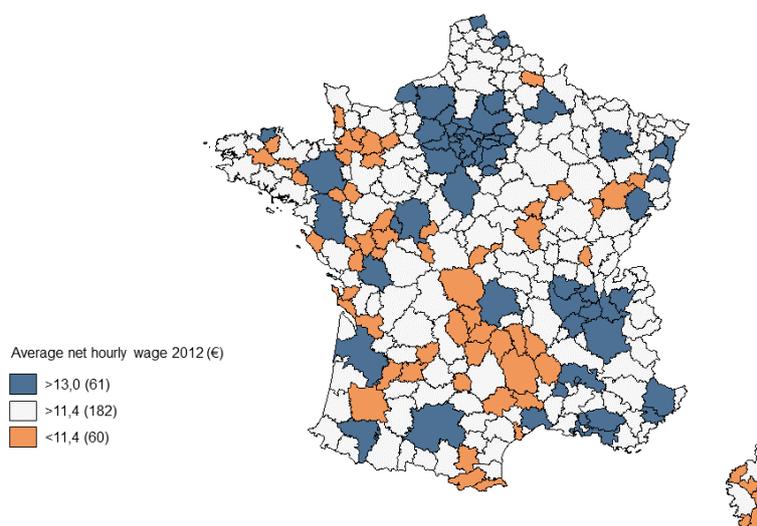


Figure 1.1 presents the heat map of this net hourly wage at the local level in 2012, based on INSEE *Déclaration annuelle de données sociales (DADS)*. In the fifth of the employment-zones where workers are the best paid on average, the average net hourly wage is greater than €13.0. In contrast, in the fifth of the zones where workers are paid the lowest, the average net hourly wage is less than 11.4 euros. Typically, the zones with the highest average wages correspond to the zones of big cities.

By averaging at the zone-level, wage figures also conceals major differences due to local composition effects. The first factor explaining the variation of wages between areas is the

<sup>7</sup>Total workforce figure for metropolitan France given by INSEE.

Table 1.1: Descriptive statistics. Average net hourly wages in 2012 at the employment-zone-level by socioprofessional group and gender.

	Mean	Standard Deviation	Median	Minimum	Maximum
<b>All</b>	12.3	1.2	12.0	10.4	19.1
Men	13.2	1.4	12.8	10.9	20.8
Women	11.0	0.9	10.7	9.7	17.0
<b>White-collar workers</b>	22.3	1.4	22.1	19.5	29.2
Men	23.7	1.5	23.5	20.3	32.1
Women	19.0	1.1	18.9	16.8	25.0
<b>Intermediate workers</b>	13.9	0.6	13.7	12.8	16.0
Men	14.7	0.8	14.5	13.3	18.0
Women	12.7	0.5	12.6	11.6	15.0
<b>Employees</b>	9.9	0.4	9.8	9.2	11.8
Men	10.5	0.4	10.5	9.6	12.6
Women	9.7	0.4	9.6	9.0	11.5
<b>Blue-collar workers</b>	10.6	0.5	10.6	9.6	12.4
Men	10.9	0.6	10.8	9.7	12.7
Women	9.4	0.4	9.4	8.5	10.8

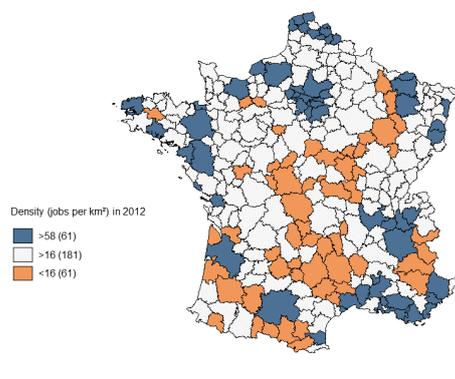
Note. Data comes from *Insee Déclaration annuelle de données sociales (DADS)*. The figures presented relate to average wages at the local-level. In 2012, the average net hourly wages for the 304 employment-zones of metropolitan France is €12.3.

difference in the respective weights of economic activities as well as in the shares of workers in the different socioprofessionnal groups in each local labor market. For example, in 2012, the share of white-collar workers – who are generally better paid – in the zone of Paris is 3 times larger than the corresponding share in the zone of Saint-Flour. Logically, the local workforce composition partly explains the considerable wage gap between these two areas. However, descriptive statistics presented in Table 1.1 show that substantial spatial differences are still observed when comparing local wages within socioprofessionnal groups and genders. The inter-zone dispersion of wages is the highest for male white-collar workers (below €23.5 in half of the zones, while it exceeds €32 in the Paris area).

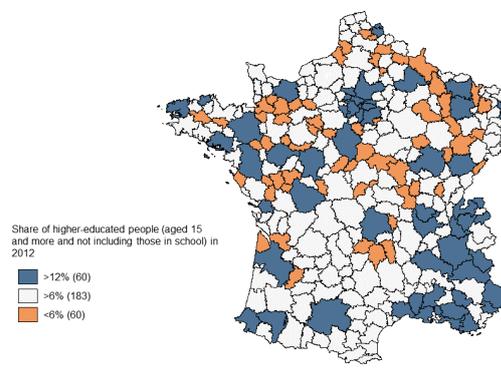
A first indication of agglomeration gains on productivity and wages is given by the heat Map 1.2a of employment density in 2012, which shows that the geographical distribution of dense areas is very similar to the relative spatial distribution of average wages (Map 1.1). As already proved by the literature this illustrates that wages are higher in denser areas. The two maps are not perfectly identical, however.<sup>8</sup> The Map 1.2b of local human capital (share of highly-educated

<sup>8</sup>Coefficient of correlation between the local average hourly wage and the employment density is about 46%.

(a) Geography of local density 2012



(b) Geography of local human capital in 2012



people - not including those in school) differs somewhat from the map of employment density and seems more similar to Map 1.1.<sup>9</sup> More specifically, the densest areas typically have a high concentration of human capital, but the opposite is less true.

Figure 1.3 plots local average wage in top 10% areas in terms of density, human capital, or both (see Annex .3 for all employment-zones belonging to these top 10% groups.). We also observed that local human capital seems to be the main driver of local wage gains. A worker in a top 10% area in terms of human capital that is not a top 10% area in terms of density earns on average 13% more than a worker in a top 10% area in terms of density only (14.4 euros compared to 12.7 euros).<sup>10</sup>

### 1.3.2 Data

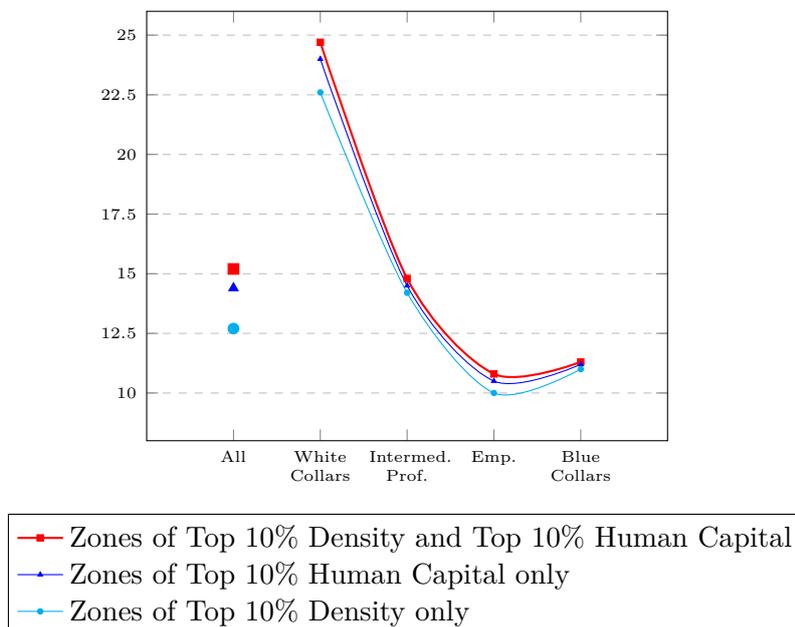
**Sources.** We exploit data from the *Panel Tous Salariés - Echantillon Démographique Permanent (EDP)*, which merges two sources of data: the *Déclarations Annuelles des Données Sociales (DADS) - Panel tous salariés* and the *Echantillon Démographique Permanent (EDP)*.

The *EDP* data gathers all the information recorded in the successive census and vital statistics data (starting with the collection in the year 1968) for people born on one of the four reference days: October 1, 2, 3 and 4 of each year. The base is therefore close to a 1/100th (4/365th) representative sample of the population residing in metropolitan France. The panel

<sup>9</sup>Coefficient of correlation between the local average hourly net wage and the human capital share is about 78%.

<sup>10</sup>Workers in areas in both top 10% have on average a wage premium of another 6% (15.2 euros).

Figure 1.3: Average net hourly wages in 2012 in top areas.



Notes. This figure plots the local average net hourly wages in 2021 by socio-professional groups in three different groups of zones. The zones of the top 10% areas in terms of density, the zones of the top 10% in terms of human capital, or the zones of both tops. Data comes from *Insee Déclaration annuelle de données sociales (DADS)*. Socioprofessionnal groups correspond to one-digit categories: White-collar workers, Intermediate professions, Employees and Blue-collar workers.

"*DADS tous salariés*" covers all employees in the private and public sectors since 1988, and employees of individual employers since 2009. It also includes unemployment benefits since 2008. After the year 2002, this panel version of the raw DADS<sup>11</sup> consists of a 1/12 representative sample of all the workers who earn a salary in France (around 2.2 million employees each year) – including all individuals born on one of the four reference days of the EDP.

Individuals found in the match are those who are present in both the *EDP* and the *Panel Tous Salariés*. The merged *Panel Tous Salariés - EDP* does not provide all the variables from either of these two sources. Most of the *Panel tous salariés* variables are retained, including information on the nature and the localization of the job, the corresponding occupation and industry, the earnings, the starting and closing dates of the period of paid work, the number of paid hours, the terms of employment (full time, part time), etc. However, only a few variables

<sup>11</sup>DADS are built on mandatory annual declarations of social data, based on a registration procedure which must be completed by any firm which employs at least one worker - or by self-employed workers. This annual document serves both fiscal and social administrative purposes.

of the *EDP* are retained, but importantly these include the level of education.

**Breaks in series.** Our version of the panel makes it possible to work on a constant-data basis between 1988 and 2015, although the sample doubles in size after 2002, with about one million workers per year. However, depending on the variables we use, we have to manage some panels' scope increases, and breaks in series, due to changes in nomenclature. Most importantly, the classification and codification of socioprofessional categories (SPC – based on occupations) underwent noticeable changes between 2008 and 2009, to improve the data.<sup>12</sup> This break in series makes difficult the direct comparison between the periods before and after 2008-2009. Similar concerns come from the changes in the economic activities classification also in 2008 (change from the classification NAF rev1 to NAF rev2) – which was done for modernization purposes, to better reflect economic developments, and to ensure a better international comparability.

**Periods of analysis.** Consequently, taking into account these different breaks, we choose to work on a constant-data basis, retaining the seven-year period 2009-2015. This period gives us a broader database field, as the employees of private households are included in the panel from 2009. We still use the period 2002-2008 to construct variables of accumulated experience in the analysis of dynamic human capital effects.

**Jobs positions and annualized observations.** In the initial panel, each observation corresponds to a specific job position (*poste*) for a given worker. A job position is defined as the sum of employment spells of a given worker in the same establishment in a given year. This may cover separate periods (for seasonal workers, for example) or identical periods (some wage premium/bonuses correspond to a separate observation). A given worker can hold several job positions during a year, which relate to different occupations, activities, earnings, etc. This is

---

<sup>12</sup>Since 2009, in the DADS, the nomenclature of occupations and socioprofessional categories of private and public sectors employees (*Nomenclatures des professions et catégories socioprofessionnelles des emplois salariés des employeurs privés et publics (PCS-ESE)*) based on four-digit codes, replaces the former nomenclature based on two-digit codes. In addition to the job label, the PCS-ESE is now based on the employee's hierarchical position and the occupation declared by the establishment. This more detailed breakdown improves the accuracy of the coding, and therefore sheds further light on the employee's true occupation and social status. For example, an accounting officer previously classified as an "Intermediate Profession" - an occupation between manager and executing-employee - can now be classified either as an employee or as a manager, depending on the information available.

especially the case for workers with part-time and low-paid jobs, or with jobs in industries with high employee turnover. For the period 2009-2015, the panel contains about 1.2 million workers each year, of which 30% have more than one job position, and 10% have at least three positions during the year.

In this paper, we define the occupation of each worker for a given year as the occupation attached to the jobs that provide at least two-thirds of her annual wage. We choose to keep a maximum of information, and therefore not to retain only the most remunerative position of the year. Indeed, the most remunerative position of the year does not necessarily reflect the main socioprofessional status of a worker during that year. This is especially an issue for a worker with multiple job positions in a given year, one with substantial wages, and the others lower-paying, but which taken together represent significant annual earnings and constitute her main activity during the year.

Practically, we first identify the two-digit occupation associated with the job positions that provide at least two-thirds of a worker's annual earnings (note that about 80% of workers draw their salary from a single two-digit occupation per year - although it may reflect several job positions). Then we add earnings from all the different job positions within this occupation to calculate the annualized wage. We finally report the characteristics of the job corresponding to this main occupation.<sup>13</sup> This specific treatment implies that some workers fall into a "mixed category," *i.e.* wherein no occupation represents at least two-thirds of annual earnings - for example two job positions which each represents half of annual earnings. We exclude individuals who fall into this mixed category, which represents around 7% of the sample each year.

**Sample restrictions.** After our wage annualizing process, we have about 7.2 million observations left in our 2009-2015 panel. We apply further restrictions to obtain our estimation sample. We first exclude some SPC that are not relevant for our analysis: interns and apprentices (2% of the sample); farmer-operators (250 observations); individuals that earn income primarily from unemployment benefits in a given year (8% of the sample). We winsorize the average net hourly

---

<sup>13</sup>In a few cases, the job positions within an occupation show distinct characteristics. For example, consider a worker that has three jobs during the year corresponding to a given occupation. Two of them in the Media & Entertainment industry and the last in Telecommunication Services. In this case, we retain the industry associated with the most remunerative job position within the specific occupation.

wage, and we exclude wages below 6.5 euros (about 2% of the sample).<sup>14</sup> We only retain workers that have worked a minimum of 150 hours during the year. We also restrict the analysis to workers aged 16 to 70, employed in metropolitan France. After deleting observations for which control variables are missing, the resulting estimation sample consists of about 740,000 unique individual identifiers, with 4 million observations for the period 2009-2015 (corresponding to an average of 570,000 workers each year). Table 1.2 reports descriptive statistics of this estimation sample.

Table 1.2: Descriptive statistics – Estimation sample.

	Mean	Std. Dev.	Median	Obs.
Net hourly wage	13.57	6.62	11.48	
SMEs business owners and CEOs	0.66%			26,319
White-collar workers	16.06%			639,077
Intermediate Professions	21.47%			854,291
Skilled Employees	20.80%			827,488
Skilled Blue-collar workers	19.43%			773,24
Unskilled Employees	13.77%			547,799
Unskilled Blue-collar workers	7.81%			310,763
Experience since 2002 (in years)	7.80	3.37	8	
Tenure at firm (in years)	4.39	4.64	3	
Agriculture	1.09%			43,564
Manufacturing, Mining and other industries	13.41%			533,669
Construction	6.49%			258,102
Wholesale/Retail Trade, Transport, Accomodation and Food	23.06%			917,49
Information and Communication	2.61%			103,904
Financial and Insurance	3.29%			131,036
Real Estate	0.97%			38,51
Professional, Scientific, and Technical Services, Admin. and Support Services	12.55%			499,423
Public administration, education, human health and social work	31.61%			1,257,736
Other service activities	4.91%			195,543
No education	11.02%			438,566
Primary education	1.98%			78,68
Lower secondary education	5.12%			203,839
Certificate of professional competence	25.97%			1,033,511
Baccalauréat	20.37%			810,468
Higher education (Bac + 2)	17.51%			696,541
Higher education (Bac+3 and more)	18.03%			717,372
Women	49.46%			1,968,061
Men	50.54%			2,010,916
Age	40.61	11.21	40	

Note. This Table presents the descriptive statistics of our estimation sample, 3,978,977 observations over the period 2009-2015.

<sup>14</sup>The minimum net hourly wage (SMIC) was about 6.9 euros in 2009.

**Local labor markets.** The panel contains information regarding location of the job, at the municipality level (smaller French administrative subdivision of *commune*). The geographical level of analysis in this paper is the employment-zone level (*zone d'emploi*). Employment zones (ZE) are commuting zones. An employment zone is a geographical area, delineated by INSEE, within which most of the labor force lives and works, and from which firms hire most of their labor force. It is therefore a geographical scale adapted to local studies, especially to evaluate the influence of local environments on individuals. There are 304 employment zones in metropolitan France.<sup>15</sup> Data at the employment-zone level comes from INSEE exhaustive databases.

## 1.4 Empirical Strategy

In this section, we discuss several identification concerns and present the main empirical strategy aiming to disentangle the effects of density and local human capital on the local wage premium across French employment-zones.

### 1.4.1 OLS two-steps model.

We exploit the panel dimension of our data to implement the two-stage empirical strategy proposed in Combes et al. (2008) with the aim of addressing the issue of spatial sorting. Indeed, people choose their work location and the wage premium is overestimated if individuals observed in densest areas (and/or well-educated areas) are inherently more productive (high unobserved ability typically) than individuals in other areas. In this case, part of the coefficients associated with local density, in a basic Mincerian wage equation, reflects the difference in workers' unobserved ability rather than agglomeration gains.

The two-step procedure consists of a first regression of log wage<sup>16</sup> on time-variant workers and jobs characteristics, zone fixed effects and individual fixed effects that control for the workers'

---

<sup>15</sup>The employment zones are defined as the geography applicable in 2012. INSEE redesigned the boundaries of employment zones in 2010. We adjusted the geography in the former period using the 2010 table of correspondence, which provides a consistent basis for comparing local outcomes over time.

<sup>16</sup>Note that we base our analysis on nominal wage, which is unadjusted for cost of living. As discussed in the literature on agglomeration economies, this is the correct dependent variable to capture the advantages of local labor markets (Combes and Gobillon, 2015). Basically, higher nominal wages reflect higher productivity, if this were not the case, firms and workers would relocate (see Acemoglu and Angrist (2001) and Moretti (2004b) for theoretical discussion).

spatial sorting across zones on unobserved characteristics:

$$\ln wage_{i,z,t} = X_{i,t}\gamma + \theta_z + \beta_t + \mu_i + \epsilon_{i,z,t} \quad (1.1)$$

where  $\ln wage_{i,z,t}$  is the logarithm of the average hourly net wage for worker  $i$  in the employment-zone  $z$  during the year  $t$ ;  $X_{i,t}$  is a vector of time-varying worker and job observed characteristics (labor market experience (since 2002), squared experience, socioprofessionnal category and industry<sup>17</sup>);  $\theta_z$  is the employment zone fixed effect;  $\beta_t$  is a year fixed effect;  $\mu_i$  is an individual fixed effect; and  $\epsilon_{i,z,t}$  an error term.

This first stage sweeps out the effects of observed and unobserved individual characteristics and isolates agglomeration effects in the  $\theta_z$ .<sup>18</sup> The second stage consists in regressing these estimated zone fixed effects on zones' characteristics to evaluate their relative importance in explaining the spatial wage difference.<sup>19</sup> We have a particular interest in the roles of employment density and local human capital:

$$\hat{\theta}_z = \rho + \beta HC_z + \delta \ln Dens_z + X_z\gamma + \mu_z \quad (1.2)$$

where  $X_z$  is a vector of zone-level controls in 2012 and  $\mu_z$  is an error term. Our coefficients of interest are the parameter  $\beta$  associated to  $HC_z$ , the local share of people with at least three years of higher education in 2012 (aged 15 and more, not including those in school), and the parameter  $\delta$  associated with  $\ln Dens_z$  the local job density in 2012 (the log of total employment per km<sup>2</sup>).

The coefficient  $\beta$ , which reflects the local human capital externalities<sup>20</sup> on wages, is the main

<sup>17</sup>Socioprofessionnal categories are based on the hierarchical classification of INSEE with 8 groups: SMEs business owners and CEOs; White-collar workers; Intermediate Professions; Skilled Employees; Skilled Blue-collar workers; Unskilled Employees; and Unskilled Blue-collar workers.

<sup>18</sup>Zone fixed effects are estimated in the first stage on the basis of migrants. As explained in [De la Roca and Puga \(2017\)](#), this can bias the estimation of zone fixed effect if migrants are not representative of the whole population or if the decision of migration depends on shocks specific to a worker-zone pair. A solution is to include in the specification enough individual controls that motivate their choice of location. With individual fixed effects and controls for labor market experience, occupation, industry, we are confident that the estimation of  $\theta_z$  does not suffer from this potential bias.

<sup>19</sup>The introduction of local level explanatory variables in a simple one-step specification would lead to inconsistent estimated coefficients, because of the correlation between the unobserved characteristics of workers and the characteristics of their working zone. See [Combes et al. \(2008\)](#) and [Bosquet and Overman \(2019\)](#) for details on the advantages of this two-step method.

<sup>20</sup>Actually, theoretically speaking, the coefficient  $\beta$  of equation (1.2) does not capture "pure" human capital

parameter of interest in this paper. We also have a particular focus on the parameter  $\delta$ , which captures the elasticity of wages with respect to local density.

We include in the  $X_z$  the two main variables commonly used in the empirical agglomeration literature: market potential and sectoral diversity. Market potential is an indicator of the spatial extent of urbanization gains – because density effects may exceed the locality of the employment zone. To control for the spillovers’ diffusion beyond the limits of the zone we introduce a market potential variable (first developed by [Harris \(1954\)](#)) constructed as the sum of each employment density for each external local labor market divided by the distance separating this market from the given employment zone:  $PM_z = \sum_{i \neq z}^n \frac{Dens_i}{distance_{i-z}}$ . Next, the sectorial diversity of the local economy is also known to have a favorable impact on productivity and wages ([Glaeser et al., 1992](#)), as it facilitates cross-fertilization between industries and protects against industry-specific negative shocks.<sup>21</sup> We control for local diversity using the inverse of a Herfindahl index calculated from the sectoral employment shares<sup>22</sup> at the local level:  $div_z = \left[ \sum_s \left( \frac{emp_{z,s}}{emp_z} \right)^2 \right]^{-1}$ .

#### 1.4.2 Instrumental variables in second step.

**Zone-level endogeneity.** While the two-step procedure takes into account spatial sorting on unobservables (with worker fixed effects), a potential identification problem remains in the second step due to the endogeneity of the size and the composition of each zone. Here, the possibility for workers to choose their work localization may induce reverse causality issues. Indeed, workers may be attracted by dense areas in which they expect higher wages and living comfort (urban amenities, cultural activities, etc.). Hence, it is not only the density or the local hu-

---

externalities. It has potentially two other components. First, as detailed in [Moretti \(2004b\)](#), it captures an additional effect coming from imperfect substitution between skilled and less-skilled workers – which is positive or negative depending on the local education level. Indeed, when the concentration of higher-educated workers in a zone increases, all workers benefit from human capital externalities, but the less-educated, who become relatively fewer, benefit from an additional effect due to the improvement of their marginal productivity. This mechanism works in the opposite direction for higher-educated workers, whose marginal productivity decreases with their concentration. A second additional component of the  $\beta$  may come from the specific demand of the higher-educated people. Skilled workers, better paid, may have high preferences for local non-traded low-skilled services ([Manning, 2004](#); [Mazzolari and Ragusa, 2013](#)), such as domestic services, catering and fast food services, etc. In the absence of significant inter-zone migrations of low-skilled workers, this again translates into a wage premium for these low categories.

<sup>21</sup>However, our estimates show a negative impact of diversity on wages (see Section 1.5). In fact, [Combes and Gobillon \(2015\)](#) indicate that the role of diversity has been shown to be not robust: sometimes positive, negative or not significantly different from zero. We keep it as a necessary control, but we will not comment the coefficients.

<sup>22</sup>We use localized data from ACOSS/URSSAF (organizations for the payment of social security and family benefit contributions) based on the classification of activities NAF 88.

man capital that increase productivity and wages. Higher wages also attract workers, especially highly-educated workers, who are generally more mobile, and create density and/or increase the education level of the local population. The causality is circular. Additional endogeneity concerns may also come from omitted variables or potential time-varying local productivity shocks that would be correlated with both local wages and the employment density or the share of higher-educated workers. Typically, some technological shocks may simultaneously boost wages and attract skilled workers. For example, the aerospace industry in the zone of Toulouse creates an ecosystem of innovations and jobs that simultaneously raises wages and attracts workers (especially higher-educated workers).

The empirical literature on agglomeration gains addresses these endogeneity concerns using instrumental variables (IV) and does not find significant difference with the OLS model (Ciccone and Hall, 1996; Moretti, 2004b; Combes et al., 2008, 2010; De la Roca and Puga, 2017). The main approach consist of using instruments based on historical lag values of local density or local education variables. In equation (1.2), the variables of density, local human capital and market potential present risks of endogeneity bias. We propose two IV models based on two different sets of instruments for these three variables. The first one (IV1) relies only on employment-zone-level data, whereas the second (IV2) also use departement-level information, which implies losing some areas due to lack of data (see below).<sup>23</sup>

**IV1.** Our first set of excluded instruments includes seven variables. Two instruments for the density variable: the historical value of local density in 1806 and the local density of religious buildings built before the 20th century. Two instruments for the local human capital variable: the change in local share of higher-educated people between 1968 and 1982 and a shift-share type instrument *à la* Moretti (2004b) based on lagged age-structure and national education levels by age group. Three instruments for market potential variable: a lagged market potential based on historical value of local density in 1806, the average distance to all other employment zone and the distance to Paris.

---

<sup>23</sup>Since our model is based on very common instruments, whose relevance is already well commented and justified in the agglomeration literature, we do not detail here the intuitions for each variable (refer to the Appendix .2 for more details).

**IV2.** We next propose a second IV model based on ten excluded instruments. The main reason for this is that we introduce alternative instruments for local human capital variable, to replace the shift-share instrument *à la* Moretti (2004b). Because of the recent debate on the relevance of Bartik shift-share instruments (see Goldsmith-Pinkham et al. (2020)), we want to ensure that our IV results do not overly depend on such instrument. We therefore drop this instrument and use instead two new instruments based on historical departmental data:<sup>24</sup> the number of secondary students in 1837 compared to departmental population and the share of conscripts with *baccalauréat* diploma in 1896. Due to lack of data, the IV2 model with departmental-level instruments is based on 290 zones only<sup>25</sup>. As we are now using data at the department-level, we also add two new instruments for local density in this IV2 model: the number of royal road kilometers in 1837 in the department and the ratio of this number to the total surface of the department. In this model, because the residuals (employment zone level) are likely to be correlated within department, we cluster standard errors by department.

## 1.5 Results

We present and discuss in this section the estimation results of the second-step regressions based on equation (1.2) reported in Table 1.3. We do not comment the first-step estimates from which we recover the estimated zone fixed effects, used as dependent variable in the second step (see Table 9 for first-step estimation results).

### 1.5.1 Main Results

Looking at the specifications which include only the log of density in second step (columns (1) and (4) of Table 1.3), the introduction of worker fixed effects in the first step causes a drop of

---

<sup>24</sup>The French territorial collectivity of *département* ("*département*") was created during the French Revolution. These jurisdictions correspond to constant spatial units with local legislation and administration – the intermediate level of government between the administrative regions and the communes. Since 1975, there are 96 departments in metropolitan France. The department is then a larger geographical unit than the employment-zone, with on average five employment-zones per department. However, this number varies depending on the size of both employment-zones and departments. For example, the employment-zone of Limoges fills the entire department of Haute-Vienne, while the big department of Seine-et-Marne contains ten small employment-zones – some straddling other neighboring departments. When an employment-zone straddles several departments, we attribute to this zone the department of its largest city of influence.

<sup>25</sup>The missing zones correspond to the areas that belonged temporarily to Germany or Italy during the 19th century.

almost half of the wage premium with respect to density (from 0.02 to 0.01). This is perfectly in line with previous empirical literature. [Combes et al. \(2008\)](#) and [Combes et al. \(2015\)](#) on French data, and [De la Roca and Puga \(2017\)](#) on US data, show a comparable 50% drop in the elasticity of wages with respect to local density when taking into account potential spatial sorting on unobservables.<sup>26</sup>

Taking into account spatial sorting bias, we estimate in column (4) an elasticity of wage with respect to density of 0.01 – significant at the 1% level. This is in the lower range of previous empirical literature. The meta-analysis of [Melo et al. \(2009\)](#) documents an elasticity range between 0.012 and 0.143 in the case of France. With comparable methods on French data, but older periods, [Combes et al. \(2008, 2010\)](#) find elasticities around 0.02-0.05, depending on the specification, and [Combes et al. \(2015, 2016\)](#) find an elasticity of about 0.01 when including additional controls for local endowments, such as market potential and sectorial specialization and/or diversity.

When we jointly estimate the influence of density and local human capital, also including our two additional controls for sectoral diversity and market potential,<sup>27</sup> the coefficient associated with density is no longer significant.<sup>28</sup> In line with [Combes et al. \(2011\)](#) and [Chauvin et al. \(2017\)](#), this suggests that wage premium in dense areas is much more linked to local human capital than pure density effects.<sup>29</sup>

Our preferred regression in column (6) of Table 1.3 results in a semi-elasticity of wages with respect to local human capital of 0.130 (significant at the 1% level).<sup>30</sup> In other words, all things

---

<sup>26</sup>While it is always substantial, note that this drop in the elasticity when including individual fixed effects is sometimes smaller, depending on the method and the data. For example, [Combes et al. \(2010\)](#) on French data show a reduction in the value of the elasticity by about one-third.

<sup>27</sup>While we may anticipate multicollinearity issues when introducing all our zone-level controls in the specification, post-estimation diagnostics indicate that these agglomeration variables are not too similar to doubt of the ability of the regression to estimate the effects separately. Each variance inflation factors (VIF) check rejects a perfect linear relationship among our local predictors – above the minimum tolerance value of 0.1 (1/VIF). VIF are 1.95 for log density, 1.61 for local HC, 1.26 for log market potential and 1.22 for log diversity.

<sup>28</sup>This coefficient associated to density also falls to 0.004 when we add only the variables of diversity and market potential, but remains significant at the 1% level – it is no longer the case when local human capital variable is included.

<sup>29</sup>[Melo et al. \(2009\)](#) have also shown that controlling for differences in the level of time-variant human capital reduces the size of the elasticity of urbanization by about 5 to 6 percentage points. Nevertheless, in our regression, density still plays an important role through the market potential effect, which reflects gains due to the proximity to other very dense local labor markets – also a measure of centrality, as the very dense areas are concentrated in the Parisian region.

<sup>30</sup>We also find a substantial drop in the coefficient associated to human capital when we introduce individual fixed effects in first step. When local human capital is introduced separately, the coefficient drops by about

being equal, a one percentage point increase in the local share of highly-educated people is associated with a 0.13% increase in the individual wage. With a gap of 11 percentage points in 2012 between the top 5% and the bottom 5% of employment-zones regarding local human capital, we expect a wage premium of 1.5% in a zone with high level of human capital compared to a zone with few graduates.

Although it is difficult to compare with previous empirical literature because of diverse data and designs, our estimated coefficients seems much smaller than usual findings on US data. In his main analysis of social return to higher education using US Census data for 1980 and 1990, [Moretti \(2004b\)](#) finds a semi-elasticity of wage with respect to the share of college graduates between 0.6 and 1.2 – but his main specifications do not include controls for other local factors such as density. On aggregate data, [Chauvin et al. \(2017\)](#) conclude on core human capital externality coefficients about 1.1 in the U.S. in 2010, when log of earnings is regressed on the share of adults with a college degree or more, controlling only for individuals' education level and age.

However, some studies find smaller human capital effect. [Bentsen et al. \(2019\)](#) find very small human capital external effect between 0.005 and 0.016 when focusing on spillovers at the workplace. [Combes et al. \(2016, 2015\)](#), presumably the most comparable data with the one in this paper, include the intra-industry share of white-collar workers (who are generally highly-educated workers) in their specifications. This share is associated with a wage semi-elasticity of about 0.04. We may think that our larger coefficient associated with local human capital suggests that education spillover effects also derive from outside the workplace (or not only from intra-industry cross-fertilization effects).

Columns (7) and (8) report the estimation results of our two IV regressions. Post estimation statistics and the first stage estimates (presented in [Table 10](#)) confirm the relevance of our two sets of instruments. While slightly higher in magnitude, the coefficients associated with local human capital are not significantly different from the OLS regression in column (6): about 0.16

---

60% when first step include individual fixed effects (0.200 in column (5) compared to 0.507 in column (2)). However, one concern remains regarding the magnitude of this coefficient due to the fact that zone fixed effects are estimated in first-step on the basis of more observations for dense and well-educated areas. This can introduce heteroscedasticity through sampling errors (see [Combes et al. \(2008\)](#) and [De la Roca and Puga \(2017\)](#) on this issue). Using Feasible Generalized Least Squares to deal with this concern leads to a very similar coefficient of 0.127 (significant at the 1% level).

with standard errors about 0.04. This is perfectly in line with the previous empirical literature showing that the potential endogeneity of local density or human capital is not an important source of bias in the estimation of agglomeration gains (Moretti, 2004b; Combes et al., 2010; De la Roca and Puga, 2017; Chauvin et al., 2017).

Table 1.3: Determinants of agglomeration gains on individual wages.

	Zone FE coeff. from first-step (sorting bias)			Zone FE coeff. from first step (control for sorting)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	OLS	OLS	OLS	IV1	IV2
Local Human Capital	-	0.507*** (0.0400)	0.296*** (0.0409)	-	0.200*** (0.0301)	0.130*** (0.0276)	0.164*** (0.0370)	0.163*** (0.0411)
ln Density	0.022*** (0.0015)	-	0.011*** (0.0018)	0.010*** (0.0012)	-	0.001n.s. (0.0012)	-0.001 n.s. (0.0024)	-0.003 n.s. (0.0018)
ln Diversity	-	-	-0.010*** (0.0058)	-	-	-0.013*** (0.0039)	-0.010** (0.0041)	-0.009** (0.0045)
ln Market Potential	-	-	0.020*** (0.0026)	-	-	0.027*** (0.0018)	0.023*** (0.0062)	0.027*** (0.0027)
Worker fixed effects in first-step	No	No	No	Yes	Yes	Yes	Yes	Yes
<i>Under id. LM stat – p-value</i>							0.000	0.001
<i>Weak id. Wald F stat.</i>							40.342	30.619
<i>Sargan/Hansen J-Stat – p-value</i>							0.000	0.353
<i>Endogeneity test – p-value</i>							0.004	0.357
Observations	304	304	304	304	304	304	304	290
R <sup>2</sup>	0.41	0.42	0.57	0.19	0.13	0.55	0.54	0.54

Notes. This Table presents estimation results of our second-step regressions based on equation (1.2). The dependent variable is the 304 estimated coefficients of zone fixed effects included in the first-step estimations based on equation (1.1). Each column is a separate regression. Columns (1) to (3) correspond to second-step estimations when the first-step does not include individual fixed effects – and therefore suffer from potential spatial sorting bias (see column (2) of Table 9 for first-step estimates). Columns (4) to (8) take into account spatial sorting in first step (with worker fixed effects, column (1) of Table 9). The first-step regression always includes controls for work experience since 2002 (and squared), tenure at firm, socioprofessional category, two-digit industry, year indicators and zone fixed effects. In addition, the first step without worker fixed effects includes controls for time-invariant worker’s characteristics: age, age-squared, gender and level of education – those variables are omitted when introducing workers fixed effects. Local human capital, density and market potential are instrumented in column (7 - IV1) and (8 - IV2). Excluded instrument for IV1 model are: density in 1806; density of historical religious buildings; the change in local human capital 1968-1982; shift-share local human capital; market potential in 1806; average distance to all other ZE; and distance to Paris. IV2 drops the shift-share local human capital instrument and add four instruments at the departement-level: the number of secondary students in 1887 compared to population; the share of conscripts with *baccalauréat* diploma in 1896; the number of royal road kilometers in 1837; and the ratio of the number of royal road kilometres in 1837 to the total surface of the departement. Due to lack of information in the historical departement-level database, IV2 regression is based on 250 observations. Reported Wald F-statistics, for weak instruments identification test, exceed the Stock and Yogo (2005) critical values for a 5% maximal IV relative bias (13.95 for IV1 and 16.80 for IV2). \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

## 1.5.2 Robustness

We present and discuss in this section several alternative specifications designed to test the sensitivity of our main results, regarding potential data, sample and control variables issues. All the estimates results are reported in the summary Table 1.4.

**Industry exclusions.** We first run sensitivity checks by excluding several industries that may bias our results due to problems with data quality or with their relevance in the analysis of agglomeration gains.

First, certain large companies group together on a single administrative declaration, work

Table 1.4: Determinants of agglomeration gains on individual wages – Robustness checks synthesis.

Dependent variable: estimated ZE fixed effects from first step equation (1.1)					
		Human Capital	ln Dens.	ln Div.	ln MP
	Obs. in first step	(1)	(2)	(3)	(4)
(a) Base Model	3,978,977	0.130*** (0.0276)	0.001n.s. (0.0012)	-0.013*** (0.0039)	0.027*** (0.0018)
(b) industry exclusions	2,456,620	0.134*** (0.0311)	0.002 n.s. (0.0014)	-0.013*** (0.0044)	0.023*** (0.0020)
(c) private sector workers	2,700,392	0.124*** (0.0306)	0.002 n.s. (0.0013)	-0.010** (0.0043)	0.023*** (0.0020)
(d) full-time workers	3,057,269	0.142*** (0.0298)	0.002 n.s. (0.0013)	-0.016*** (0.0042)	0.027*** (0.0019)
(e) women	1,968,061	0.131*** (0.0305)	-0.002 n.s. (0.0013)	-0.015*** (0.0043)	0.023*** (0.0020)
(f) men	2,010,916	0.125*** (0.0356)	0.003* (0.0015)	-0.011** (0.0050)	0.030*** (0.0023)
(g) without immigrants	3,654,439	0.145*** (0.0283)	0.001n.s. (0.0012)	-0.013*** (0.0040)	0.027*** (0.0018)
(h) without Parisian areas	3,234,437	0.179*** (0.0369)	-0.000 n.s. (0.0013)	-0.016*** (0.0045)	0.023*** (0.0027)
(i) zone x year fixed effects	2,177,567	0.125*** (0.0153)	0.001 n.s. (0.0006)	-0.015*** (0.0021)	0.026*** (0.0009)
(j) database most paid job position	3,460,847	0.125*** (0.0299)	0.001 n.s. (0.0013)	-0.013*** (0.0042)	0.028*** (0.0019)
(k) without SPC indicators in first step	3,978,977	0.128*** (0.0291)	0.001n.s. (0.0013)	-0.013*** (0.0019)	0.031*** (0.0019)
(l) controls for lower HC shares	3,978,977	0.138*** (0.0283)	0.001 n.s. (0.0015)	-0.011*** (0.0042)	0.027*** (0.0019)
(m) additionnal local controls	3,978,977	0.134*** (0.0340)	0.001 n.s. (0.0014)	-0.010** (0.0049)	0.027*** (0.0019)

Notes. This Table presents robustness estimation results for several alternative second-step regressions based on equation (1.2). The dependent variable is the 304 estimated coefficients of zone fixed effects included in the first-step estimations based on equation (1.1). Each line is a separate regression. Each column corresponds to second-step coefficients associated to one of our four zone-level variables. The first step regression always includes controls for work experience since 2002 (and squared), tenure at firm, socioprofessional category (except line(k)), two-digit industry, year indicators, worker fixed effects and zone fixed effects. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

positions that are actually localized in several establishments of the firm. This creates a bias in locating workers at their place of work (about 5% of workers). Since 2003 validity, INSEE has adjusted this bias using the information system CLAP (*Local knowledge on establishment production*) to unbundle the work periods and reassign workers to their correct establishments (in particular by using workers' place of residence). However, a few gaps remain for about 0.5%<sup>31</sup> of the workforce because of the complexity of the unbundling algorithm. Workers may still be located at the site of the headquarters (and not at the correct work place), especially in financial sectors or public companies such as SNCF and La Poste. To ensure the robustness

<sup>31</sup>Information provided in *INSEE DADS Grand Format - Guide Utilisateur Validité 2014*.

of our results, we exclude Finance and insurance and Public administration from our sample. In addition, we follow other recommendations on data discussed in [Combes et al. \(2008, 2011\)](#), by excluding some industries that are either irrelevant for the study of agglomeration gains, or present information issues. This concerns Agriculture, forestry and fisheries, Spatial transport and Extraction of uranium and metals.

Line (b) of Table 1.4 shows the estimates for this specification with excluded industries (Finance and insurance; Public administration; Agriculture, forestry and fisheries; Spatial transport; Extraction of uranium and metals). Coefficients show no significant difference compared to our base model – although the coefficient associated with market potential is slightly smaller in value. We also report in line (c) a specification that keeps all industries but excludes all the public-sector workers. While we expect less agglomeration gains in these regulated public industries with specific career paths, we see no significant difference compared to our main results.

**Full-time workers.** While the dependent variable in the first-step regression is the hourly wage (annual wage divided by the number of hours worked), we run an alternative specification excluding part-time workers, who may be a source of identification bias. Line (d) of Table 1.4 shows no significant difference with the base model, which includes both full-time and part-time workers.

**Women and Men sub-samples.** We next explore whether there is heterogeneity in agglomeration gains depending on men or women sub-samples. Lines (e) and (f) of Table 1.4 show no significant difference, except regarding the wage gain associated with market potential, which seems to be larger for men workers (0.030 vs. 0.023). The density also shows a larger effect for men, but this effect is not very significant in view of the standard errors (10% threshold).

**Immigrants.** If immigrants workers, who generally experience discrimination in the labour market, are not equally dispersed over employment zones (typically more localized in big cities), agglomeration gains may be underestimated. Line (g) of Table 1.4 presents an alternative specification that excludes immigrant workers from our base sample. We do not find significant

difference with our base model.

**Excluded zones around Paris.** We next run a robustness check to test the sensitivity of our main results to the exclusion of very dense, and highly-educated, zones around Paris, in the Ile-de-France region.

The Ile-de-France region is characterized by a very strong polarization of employment and numerous migration flows between municipalities. Consequently, in the case of Ile-de-France, INSEE's method used to define other employment-zones would lead to create only one big employment zone for the whole region. More detailed work on commuting flows and distances results on the actual set of nineteen Parisian zones<sup>32</sup> that are more coherent with the basic definition. However, because of their specificity regarding labor market dynamics and the complexity of local interactions between these different areas, and given the large number of individuals in our sample working in them, the presence of these areas may overly influence our results, if only because of their distinctive design algorithm.

Line (h) of Table 1.4 reports estimation results of the specification when we exclude the nineteen Parisian zones of our sample. We find here a substantial increase in the coefficient associated to local human capital variable – while other coefficients remain relatively stable compared to our base model.

**Zone  $\times$  year fixed effects.** We run an alternative model in which we include a zone-year fixed effect in first step equation (1.1) – instead of year and zone fixed effects included separately – in order to explore the possibility of agglomeration gains being time-variant. The second step is to regress these estimated zone-year fixed effects on our local variables of interests (time-varying), adding year indicators. Since we only have information about the number of highly-educated (Bac +3 and more) individuals up to 2012, this model is restricted to the period 2009-2012. Results presented in line (i) of Table 1.4 show no significant difference with our base model.

**Alternative wage annualization.** We also assess the robustness of our results to a change in the computation method of annualized nominal wage. We recreate an alternative database

---

<sup>32</sup>Paris, Marne-la-Vallée, Coulommiers, Meaux, Melun, Montereau-Fault-Yonne, Nemours, Provins, Houdan, Mantes-la-Jolie, Poissy, Rambouillet, Plaisir, Étampes, Évry, Saclay, Créteil, Orly and Cergy.

using a more common methodology, which consists in keeping only the position corresponding to the highest annual remuneration, when the individual has several work positions during the year. In this procedure, we retain only full-time workers. Line (j) of Table 1.4 reports estimation results using this alternative database. Estimates are very similar to the base model.

**Endogeneity of occupations.** The socioprofessionnal category (mainly based on occupation, but not only) of one worker is likely to be influenced by the characteristics of her local labor market. In particular, local human capital externalities increase the chance of promotion over the working life (see chapter 2 of these thesis). Our main model, which includes SPC indicators in first step, may suffer from endogeneity since occupations are attached to jobs and are jointly determined with wages (Combes and Gobillon, 2015; Combes et al., 2016). Therefore, this inclusion of socioprofessionnal-category indicators may be a "bad control" (see Angrist and Pischke (2008)) because agglomeration effects on wages are likely to come, in part, from the fact that workers can access better jobs in dense or highly educated areas. Line (k) of Table 1.4 presents estimation results when we do not include SPC indicators in the first step regression. We do not find significant difference compared to the base model.

**Controls for lower human capital shares.** In our main model, the second step regression includes the share of highly-educated people, which refers to individuals with at least 3 years of education after the *baccalauréat*. We do not include the shares of lower educated groups because we seek to measure the effect of the concentration of high human capital, regardless of the composition of the rest of the local population (this also facilitates the interpretation of the coefficients and the comparison with the literature after Moretti (2004b)). However, geographical difference in wages may also be influenced by the difference in the share of people with lower level of education.

Line (l) of Table 1.4 presents alternative estimates of our second-step regression when we introduce additional controls for the local share of people with baccalauréat degree (aged 16 and more, not including those in school), and the local share of people with one or two years (at most) of higher education (aged 16 and more, not including those in school). The coefficient of our main HC variable (share of people with at least three years of higher education) is very

robust to the inclusion of the controls for lower HC – which do not significantly influence wages.

**Additional zone-level controls.** Finally, we check the sensitivity of our main results to controlling for additional zone-level variables. Line (m) of Table 1.4 present alternative form of our second-step regression where we add controls for the surface ( $\text{km}^2/1000$ ) and the industry specialization, i.e. the employment shares of the different aggregate sectors in the total employment of the zone (sectors: Construction, Industry and Manufacturing – the share of the category Trade and Services is omitted). Our coefficients of interest do not significantly change with the introduction of these additional controls.

\*\*\*

**Summary.** We find strong evidence that agglomeration gains on wages, over the period 2009-2015 across French employment-zones, are more related to human capital effects, rather than pure urbanization effects – although the proximity to other very dense areas, Parisian region in particular, remains an important determinant of wage gains.

## 1.6 "Learning by working in well-educated labor markets"

In the previous section, we investigated the impact on wages of working in employment-zones where the human capital is abundant. We have taken into account the spatial sorting of workers by including worker fixed effects in our wage equation. But in doing so, we risk to control not only for the intrinsic unobserved ability of workers, but also for the ability acquired over time, with the experience that workers accumulate in some specific zones (especially the zones where the human capital is abundant). This would bias downward our estimation of the overall human capital externalities. The aim of this section is to re-introduce the advantage of experience accumulated in highly-educated areas and see how it changes our estimates of local human capital externalities.

We follow the methodology developed by [De la Roca and Puga \(2017\)](#), who show that when introducing worker fixed effects in the first-step wage equation, the drop – in second step – in the estimated elasticity of wage with respect to local density can be explained by two mechanisms: the workers' sorting on unobservables or the dynamic agglomeration gains in denser areas (in fact

a combination of both). We suspect that similar mechanisms operate in the case of wage premium with respect to local human capital. First, at the same level of education, workers who sorted in well-educated areas may have higher intrinsic unobserved ability (sorting component). Next, because well-educated areas offer more settings for learning, experimentation and knowledge spillovers, workers may accumulate more valuable experience that remains an advantage when they relocate in other area (experience component).

### 1.6.1 Accumulated experience in highly-educated areas

The basic idea is to allow the effect of labor market experience to vary depending on where it is acquired. To this end, we construct two new variables that represent the experience accumulated by each worker in two groups of "top areas" since 2002. The first group of "top areas" is the 5% most highly-educated areas in 2009 (the 15 employment-zones where local human capital ratio is the highest); the second group is the next top 5% most highly-educated in 2009 (the 15 next employment-zones in the distribution).<sup>33</sup> We include these variables of experience in top areas (and their square) in the first-step equation (1.1):

$$\ln wage_{i,z,t} = \sum_{g=1}^2 [\sigma_g Exp_{i,g,t} + \omega_g Exp_{i,g,t}^2] + X_{i,t}\gamma + \rho Exp_{i,t} + \tau Exp_{i,t}^2 + \theta_z + \beta_t + \mu_i + \epsilon_{i,z,t} \quad (1.3)$$

$Exp_{i,g,t}$  is the experience accumulated by worker  $i$  until time  $t$  in a zone belonging to the group  $g$  (either the top 5% areas or the next top 5% areas in the distribution of zones regarding local human capital in 2009).  $Exp_{i,g,t}^2$  is the square of this variable. We keep all the control included in the equation (1.1), in particular, we still include the overall experience in the specification ( $Exp_{i,t}$  and  $Exp_{i,t}^2$  – listed here outside of the  $X_{i,t}$ ).

Table 1.5 column (1) present the estimated results of this first-step regression. Accumulated experience in areas with abundant human capital has a positive and significant impact on workers' wages. One year of experience in a top 5% area raises wages by 0.7% – relative to having worked one year in a zone below the top 5% areas. One year of experience in the next 5%

---

<sup>33</sup>The list of employment-zones in 2009, sorted by human capital intensity. TOP 5% : Lille, Lyon, Paris, Saclay, Strasbourg, Marne-la-Vallée, Poissy, Montpellier, Plaisir, Grenoble, Toulouse, Nice, Aix-en-Provence, Houdan and Rambouillet. NEXT TOP 5% : Marseille - Aubagne, Créteil, Orly, Cergy, Nantes, Cannes - Antibes, Bordeaux, Rennes, Dijon, Nancy, Annecy, Besançon, Corte, Genevois Français and Melun.

top-educated areas raises wages by only 0.3%.

We next recover the estimated coefficients of the zone fixed effects  $\hat{\theta}_z$  and run the second-step regression based on equation (1.2) aiming at estimating the static agglomeration gains. The new semi-elasticity of wages with respect to the local human capital is about 0.11 (see columns (3) of Table 1.5). In view of the standard errors, this is very similar to the previous estimates in column (6) of Table 1.3 (which include the overall experience in first-step, regardless of where it is acquired). As detailed in De la Roca and Puga (2017), when we do not control specifically for accumulated experience in top areas, the bias in the static fixed effects estimates is expected to be small if the direction of migration flows are balanced – which is the case in our sample – and the learning benefits of top areas are portable.

Table 1.5: Human capital externalities – accumulated experience and medium term gains.

	First-step		Second-step						
	ln Wage		Zone FE coef.	Medium-term gains			Medium-term gains		
			from (1)	from (1)			from (2)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	OLS	OLS	IV1	IV2	OLS	IV1	IV2
Local Human Capital			0.109*** (0.0272)	0.334*** (0.0289)	0.347*** (0.0358)	0.350*** (0.0451)	0.267*** (0.0281)	0.286*** (0.0348)	0.289*** (0.0432)
Exp. in top 5% HC	0.010*** (0.0019)	0.007*** (0.0013)							
(Exp. in top 5% HC) <sup>2</sup>	-0.0005*** (0.0001)	-0.0002*** (0.0001)							
Exp. in next top 5% HC	0.005*** (0.0007)	0.004** (0.0008)							
(Exp. in next top 5% HC) <sup>2</sup>	-0.0003*** (0.0000)	-0.0002*** (0.0000)							
Exp. in top 5% Dens.	-	0.006*** (0.00151)							
(Exp. in top 5% Dens.) <sup>2</sup>	-	-0.0003*** (0.0001)							
Exp. in next top 5% Dens.	-	0.004*** (0.0009)							
(Exp. in next top 5% Dens.) <sup>2</sup>	-	-0.0003*** (0.0000)							
Experience	0.031*** (0.0010)	0.031*** (0.0010)							
Experience <sup>2</sup>	-0.0002*** (0.0000)	-0.0002*** (0.0000)							
Worker fixed effect	Yes	Yes							
Observations	3,978,977	3,978,977	304	304	304	290	304	304	290
R2	0.92	0.92	0.53	0.68	0.68	0.67	0.64	0.63	0.64

Notes. This table presents the estimation results of dynamic effects associated with accumulated work experience in top-educated areas since 2002. Variables of experience are computed during the period 2002-2015. Each column is a separate regression. Columns (1) and (2) correspond to the first step equation (1.3), including controls for tenure at firm, socioprofessional category, two-digit industry, year indicators, worker fixed effects and zone fixed effects. Second step in column (3) regresses the estimated coefficients of zone fixed effects from column (1) on the local human capital variable. Columns (4) to (9) present the estimated medium-term human capital externalities (for an average period in one zone of 6.2 years). All second step regressions include the additional controls for log density, log market potential and sectoral diversity.

IV1 and IV2 are the instrumental variable estimations where local human capital, density and market potential are instrumented (see previous section 1.4 for list of the excluded instruments and Table 10 for 2SLS first stage estimations).

Coefficients are reported with clustered standard errors at the zone level in parenthesis. IV2 regression are clustered at the department-level. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

### 1.6.2 Medium-term human capital externalities

We now explore the size of the dynamic human capital externalities. Following [De la Roca and Puga \(2017\)](#), we consider a medium-term period and reintegrate in each estimated zone fixed effect  $\hat{\theta}_z$ , the dynamic component of human capital externalities, that is the advantages coming from accumulated experience in top-educated areas. Practically, taking the estimated fixed effect  $\hat{\theta}_z$  of one zone  $z$  from equation (1.3), we add the estimated gain  $\hat{\rho}$  from overall experience, plus the additional gain  $\hat{\sigma}_g$  if the zone  $z$  belonging to the group  $g$ , considering the average experience in a single location for a worker in our sample (6.2 years). For example, considering a zone of the top 5% in terms of local human capital, we obtain a medium-term zone's effect taking the following form:

$$MTeffect_{z,top5\%} = \hat{\theta}_z + \underbrace{(6.2 \times \hat{\sigma}_{top5\%}) + (6.2^2 \times \hat{\omega}_{top5\%}) + (6.2 \times \hat{\rho}) + (6.2^2 \times \hat{\tau})}_{\text{wage gains from accumulated experience in } z}$$

**Separating accumulated experience gains from sorting on intrinsic ability.** Column (4) of Table 1.5 shows that the local human capital concentration naturally increases the medium-term effect of the zone on wages. Compared to the static gains in column (3), the value of the coefficient associated to local human capital in column (4) is about three times larger. This implies that one-third of our measurement of human capital externalities comes from static effects, whereas the other two-thirds are due to dynamic effects. Moreover, the coefficient of 0.334 in column (4) of Table 1.5 is not significantly different from the 0.296 coefficient in column (3) of Table 1.3 – i.e. when we do not control for unobserved ability of workers in the first-step regression. This suggests that the drop in the coefficient when introducing workers fixed effects in first-step comes essentially from dynamic effects.

A last test proposed by [De la Roca and Puga \(2017\)](#) to corroborate this results is to regress equation (1.3) without worker fixed effects  $\mu_i$ . After recovering the estimated  $\hat{\theta}_z$  and regress them on the local human capital variable, we obtain a semi-elasticity with respect to local human capital of 0.129 (significant at the 1% level). Very similar to the semi-elasticity of 0.130 in column (3) of Table 1.3. Therefore, including individual fixed effects or including controls for accumulated experience in top areas, leads to the same drop in the estimated value of human

capital externalities – suggesting again that the drop was primarily due to dynamic effects, rather than sorting on intrinsic ability.

**Quantifying overall human capital externalities over a medium-term period.** We have reintegrated both the static and the dynamic effects into the medium-term zone effect  $MTeffect_{z,g}$ . In section 1.5, we considered that the semi-elasticity of wages with respect to local human capital was about 0.13. This reflected purely-static advantage and therefore underestimated the overall benefits of working in well-educated labor markets. Now, the joint estimation of static and dynamic components shows a semi-elasticity of wages with respect to local human capital of 0.33 – considering a period of 6.2 years. In other words, considering that workers stay about 6 years in their employment-zone, one percentage point difference in the local share of graduates is associated with a 0.33% increase in their wages in the medium-term. With a gap of 11 percentage points in 2012 between the top 5% and the bottom 5% of employment-zones with respect to local human capital, we expect a wage premium of about 4% for a worker that stays six years in a well-educated area, compared to a worker in a zone with few graduates.

**Disentangling the experience in well-educated areas from experience in denser areas.**

One concern remains in the estimation of the medium term human capital effects. We know that top areas in terms of local human capital are also frequently very dense employment-zones (see Table 20 in Annex .3). Therefore it is not totally clear that our variables of accumulated experience in top 5% (and next top 5% areas) capture only learning effects that are specific to highly-educated zones. It may also capture dynamic advantages by working in very dense cities. To disentangle those effects, we include in equation (1.3) additional variables of accumulated experience in top areas with respect to density. We therefore estimate the coefficients  $\sigma_g$  and  $\omega_g$  at a given level of accumulated experience in the densest areas. Estimation results are presented in column (2) of Table 1.5. As expected, the coefficients associated to the accumulated experience in highly-educated areas decrease (from 0.010 to 0.007 for the top 5% areas and from 0.005 to 0.04 for the next top 5%). Finally, column (7) reports the estimates of the second-step regression of the medium-term zone effect purged of accumulated experience in denser areas. The magnitude

of human capital externalities over a medium-term period is slightly reduced (with a coefficient of 0.267), but our conclusions remain essentially the same.

\*\*\*

**Summary.** When we estimate jointly the static and dynamic local human capital externalities, we find a substantial medium-term wage premium. Overall, over a period of six years, one percentage point difference in the local share of graduates is associated with a 0.33% increase in their wages. Purely static human capital externalities account for about one third of this overall effect (0.11), the rest comes from dynamic learning effects.

## 1.7 Local human capital externalities: separating private and public-sector higher-educated workers

Public administration is a major fraction of employment in France. In our sample the share of public sector workers is about 30%. Hence, the geography of public employment has emerged in the public debate as an important issue for the attractiveness and the local development of territories. There are two dimensions here. The first dimension concerns equal access to public services (education, healthcare, etc.) for individuals, regardless their localization. The second one concerns the possibility to (re)locate some general purpose public administration in some specific disadvantaged areas, with the hope to sustain local economic activities while reducing the costs of part of the public administration (in particular real estate costs). Both these policies have an impact regarding human capital concentration as public services employ a larger share of higher-educated workers than the rest of the economy.

In this section, we explore the role of local human capital externalities according to whether they come from private or public highly-educated workers. The Panel *tous salariés-EDP* includes all type of salaried workers and gives information on their education attainment. Despite representativeness problems at the employment-zone level, this gives us the possibility to reconstruct a local human capital ratio variable from our sample, distinguishing between public and private sector workers.

We first compute, from our sample in the year 2012, a new ratio of higher-educated workers

to the total employment at the local level. This new local human capital variable is relatively higher on average (11% – see Table 1.6), but very correlated with the previous variable computed from INSEE exhaustive data (correlation coefficient of 84% – see Table 1.7). While not perfect, we therefore believe that the reconstruction of human capital ratios at the local level with our sample gives a good approximation of the true local compositions. Nonetheless, because of this approximation, we estimate a slightly lower coefficient associated to local human capital compared to our main model (about 0.09, see column (1) of Table 1.8), but still significant at the 1% level.

Table 1.6: Comparison of local Human Capital variables.

	Mean	Std. Dev.	Min.	Max.
Local HC (main)	0.09	0.04	0.04	0.33
Local HC (sample)	0.11	0.03	0.06	0.30
Local HC public workers	0.06	0.02	0.02	0.28
Local HC private workers	0.05	0.03	0.01	0.22

Notes. This table presents the means, standard deviation, minimum and maximum of our different measures of local human capital. HC (main) is our previous local human capital variable (computed from exhaustive INSEE data); HC (sample) is the local human capital variable computed from our sample; HC public is the ratio of higher-educated public-sector workers in the total employment of the zone (computed from our sample); HC private is the ratio of higher-educated private-sector workers in the total employment of the zone (computed from our sample).

We then split this measure of human capital concentration into two components: public-sector component and private-sector component. In practice, for each zone, we compute the ratio of the number of highly-educated worker (Bac + 3 and more) in the public sector (respectively private) to total employment of the zone. Table 1.6 statistics show that, on average, highly-educated public-sector workers represent 6% of the local workforce, whereas highly-educated workers in the private sector represent on average 5%.

Statistics reported in Table 1.6 also show that, the ratio of private HC is much more dispersed (coefficient of variation about 60%<sup>34</sup>) than the ratio of public HC (coefficient of variation of about 30%). This is in line with [Dherbécourt and Deshard \(2019\)](#), who show that the share of French public employment in total employment is quite homogeneous and does not vary much with the

<sup>34</sup>Standard deviation divided by mean.

density or the size of the employment-zone. Moreover, the two ratio are very weakly correlated (their coefficient of correlation is about 0.05 – see Table 1.7) and the variance of the total local human capital ratio across employment-zones is explained for 2/3 by the variance of private workers ratio.

Table 1.7: Correlations between local determinants of agglomeration.

	Local HC (main)	Local HC (sample)	Local HC public	Local HC private	ln Dens.	ln Div.	ln MP
Local HC (main)	1						
Local HC (sample)	0.84	1					
Local HC public	0.31	0.59	1				
Local HC private	0.79	0.82	0.05	1			
ln Dens.	0.60	0.52	-0.04	0.64	1		
ln Div.	0.35	0.25	-0.02	0.28	0.39	1	
ln MP	0.21	0.17	-0.08	0.24	0.44	0.22	1

Notes. This table presents the coefficients of correlation between the different local variables included in our model. HC (main) is our previous local human capital variable (computed from exhaustive INSEE data); HC (sample) is the local human capital variable computed from our sample; HC public is the ratio of higher-educated public-sector workers in the total employment of the zone (computed from our sample); HC private is the ratio of higher-educated private-sector workers in the total employment of the zone (computed from our sample).

We now turn to the regression analysis where the ratio of higher-educated workers depends on their sector of employment (public or private). Columns (3) to (5) of Table 1.8 reports estimation results when we replace our previous local human capital variable by the two public/private ratios in the second step regression based on equation (1.2). We find no significant influence of the concentration of public sector higher-educated workers on local wages, whereas the ratio of private sector human capital shows a very significant semi-elasticity of 0.15%. This suggests that local human capital externalities are potentially underestimated when we do not distinguish the sector of employment (public or private) of higher-educated local population. In other words, geographic wage inequalities stem more from differences in human capital in the private sector than in the public sector.

However, since public workers are equitably spread across zones, it is not possible to infer from results of Table 1.8 that higher-educated workers in the public sector are not a source of externalities. Another way of seeing these results is that the dispersion of highly-educated public-sector workers across zones is not a source of geographical wage inequalities. Indeed, if those workers were geographically spread across France according to the – unbalanced – geographical distribution of higher-educated workers in the private sector, spatial wages inequalities would be larger than they currently are. And these greater inequalities would come not only from a direct composition effect (more graduates with higher wages in zone which are already concentrated

Table 1.8: Influence of higher-educated workers: private and public sectors.

	Zone FE coef. from first step			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
Local HC (sample)	0.087*** (0.02969)	-	-	-
Local HC public workers	-	0.018 n.s. (0.0456)	-	0.004 n.s. (0.0449)
Local HC private workers	-	-	0.150*** (0.0420)	0.150*** (0.0422)
ln Density	0.003* (0.0015)	0.004*** (0.0011)	0.002 n.s. (0.0013)	0.002 n.s. (0.0013)
ln Diversity	-0.011*** (0.0039)	-0.010** (0.0040)	-0.011*** (0.0039)	-0.011*** (0.0039)
ln Market Potential	0.026*** (0.0018)	0.026*** (0.0018)	0.026*** (0.0018)	0.026*** (0.0018)
Worker fixed effects in first step	Yes	Yes	Yes	Yes
Observations	304	304	304	304
R <sup>2</sup>	0.53	0.41	0.42	0.57

Notes. This Table presents estimation results of our second-step regressions based on equation (1.2). The dependent variable is the 304 estimated coefficients of zone fixed effects included in first-step estimation based on equation (1.1) (see estimation results in annex .1). Each column is a separate regression. Local HC (sample) is the share of higher-educated workers in total employment computed from our panel. Local HC public (resp. ratio HC private) is the number of higher-educated worker (Bac + 3 and more) in the public sector (resp. private) to total employment of the zone.

\*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

in human capital) but also from the externality effect we measure in this chapter.

For many years, French governments have tried to relocate some part of public services outside of the Paris region, with the aim of alleviating congestion in this region and stimulating activity in less central areas. For example, the central civil status service (a service of the Ministry of Foreign Affairs) were decentralized in Nantes in 1985.

Generalization of information and communication technology increases the opportunities of further decentralization. It is no longer the large provincial cities that are the recipients of relocation, but rather smaller cities. For example, since 2020 there is a plan to relocate tax services to nearly 70 medium and small towns. This recent movement follows other decisions that have resulted in a reduced presence of some public services in less dense areas (and the

educated civil servants they employ) in the name of efficiency (reform of the judicial map, closure of smaller hospitals...). To our knowledge, there is no impact evaluation of all these decisions on the concentration of higher-educated people in smaller areas.

## 1.8 Conclusion

Using a rich panel of French administrative data over the period 2009-2015, we find strong evidence that local concentration of human capital is a major agglomeration force, which superseded the role of pure density. When we jointly estimate the respective influence on wages of local density, human capital, diversity and market potential, the influence of density is not significantly different from zero. The results suggest that wage gains in dense areas rather involve human capital effects than pure urbanization effects. However, the role of density appears to remain important on a broader spatial extent, as the impact of market potential is also significant and substantial.

We estimate jointly the static and dynamic local human capital effects on wages by separating the intrinsic and experience components of workers' unobserved ability captured in individual fixed effects (following the methodology of [De la Roca and Puga \(2017\)](#)). We find that, over a period of six years, one percentage point difference in the local share of highly-educated people (i.e. with a minimum of three years of study after the *baccalauréat* – and not including those in school) is associated with a 0.33% increase in the wages of workers in the employment-zone. Purely static human capital externalities account for about one third of this overall effect (0.1%), the rest comes from dynamic learning effects.

Finally, exploring the role of local human capital externalities according to whether they come from private or public higher-educated workers, we find that geographic wage inequalities appear to stem more from geographical differences in the private sector.

These results are instructive for local public policies, particularly with regard to the questions we asked in the introduction to this chapter. Indeed, given the importance of human capital effects compared to pure urbanization effects, it seems advisable to give priority on education, training and qualifications, especially since high density may also lead to crowding and congestion negative effects ([Duranton and Puga, 2020](#)). This is not intended to imply that big cities should

unquestionably reduce the density of their economic activities, as it can have spillover effects in the surrounding areas.

Decentralization policies, such as efforts to ensure the equitable distribution of public-sector higher-educated workers throughout the country, appear to be effective in addressing geographic inequalities. On the one hand, they are a catalyst for local human capital externalities, and on the other hand, they help to prevent further territorial imbalances due to the already unequal distribution of private-sector graduates. The objective of these territorial policies is also to guarantee the proximity of public service, more focused on the quality of life and the well-being of local populations.

The next chapter of this thesis will focus on another aspect of geographic inequalities, which has consequences for both wages and life satisfaction: the chances of upward social mobility during working life.

## .1 First step estimation results

Table 9: First step wage regressions based on equation (1.1).

	ln wage (1)	ln wage (2)
Tenure at firm	0.003*** (0.0001)	0.005*** (0.0002)
Experience	0.035*** (0.0023)	-0.007*** (0.0014)
Experience <sup>2</sup>	-0.0003*** (0.0000)	0.001*** (0.0000)
SPC indicators (ref. SMEs business owners & CEOs):		
White-collar workers (managers and higher intellectual professionals)	-0.065*** (0.0115)	-0.053*** (0.0098)
Intermediate Professions	-0.204*** (0.0104)	-0.396*** (0.0098)
Skilled-Employees	-0.259*** (0.0098)	-0.568*** (0.0093)
Skilled-Blue-collar-Workers	-0.278*** (0.0092)	-0.623*** (0.0081)
Unskilled-Employees	-0.309*** (0.0089)	-0.684*** (0.0088)
Unskilled Blue-collar-Workers	-0.304*** (0.0090)	-0.693*** (0.0093)
Industry indicators (ref. Agriculture):		
Manufacturing, Mining and other industries	0.063*** (0.0041)	0.107*** (0.0080)
Construction	0.069*** (0.0038)	0.108*** (0.0081)
Wholesale/Retail Trade, Transport, Accommodation and Food	0.035*** (0.0035)	0.019** (0.0085)
Information and Communication	0.043*** (0.0043)	0.005 n.s. (0.0125)
Financial and Insurance	0.086*** (0.0050)	0.132*** (0.0077)
Real Estate	0.057*** (0.0050)	0.016** (0.0077)
Professional, Scientific, and Technical Services, Admin. and Support Services	0.062*** (0.0043)	0.028*** (0.0066)
Public administration, education, human health and social work	0.012*** (0.0045)	0.019** (0.0095)
Other service activities	0.005** (0.0038)	-0.051*** (0.0081)
Year indicators (ref. 2009):		
2010	-0.006*** (0.0011)	0.009*** (0.0009)
2011	0.001 n.s. (0.0025)	0.031*** (0.0019)
2012	0.001 n.s. (0.0036)	0.035*** (0.0033)
2013	0.008 n.s. (0.0051)	0.060*** (0.0041)
2014	0.001 n.s. (0.0059)	0.061*** (0.0046)
2015	-0.006 n.s. (0.0072)	0.062*** (0.0056)
Men	-	0.107*** (0.0017)
Age	-	0.013*** (0.0009)
Age <sup>2</sup>	-	-0.0001*** (0.0000)
Level of education (ref. Baccalauréat)		
No education	-	-0.103*** (0.0054)
Primary education	-	-0.138*** (0.0029)
Lower secondary education	-	-0.044*** (0.0015)
Certificate of professional competence	-	-0.051*** (0.0015)
Higher education (Bac + 2)	-	0.038*** (0.0008)
Higher education (Bac+3 and more)	-	0.089*** (0.0061)
Constant	7.669*** (0.2781)	13.136*** (0.4834)
Individual fixed effects	Yes	No
Zone fixed effects	Yes	Yes
Observation	3,978,977	3,978,977
R <sup>2</sup>	0.92	0.63
Within R <sup>2</sup>	0.22	0.60

Note. This table presents the first-step estimation results of wage equation (1.1) for the period 2009-2015. Column (1) include worker fixed effects, column (2) does not. From these regressions, we recover the estimated coefficients of zone fixed effects to run the second-step estimations based on equation (1.2). Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

## .2 Instrumental variables.

**Instruments for local density.** Following Ciccone and Hall (1996), the first instrument is the historical value of local density. Due to the secular trend of urbanization, the current density is very correlated to past density. Additionally, it is unlikely that very old density is a determinant of current productivity and wage levels, which rather result from recent shocks occurred during the 20th century (such as rural exodus, world wars, women's participation in the labor market, service sector expansion, etc.). Thus, assuming that density at the very beginning of the 19th century does not influence contemporary wages (other than the effect that passes through the determination of current density), we use the local population density in 1806 (from the database Cassini of EHESS<sup>35</sup>) as an instrument of current local density.

Similarly, we also instrument for current density with information on historical monuments from the Ministry of Culture. The number of religious buildings in an area (catholic churches for the vast majority) classified as "historic monuments" reflects the localization of historical populations. Hence, with the same logic than the first instrument, the density of religious buildings (that existed prior the 20th century) per km<sup>2</sup> is related to density observed in 2012 but does not directly affect the current productivity and wages at the local level.<sup>36</sup>

In our second set of excluded instrument (IV2), we use two more instruments for local density based on historical departmental data from the *Centre de recherches historiques, EHESS-CNRS*:<sup>37</sup> the number of royal road kilometres in 1837 and the ratio of the number of royal road kilometres in 1837 to the total surface of the departement. Those instruments are based on the similar idea that the existence of royal roads in the early 19th century indicates the presence of populations (or facilitates settlement of future ones) and creates inertia in the local density.

**Instruments for local human capital.** Regarding the local human capital, we use as an instrument the past change in the share of higher-educated people between 1968 and 1982.

<sup>35</sup>Thanks to Claude Motte from EHESS for providing the data.

<sup>36</sup>We must verify that this density of religious historic monuments does not excessively influence local current economic activities – in particular, it may be the case regarding tourism (Durieux et al., 2015). The weak correlations between the instrument and the employment shares in sectors linked to tourism (accommodation -0.1, catering 0.09, and travel agencies 0.09) suggest that this is not the case.

<sup>37</sup>Source: « *Territoire et population de 1800 à 1890* », *L'Atelier du Centre de recherches historiques [En ligne]*, *La Statistique Générale de la France, Les données traitées par l'INSEE, 2011*. URL : <http://journals.openedition.org/acrh/3410>.

Chauvin et al. (2017) already introduce such instrument using the 1980 value of their local education variable. In our case, the relevance of this lagged value comes from the fact that France is characterized by a long term trend of geographic inequality in access to higher education (Dherbécourt, 2015). Areas with large current share of higher-educated people are the same ones that experienced relatively strong human capital accumulation fifty years ago.

A second instrument for local education follows the shift-share type variable proposed by Moretti (2004b), which use lagged demographic structure of employment zones. In practice, we predict the local share of higher-educated people that would have been observed in 2012, if the 1999 local age structure has remained and if each individual in the area had the average national education level of her age group. Our predicted share is therefore a weighted average of the national education attainment from each age-group, with weights depending on the initial distribution of age-groups in the local population:

$$IVHC_z = \sum_a W_{a,z} HC_a^{national}$$

with  $W_{a,z}$  the initial share of the age-group  $a$  in the population of the zone  $z$  in 1999 and  $HC_a^{national}$  the share of higher-educated people within the age-group  $a$  at the national level in 2012. We use the lagged age structure in 1999 to take into account potential endogeneity of the age distribution of zones – it may reflect expected changes in the local economy.

This instrument is based on the idea that local labor force structure results from the long-term trend of increasing education: on average, younger entrants have more degrees than older workers. Thus, differences in age structures drive differences in the local shares of higher-educated people. In contrast, the increase in the share of higher-educated people (determined by the demographic structure of employment areas about ten years earlier) is likely exogenous to local factors influencing demand for skilled labor in 2012.

There is an ongoing debate about the relevance of Bartik shift-share instruments since the work of Goldsmith-Pinkham et al. (2020). To verify that our IV results do not rely entirely on this instrument, we develop a second IV model that replaces it with two instruments at the departement-level: *Source: Centre de recherches historiques, EHESS-CNRS*.<sup>38</sup> the number

<sup>38</sup>Source: "L'enseignement primaire et secondaire de 1865 à 1906 », L'Atelier du Centre de recherches

of secondary students in 1887 compared to departemental population and the share of conscripts with *baccalauréat* diploma in 1896. These instruments are based on the same idea of inertia in the spatial distribution of human capital across France.

**Instruments for market potential.** In the same vein as our instruments for density, the first instrument for market potential is simply the market potential in 1806 – therefore based on local population densities in 1806. Next, following [Combes et al. \(2011\)](#), we use two peripherality measures of each zone by computing the average distance to all the other employment zones, and the average distance to Paris.

Table 10: Instrumental variable model – First stage regressions.

2SLS IV models – First stage estimation results	IV1			IV2		
	ln Density	Human capital	ln MP	ln Density	Human capital	ln MP
Var. HC 68-82	15.790*** (3.247)	2.203*** (0.0963)	9.084*** (1.0994)	18.134*** (3.9325)	2.319*** (0.1779)	8.461*** (1.7321)
Shift-share HC	3.037*** (0.4499)	0.099*** (0.0133)	0.404*** (0.1523)	-	-	-
Cons. Bac 1896	-	-	-	3.913 n.s. (9.815)	0.415*** (0.1942)	4.411 n.s. (3.5234)
Sec. Students / Pop. 1837 (department)	-	-	-	150.47*** (55.230)	2.890** (1.3630)	31.080** (15.410)
Dens. Religious buildings	6.99*** (0.436)	0.441*** (0.0677)	3.537*** (0.7722)	2.757 n.s. (3.5921)	0.325*** (0.1233)	0.563 n.s. (1.6431)
ln Density 1806	0.782*** (0.0828)	0.003 n.s. (0.0025)	-0.132*** (0.0280)	0.791*** (0.0952)	0.004 n.s. (0.0046)	-0.134*** (0.0424)
Km royal roads 1837 (department)	-	-	-	-0.001* (0.0004)	-0.000 n.s. (0.0000)	-0.000 n.s. (0.0002)
Dens. Royal roads 1837 (department)	-	-	-	4.603*** (1.6718)	0.042 n.s. (0.0458)	3.159*** (0.7709)
ln MP 1806	1.022** (0.5180)	-0.030* (0.0154)	1.208*** (0.1754)	1.257* (0.6739)	-0.019 n.s. (0.0170)	0.998*** (0.2730)
Average distance other zones	0.003** (0.0012)	-0.0001*** (0.0000)	0.003*** (0.0004)	0.004*** (0.0015)	-0.000 n.s. (0.0001)	0.003*** (0.0006)
Distance to Paris	-0.000 n.s. (0.0003)	0.0001*** (0.0000)	-0.002*** (0.0001)	-0.000 n.s. (0.0004)	0.000 n.s. (0.0000)	-0.002*** (0.0001)
ln Diversity	0.494*** (0.1720)	0.003* (0.0051)	-0.160*** (0.0582)	0.690*** (0.2335)	0.009 n.s. (0.0080)	-0.122** (0.0599)
Observations	304	304	304	290	290	290
Cragg-Donald F-Stat.	58.52	144.13	250.55	122.07	121.48	240.21
p-value	0.000	0.000	0.000	0.000	0.000	0.000

Notes. This table presents first-stage estimation results of the 2SLS IV models. Endogenous regressors are the local density, the local human capital and market potential. In the IV1 model, excluded instrument are: log density in 1806; log density of historical religious buildings; the change in local human capital 1968-1982; shift-share local human capital; market potential in 1806; the distance to Paris and the average distance to other zones. In the IV2 model, we drop the shift-share instrument and add four additional excluded instruments: the number of secondary students in 1887 compared to population; the share of conscripts with *baccalauréat* diploma in 1896; the number of royal road kilometers in 1837; and the ratio of the number of royal road kilometres in 1837 to the total surface of the département. Due to lack of information in the historical department-level database, IV2 regression is based on 250 zones. Reported Wald F-statistics, for weak instruments identification test, exceed the Stock and Yogo (2005) critical values for a 5% maximal IV relative bias (19.86 for IV1 and 20.74 for IV2).

Coefficients are reported with clustered standard errors at the zone level in parenthesis for IV1 and at the zone-department level for IV2. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

### .3 Zones ranking depending on local density and human capital

Table 11: Top areas depending on local density or human capital in 2012.

	Local Human Capital	Density
Lille	Top 5%	Top 5%
Lyon	Top 5%	Top 5%
Paris	Top 5%	Top 5%
Saclay	Top 5%	Top 5%
Strasbourg	Top 5%	Top 5%
Marne-la-Vallée	Top 5%	Top 5%
Poissy	Top 5%	Top 5%
Montpellier	Top 5%	Next Top 5%
Plaisir	Top 5%	Next Top 5%
Grenoble	Top 5%	-
Toulouse	Top 5%	-
Nice	Top 5%	-
Aix-en-Provence	Top 5%	-
Houdan	Top 5%	-
Rambouillet	Top 5%	-
Marseille - Aubagne	Next Top 5%	Top 5%
Créteil	Next Top 5%	Top 5%
Orly	Next Top 5%	Top 5%
Cergy	Next Top 5%	Top 5%
Nantes	Next Top 5%	Next Top 5%
Cannes - Antibes	Next Top 5%	Next Top 5%
Bordeaux	Next Top 5%	-
Rennes	Next Top 5%	-
Dijon	Next Top 5%	-
Nancy	Next Top 5%	-
Annecy	Next Top 5%	-
Besançon	Next Top 5%	-
Corte	Next Top 5%	-
Genevois Français	Next Top 5%	-
Melun	Next Top 5%	-
Évry	-	Top 5%
Lens - Hénin	-	Top 5%
Roubaix - Tourcoing	-	Top 5%
Valenciennes	-	Top 5%
Roissy - Sud Picardie	-	Next Top 5%
Toulon	-	Next Top 5%
Béthune - Bruay	-	Next Top 5%
Boulogne-sur-mer	-	Next Top 5%
Douai	-	Next Top 5%
Dunkerque	-	Next Top 5%
Istres - Martigues	-	Next Top 5%
Le Havre	-	Next Top 5%
Mulhouse	-	Next Top 5%
Perpignan	-	Next Top 5%
Sète	-	Next Top 5%

Notes. This table presents the employment zones ranking among the 'top areas' in terms of local human capital or density in 2012. For example, the zone of Lille is both ranked in the top 5% of zones in terms of density or local human capital, whereas the zone of Montpellier is ranked in the top 5% in term of local human capital, but not in terms of density (next top 5% or 19th vintile).

## Chapter 2

# The Geography of Social Mobility During Working Life

### Abstract

This chapter examines the influence of local labor markets on individuals' social mobility during their working life. In France, over a period of six years, individuals working in dense areas and areas with abundant human capital are more likely to be promoted to a higher socioprofessional status – a promotion associated with a substantial wage increase and/or better working conditions. Applying an empirical strategy to remove spatial sorting bias in a sample of about 350,000 workers in metropolitan France over the period 2009-2015, we find that density, local human capital, and labor market size significantly increase the likelihood of being promoted, but have no impact on demotions. In addition, we show that local density increases both internal (within the same firm) and external (in another firm) promotions, with the impact of density being even larger on external promotion. Finally, experience accumulated in the densest areas is portable, and increases the chances of promotion after relocation to less dense areas.

## 2.1 Introduction

This paper examines the influence of local labor markets on intragenerational social mobility, i.e. the mobility that takes place during workers' careers. The question we ask is whether socioprofessional mobility during working life depends on local factors, suspecting that denser labor markets and areas with high levels of human capital make social promotion easier.

For individuals, social mobility is one of the most important drivers of wage gains. [Bayer and Kuhn \(2019\)](#) recently showed that moving up the job ladder accounts for half of average wage growth, and virtually all of the increase in wage dispersion over the life cycle, in Germany and the United States. In our sample, over the period 2009-2015, real wages increased by 27% for promoted workers, and by only 11% for workers staying at the same job level.<sup>1</sup> Moreover, social mobility reflects important changes for individuals that go far beyond a higher wage. Finding a better job (or the mere possibility of it) has been shown to be an important component of individual well-being and life satisfaction ([Abel et al., 2018](#); [Senik, 2008, 2020](#)). In France, several accounts of the fears and joys of upward social mobility have been published with great success, including that of the sociologist [Eribon \(2009\)](#) and the novelist [Ernaux \(2008\)](#), who also evoke the role of geography in their own life paths. Indeed, the chances of social promotion are not the same throughout France. Our raw data show that, on average, a worker had a 17% chance to climb the job ladder over the six-year period 2009-2015, but that this ranged from 10% to 27% depending on his commuting zone in 2009.

Studying the geographic dimension in promotion opportunities contributes to our understanding of the sources of agglomeration effects and how they benefit individuals. In addition, the impact of local environment on social mobility cannot be ignored in the implementation of place-based policies, especially those aiming at the revival of places that have lost their manufacturing industry and been deserted by the young and educated. While the European Union and European states are spending large amounts of money to ensure the economic convergence of regions, and the maintenance of activities in sparsely populated areas, the question of promotion opportunities for those who work in these areas is crucial. If career progression is much slower or

---

<sup>1</sup>As explained later in Section 2.3, our definition of promotion is based on the hierarchical typology of socioprofessional categories (SPC) proposed by INSEE. This typology allows us to specify upward and downward mobility, respectively understood as a transition from a given category to a "better" or a "lower" one.

even absent in these areas, emigration will reduce the efficiency of these policies. Also, in these areas, low opportunities can generate discouragement and fuel a feeling of injustice for those who do not leave, as the "*gilets jaunes*" ("yellow vests") movement has shown in France. Finally, at a time where policy makers are concerned with low social mobility, it may be judicious to design policies that tackle not only low intergenerational mobility, but also help boost upward social mobility in the labor market, as this might bring faster results.

To our knowledge, despite its relevance, the question of hierarchical social mobility in local labor markets has never been explicitly addressed. That said, we do not start from scratch, but instead draw on three different streams of literature. The first one examines the individual determinants of workers' ability to find a better job or to climb in the wage distribution during their careers - with no focus on the role of local environment (e.g. [Abel et al. \(2018\)](#), [Bayer and Kuhn \(2019\)](#)). The second one focuses on the geographical differences in wages and job switches within the body of agglomeration economies research (see [Combes and Gobillon \(2015\)](#) for a review of the literature) - but without any focus on the social mobility dimension. This literature emphasizes the role of local density on churning and the job switching of workers (e.g. [Wheeler \(2008\)](#); [Bleakley and Lin \(2012\)](#); [Andersson and Thulin \(2013\)](#)). The third stream of literature on which we draw is that of the organization of firms, in particular the growing research on the hierarchical structure of firms (e.g. [Baker et al. \(1994\)](#); [Caliendo et al. \(2015\)](#); [Spanos \(2019\)](#)). Firms' organization and their internal labor markets may differ across areas, with potential consequences for opportunities for upward mobility.

As always with regional and urban economics, the key challenge is to identify the causal impact of local characteristics on individual outcomes. We design an empirical strategy robust enough to make sure that any positive relationship between local characteristics and social mobility is not the result of an endogenous spatial sorting of workers; workers who are inherently more ambitious and capable may choose to locate in denser areas, increasing the average chances of promotion in those labor markets because of this sorting. As individual promotion is a 1/0 rare event with a negative time auto-correlation, we choose to consider promotion over a six-year period, and not annual promotion. A given individual is considered as promoted (or demoted) if his socioprofessionnal category (SPC) at the end of the six-year period is higher (or lower)

than his SPC at the beginning at the six-year period. Our approach is thus cross-sectional and does not allow for individual fixed effects. To circumvent the issue of individual non-observable characteristics, we make use of the panel dimension of our database to estimate worker fixed effects (controlling for unobserved characteristics) through a standard Mincerian wage equation on annual data. We then use this individual fixed effect as an independent variable among others to explain the probability of being promoted, in an equation that also includes the characteristics of the commuting zone, as well as observed characteristics of the individuals. We are mainly interested in the effect of the density, human capital concentration, and size of the commuting zone on the chances of being promoted. To address any concerns regarding potential reverse causality between chances of being promoted and these commuting zone characteristics, we use an IV model. We allow for the non-linear impact of zone characteristic on promotion, and perform a variety of robustness tests. We also develop additional empirical models to explore the plausible mechanisms underlying a local promotion premium.

Applying this strategy to a sample of about 350,000 workers in metropolitan France over the period 2009-2015, we find that density, local human capital, and size significantly increase the chances of promotion, but have no impact on demotion. The impact of these local characteristics on the probability of promotion is not linear. Doubling density at a low level of density has no effect on the chances of being promoted, whereas doubling density at a higher level of density increases the probability of promotion by two percentage points. Results are similar when considering the size of local labor markets or human capital concentration. While density and size explain part of the promotion premium, it seems that the effect is channeled through human capital externalities. In addition, we show that the effect of local characteristics play a significant role in both internal (within the firm) and external promotions, with the impact on external promotion being the largest. Finally, experience accumulated in the densest areas, or the richest in term of human capital, appears to be portable, as we show that it increases the chances of promotion after relocation to less dense/educated areas.

The rest of the paper is organized as follows. Section 2.2 discusses the related literature. Section 2.3 discusses the definition of social mobility used in this paper, and presents data sources, as well as main stylized facts, about professional and social promotion and demotion

in France over the periods 2002-2008 and 2009-2015. Section 2.4 presents our identification strategy. Section 2.5 contains the main results. Section 2.6 presents some robustness tests. Section 2.7 examines plausible mechanisms underlying the local promotion premium. Section 2.8 concludes.

## 2.2 Related Literature

To our knowledge, this paper is the first to investigate the geographical dimension of social mobility over workers' careers. There is a growing empirical literature that investigates earning dynamics over workers' life cycles (Bayer and Kuhn, 2019; Guvenen et al., 2021). It shows that there is substantial heterogeneity in the cumulative growth rates of earnings between individuals, which Bayer and Kuhn, 2019 explain by heterogeneity in promotions over workers' careers, with some workers being promoted while others are not. Regarding promotions, Abel et al. (2018) focus on the extent to which low-wage workers move up the job ladder. They define the transition into a better job based on a metric of job quality to identify low-wage jobs, which depends on the average hourly pay, the occupation's prestige, and the working conditions and benefits associated with the job. Based on US data for the period 2011-2017, their empirical findings show that few workers successfully escaped low-wage work within a 12-month period (5%). Finding a better job is correlated with educational attainment, as well as with the overall labor market situation.

None of these papers investigate the role of local factors. This is in contrast with recent papers on intergenerational mobility - a child's chance of moving up the social ladder relative to her parents - that show the relevance of the geographical dimension. Chetty et al. (2014) and Chetty and Hendren (2018) show that in the US, the neighborhood one lives in during childhood has a causal impact on long-term outcomes. In France, Dherbécourt (2015) shows that the chances that children from the working classes will find high-level jobs vary by a factor of two, depending on their place of birth.

To investigate the link between location and promotion, we can draw on a large set of empirical studies.

Firstly, since social promotion is - in our definition - always associated with a change in job (within a firm or not), it is interesting to recall some important and consensual findings

regarding the link between job switching and local density. Geographical economists consider mobility of workers between jobs in dense areas as an important source of agglomeration economies through learning and/or matching (see Rosenthal and Strange (2004); Duranton and Puga (2004); Combes and Gobillon (2015); Duranton and Puga (2004)). This consideration is based on a large body of empirical evidence obtained in a variety of countries, using different methods and datasets. Inter-firm mobility has been shown to be higher in dense and/or large job markets in the US, at least for younger individuals (Wheeler (2008); Finney and Kohlhase (2008); Bleakley and Lin (2012)), as well as in Italy (Andini et al. (2013)), Sweden (Andersson and Thulin (2013)), and Brazil (Amarante et al. (2019)). Beyond the labor pooling advantage of dense labor markets, urban economics literature also underlines the role played by assortative matching (Becker, 1973; Abowd et al., 1999; Shimer and Smith, 2000; Shimer, 2005) between workers and plants in explaining geographical wage differences. The idea is that density also facilitates the quality of matches: in denser areas, high-quality workers are more likely to work for high-quality firms (Andersson et al., 2007; Dauth et al., 2022). In this literature, no attention is paid to social mobility, to whether workers that switch jobs also move up the job ladder significantly, or to whether assortative matching leads to better chances of promotion.

Secondly, the way firms organize their production may have an impact on the likelihood of promotion. Caliendo and Rossi-Hansberg (2012) propose a model of the firm wherein production requires a combination of labor and knowledge. Firms are production hierarchies, and organize workers into layers of different sizes. A firm's number of layers is determined by its profit maximizing output, which in turn depends on the demand for its products, and the degree of competition between firms. Based on this model, Caliendo et al. (2015) provide stylized facts showing that in the French manufacturing sector, faster-growing firms are more likely to add layers to their hierarchies, while shrinking firms are more likely to reduce the number of hierarchical layers. In addition, empirical literature provides evidence that the internal market of firms is at least partly consistent with the Doeringer and Piore (1971) view, in which workers enter an internal labor market at entry level, and climb up the job ladder within the firm. Available data on private sector firms show that the share of external hires declines with the hierarchical level (Baker et al., 1994; Lazear and Oyer, 2004; Huitfeldt et al., 2021). Turning

to the geographical heterogeneity of firm organization, [Spanos \(2019\)](#) introduces geographical heterogeneity between firms, in the model proposed by [Caliendo and Rossi-Hansberg \(2012\)](#). Bringing his model to the data, he shows that in the French service industry, firms in denser labor markets organize into a greater number of layers. Also, [Duranton and Jayet \(2011\)](#) provide evidence that a larger local market increases the division of labor offerings and the number of available occupations. These theoretical and empirical results provide one channel between density and promotion probabilities. More hierarchical layers within firms in denser areas means higher promotion opportunities for workers both within the firm and at the commuting-zone level. Also, growth at the local level is likely to be associated with the overall promotion rate.

Thirdly, it has been shown that larger or denser labor markets facilitate the transmission and accumulation of skills ([Glaeser and Maré, 2001](#); [Duranton and Puga, 2004](#); [De la Roca and Puga, 2017](#); [Duranton and Puga, 2020](#)). As a consequence, experience in these labor market makes people more productive, and may increase their chances of being promoted. Contrary to the two previous channels we discussed (job mobility and firm organization), this human capital effect is attached to the worker. As such, once acquired, additional human capital or ability resulting from working in dense labor markets increases a worker's probability of being promoted, even after relocation to less dense labor markets. Since the pioneering work of [Fielding \(1989, 1993\)](#), economic geographers have also investigated such dynamic effects of location and migration on individuals' social progression, giving the name of "escalator" effects to the advantages of experience accumulated in big cities/regions ([Findlay et al., 2009](#); [Newbold and Brown, 2012](#); [Gordon et al., 2015](#)).

## **2.3 Social mobility over the life cycle: definition, data, and stylized facts**

### **2.3.1 Social mobility**

There are many ways to define the social position of individuals, and thus many ways to define social mobility. One way is to consider the individual's position within the income distribution, and treat mobility as moving from one decile (or one quartile) to another. This approach is the

most popular to measure intergenerational mobility. However, the social status of individuals is not only related to their labor income, but also to the quality of life their jobs allow (stability, work hours, and benefits), and the prestige associated with it. For example, a young doctor, or a young assistant professor, in highly prestigious occupations, may not earn a very high wage, but may still be considered to belong to the top social category. Indeed, besides hourly wage, to rank jobs, [Abel et al. \(2018\)](#) use a broad range of indicators related to the nature of one's job.

We consider social status and social mobility, accordingly, from the point of view of the INSEE classification of Socioprofessional Categories (SPC). It classifies the population by a combination of profession, economic activity, hierarchical position, qualification and status (salaried employee or otherwise). Following [Razafindranovona \(2017\)](#), we use the socioprofessional categories, divided into 42 items, to build a hierarchical classification of occupations, with four socioprofessional categories that cover more than 90% of the working population, and we define upward (or downward) mobility as a transition from a given SPC to a "higher" one (or "lower" one). Because we consider four broad categories, moving up or down this social ladder reflects important changes for any individual that go far beyond a higher wage or a small job promotion.

The four broad categories we consider are the following: 1. White-collar workers (20.4% of workers in 2020<sup>2</sup>); 2. Intermediate Professions (26% of workers in 2020); 3. Skilled Employees (13.5%) and Skilled Blue-collar workers (12.7%); and 4. Unskilled Employees (12.3%) and Unskilled Blue-collar workers (6.5%).<sup>3</sup> We exclude SME business owners and CEOs (6.8%) and farmers (1.4%). This approach, of broad categories, is similar to the one followed by [Caliendo and Rossi-Hansberg \(2012\)](#) and [Spanos \(2019\)](#). As there was a slight change in the INSEE classification in 2009,<sup>4</sup> we will consider social mobility over two different six-year periods: 2002-2008 and 2009-2015.

There is an ongoing debate about the relevance of the INSEE SPC classification. The first concern is related to the profound changes since the end of the 20th century in the tasks and skills required in a given occupation. [Askenazy et al. \(2019\)](#) question the relevance of statistics

---

<sup>2</sup>INSEE Socioprofessional category by gender and age. Annual data 2020.

<sup>3</sup>See Table 14 for more details on the composition of these four broad SPC.

<sup>4</sup>For example, an accounting officer previously classified as an "Intermediate Profession" - an occupation between manager and executing-employee - is now classified either as an "Employee" or as a "White-collar worker" (manager), depending on the actual level of the job.

by SPC in France for long-term longitudinal studies – two occupations that seemed similar forty years ago may have diverged, while remaining in the same SPC. However, this is less a concern for our study since we assess the socioprofessional mobility over a relatively short period of time (six years), during which any evolution in a particular occupation is less likely to occur. A second concern is linked to the quality of data, as each firm declares the SPC of its employees (within the 4-digit classification), and since there are no incentives to provide accurate information, firms can change the SPC number of a given employee from one year to another, while neither the job nor the salary of the worker actually changes. This quality issue would be worrisome if we focused on incremental changes in social status, but is less of a concern given the broad SPC we consider in this study.

Nevertheless, for robustness purposes we will also consider three alternative definitions of promotion. The first one is based on a classification of occupations in the French labor market proposed by [Le Moigne \(2021\)](#). Her hierarchical classification applies to French data the insights developed by [Baker et al. \(1994\)](#). The main difference with our classification is that Le Moigne’s regroups our two lower categories into a single one and adds a top category: that of SME business owners and CEOs, a category we do not consider in our classification given its heterogeneity and its small size (6.8%). Promotion is then defined as moving up from one level to the one above. The second alternative definition of promotion we consider is based on wage gain, and we consider an individual to be promoted when his nominal wage increases by at least 35% between 2009 and 2015. This 35% allows us to match the same promotion rate as in our classification based on SPC. The third one is more restrictive and considers changes in SPC and wages. A transition to a better SPC is not a promotion if it is not associated with a wage increase of more than 35% over the six-year period.

### 2.3.2 Data

Our main dataset is *Panel Tous Salariés - Echantillon Démographique Permanent (EDP)*, which is produced by INSEE. It follows some workers (those born in early October) during their careers on an annual basis between 1967 and 2015, with new workers being added to the panel every year so to ensure its representativeness of the overall population of workers. It is therefore the

ideal database to measure social mobility over workers' careers. This panel provides information on the nature and localization of the work, occupation and industry, earnings, number of paid hours, terms of employment (full-time, part-time), level of education, etc. The same data sample used for chapter one is also exploited in this second chapter. See section 1.3.2 of the first chapter for more detail on the data and the construction of the estimation sample.

One of the advantages of this panel is the completeness of workers' job positions at a given time. However, a given worker can hold several jobs during a single year, relating to different occupations. This is especially the case for workers with part-time and low-paid jobs, or jobs in industries with high employee turnover. We thus need to define what the annual principal occupation for each individual is. As in the first chapter of this thesis, in order to keep a maximum of the information provided by the panel, we define the occupation of each worker for a given year as the occupation attached to the jobs that provide at least two-thirds of her annual wage.

As a matter of fact, using this method, a worker who has one retail cashier job at an annual wage of 2,000 EUR, combined with many home carer jobs with a cumulative annual wage of 10,000 EUR, but not exceeding 1,000 EUR each, will be classified in the occupation "domestic services" - whereas she would have been classified as a "retail employee" if we used only the most remunerative job. This specific treatment implies that some workers fall into a "mixed category," *i.e.* where no single occupation represents at least two-thirds of annual earnings (for example two jobs which each represents half of annual earnings). We exclude individuals who fall into this mixed category, which represents around 14% of the observations for the year 2002, and 7% for the year 2009.<sup>5</sup> After sample restrictions (see chapter one Section 1.3.2), we are left with about 4 million observations for the period 2009-2015 and 4,5 million observations for the period 2002-2008. Considering our cross-sectional empirical strategy based on the probability of promotion (or demotion) over a six-year period (see section 2.4), our estimation samples correspond to about 350,000 workers observed in 2009 (and 2015) and about 430,000 workers observed in 2002 (and 2008).

---

<sup>5</sup>For comparison with previous literature and for robustness reasons, we also use a database that keeps only the information of the job corresponding to the highest annual remuneration for each worker. In this case, we exclude part-time workers and unemployed individuals.

The panel contains information regarding location of the job, at the city level. We evaluate the impact of local characteristics on social mobility over the life cycle at the employment-zone level. An employment zone, or commuting zone, is a geographical area, delineated by INSEE, within which most of the labor force lives and works, and from which firms hire most of their labor force. It is therefore a geographical scale adapted to local studies, especially to evaluate the influence of local environments on individuals. There are 304 employment zones in metropolitan France.<sup>6</sup>

### 2.3.3 Stylized facts on social mobility during working life

Based on our panel data, between 2002 and 2008 an individual had on average a 22% chance of being promoted to a higher social and professional status than the one she had in 2002. The probability of promotion over the 2009-2015 period is 18% (see Table 2.1). Over these two periods, the probability of promotion is four to five percentage points higher for men than for women. Moving downward is rarer. In our data, over the period 2002-2008, 13% of workers descended the social ladder, and 10% over the period 2009-2015, with no significant difference between men and women (see Table 2.2). These patterns for job promotion and demotion are in line with previous findings for France by Monso (2006), using different data sources.<sup>7</sup>

Table 2.3 reports the average real wage growth, depending on the social mobility status (demotion, no change, or promotion), over our two periods. Promoted workers benefit from a sizable wage premium. On average, compared to the "no change" situation, promotion is associated with a larger increase in real wages (of about 17 percentage points for the period 2009-2015, and 20 point for 2002-2008). This is true for both men and women.<sup>8</sup> There is no substantial difference across SPC in the wage premium associated with promotion.

Figure 2.1 presents the heat maps of upward social mobility between 2009 and 2015 (left) and of employment density in 2009 (right). There is substantial geographical heterogeneity

---

<sup>6</sup>The employment zones are defined as the geography applicable in 2014. INSEE redesigned the boundaries of employment zones in 2010. We adjusted the geography in the former period using the 2010 table of correspondence, which provides a consistent basis for comparing local outcomes over time.

<sup>7</sup>Training and vocational skills survey, 1985, 1993, and 2003.

<sup>8</sup>The gender gap in wages is not within the scope of this paper. Yet our raw data suggest that gender wage gaps may be less a problem of growth pace than the consequence of fewer promotions for women in the labor market. Compared to men, their wages equally increase when they climb the social ladder, but the fact is that women face greater difficulties in doing so.

Table 2.1: Upward Social Mobility in 2002-2008 and 2009-2015.

	2002-2008			2009-2015		
	Total	Men	Women	Total	Men	Women
<b>All</b>	22%	24%	20%	18%	20%	15%
<b>Intermediate Professions</b>	14%	18%	10%	12%	16%	9%
<b>Skilled-Employees</b>	21%	25%	19%	16%	17%	16%
<b>Skilled-Blue-collar-Workers</b>	12%	13%	10%	12%	12%	10%
<b>Unskilled-Employees</b>	37%	52%	31%	25%	39%	21%
<b>Unskilled Blue-collar-Workers</b>	50%	57%	36%	42%	48%	29%

Note. Data comes from *Insee Panel tous salariés-EDP* with specific treatment to determine the annual principal socioprofessional category (see Section 2.3). We present here average statistics for workers in our estimation samples (in particular we keep only workers who can experience an upward mobility over the period). This represents a total of 346,690 workers for the 2009 - 2015 period and 430,119 workers for 2002 - 2008. Reading: 12% of intermediate professions workers in 2009, and still in employment in 2015, are in an socio-professional category with a higher status in 2015 (in this case, it can only be white-collar SPC).

Table 2.2: Downward Social Mobility in 2002-2008 and 2009-2015.

	2002-2008			2009-2015		
	Total	Men	Women	Total	Men	Women
<b>All</b>	13%	13%	14%	10%	10%	10%
<b>White-collar-Workers</b>	17%	15%	22%	13%	11%	14%
<b>Intermediate Professions</b>	18%	17%	20%	15%	15%	14%
<b>Skilled-Employees</b>	6%	7%	6%	5%	4%	6%
<b>Skilled-Blue-collar-Workers</b>	10%	9%	22%	9%	8%	15%

Note. Data comes from *Insee Panel tous salariés-EDP* with specific treatment to determine the annual principal socioprofessional (see Section 2.3). We present here average statistics for workers in our estimation samples (we keep only workers who can experience a downward mobility over the period). This represents a total of 325,872 workers for the 2009 - 2015 period and 416,512 workers for 2002 - 2008. Reading: 13% of white-collar workers in 2009, and still in employment in 2015, are in an socio-professional category with a lower status in 2015.

in upward mobility.<sup>9</sup> A simple comparison with the heat map of employment density at the local level reveals a notably good match. This is partly due to composition effect, as the socioprofessional categories with greater probability of upward mobility are over-represented in denser areas. However, the positive relationship between upward mobility and density remains true when separating out the probability by SPC-gender. Table 2.4 presents these probabilities depending on the position of a worker's commuting zone in the distribution of employment density. We see that, on average, the probability of upward mobility is almost always greater in denser areas – except for female unskilled blue-collar workers and male unskilled-employees.

<sup>9</sup>Same pattern for the 2002-2008 period.

Table 2.3: Average change in real hourly wages depending on social mobility status.

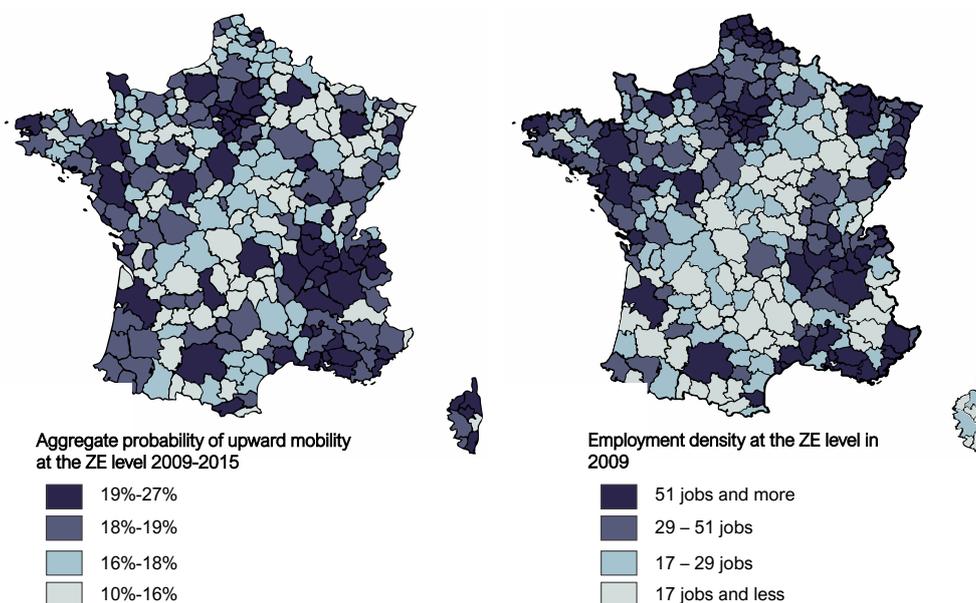
	2002-2008				2009-2015			
	Downward	No change	Upward	Total	Downward	No change	Upward	Total
<b>All</b>	4%	15%	35%	18%	2%	8%	25%	11%
<b>White-collar</b>	-7%	20%	-	15%	-5%	14%	-	12%
<b>Intermediate Professions</b>	4%	15%	43%	17%	4%	11%	32%	12%
<b>Skilled-Employees</b>	8%	11%	33%	16%	-4%	9%	27%	11%
<b>Skilled-Blue-collar</b>	10%	14%	35%	16%	6%	9%	26%	11%
<b>Unskilled-Employees</b>	-	14%	35%	16%	-	13%	31%	18%
<b>Unskilled-Blue-collar</b>	-	11%	24%	17%	-	10%	20%	14%

Note. Data comes from *Insee Panel tous salariés-EDP* with specific treatment to determine the annual principal socioprofessional category (see Section 2.3). We present here average statistics for workers in our estimation samples. Real wage changes are, for each group, the average difference in individual hourly wage (winsorized at 1th and 99th percentiles) over the period, adjusted for inflation (12% over 2002 - 2008, and 7% over 2009 - 2015), by socioprofessional group and mobility status.

Reading: over 2009 - 2015, intermediate profession workers experienced on average a 12% real wage increase. This change in real wage reaches 32% considering workers who have been promoted over the period, almost three times the average change for workers who stay in the same socioprofessionnal category over the 2009 - 2015 period (11%).

The same figures for downward social mobility are reported in Table 2.5. Agglomeration gains are less clear in the case of demotions, except for white-collar workers over 2002-2008.

Figure 2.1: Geography of upward mobility 2009-2015 (comparison with density heat map)



Note. Data on socioprofessionnal mobility comes from *Insee Panel tous salariés-EDP* with specific treatment to identify the annual main SPC (see Section 2.3). Data on employment density comes from INSEE population census. The maps are constructed by grouping employment zones into four quartiles and shading the areas so that lighter colors correspond to lower aggregate probability of upward social mobility or job density.

Table 2.4: Upward mobility depending on worker's commuting zone position in the distribution of jobs density.

<i>Position in the ZEs' density distribution:</i>	2002-2008			2009-2015		
	Bottom 10%	Middle 50%	Top 10 %	Bottom 10%	Middle 50%	Top 10 %
<b>All</b>	19%	21%	24%	14%	16%	22%
<b>Men</b>	22%	23%	26%	16%	17%	24%
<b>Women</b>	16%	19%	22%	13%	13%	20%
<b>Intermediate Professions</b>	9%	12%	17%	8%	10%	16%
<b>Skilled-Employees</b>	14%	18%	26%	11%	15%	20%
<b>Skilled-Blue-collar</b>	10%	11%	15%	8%	11%	16%
<b>Unskilled-Employees</b>	36%	35%	40%	22%	23%	29%
<b>Unskilled-Blue-collar</b>	49%	49%	52%	36%	41%	45%

Note. Data comes from *Insee Panel tous salariés-EDP* with specific treatment to determine the annual principal socioprofessional category (see Section 2.3). We present here average statistics for workers in our estimation samples (we keep only workers who can experience an upward mobility over the period). We classify workers depending on the position of their employment-zone in the distribution of zones with respect to density at the beginning of the period. Reading: 16% of intermediate profession workers working in a zone of the top 10% areas in terms of density in 2009, and still in employment in 2015, are in a SPC with a higher status in 2015. This proportion fall to 8% considering only individuals working in a zone of the bottom 10% in terms of density in 2009.

Table 2.5: Downward mobility depending on worker's commuting zone position in the distribution of jobs density.

	2002-2008			2009-2015		
	Bottom 10%	Middle 50%	Top 10 %	Bottom 10%	Middle 50%	Top 10 %
<b>All</b>	13%	13%	12%	10%	10%	10%
<b>Men</b>	14%	12%	11%	10%	10%	10%
<b>Women</b>	13%	13%	13%	10%	10%	10%
<b>White-collar</b>	26%	20%	15%	16%	14%	12%
<b>Intermediate Professions</b>	22%	19%	17%	14%	15%	15%
<b>Skilled-Employees</b>	7%	6%	2%	5%	5%	5%
<b>Skilled-Blue-collar</b>	11%	10%	9%	9%	9%	8%

Note. Same as Table 2.4 in the case of downward social mobility. Reading: 15% of intermediate profession workers working in a zone of the top 10% in terms of density in 2009, and still in employment in 2015, are in a SPC with a lower status in 2015. This proportion very slightly differs (14%) when considering only individuals working in a zone of the bottom 10% in terms of jobs density in 2009.

## 2.4 Empirical strategy

Promotion (or demotion) is a rare 1/0 event, with strong negative time auto-correlation: after a promotion, the chances of being promoted again in the following years are low. With such a dependent variable, we cannot estimate an equation with a worker fixed effect that would capture the workers' non-observable characteristics. Our empirical analysis then relies on cross-sectional variation in the chances of being promoted (or demoted) in the next six years. We seek to measure the role of working-zone characteristics on the probability of promotion (or demotion) of an individual over a period of six years. So we are interested in the parameter  $\gamma$  in equation (2.1).

$$\text{Proba}(\text{promotion}_{i,z}) = \alpha + \gamma C_z + x_i \beta + \varepsilon_{i,z} \quad (2.1)$$

with  $Proba(promotion_{i,z})$  a variable taking the value 1 (0 otherwise) if the worker  $i$  experiences a promotion in the next six years,  $C_z$  is either a variable of job density (in log), the concentration of highly-educated individuals (the share of people with at least three years of higher education – aged 15 and older, not including those in school), or the size of the labor market (log of total employment). Our hypothesis is that promotion is more likely in zones where these variables are higher, and will be validated if  $\gamma$  is positive.  $x_i$  includes observable characteristics of individual  $i$  (age, age<sup>2</sup>, work experience since 2002 (and squared), tenure at firm, industry indicators, SPC indicators, level of education), and  $\varepsilon_{i,z}$  is the error term.

There are potentially two issues in estimating directly this equation in our sample. Firstly, in the absence of a control for unobserved individual characteristics, equation (2.1) does not take into account the spatial sorting of individuals, i.e. the fact that the intrinsically most ambitious and able individuals, or those most likely to be promoted, are more likely to be found in specific zones (in particular, the densest or the more abundant in local human capital). Estimating equation (2.1) will then probably result in a positive bias in parameter  $\gamma$ . Secondly, there might also be some reverse causality, by which a higher chance of promotion in a given zone would attract workers, thus increasing the density and potentially the size and the human capital concentration of the zone. This would again generate a positive bias in the estimation of parameter  $\gamma$ .

**Spatial sorting.** Our strategy for dealing with spatial sorting consists in recovering individual fixed effects coefficients from a preliminary standard Micerian wage equation over the six years period. In practice, the idea is to estimate the following equation:

$$w_{(i,z,t)} = \eta + x_{it}\tau + \delta_z + \theta_t + a_i + \nu_{i,z,t} \quad (2.2)$$

with  $w_{(i,t,z)}$  the logarithm of the annual net hourly wage,  $\delta_z$  a zone fixed effect,  $x_{it}$  a vector of time-varying individual characteristics (work experience since 2002 (and its square), experience in the firm, socioprofessional category, sector of employment),  $\theta_t$  a year fixed effect,  $\nu_{i,z,t}$  an error term, and  $a_i$  capturing the effect on wages of unobserved (talent, ambition, etc.) and observed but constant (diploma, gender) workers' characteristics. Note that in this wage equation, we

can use the time dimension of our panel, and this is indeed what allows us to recover the worker characteristics  $\hat{a}_i$  that will help us reduce the bias on  $\gamma$ .

Our hypothesis is that the intrinsic characteristics of the worker that affect her wage also affect her chances of promotion or demotion. Therefore estimated  $\hat{a}_i$  can be included in equation (2.1) as a proxy for individual fixed effects on the probability of promotion (or demotion), allowing us to identify the impact of the characteristic of the zone, purged of potential spatial sorting bias. Thus we can estimate an unbiased effect of local characteristics on promotion (or demotion) chances. We therefore rewrite equation (2.1) as follows:

$$\text{Proba}(\text{promotion}_{i,z}) = \alpha + \gamma C_z + \phi \hat{a}_i + x_i \beta + \varepsilon_{i,z} \quad (2.3)$$

Note that the individual coefficients  $\hat{a}_i$  are drawn from a wage equation (2.2) that controls for socioprofessional category. If this were not the case, the effect for individuals promoted over the period would naturally be high (and vice-versa for those who move to a lower SPC), and therefore would bias both  $\phi$  and  $\gamma$  in equation (2.3).

**Reverse causality.** The strategy delineated above takes care of spatial sorting, but not of reverse causality. Equation (2.3) still has a potential identification problem, derived from the possibility for workers to choose their work localization. Indeed, even if we controlled for the workers' spatial sorting on unobservables with our proxy for individual fixed effects, individuals may be attracted by areas in which they expect higher chances of socioprofessional progression. Therefore this expected higher chance of promotion attracts workers (especially higher educated workers who are more mobile) and creates density and/or concentration of human capital: the causality is reversed.

To address the potential reverse causality we use the same IV approach applied in the first chapter of this thesis, which is very common in the economic geography literature. Recall that the literature found that reverse causality might be considered as a minor issue (Ciccone and Hall, 1996; Moretti, 2004b; Combes et al., 2010; De la Roca and Puga, 2017; Combes and Gobillon, 2015). We propose the same two IV models, IV1 and IV2, of chapter one to instrument the variables of density, local human capital and size (see section .2 of chapter one for details

on the IV models).

**IV1.** Excluded instruments for the density variable are the historical value of local employment density in 1806 and the local density of religious buildings built before the 20th century. We use these two instruments in level (i.e. total number not divided by the surface area) for the variable of size. The two instruments for the local human capital variable are the change in local share of higher-educated people between 1968 and 1982 and a shift-share type instrument *à la* [Moretti \(2004b\)](#) based on lagged age-structure and national education levels by age group.

**IV2.** In this model, we have three excluded instruments for local human capital variables: the number of secondary students in 1837 compared to population at the department-level, the share of conscripts with *baccalauréat* diploma in 1896 at the department-level and the change in local share of higher-educated people between 1968 and 1982. For the variables of density and size, this IV2 model includes two additional instruments compared to IV1: the number of royal road kilometers in 1837 in the department, the ratio of this number to the total surface of the department.

**Non linearity.** Finally, to take into account the fact that the effect might not be linear with respect to the level of  $C_z$ , we also use the decile of  $C_z$  instead of the value of  $C_z$

$$\text{Proba}(\text{promotion}_{i,z}) = \alpha + \sum_{j=2}^{10} \gamma_j \text{Decile}C_{j,z} + \phi \hat{a}_i + x_i \beta + \varepsilon_{i,z} \quad (2.4)$$

with  $\text{Decile}C_{j,z}$  taking the value 1 (0 otherwise) if the zone  $z$  is in the decile  $j$  regarding the distribution of the variable  $C_z$  (local density, human capital, or size).

## 2.5 Main results

We apply the identification strategy to the main sample of workers who have the potential to be promoted over the period 2009-2015 (346,690 workers). Results are presented in [Table 2.6](#). When we do not control for spatial sorting of workers, we estimate a semi-elasticity of the probability of upward mobility with respect to density of 1.2% (column (1) of upper part of

Table 2.6).

Table 2.6: Regression results - Influence of local characteristics on workers' promotion between 2009 and 2015 (density, human capital and size).

Dependent variable : Probability of upward mobility 2009-2015				
	(1)	(2)	(3)	(4)
	OLS	OLS	IV1	IV2
Log Density	0.012*** (0.0006)	0.009*** (0.0004)	0.008*** (0.0006)	0.008*** (0.0006)
Proxy for workers' ability	-	0.330*** (0.0113)	0.330*** (0.0112)	0.330*** (0.0063)
Observations	346,690	346,690	346,690	330,735
R <sup>2</sup>	0.15	0.18	0.18	0.18
Local Human Capital	0.301*** (0.0220)	0.228*** (0.0171)	0.238*** (0.0222)	0.233*** (0.0154)
Proxy for workers' ability	-	0.330*** (0.0111)	0.330*** (0.0111)	0.329*** (0.0061)
Observations	346,690	346,690	346,690	330,735
R <sup>2</sup>	0.15	0.18	0.18	0.18
Log Size	0.014*** (0.0008)	0.010*** (0.0009)	0.010*** (0.0013)	0.010*** (0.0013)
Proxy for workers' ability	-	0.330*** (0.0113)	0.330*** (0.0113)	0.330*** (0.0065)
Observations	346,690	346,690	346,690	330,735
R <sup>2</sup>	0.15	0.18	0.18	0.18

Notes. This table presents estimation results based on equation (2.3) for the period 2009-2015. Each column is a separate regression. All specifications include controls for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socio-professional category, two-digit industry. Each part of the Table corresponds to a specific local characteristic of interest (introduced separately): log density, local human capital or log size. The proxy for workers' unobserved ability corresponds to the coefficients of individual fixed effects estimated in a first equation (2.2) on wages. Columns (3) and (4) present the IV estimates. See Table 16 for first stage IV estimates and statistics.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Recall that the average chances of promotion in our sample is 18% over 2009-2015 period.

As previously noted, spatial sorting of workers on unobservable characteristics is likely to lead to a positive bias for this estimated semi-elasticity. Indeed, when we introduce the proxy for workers' ability, the semi-elasticity of promotion to density drops to 0.9% (see column (2)), and to 0.8% when using IV (see column(3) and (4)). This means that, all other things being

equal, a doubling of density is associated with a 0.8-0.9 percentage point higher probability of promotion. With an interdecile difference of a factor of 8 in 2009 between the top 10% and the bottom 10% in density, we expect a promotion premium of 1.6 - 1.8 points<sup>10</sup> in a very dense area compared to a relatively sparse area. Also note that the coefficient associated with our proxy for unobserved workers' ability estimated from the wage equation is, as expected, significant and positive with the same value in OLS or IV estimations, and results in a higher  $R^2$ . The results for the wage equation from which the proxy for workers' ability is obtained, and estimation results of the first stages IV, are presented in Tables 15 and 16. We do not comment on these tables as these specifications and results are in line with the literature (Combes and Gobillon, 2015; Combes et al., 2010; De la Roca and Puga, 2017; Moretti, 2004b) – see the first chapter of this thesis for more detail on the IV models. The same approach is replicated twice to study the impact on the chances of being promoted of the local human capital concentration, and the impact of the size of the local labor market. Once we correct for spatial sorting of workers, we find that the impact of an increase of 1 point in the ratio of higher-educated people increases the chance of being promoted by around 0.23 percentage point (0.228 with OLS and 0.238-0.233 with IV estimates, see middle part of Table 2.6 columns (2), (3) and (4)). With an interdecile gap of 8 points, this translates into a 1.8 points increase in the chance of being promoted when working in a high human capital concentration area, compared to an area with a relatively low concentration of human capital. Effects are qualitatively similar when using the size of the zone as an explanatory variable (compared to the regression on log density). The semi-elasticity of the probability of upward mobility with respect to a zone's size is 0.1%. Table 17 reports similar results for upward mobility between 2002 and 2008. Estimated coefficients are the same as those estimated for the period 2009-2015, and therefore in the remainder of this paper we focus only on the period 2009-2015.

**Spatial mobility of workers.** Before moving forward in our analysis, we need to check that our results regarding the effect of density, human capital concentration, and size, are stable when we restrict our sample to those individuals that stay in the same commuting zone over the period. We need to do that because while our strategy solves the issue of spatial sorting

---

<sup>10</sup> =  $\ln(8) \times 0.8$  or =  $\ln(8) \times 0.9$

of individuals until the beginning of our observation period (2009), it is not enough to deal with workers moving to another commuting zone from 2010 on. To illustrate the potential bias this may generate, we can consider the case of an individual working in a disadvantageous zone (in terms of density, human capital, or size) in 2009. If she moves to a denser zone in 2010, where she benefits from agglomeration gains that contribute to her promotion over the period 2010-2015, then the influence of the initial zone in 2009 would be overstated (biasing downward agglomeration effects on promotion). The opposite case may also lead to an underestimation of agglomeration gains on promotion. Suppose an individual working in an advantageous zone (in terms of density, human capital, or size) in 2009. If she moves to a less favorable zone in 2010, where she does not benefit from agglomeration gains, which blocks her promotion over the period 2010-2015, then the influence of the initial dense zone in 2009 would be reduced, which would also bias downward the coefficient of interest. Another related issue may come from individuals that move back and forth between different commuting zones (for example, from a low-density zone to a dense zone to be promoted, and then back to a less dense zone, which again would bias downward the effect of the agglomeration effect on promotion). To account for the fact that individuals are likely to change commuting zones between 2009 and 2015, we run the equation (2.3) on two restricted samples that keep only workers that are observed in the same zone in 2009 and in 2015 (275,420 individuals), and workers that never change zones over the whole period (225,025 individuals). Results presented in Table 2.7 are from the OLS estimation, and include the proxy for workers' ability as well as the same control variables as before. They suggest that estimates based on the full sample were indeed slightly underestimated. The coefficients associated with log density increase to 0.011 and 0.009; those of local human capital to 0.287 and 0.259; and those of log size to 0.014 and 0.013.

**Decile model.** Equation (2.3) allows only for a linear impact of the zone characteristics (size, density, human capital concentration) results on the effect of density on promotion chances, which means that results presented in Table 2.6 are estimated under the assumption that density (or rather its log) plays the same role at all density levels. It is possible, however, that density affects promotions only at sufficient density levels, and little or not at all at low density levels. To put it another way, a doubling of density at a low level of density (e.g. from Pau (36 jobs

Table 2.7: Promotion premium with respect to density, human capital or size. Stayers sub-samples.

Dependent variable: Probability of upward mobility between 2009 and 2015			
	(1)	(2)	(3)
	OLS	OLS	OLS
	All	Same ZE in 2009 and 2015	Same ZE over 2009-2015
Log Density	0.009*** (0.0004)	0.011*** (0.0008)	0.009*** (0.0007)
Observations	346,690	275,420	225,025
R <sup>2</sup>	0.18	0.15	0.14
Local Human Capital	0.228*** (0.0171)	0.287*** (0.0133)	0.259*** (0.0154)
Observations	346,690	275,420	225,025
R <sup>2</sup>	0.18	0.15	0.14
Log Size Proxy for workers' ability	0.010*** (0.0009)	0.014*** (0.0011)	0.013*** (0.0008)
Observations	346,690	275,420	225,025
R <sup>2</sup>	0.18	0.15	0.14

Notes. This table presents estimation results based on equation (2.3) for two sub-samples depending on workers' migration paths over the period 2009 - 2015. Column (1) is the OLS base case on the full sample already presented in column (2) of Table 2.6. All specifications include controls for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socio-professional category, two-digit industry, our proxy for worker's unobserved characteristics. Columns (2) presents the estimation results on 275,420 workers who are in the same zone in 2009 and in 2015. Columns (3) presents the estimation results on the 225,025 workers who stay in the same zone over the whole period. Each part of the Table corresponds to a specific local characteristic of interest (introduced separately): log density, local human capital or log size.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

per km<sup>2</sup>) to Saint-Nazaire (71 jobs per km<sup>2</sup>) does not necessarily have the same effect on the chances of promotion as a doubling of density at high levels of density (e.g. from Saint-Nazaire to Montpellier 130 jobs per km<sup>2</sup>). And intuition suggests that density effects are much stronger at high density levels. To get around the linearity assumption in equation (2.3), we replace the level of density (or human capital, or size) with the decile of density (or human capital, or size) as in equation (2.4).

The results are presented in column (1) of Table 2.8. Compared to the first density decile,

Table 2.8: Influence of local characteristics on workers' promotion between 2009 and 2015 (density, human capital and size). Decile Model.

Dependent variable : Probability of upward mobility between 2009 and 2015			
Variable of interest:	Log Density	Human Capital	Log Size
	(1)	(2)	(3)
Working zone decile (ref. first decile)			
#2	-0.004 n.s. (0.0057)	0.004 n.s. (0.0043)	-0.006 n.s. (0.0091)
#3	-0.006 n.s. (0.0063)	0.006 n.s. (0.0049)	-0.003 n.s. (0.0073)
#4	-0.000 n.s. (0.0055)	0.009* (0.0048)	0.000 n.s. (0.0080)
#5	0.006 n.s. (0.0064)	0.008* (0.0045)	-0.001 n.s. (0.0075)
#6	0.005 n.s. (0.0055)	0.013*** (0.0043)	-0.003 n.s. (0.0070)
#7	0.011* (0.0062)	0.013*** (0.0045)	-0.004 n.s. (0.0071)
#8	0.002 n.s. (0.0054)	0.019*** (0.0043)	0.001 n.s. (0.0070)
#9	0.012** (0.0056)	0.021*** (0.0042)	0.009 n.s. (0.0074)
#10	0.027*** (0.0076)	0.039*** (0.0059)	0.023*** (0.0079)
Observations	346,690	346,690	346,690
R <sup>2</sup>	0.18	0.18	0.18

Notes. This table presents estimation results for the OLS specification based on equation (2.4) for the period 2009-2015. We reproduce regressions of column (2) of Table 2.6 (including the proxy for worker's ability), but changing the continuous explanatory variables of interest into categorical decile variables. The reference is the first decile of the working zones' distribution in 2009 (with respect to density, human capital or size). Each column is a separate regression. All specifications include controls for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socio-professional category, two-digit industry, proxy for worker's unobserved characteristics. Column (1) corresponds to the regression with density deciles. Column (2) correspond to the regression on local human capital deciles. Column (3) corresponds to the regression on zone's size deciles.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Reading. All things being equal, the probability of promotion between 2009 and 2015 is increased by 2.7 percentage points for one worker working in a zone of the top 10% in terms of density in 2009, compared to a worker in a zone of the bottom 10%.

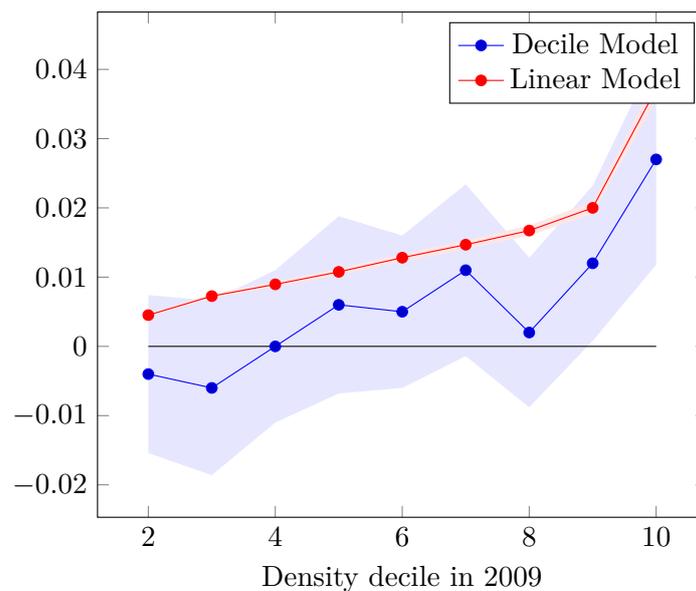
only the last decile is characterized by a significantly higher chance of promotion (at a 1% level).

Only individuals working in the most dense areas would therefore have a promotion premium.

This would be of the order of 2%, i.e. the chances of being promoted over a period of six years are

two percentage points higher for individuals working in the densest areas than for those working in the areas of the eight least dense deciles, and one percentage point higher than for those in the areas belonging to the 9th density decile. The Figure 2.2 represents the promotion premium (in relation to the first density decile) estimated under the assumption of a linear model (see Table 2.6 , column (2)), and that resulting from the discrete model (by density decile) (see Table 2.8, column (1)). While the general shape of the curve is the same in both cases, the lessons that can be drawn from it are quite different. With the linear model (red curve) it appears that the chances of promotion increase regularly with the level of density (significant at 1% even for low levels of density), whereas this is no longer the case in the discrete model (blue curve), in which the promotion premium is only significant for the last decile.

Figure 2.2: Promotion premium 2009-2015 with Density decile (compared to 1st decile)



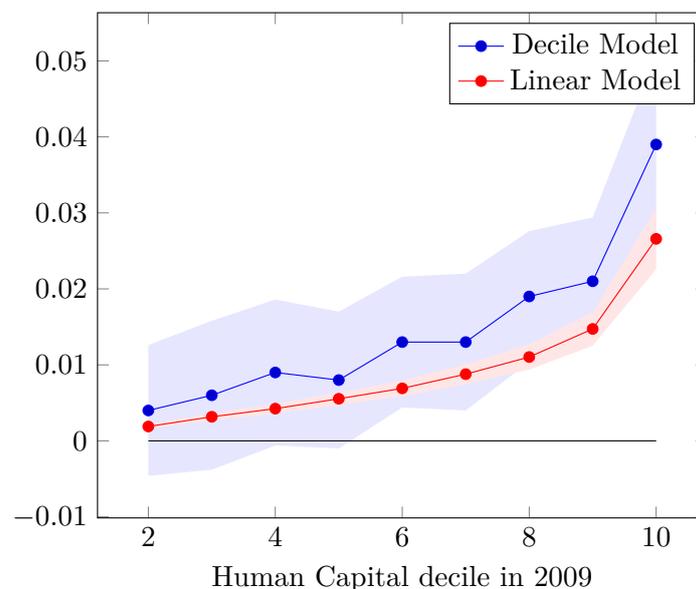
Note. This Figure represents the promotion premium with respect to an average zone in the first density decile. Red plots correspond to the promotion premium estimated under the assumption of a linear model (see Table 2.6 , column (2)), and blue plots result from the discrete model (by density decile) (see Table 2.8, column (1)).

Reading: considering the linear model in red, all other things being equal, workers in a zone of the 6th decile in terms of density have about 1.3 percentage points higher probability of promotion between 2009 and 2015 compared to workers in a 1st decile zone. Considering the decile model in blue, this promotion premium is about 0.5 percentage points, but not significantly different from zero (regarding the standard errors represented by the blue area).

The same steps can be implemented according to the decile of human capital concentration. Qualitatively, the conclusions are more or less the same. It appears that the estimated average effect of human capital on the chances of promotion with the linear model may be misleading

as to the effects to be expected in areas with few higher educated individuals. It is only in areas with relatively high levels of human capital that the chances of promotion are significantly higher. The results presented in column (2) of Table 2.8, show an effect from the 6th decile of human capital (i.e. from a ratio of higher education graduates of 8%). Compared to the first decile, the promotion premium in the areas of the 6th decile of human capital would be one percentage point, around two percentage points for the 8th and 9th deciles, and four percentage points for the 10th decile. The differences between the linear and discrete models are shown in Figure 2.3. In the case of human capital, the decile model shows higher point estimates compared to that of the linear model. This was not the case with the model with density deciles.

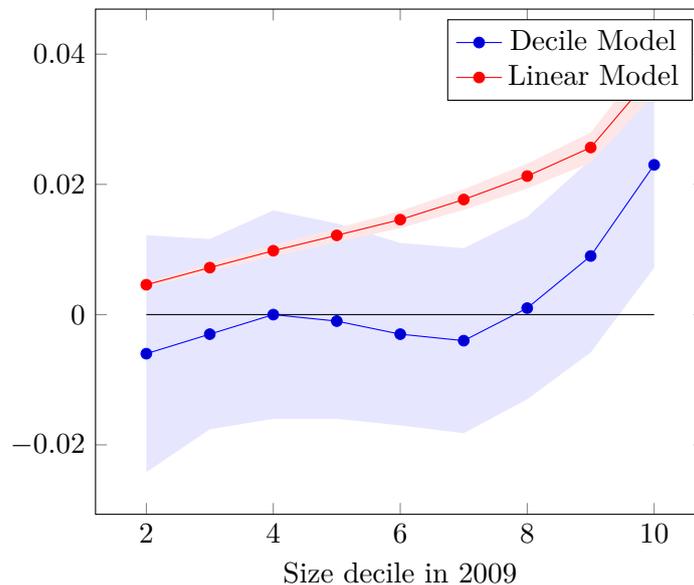
Figure 2.3: Promotion premium 2009-2015 with Human Capital decile (compared to 1st decile)



Note. This Figure represents the promotion premium with respect to an average zone in the first local human capital decile. Red plots correspond to the promotion premium estimated under the assumption of a linear model (see Table 2.6, column (2)), and blue plots result from the discrete model (by local human capital decile) (see Table 2.8, column (2)). See Figure 2.2 for reading example.

We now apply the identification strategy to the sample of workers who have the potential to be demoted over the period 2009-2015 (325,872 workers). Taking into account potential sorting bias, that leads to a non-significant impact of local density (see last column of Table 18). This result is robust to the estimations based on restricted samples of stayers. Unlike upward social mobility, downward mobility does not depend on the local density of jobs. Also, the influence of local human capital and size is no longer statistically significant.

Figure 2.4: Promotion premium 2009-2015 with Size decile (compared to 1st decile)



Note. This Figure represents the promotion premium with respect to an average zone in the first size decile. Red plots correspond to the promotion premium estimated under the assumption of a linear model (see Table 2.6 , column (2)), and blue plots result from the discrete model (by size decile) (see Table 2.8, column (3)). See Figure 2.2 for reading example.

**An encompassing model.** We have successively estimated the impact on social mobility of three local-level characteristics related to agglomeration economies: density, size, and human capital. We can infer from our results that workers' chances of socioprofessional upward mobility are higher in agglomerated areas (in terms of jobs and skills). This promotion premium in these types of zones may arise from a number of different mechanisms, especially search, and matching or learning effects (see section 2.2).

We now simultaneously introduce different zone-level factors in our specification in order to estimate the magnitude of each local determinant with respect to the others, and give some indications to gauge the role of different agglomeration channels on the chances of climbing the social ladder during one's career. In addition to the three variables of density, human capital, and size, we consider the role of two other common local agglomeration determinants (Combes and Gobillon, 2015): market potential and industrial diversity. We also introduce the local employment growth between 2009 and 2015, to take into account the fact that firms that expand substantially also add layers (Caliendo et al., 2015), which may facilitate workers' promotion.

The results are seen in columns (1) to (4) of Table 2.9.<sup>11</sup> The new variable of market potential shows a significant positive impact (coefficient of 0.008-0.009) on the probability of promotion between 2009 and 2015, whereas local sectoral diversity is significant only at the 10% level. Local employment growth does not significantly influence the chance of promotion. Principally, the inclusion of all local factors in the same regression does not significantly affect the coefficient associated with human capital, whereas the elasticities of density and size are no longer significantly different from zero.

Table 2.9: Local determinants of agglomeration effects on promotion between 2009 and 2015.

Dependent variable:	Probability of upward mobility 2009-2015				Zone FE coefficients		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	IV1	IV2	OLS	IV1	IV2
Log of zone Density	0.001 n.s. (0.0013)	-0.001 n.s. (0.0015)	-0.002 n.s. (0.0017)	-0.002 n.s. (0.0018)	-0.002 n.s. (0.0021)	-0.003 n.s. (0.0025)	-0.003 n.s. (0.0025)
Local Human Capital	0.210*** (0.0303)	0.203*** (0.0313)	0.227*** (0.0422)	0.248*** (0.0274)	0.255*** (0.0499)	0.251*** (0.0719)	0.292*** (0.0687)
Log of zone Size	-0.001 n.s. (0.0015)	-0.002 n.s. (0.0019)	-0.002 n.s. (0.0024)	-0.004* (0.0021)	0.002 n.s. (0.0025)	0.004 n.s. (0.0037)	-0.002 n.s. (0.0037)
Log Diversity	-	0.013* (0.0073)	0.015* (0.0090)	0.018** (0.0077)	-0.013* (0.0077)	-0.016 n.s. (0.0101)	-0.006 n.s. (0.0144)
Log Market Potential	-	0.008*** (0.0019)	0.009*** (0.0019)	0.009*** (0.0022)	0.010*** (0.0028)	0.011*** (0.0030)	0.010*** (0.0039)
Employment growth 09-15	-	0.004 n.s. (0.0291)	0.004 n.s. (0.0289)	0.047 n.s. (0.0325)	0.016 n.s. (0.0339)	0.017 n.s. (0.0352)	0.015 n.s. (0.0419)
Observations	346,690	346,690	346,690	330,735	304	304	290
R <sup>2</sup>	0.18	0.18	0.18	0.18	0.20	0.20	0.19
Weak id Wald F stat.	-	-	16.998	19.816	-	20.704	28.762
Stock-Yogo critical values (5% max)	-	-	12.20	16.10	-	12.20	16.10
P-value Hansen J Stat	-	-	0.331	0.102	-	0.032	0.482
P-value LM Stat	-	-	0.000	0.005	-	0.000	0.003

Notes. This table presents estimation results based on equation (2.3) when all our local factors related to agglomeration economies are simultaneously introduced. Columns (3) and (4) present IV models where density, size and human capital are instrumented. See chapter one of this thesis for detail on IV1 and IV2 models. 2SLS first-stage estimates and statistics are reported in Table 19. Specifications in columns (1) to (4) include controls for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at the firm, socio-professional category, two-digit industry, our proxy for worker's unobserved characteristics.

Columns (5) to (7) present the second stage estimation results where we regress the 304 coefficients of zone fixed effects estimated in first-stage on local characteristics. This two-stage model aims at taking into account the overstatement of local characteristics multicollinearity in the regression on 346,690 individuals (see section 2.5).

Coefficients are reported with clustered standard errors at the zone level in parenthesis (and zone-department level for IV2 models). \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

One problem with having more than one zone-level characteristic in the econometric model is that these characteristics are likely to be highly correlated, leading to bias coefficients. A simple

<sup>11</sup>Columns (3) and (4) present IV models where density, size and human capital are instrumented. See chapter one of this thesis for detail on IV1 and IV2 models. Note that we could also instrument the market potential variable, as in chapter one. Since we already have three instrumented variables and chapter one has already shown that the endogeneity bias from market potential appears to be of small importance, we do not add this treatment here, for simplification reasons. Nevertheless, when we do so, the results are not significantly different.

post-estimation variance inflation factors (VIF) check rejects a perfect linear relationship among our local predictors. However, these VIF seem large for log size (VIF value about 9), log density (VIF value about 8), and local human capital (VIF value about 6), although they remain above the minimum tolerance value of 0.1 (1/VIF). Indeed, correlation coefficients between density, size, and human capital are about 90% in our sample. These correlations between zone-level predictors are overestimated in our sample of individual observations. Computing correlation in the sample of the 304 zones (rather than individuals), these correlations appear to be smaller (around 60%-70%). We thus run a two-stage strategy where we first introduce zone fixed effects  $\sigma_z$  in equation (2.3) (therefore omitting the variables of local characteristics  $C_z$ ):

$$\text{Proba}(\text{promotion}_{i,z}) = \alpha + \phi \hat{a}_i + x_i \beta + \sigma_z + \varepsilon_{i,z} \quad (2.5)$$

We then regress the 304 estimated zone fixed effects coefficients  $\hat{\sigma}_z$  in a second-stage specification on our local predictors:

$$\hat{\sigma}_z = \alpha + \gamma C_z + \epsilon_z \quad (2.6)$$

As expected, this second-stage specification on 304 zone-level observation shows much lower VIF values (4, 3 and 2 for log size, log density, and human capital respectively). Coefficients presented in columns (5) to (7) of Table 2.9 are very similar to our base model.<sup>12</sup>

These results suggest that the agglomeration effects on upward mobility in dense and big areas come primarily from human capital externalities and proximity to other dense markets, rather than pure urbanization and scaled effects.

## 2.6 Robustness

In this section, we report several alternative specifications designed to test the robustness of the main results described above. All estimation results are reported in the summary Table 2.10.

---

<sup>12</sup>While it helps to resolve multicollinearity concerns, this two-stage estimation strategy raises other consistency issues related to potentially less precise estimates of zone fixed effects for areas with few observations in the first stage. These potential measurement errors may create some heteroscedasticity that biases second-stage estimations.

**Women and Men sub-samples.** We first explore whether there is heterogeneity in the advantages from dense areas depending on gender. Applying the identification strategy to gender sub-samples over the period 2009-2015, we detect slight difference in coefficients associated with local characteristics, with larger impacts on the chances of promotion for women than for men. Lines (d) and (e) of Table 2.10 show that the density of the working zone increases the chances of promotion by 1 point for women and 0.8 point for men. In the regressions on local human capital, the coefficients are 0.252 for women versus 0.212 for men. In the regressions on log size, the coefficients are 0.011 for women versus 0.010 for men. However, in view of standard errors, the coefficients in the two sub-samples do not appear to be significantly different.

**Immigrants.** Immigrant workers face discrimination in job markets (OCDE, 2013). If they are not equally dispersed over employment zones, this may bias the results concerning the geography of social mobility. Indeed, the over-representation of immigrant workers in denser zones (compared to the national average) may underestimate the influence of density because of the barriers to upward mobility faced by these workers. We check this hypothesis by removing immigrant workers from our base sample. We find that coefficients associated with local characteristics substantially increased to 0.011, 0.281, and 0.012 for log density, local human capital, and log size respectively (see line (f) of Table 2.10).

**Industry specificities and industry fixed effects.** We also run sensitivity checks, dropping specific industries from our estimation sample. As explained in the first chapter of this thesis, several large firms group together on a single administrative declaration, work positions that are actually localized in several establishments of the firm. This creates a bias in locating workers at their place of work (about 5% of workers). Since 2003 validity, INSEE has adjusted this bias using the information system CLAP (*Local knowledge on establishment production*) to unbundle the work periods and reassign workers to their correct establishments (in particular by using workers' place of residence). However, a few gaps remain for about 0.5%<sup>13</sup> of the workforce because of the complexity of the unbundling algorithm. Workers may still be located at the site of the headquarters (and not at the correct work place), especially in financial sectors or public

---

<sup>13</sup>Information provided in *INSEE DADS Grand Format - Guide Utilisateur Validité 2014*.

companies such as SNCF and La Poste. To ensure the robustness of our results, we exclude these industries from our sample.

Line (g) of Table 2.10 shows results excluding the financial and insurance sectors, with no significant difference in the results compared to our base model. Line (h) of Table 2.10 presents results excluding all the public sector (thus the SNCF and La Poste jobs), within which workers may experience specific career dynamics - possibly less affected by agglomeration effects. Again, we see no significant difference compared to our main results.

So far, all our specifications include industry fixed effects. We know however, that industries are not randomly distributed across employment zones. If agglomeration gains occur through industrial composition of local labor markets we underestimate the promotion premium when we include industry fixed effects. Indeed, line (i) of Table 2.10 shows that when we removed the industry fixed effects the coefficient associated with local characteristics increases by at least 40%.

**Alternative definitions of upward mobility.** We test the robustness of our results with alternative definitions of upward socioprofessionnal mobility (see section 2.3). Using the hierarchical classification of Le Moigne (2021), we find the same positive and significant influence of local factors on the chances of social promotion, with lower value of coefficients (see line (j) of Table 2.10). Using our second alternative definition of upward mobility based on wage gain (at least 35% between 2009 and 2015), we find results similar to our base model (see line (k) of Table 2.10). Finally, our results are robust to a third more restrictive definition of upward socioprofessionnal mobility, taking the value 0 if a promotion in terms of SPC (the main definition based on hierarchical layers, from INSEE) is not associated with a wage increase of more than 35% over the 2009-2015 period (see line (l) of Table 2.10). Using this definition, only 7% of the sample is promoted over the period.

**Alternative database.** Line (m) of Table 2.10 presents estimation results using data based on an alternative identification of the annual occupation of one worker. For each individual of the panel, we characterize her main job position each year as the most remunerative position of that year (see section 2.3). Using this alternative data estimation, results based on equation

(2.3) are very similar to the base model.

Table 2.10: Influence of local characteristics on workers' promotion between 2009 and 2015. Robustness checks synthesis.

Dependent variable : Probability of upward mobility between 2009 and 2015				
		Log Density	Human Capital	Log Size
	Obs.	(1)	(2)	(3)
(a) Base Model	346,690	0.009*** (0.0004)	0.228*** (0.0171)	0.010*** (0.0009)
(b) Same zone in 2009 and 2015	275,420	0.011*** (0.0008)	0.287*** (0.0133)	0.014*** (0.0011)
(c) Same zone over 2009-2015	225,025	0.009*** (0.0007)	0.259*** (0.0154)	0.013*** (0.0008)
(d) Women	177,585	0.010*** (0.0005)	0.252*** (0.0189)	0.011*** (0.0011)
(e) Men	169,105	0.008*** (0.0006)	0.212*** (0.0138)	0.010*** (0.0011)
(f) without Immigrants	320,698	0.011*** (0.0005)	0.281*** (0.0163)	0.012*** (0.0014)
(g) without Finance	336,961	0.009*** (0.0004)	0.223*** (0.0181)	0.010*** (0.0009)
(h) without Public Administration	237,816	0.008*** (0.0006)	0.218*** (0.0240)	0.009*** (0.0009)
(i) without Industry FE	346,690	0.013*** (0.0004)	0.314*** (0.0154)	0.015*** (0.0013)
(j) Alternative SPC classification	332,985	0.006*** (0.0005)	0.142*** (0.0091)	0.007*** (0.0009)
(k) Promotion based on wage growth	345,741	0.010*** (0.0005)	0.233*** (0.0109)	0.010*** (0.0016)
(l) SPC change and 35% wage growth	346,690	0.007*** (0.0003)	0.170*** (0.0073)	0.008*** (0.0009)
(m) Database most paid job position	395,776	0.009*** (0.0005)	0.235*** (0.0187)	0.011*** (0.0008)

Notes. This table presents estimation results for our different robustness checks. Each entry is a separate regression.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

## 2.7 Plausible mechanisms

After controlling for potential spatial sorting, the upward mobility premium combines two broad types of effect. A direct effect from agglomeration gains that is external to the worker (linked to

her immediate environment), and an internal effect from experience accumulated in dense areas. While the external effect is likely to vanish if the worker relocates to a new area, the internal effect remains. In this section, we successively examine these two different effects, with regard to several potential mechanisms underlying the promotion premium in dense areas.

### 2.7.1 External vs. internal promotion

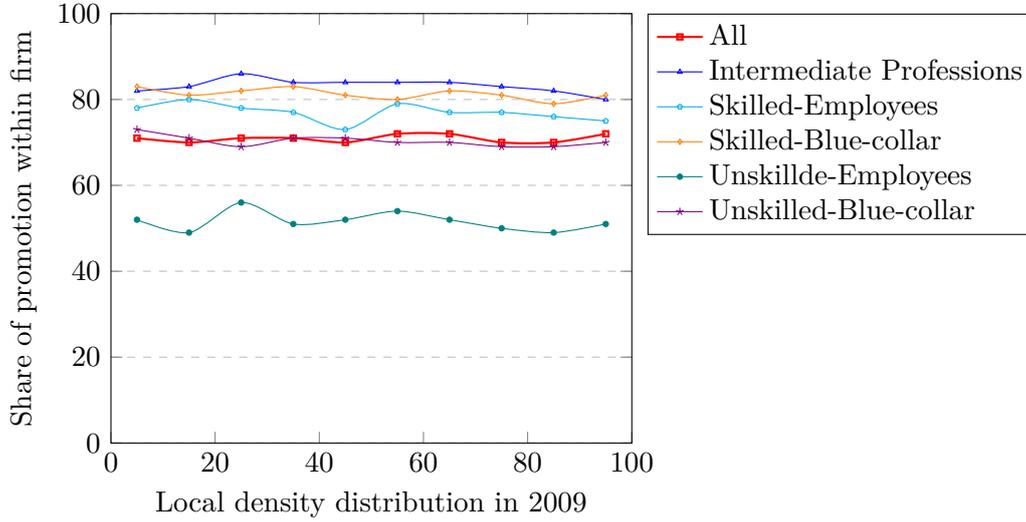
To study the mechanisms that favor promotion in denser areas, we can draw on the main results of the empirical literature on the origin of agglomeration gains on productivity and wages. On the one hand, the size of the labor market and the turnover of workers between firms and industries are at the origin of higher wages in dense areas (Duranton and Puga, 2004; Combes and Gobillon, 2015). On the other hand, it appears that firms in dense areas have different characteristics from those in less dense areas. There are more hierarchical levels in firms in dense areas (Spanos, 2019), and firms are more innovative, which is associated with better wages, including at the bottom of the ladder (Aghion et al., 2019). Intuition suggests that these mechanisms that lead to higher wages can also be put forward to account for higher chances of promotion within firms.

Here we seek to identify whether the higher probability of promotion in dense areas stems from the size of the labor market (which favors mobility), or from the specificities of firms in these areas. A simple way to look at this is to calculate the share of intra-firm promotions in total promotions, and in the share of promotions that are coupled with a change of firm. If promotions are more related to the specificities of firms that favor internal promotion in a given area, then we should observe that the share of internal promotions is relatively higher in dense areas. If, on the contrary, promotions are more related to the size of the market and the opportunities it offers, then we should observe that internal promotions are relatively less significant in dense areas.

Figure 2.5 shows no distortion in the relative probabilities of internal and external promotion in dense areas compared to less dense areas, regardless of initial socioprofessional category. These statistics suggest that immediate environment effects play as much a role as firm-specific effects in the higher promotion rate in dense areas. This remains true if we classify areas not by their

density but by their human capital concentration.

Figure 2.5: Share of promotion within firm by decile of working zone's density or human capital.



A more precise way to measure the effect of density on the chances of external promotion relative to internal promotion is to concentrate on individuals that have been promoted over the period 2009-2015, such that only individuals that are promoted appear in the sample. An individual who is promoted twice (say in 2010 and in 2013) appears twice in the sample. The equation we estimate is the following:

$$\text{Proba}(\text{EXTpromotion}_{i(z,t)}/\text{promotion}_{i(z,t)} = 1) = \alpha + \theta C_z + \gamma_t + \phi \hat{a}_i + x_{i(z,t)}\beta + \varepsilon_{i(z,t)} \quad (2.7)$$

$x_i$  contains the same observable characteristics of the individual  $i$  in the year he is promoted (age, age<sup>2</sup>, work experience since 2002 (and squared), tenure at firm, SPC, the sector of employment, gender and level of education);  $a_i$  is, as before, the proxy for the worker's ability,  $\theta_t$  a year fixed effect, and  $\varepsilon_{i(z,t)}$  an error term.  $C_{z,t}$  is either a variable of job density (in log), the concentration of highly-educated individuals (the share of people with at least three years of higher-education – aged 15 and older, not including those in school), or the size of the labor market (log of total employment).

We are interested in the coefficient  $\theta$ . A positive value of  $\theta$  means that the density, size, or human capital favor external promotion, whereas a negative value of  $\theta$  would be interpreted as the density favoring internal promotion over external promotion.

Table 2.11: Estimated elasticity of promotion outside the firm (vs. intra-firm promotion) with respect to local density, human capital or size.

Dependent variable: Probability of upward mobility outside the firm (2009-2015)		
	(1)	(2)
	OLS	OLS
Log Density	0.002** (0.0009)	0.002*** (0.0009)
Size of firm (ref. less than 10 workers)		
10 to 199 workers	-	-0.031*** (0.0036)
200 workers and more	-	-0.036*** (0.0053)
Observations	154,442	154,442
R <sup>2</sup>	0.15	0.15
Local Human Capital	0.051*** (0.0191)	0.056*** (0.0185)
Size of firm (ref. less than 10)		
10-199 workers	-	-0.031*** (0.0036)
200 workers and more	-	-0.036*** (0.0053)
Observations	154,442	154,442
R <sup>2</sup>	0.15	0.15
Log Size	0.003** (0.0011)	0.003*** (0.0011)
Size of firm (ref. less than 10)		
10-199 workers	-	-0.031*** (0.0036)
200 workers and more	-	-0.036*** (0.0053)
Observations	154,442	154,442
R <sup>2</sup>	0.15	0.15

Notes. This table presents estimation results based on equation 2.7 regressing the probability for a worker to be promoted outside her firm. The period of analysis is 2009-2015. All specifications include controls for age, age-squared, gender, level of education, work experience since 2002 (and squared), tenure at firm, socio-professional category, two-digit industry. Each part of the Table corresponds to a specific local characteristic of interest (introduced separately): log density, local human capital or log size. Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Regardless of the specification chosen, we obtain a positive value for  $\theta$ , which means that, on average, promotions are more external in the densest areas after controlling for firms' and workers' characteristics. More precisely, when controlling for the size of the firm in which the worker was working before being promoted, the effect of density on external promotion is clearly positive (see Table 2.11 column (2)). This positive effect is smaller or less significant when we do not control for firm size (column (1)). This is what one would expect: firms are on average larger in denser areas, which would be likely to facilitate internal promotion over external promotion, and thus decreases the coefficient of density when we do not control for a firm's size. Indeed, the fixed effects that control for firm size are negative for larger firms (cf. column (2)), meaning that the chances of internal promotion are higher in larger firms. The same is true when the independent variable is the human capital concentration or the zone size.

All in all, if both internal and external promotion rates are higher in denser zones, the effect of density and/or local human capital on promotion is higher for external promotions than for internal promotion.

### **2.7.2 The role of work experience in top areas**

Another way of understanding the impact that zone characteristics have on promotions is to study the impact of the experience individuals accumulate by working in zones that are either dense or rich in human capital, compared to experience acquired in other zones. If the experience accumulated by workers in dense areas is portable, then these workers will be more likely to be promoted even after leaving these dense areas. We now focus on this internal component of the promotion premium by exploring whether the experience acquired in a dense area – or an area with abundant human capital – increases the probability of being promoted after relocation to less dense areas.

To this end, we construct a new variable that represents the experience accumulated by each worker in a "top area" over the period 2002-2008. "Top areas" are the 5% densest areas in 2009 (15 commuting zones) or the next top 5% densest areas in 2009 (next 15 commuting zones). In addition, we also consider "top areas" to be the top 5% and the next top 5% areas in terms of human capital in 2009 (share of higher-educated people, not in school), and in terms of size

(total number of jobs). The list of commuting zones that appears in these top 5% and next top 5% lists is presented in Table 20 in the annex. One can see that these three lists do not overlap, as some very dense areas (Roubaix - Tourcoing, or Evry) are neither among the larger or the richest in term of human capital, while Aix-en-Provence and Rambouillet, which are in the top 5% for human capital, are not in the top 10% lists for density, or for size.

To measure the role played by experience in top areas, we conduct the following experiment. We restrict our sample to include only those workers that were not in a top 10% area in either 2009 or 2015, and examine their chances of being promoted according to their previous experience in a top area. This restriction therefore excludes workers that still work in the top 10% areas in 2009, and benefit from a direct substantial promotion premium regardless of their previous work experience. This restriction allows us to eliminate the cases of promotions of individuals who, although they had no work experience in a top area before 2009, were in a dense zone in 2009 and thus benefited *de facto* from the promotion opportunities specific to these zones. This restriction also allows us to eliminate individuals who joined a dense area after 2009, in which they were promoted.

The unconditional probabilities of promotion for individuals not working in a dense area in either 2009 or 2015 are given in Table 2.12. Individuals who had (before 2009) experience in a dense area (even of only one year), have a probability of promotion of around 20%, whereas it is only 15% for those who had no experience of working in a dense area in 2009. The same differential is found when considering past experience in areas with high human capital (column 2 of Table 2.12), or in the largest areas (column 3 of Table 2.12). Behind these differences in probability of promotion may be compositional effects (age, gender, experience) or spatial sorting.

To verify that density is the determining factor, we re-estimate specification (2.3), adding the experience (in years) accumulated in "top areas" over the period 2002-2008:

$$Proba(prom_{i,z}) = \alpha + \sigma_z + \phi \hat{a}_i + \sum_{j=1}^2 (\zeta_j Exp_{i,j} + \rho_j Exp_{i,j}^2) + \zeta_0 Exp_i + \rho_0 Exp_i^2 + x_i \beta + \varepsilon_{i,z} \quad (2.8)$$

where  $Exp_{i,j}$  is the experience accumulated since 2002 by worker  $i$  until 2009 in a zone belonging to the group  $j$  in the distribution of zones regarding local density in 2009. We charac-

Table 2.12: Unconditional probability of promotion depending on previous experience in top 10% areas.

Top zones with respect to:	Density	Human Capital	Size
Promotions between 2009 and 2015			
All workers	17%	16%	16%
if exp. Top 10% 2002-2008 = 0	17%	16%	15%
if exp. Top 10% 2002-2008 > 0	20%	20%	19%

Notes. This table presents the probability of upward mobility between 2009 and 2015 if workers already had experience in top areas or not. Probabilities are computed on a sub-sample of workers who are neither in a top 10% area in 2009 nor in 2015. Variables of experience are computed during the period 2002-2008. It takes the value one (zero otherwise) if individual has worked at least one year in a top area before 2009. "Top areas" are the top 10% first areas in 2009 (30 commuting zones) in terms of density (first column), human capital (second column) or size (third column).

Reading. On average, individuals who are neither in a top 10% area in 2009 nor in 2015 have 17% chances to be promoted between 2009 and 2015. This probability is 20% if they have worked in a top area between 2002 and 2008.

terize two groups  $j$ : the top 5% areas (15 employment zones) and the next 5% (15 employment zones between 90% and 95%). We keep the overall experience in the specification (with  $\zeta_0$  the associated coefficient).

Table 2.13 reports estimated results with standard errors clustered at the commuting-zone level. According to these results, experience accumulated in the densest areas remains attached to the worker, and gives better chances of being promoted than experience accumulated elsewhere. One year of experience in the top 5% of areas increases the odds of promotion by 1.5 percentage points.<sup>14</sup> The effect of experience in a dense area on the odds of promotion is at its highest for experience of four and five years (with an additional chance of promotion of 3.5%). This effect of experience in dense areas adds to the effect of overall experience. This means that one year of experience in a dense area almost doubles the effect of overall experience on the chance of being promoted over the following five years. We also included experience in the next densest 5% areas to measure whether the effect is specific to very dense areas, or also exists in slightly less dense areas. One year of experience in a zone belonging to the next densest 5% does not increase the chances of promotion.

From these results we can conclude that the experience accumulated in dense areas is portable, and increases the chances of promotion even in less dense areas. The effect is statisti-

<sup>14</sup> $0.017 - 0.002 = 0.015$ .

Table 2.13: The benefits of previous work experience in top zones.

Dependent variable : Probability of upward mobility between 2009 and 2015			
Top zones with respect to:	Density	Human Capital	Size
	(1)	(2)	(3)
<b>Exp. in top 5%</b>	0.017*** (0.0029)	0.014*** (0.0031)	0.012*** (0.0027)
<b>(Exp. in top 5% )<sup>2</sup></b>	-0.002*** (0.0006)	-0.002*** (0.0007)	-0.002*** (0.0005)
<b>Exp. in next top 5%</b>	0.002 n.s. (0.0051)	0.011** (0.0047)	0.012** (0.0051)
<b>(Exp. In next top 5% )<sup>2</sup></b>	0.000 n.s. (0.0010)	-0.001 n.s. (0.0010)	-0.001 n.s. (0.0011)
Experience (overall)	0.021*** (0.0021)	0.019*** (0.0023)	0.021*** (0.0023)
Experience <sup>2</sup> (overall)	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)
Observations	213,385	196,549	167,979
R <sup>2</sup>	0.17	0.17	0.16
<b>Prob. of promotion 2009-2015</b>	<b>17%</b>	<b>16%</b>	<b>16%</b>

Notes. This table presents the effect of accumulated work experience in top areas on the probability of upward mobility between 2009 and 2015. Variables of experience are computed during the period 2002-2009. Each column is a separate regression. "Top areas" are the 5% first areas in 2009 (15 commuting zones) or the next top 5% areas in 2009 (15 commuting zones) in terms of density (column (1)), human capital (column (2)) or size (column (3)). All specifications include controls for age, age-squared, gender, level of education, work experience since 2002 (and squared), tenure at firm, socio-professional category, two-digit industry, zone fixed effects. The estimation sample excludes workers who are in the top 10% areas in 2009.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

cally significant. Note that these results were obtained even though we controlled for unobserved characteristics of the individuals (proxy), as well as for the industry in which the individual works. The impact we measured is not the result of spatial sorting of workers with experience in the densest areas to low density areas, where the likelihood of promotion is nevertheless high.

We repeat the same estimations, but while categorizing top areas according to either local human capital concentration or size. The results shown in columns (2) and (3) of Table 2.13 lead to similar conclusions.

## 2.8 Conclusion

This paper focuses on social promotion in the labor market, and its geographical dimension. Considering the uneven geography of individuals' social mobility over their working life in France, this paper examines the role of local environment in the chances of social promotion or demotion. Using rich administrative panel data, we estimate the determinants of the chances for an individual to be promoted over a six-year period (2009-2015). We design a methodology that allows us to remove the spatial sorting bias, as well as the potential reverse causality.

We show a strong effect of work location on the chances of promotion, while social downgrading appears to be independent of geography. Quantitatively, the effects are not negligible. Among the local factors that are likely to favor social promotion (the density of the employment zone, its size, or the concentration of graduates), it is the concentration of graduates that appears predominant. While density and size explain part of the promotion premium in our estimates, it seems that the effect is channeled by the concentration of graduates in these areas. Our estimated effects are robust to changes in the identification method, the definition of promotion, and the data samples.

An increase of 1 point in the ratio of higher-educated workers increases the chance of being promoted over a six-year period by about 0.23 percentage point. However, we also show that the influence of local characteristics on the chances of promotion is likely to be non-linear: agglomeration effects only appear at relatively high levels of local human capital, density, or size. Thus, an individual who moves from a low-density or low-educated area to a medium-density or highly educated area does not see his or her chances of promotion increase significantly. In order to see his or her chances of promotion increase, this individual would have to move to an area that is among the most dense, or among the richest in higher-educated people. Working in a zone of the 10th decile with respect to local human capital increases the chance of promotion by four percentage points, compared to working in a zone in the five first deciles.

We also find that the size of the firms, on average larger in the dense or highly-educated areas, is not sufficient to explain the promotion premium in these areas. Moreover, it seems that the promotion premium in dense areas is even greater for external promotions than for internal promotions. Finally, we show that the experience acquired in dense areas increases the chances

of promotion even after leaving them.

All these results seem to us to shed light on the feeling of injustice linked to a "two-tier" society, in which success depends on where one lives and works. This feeling of injustice in France has fueled, among other things, the "yellow vests" movement. Local development policies are often focused on maintaining or growing employment, but without always considering the range of qualifications required for these jobs. [Charnoz et al. \(2018\)](#) show that the arrival of high-speed rail in medium-sized cities can result in a transfer of managerial jobs to headquarters located in big cities. In the end, it is possible that the arrival of high-speed rail, presented as a tool for local development, may in fact harm the chances of promotion for the area's population. Other local development policies can increase both employment and the range of qualifications, including the highest qualifications. As is the case, for example, with the creation of universities, as we will see in the next chapter of this thesis.



## .1 Additional Tables

Table 14: Hierarchy of SPC Mobility

Upward Social Mobility ↑	<b>1. White-collar workers (senior staff and top management positions)</b>	<ul style="list-style-type: none"> <li>Liberal medical or lawyer professions (salaried status)</li> <li>Public service executives</li> <li>Professors and scientific professions</li> <li>Information, arts and entertainment managers or artists</li> <li>Administrative and Commercial Managers</li> <li>Engineers and technical managers</li> </ul>
	<b>2. Intermediate Professions</b>	<ul style="list-style-type: none"> <li>Primary School and Early Childhood Teachers and related professions</li> <li>Health and social intermediate professionals</li> <li>Clerics, religious</li> <li>Public services and administration intermediate professionals</li> <li>Business administrative and commercial intermediate professionals</li> <li>Technicians (except services sector)</li> <li>Foremen, supervisors, with hierarchical authority (excluding administrative supervisors)</li> </ul>
	<b>3. Skilled Employees and Skilled Blue-collar workers</b>	<ul style="list-style-type: none"> <li>Civilian employees and support workers of the public service</li> <li>Police, military and security workers</li> <li>Administrative support employees</li> <li>Skilled craft and related workers</li> <li>Drivers</li> <li>Industrial skilled workers</li> <li>Transport and storage skilled workers</li> </ul>
	<b>4. Unskilled Employees and Unskilled Blue-collar workers</b>	<ul style="list-style-type: none"> <li>Sales workers</li> <li>Direct personal services workers</li> <li>Industrial unskilled workers</li> <li>Craft and related unskilled workers</li> <li>Agricultural, forestry and fishery labourers</li> </ul>

Table 15: Wage regression based on equation (2.2) over the periods 2009-2015 and 2002-2008.

Dependant variable: log net hourly wage	2009-2015	2002-2008
	(1)	(2)
Tenure at firm	0.003*** (0.00014)	-
Experience	0.035*** (0.0023)	-
Experience <sup>2</sup>	-0.0003*** (0.0000)	-
SPC indicators:		
Intermediate Professions	-0.138*** (0.0022)	-0.191*** (0.0021)
Skilled-Employees	-0.192*** (0.0035)	-0.276*** (0.0035)
Skilled-Blue-collar-Workers	-0.211*** (0.0042)	-0.290*** (0.0038)
Unskilled-Employees	-0.242*** (0.0047)	-0.326*** (0.0046)
Unskilled Blue-collar-Workers	-0.237*** (0.0045)	-0.316*** (0.0051)
Industry indicators:		
Manufacturing, Mining and other industries	0.063*** (0.0041)	0.054*** (0.0031)
Construction	0.069*** (0.0038)	0.051*** (0.0034)
Wholesale/Retail Trade, Transport, Accommodation and Food	0.035*** (0.0035)	0.019*** (0.0033)
Information and Communication	0.043*** (0.0043)	0.052*** (0.0039)
Financial and Insurance	0.086*** (0.0050)	0.072*** (0.0041)
Real Estate	0.057*** (0.0050)	0.032*** (0.0039)
Professional, Scientific, and Technical Services, Administrative and Support Services	0.062*** (0.0043)	0.054*** (0.0044)
Public administration, education, human health and social work	0.012*** (0.0045)	0.006*** (0.0035)
Other service activities	0.005** (0.0038)	-0.005 n.s. (0.0036)
Year indicators:		
2010 (1) / 2003 (2)	-0.006*** (0.0010)	0.033*** (0.0007)
2011 (1) / 2004 (2)	0.002 n.s. (0.0023)	0.068*** (0.0005)
2012 (1) / 2005 (2)	-0.001 n.s. (0.0034)	0.010*** (0.0007)
2013 (1) / 2006 (2)	-0.009* (0.0049)	0.127*** (0.0008)
2014 (1) / 2007 (2)	-0.002 n.s. (0.0055)	0.170*** (0.0009)
2015 (1) / 2008 (2)	-0.005 n.s. (0.0067)	0.211*** (0.0014)
Constant	7.594 (0.2783)	2.481 (0.0044)
Individual fixed effects	Yes	Yes
Zone fixed effects	Yes	Yes
Observations	3,951,560	4,476,463
R <sup>2</sup>	0.92	0.90

Note. This table presents the estimation results of wage equation (2.2) for the periods 2009-2015 and 2002-2008. Due to a lack of data, regressions for the 2002-2008 period do not include controls for previous experience. We recover the coefficients of individual fixed effects to build our proxy for workers' unobserved ability included in the equation (2.3). Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Table 16: First stages estimation results of 2SLS IV models.

	Log Density		Local Human Capital		Log Size	
	IV1	IV2	IV1	IV2	IV1	IV2
	(1)	(2)	(3)	(4)	(5)	(6)
Dens. Religious buildings	8.018** (3.7539)	0.051 n.s. (5.9984)	-	-	-	-
log Density 1806	1.178*** (0.1684)	1.139*** (0.1911)	-	-	-	-
Km royal roads 1837 (department)	-	-0.079 n.s. (0.0770)	-	-	-	-0.043 n.s. (0.0517)
Dens. Royal roads 1837 (department)	-	5.711*** (1.5820)	-	-	-	3.088** (1.2625)
Var. HC 68-82	-	-	3.169*** (0.3352)	3.375*** (0.3826)	-	-
Shift-share HC	-	-	0.163*** (0.0402)	-	-	-
Sec. Students / Pop. 1837 (department)	-	-	-	4.153** (1.7378)	-	-
Cons. Bac 1896	-	-	-	1.075*** (0.2436)	-	-
Religious historical buildings	-	-	-	-	0.001 n.s. (0.0022)	0.001 n.s. (0.0017)
Size 1806	-	-	-	-	0.005*** (0.0004)	0.005*** (0.0008)
Weak id. Cragg-Donald F-Stat.	571.892	589.42	46.266	182.45	392.31	330.74
Cragg-Donald F-Stat. – p-value	0.000	0.000	0.000	0.000	0.000	0.000
Under id. LM stat – p-value	0.008	0.019	0.066	0.004	0.006	0.001
Sargan/Hansen J-Stat – p-value	0.002	0.134	0.428	0.897	0.256	0.087
Endogeneity test – p-value	0.708	0.343	0.072	0.268	0.752	0.357
Observations	346,690	330,735	346,690	330,735	346,690	330,735

Notes. This table presents first stage estimation results of the IV models. Second stage results are presented in Table 2.6. Each column is a separate regression. All specifications contain controls (included instruments) for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socio-professional category, two-digit industry, our proxy for workers' unobserved ability. Depending on the model and the specification, excluded instrument are: log density in 1806; density of historical religious buildings; the change in local human capital 1968-1982; shift-share local human capital; the number of secondary students in 1887 compared to population; the share of conscripts with *baccalauréat* diploma in 1896; the number of royal road kilometers in 1837; and the ratio of the number of royal road kilometres in 1837 to the total surface of the department. Due to lack of information in the historical department-level database, IV2 regression is based on 250 zones and 330,735 workers. Reported Wald F-statistics, for weak instruments identification test, exceed the [Stock and Yogo \(2005\)](#) critical values for a 5% maximal IV relative bias (19.93 for all IV1; 16.85 for the specifications on density IV2 and size IV2; and 13.91 for the specification on human capital IV2).

Coefficients are reported with clustered standard errors at the zone level in parenthesis for IV1 and at the zone-department level for IV2. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Table 17: Regression results – Influence of local characteristics on workers' promotion between 2002 and 2008 (density, human capital and size).

Dependent variable : Probability of upward mobility 2002-2008			
	(1)	(2)	(3)
	OLS	OLS	IV1
Log Density	0.011*** (0.0006)	0.009*** (0.0005)	0.008*** (0.0008)
Proxy for workers' ability	-	0.313*** (0.0100)	0.314*** (0.0098)
Observations	430,119	430,119	430,119
R <sup>2</sup>	0.18	0.20	0.20
Local Human Capital	0.289*** (0.0149)	0.232*** (0.0115)	0.232*** (0.0154)
Proxy for workers' ability	-	0.313*** (0.0100)	0.313*** (0.0100)
Observations	430,119	430,119	430,119
R <sup>2</sup>	0.18	0.20	0.20
Log Size	0.012*** (0.0013)	0.009*** (0.0014)	0.009*** (0.020)
Proxy for workers' ability	-	0.314*** (0.0100)	0.314*** (0.0098)
Observations	430,119	430,119	430,119
R <sup>2</sup>	0.18	0.20	0.20

Notes. This table presents estimation results based on equation (2.3) for the period 2002-2008. Each column is a separate regression. All specifications include controls for age, age-squared, gender, level of education, socio-professional category, two-digit industry. Due to a lack of data, regressions for the 2002-2008 period do not include controls for previous work experience.<sup>a</sup> Each part of the table corresponds to a specific local characteristic of interest (introduced separately): log density, local human capital or log size. The proxy for workers' unobserved ability corresponds to the coefficients of individual fixed effects estimated in a first equation (2.2) of wages (see Table 15).

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

<sup>a</sup>When including the controls for experience since 1994 and tenure at firm, the sample is reduced to 59,943 workers with coefficients for column (2) of 0.008\*\*\*, 0.226\*\*\* and 0.010\*\*\* associated respectively to log density, local human capital and log size.

Table 18: Regression results – Influence of local characteristics on workers’ demotion between 2009 and 2015 (density, human capital and size).

Dependent variable : Probability of downward mobility between 2009 and 2015			
	(1)	(2)	(3)
	OLS	OLS	IV
Log Density	-0.003*** (0.0010)	-0.000 n.s. (0.0007)	0.000 n.s. (0.0006)
Proxy for workers’ ability	-	-0.197*** (0.0032)	-0.198*** (0.0032)
Observations	325,872	325,872	325,872
R <sup>2</sup>	0.06	0.08	0.08
Local Human Capital	0.073*** (0.0202)	0.002 n.s. (0.0166)	-0.013 n.s. (0.0226)
Proxy for workers’ ability	-	-0.198*** (0.0032)	-0.197*** (0.0032)
Observations	325,872	325,872	325,872
R <sup>2</sup>	0.06	0.08	0.08
Log Size	-0.004*** (0.0009)	-0.000 n.s. (0.0010)	0.000 n.s. (0.0010)
Proxy for workers’ ability	-	-0.197*** (0.0032)	-0.198*** (0.0032)
Observations	325,872	325,872	325,872
R <sup>2</sup>	0.06	0.08	0.08

Notes. This table presents estimation results based on equation (2.3) in the case of downward socio-professional mobility for the period 2009-2015. Each column is a separate regression. All specifications include controls for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socioprofessional category, two-digit industry, our proxy for worker’s unobserved characteristics. Each part of the Table corresponds to a specific local characteristic of interest (introduced separately): log density, local human capital or log size.

Coefficients are reported with clustered standard errors at the zone level in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Table 19: First stages estimation results of 2SLS IV models: encompassing model.

	IV1			IV2		
	Density	Human Capital	Size	Density	Human Capital	Size
	(1)	(2)	(3)	(4)	(5)	(6)
Shift-share HC	2.745*** (0.606)	0.050** (0.024)	1.646*** (0.544)	-	-	-
Sec. Students / Pop. 1837 (department)	-	-	-	162.191*** (43.633)	4.334*** (1.453)	122.207*** (39.734)
Cons. Bac 1896	-	-	-	3.604 n.s. (8.283)	0.358 n.s. (0.243)	-1.816 n.s. (7.534)
Var. HC 68-82	6.052 n.s. (5.263)	2.203*** (0.267)	11.359** (5.305)	12.985*** (4.684)	2.207*** (0.221)	17.640*** (3.969)
Km royal roads 1837 (department)	-	-	-	-0.543** (0.239)	-0.011 n.s. (0.009)	-0.734*** (0.214)
Dens. Royal roads 1837 (department)	-	-	-	1.930 n.s. (1.514)	-0.142** (0.058)	3.992** (1.568)
Dens. Religious buildings	7.441*** (2.878)	0.494*** (0.108)	-4.217 n.s. (2.717)	0.964 n.s. (2.219)	0.372*** (0.104)	-9.882*** (2.328)
log Density 1806	0.995*** (0.146)	-0.002 n.s. (0.004)	0.315*** (0.121)	1.059*** (0.083)	0.001 n.s. (0.004)	0.350*** (0.097)
Size 1806	-0.152*** (0.049)	0.008*** (0.002)	0.226*** (0.049)	-0.152*** (0.041)	0.010*** (0.002)	0.180*** (0.052)
Religious historical buildings	-2.080 n.s. (1.522)	-0.156*** (0.041)	2.948 n.s. (1.914)	-2.315** (1.022)	-0.182*** (0.036)	3.104*** (1.011)
Observations	346690	346690	346690	330735	330735	330735
Cragg Donald Fstat	149.51	347.84	71.97	251.14	575.08	85.19
Sanderson-Windmeijer Fstat	88.65	26.38	36.15	161.87	38.49	30.26

Notes. This table presents first stage estimation results of the IV models. Second stage results are presented in Table 2.9. Each column is a separate regression. All specifications contain controls (included instruments) for age, age-squared, gender, level of education, work experience between 2002 and 2009 (and squared), tenure at firm, socio-professional category, two-digit industry, our proxy for workers' unobserved ability. Depending on the model and the specification, excluded instrument are: log density in 1806; log density of historical religious buildings; the change in local human capital 1968-1982; shift-share local human capital; market potential in 1806; the number of secondary students in 1887 compared to population; the share of conscripts with *baccalauréat* diploma in 1896; the number of royal road kilometers in 1837; and the ratio of the number of royal road kilometers in 1837 to the total surface of the département. Due to lack of information in the historical department-level database, IV2 regression is based on 250 zones and 330,735 workers. Reported Wald F-statistics, for weak instruments identification test, exceed the Stock and Yogo (2005) critical values for a 5% maximal IV relative bias (19.93 for IV1 and 20.53 for IV2).

Coefficients are reported with clustered standard errors at the zone level in parenthesis for IV1 and at the zone-department level for IV2. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.10 levels.

Table 20: Top areas depending on density, human capital or size in 2009.

	Density	Humanc Capital	Size
Lille	Top 5%	Top 5%	Top 5%
Lyon	Top 5%	Top 5%	Top 5%
Paris	Top 5%	Top 5%	Top 5%
Saclay	Top 5%	Top 5%	Top 5%
Strasbourg	Top 5%	Top 5%	Top 5%
Marne-la-Vallée	Top 5%	Top 5%	-
Poissy	Top 5%	Top 5%	-
Marseille - Aubagne	Top 5%	Top 10%	Top 5%
Créteil	Top 5%	Top 10%	Top 10%
Orly	Top 5%	Top 10%	Top 10%
Cergy	Top 5%	Top 10%	-
Évry	Top 5%	-	-
Lens - Hénin	Top 5%	-	-
Roubaix - Tourcoing	Top 5%	-	-
Valenciennes	Top 5%	-	-
Montpellier	Top 10%	Top 5%	Top 5%
Plaisir	Top 10%	Top 5%	-
Nantes	Top 10%	Top 10%	Top 5%
Cannes - Antibes	Top 10%	Top 10%	Top 10%
Roissy - Sud Picardie	Top 10%	-	Top 5%
Toulon	Top 10%	-	Top 10%
Béthune - Bruay	Top 10%	-	-
Boulogne-sur-mer	Top 10%	-	-
Douai	Top 10%	-	-
Dunkerque	Top 10%	-	-
Istres - Martigues	Top 10%	-	-
Le Havre	Top 10%	-	-
Mulhouse	Top 10%	-	-
Perpignan	Top 10%	-	-
Sète	Top 10%	-	-
Grenoble	-	Top 5%	Top 5%
Toulouse	-	Top 5%	Top 5%
Nice	-	Top 5%	Top 10%
Aix-en-Provence	-	Top 5%	-
Houdan	-	Top 5%	-
Rambouillet	-	Top 5%	-
Bordeaux	-	Top 10%	Top 5%
Rennes	-	Top 10%	Top 5%
Dijon	-	Top 10%	Top 10%
Nancy	-	Top 10%	Top 10%
Annecy	-	Top 10%	-
Besançon	-	Top 10%	-
Corte	-	Top 10%	-
Genevois Français	-	Top 10%	-
Melun	-	Top 10%	-
Rouen	-	-	Top 5%
Saint-Étienne	-	-	Top 5%
Angers	-	-	Top 10%
Avignon	-	-	Top 10%
Brest	-	-	Top 10%
Caen	-	-	Top 10%
Clermont-Ferrand	-	-	Top 10%
Metz	-	-	Top 10%
Orléans	-	-	Top 10%
Tours	-	-	Top 10%

Notes. This table presents the commuting zones ranking among the 'top areas' in terms of density, human capital or size distributions in 2009. For example, the zone of Lille is both ranked in the top 5% of zones in terms of density, as well as the top 5% in terms of human capital or size, whereas the zone of Evry is ranked in the top 5% in term of density, but not in terms of human capital or size. The indication 'Top 10%' means that the zone is not in the top 5%, but still in the top 10% (what we called the 'next top 5%' in Section 2.7.2.)



## Chapter 3

# The Local Impact of University Decentralization in France

### Abstract

We measure the effects on human-capital accumulation at the local level of the "new universities" created as part of the U2000 Plan implemented in France in the early 1990s. Established in 1990, this national program resulted in the creation of eight universities (spread over 15 sites), including four outside the Paris region (over 10 sites). Using the synthetic control method, we show that the opening of "new universities" has led to a significant increase in the local share of higher-educated people (not including those in school). On average, the creation of "new universities" increased this share by 4 p.p. within 27 years, though the effect differs across cases. Our estimates provide reliable results in six of the treated-zones, with positive effects between 2 and 7 p.p. within 27 years (which represents around 8,000 more higher-educated individuals on average, in comparison with the counterfactual). Moreover, exploring the employment implications of "new universities" creation, we found credible evidence that, on average, human-capital gains co-occurred with gains in skilled employment.

### 3.1 Introduction

Since [Marshall \(2013\)](#), the question of the influence of human-capital accumulation on regional economic development has generated intensive research. Concentration of human capital within a region is widely acknowledged to increase local growth, wages and productivity, innovation, or population and employment agglomeration ([Marshall \(2013\)](#); [?](#); [Barro \(1991\)](#); [Glaeser et al. \(1995, 2004\)](#); [Florida et al. \(2008\)](#) - see also chapters 1 and 2 of this thesis). The agglomeration of population with some higher education is a source of local knowledge spillovers ([Rauch, 1993](#); [Acemoglu and Angrist, 2001](#); [Moretti, 2004a,b](#); [Rosenthal and Strange, 2004](#)). In light of this common understanding, the question of how to create, attract and retain skilled population has become a central issue for local policymakers. In this paper, we focus on the most natural political response: the creation of local research universities.

Medieval cities played a crucial role in the creation and development of the first universities in France, as local social elites expected a great deal from these institutions ([Verger, 1986](#)). However, it was not until the second half of the 20th century that governmental -education policies took into account the role of universities in regional development. Most modern universities were reformed or created in the late 1960s. Afterwards, in the late 1980s, French higher education received a new impulse with the opening of smaller local universities, as a result of local authorities' initiatives. As these efforts were somewhat scattered and disorganized, the central government announced a national investment and modernization program in 1990, known as "University 2000 Plan" or "U2000," to support these local initiatives. Notably, the program led to the creation of eight "new universities" (spread over fifteen sites), including four outside the Paris region. We use eight campus of these "new universities" to study the role of such creations on local economies. U2000 program constitutes an ideal natural experiment for two main reasons. First, its top-down, centralized designation process based on the need for re-balancing the map of higher-education, and its composite financing system avoids potential endogeneity arising from the fact that "better" or "richer" localities may have greater chance to be selected for the program.<sup>1</sup> Second, because new research universities were opened during the same pe-

---

<sup>1</sup>In fact, the treated areas are located in regions with a GDP per capita below the national average in 1990 - around 80% of the national average. Moreover, their average budget per capita at the departmental-level does not exceed the national average - on average 2,905 francs per capita compared to 2,920 francs in 1990.

riod in different sites and different local contexts, our estimated average effects suffer less from external validity issues, while the specific effects for each case give the opportunity to identify and analyze potential heterogeneity across location.

The enthusiasm and heavy involvement of local authorities at the time of the program reveals their high expectations regarding the impact of these universities on local development. Central and local government archives from that time reflect these expectations, notably in regard to job creation and the attraction and retention of a skilled population – and the spillovers that go with them.<sup>2</sup> While such effects are not surprising, it is not clear that they actually exist. There may also be conceivable scenarios in which a local university is only a stepping stone towards major cities for graduates or pre-graduates, making them more prone to move and gradually depopulate small university town. The creation of a local university may therefore impacts exclusively individual paths,<sup>3</sup> but with no substantial effect in terms of local agglomeration and human capital gains, failing to fit into its local socio-economic environment. Very recently, Lee (2021) show that the opening of a second research university in Ulsan (Korea) did not significantly change the local share of college graduates.<sup>4</sup>

Thirty years later, we use the natural experience of U2000 Program to precisely measure the local effect local effects of the creation of "new universities", taking the advantage of several openings in different areas which can be considered as "empty" in terms of higher-education before the program. In particular, did the university's creation lead to an increase in the concentration of human capital in the employment-zone of the site, compared to the counterfactual situation?

Using the synthetic control method, this paper estimates a significant increase in the local share of higher-educated people (not including those in school) caused by the "new universities" creation in the 1990s as part of the French U2000 program. Furthermore, exploring the employ-

---

<sup>2</sup>See in particular Datar (1998), Duport (2008), Poulain (1997), and the report of the *Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel* (CNE, 1997, 1996a, 1999, 1996b).

<sup>3</sup>Litterature on the geography of educational and social inequalities show that better access to higher-education, and specifically the distance from university, plays a major role in terms of university attendance Frenette (2009); Spiess and Wrohlich (2010), and promotes equality of opportunity and social mobility for individuals (Dherbécourt, 2015).

<sup>4</sup>Although the autor recalls that analyzing only a single university in this specific location has limitations in terms of external validity, this encourages cautious about the self-evident reality that the creation of an university leads to local human capital gains.

ment implications of "new universities" creation, we found credible evidence that, on average, human-capital gains co-occurred with gains in skilled employment.

**Related literature.** A large literature has studied the potential influence of universities on local economies. [Drucker and Goldstein \(2007\)](#) provide a review of this literature, mostly based on US cases. Their survey points to some evidence of a positive correlation between universities and regional economic development, but, given the paucity of data, endogeneity issues are generally not fully addressed. More recently, drawing on the consequences of a university decentralization program in Sweden in the late 20th century, [Andersson et al. \(2004, 2009\)](#) show that education funding influenced regional development, in terms of innovation and productivity.

Regarding the impact of universities on local human-capital levels, [Blackwell et al. \(2002\)](#); [Huffman and Quigley \(2002\)](#); [Winters \(2011\)](#); [Abel and Deitz \(2012\)](#) all show a positive role for higher-education institutions in the creation, attraction, and retention of human capital. At a global level, [Valero and Van Reenen \(2019\)](#) develop a worldwide database on universities and show that increases in the number of universities have positive impacts on future regional growth, with some of the effects coming via human capital and innovation channels.

Generally, analyses of the local or regional effects of universities are based on university activities indicators (such as the number of publications, students or professors, or R&D expenditures, patent citations, etc.) rather than on the creation of new sites, and thus do not allow for the identification of a counterfactual without university establishment – comparisons between regions are made on the basis of the size of the universities they host. Moreover, the mentioned papers do not provide evaluation methods that sufficiently address the endogeneity issues, due to unobservable characteristics and the coincidence of economic development and university evolution or enlargement, to permit the identification of causal effects.

To our knowledge, while they do not focus on local human-capital accumulation, a few recent studies use causal inference methodology to estimate the influence of universities on local economic development – i.e. using estimates of a counterfactual situation. Using a Swedish university reform in 1999, which granted "research university" status to some universities, [Bonander et al. \(2016\)](#) implement the synthetic control method and find no convincing evidence of an effect on regional growth and development. Regarding the US, also using the synthetic

control method and event study analysis, Liu (2015) examines the designation of US land-grant universities in the 1860s and finds a positive effect on local population density and local manufacturing output per worker. Lastly, Lee (2019, 2021) uses synthetic control method to analyze two openings of new universities, one in the US and one in Korea. He finds a positive impact on local employment ten years after the 2005 opening of the Merced campus of the University of California, and he shows that the opening of the Ulsan National Institute of Science and Technology in 2009 increased manufacturing employment through the entry of new firms.

Research on the decentralization of universities in France is rare. Several reports and political or sociological studies have examined the implementation, evolution, and implications of the higher-education decentralization period of the 1990s, but we know of no empirical investigation of the local impact of the creation of "new universities."

**Statistical evidence.** Figure 3.1 shows the location of French universities within metropolitan France. The country is divided into 304 employment-zones (or commuting-zones).<sup>5</sup> The green dots indicate the locations of old universities (the 43 cities with a least one university that existed before the U2000 Plan). Most of these universities are located in Paris or in large cities.<sup>6</sup>

From 1990 to 1995, the U2000 plan led to the creation of eight "new universities" ("*Universités nouvelles*") located in 15 employment-zones – identified by red squares in Figure 3.1. These "new universities" are full-service, autonomous, and multidisciplinary institutions with serious ambitions regarding academic research.

In addition, the program led to the creation or enlargement of around 50 small secondary universities<sup>7</sup> (decentralized satellite universities that depends on a larger one and in which the

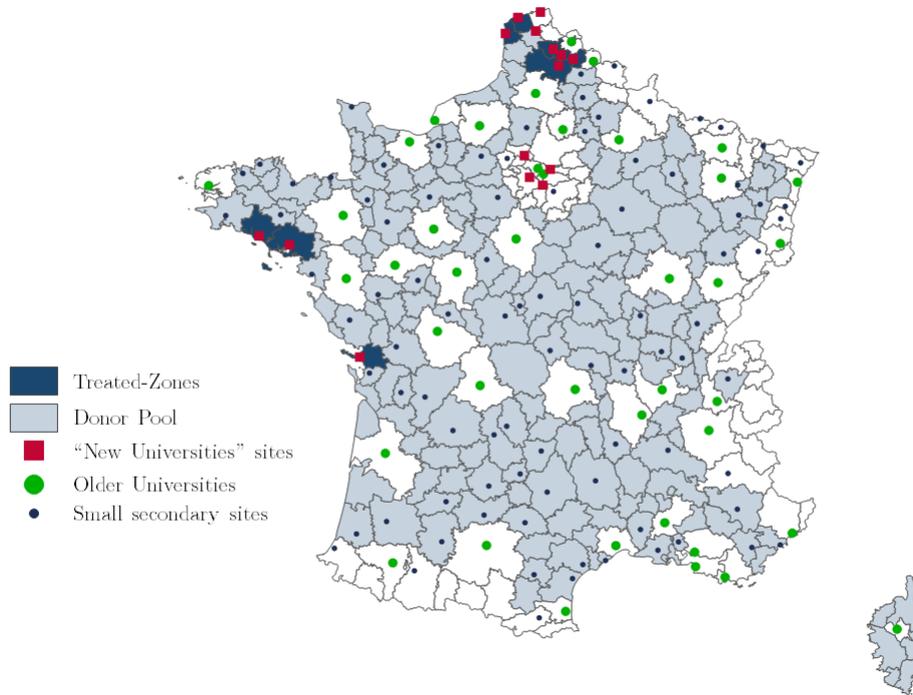
---

<sup>5</sup>According to INSEE, an employment-zone (also called a commuting-zone) is a geographical area within which most of the labor force lives and works, and in which establishments can find the majority of the labor force necessary to fill the jobs on offer. It is therefore a spatial scale adapted to local studies, especially for evaluating the economic development or population-composition of localities.

<sup>6</sup>34 are located in the 40 largest employment-zones in terms of total population in 1982. The list of universities in 1982, sorted by population size of the employment-zone (with their ranking among the 304 French employment-zones in parentheses): Paris (1); Lyon (2); Marseille (4); Orsay (5); Bordeaux (6); Toulouse (7); Nantes (8); Rouen (9); Lille (10); Grenoble (11); Saint-Étienne (12); Rennes (13); Nice (14); Nancy (15); Toulon (16); Strasbourg (17); Clermont-Ferrand (18); Metz (19); Tours (21); Créteil (22); Brest (23); Le Havre (25); Caen (26); Montpellier (27); Orléans (28); Mulhouse (29); Avignon (31); Le Mans (32); Angers (34); Dijon (35); Amiens (36); Valenciennes (37); Limoges (38); Reims (39); Pau (42); Poitiers (46); Aix (50); Besançon (57); Perpignan (60); Chambéry (70); Compiègne (91); Saint-Denis (143); Corte (303).

<sup>7</sup>Many other creations of this type emerged in the 2000s. In total, around a hundred zones had a small secondary university in 2014.

Figure 3.1: Location of Universities in France, 1968-2014



course offerings and possibilities for research are limited). They are identified by small blue dots on Figure 3.1.

Due to more accessible data and information about the decision and creation processes, we chose to focus on the impact of "new universities" only. In addition, we chose to exclude the Parisian "new universities" from our analysis because of their close proximity to the large and historical Parisian universities, situated in a region where skilled population flows are very significant, with all the potential spatial spillovers into adjacent areas that they entail. We therefore focus on: the University of Artois,<sup>8</sup> the Bretagne-Sud University,<sup>9</sup> the University of Littoral Opal Coast,<sup>10</sup> and the University of La Rochelle.<sup>11</sup> Except for the latter, the campuses of each of these universities are located in at least two cities in different employment-zones (as can be seen in Figure 3.1). We exclude two campuses of the University of Littoral Opal Coast of our main analyze: Saint-Omer, because this satellite did not receive university students before the 2000's and should rather be regarded as a decentralized secondary university satellite; and Dunkerque,

<sup>8</sup>Located in Arras, Béthune, Douai and Lens.

<sup>9</sup>Located in Lorient and Vannes.

<sup>10</sup>Located in Boulogne-sur-mer, Calais, Dunkerque and Saint-Omer.

<sup>11</sup>Located in La Rochelle.

because of its direct proximity to the Belgian border, which implies potential spillovers and migration flows that are difficult to take into account.<sup>12</sup> We end up with nine treated-zones identified by dark blue areas in Figure 3.1.

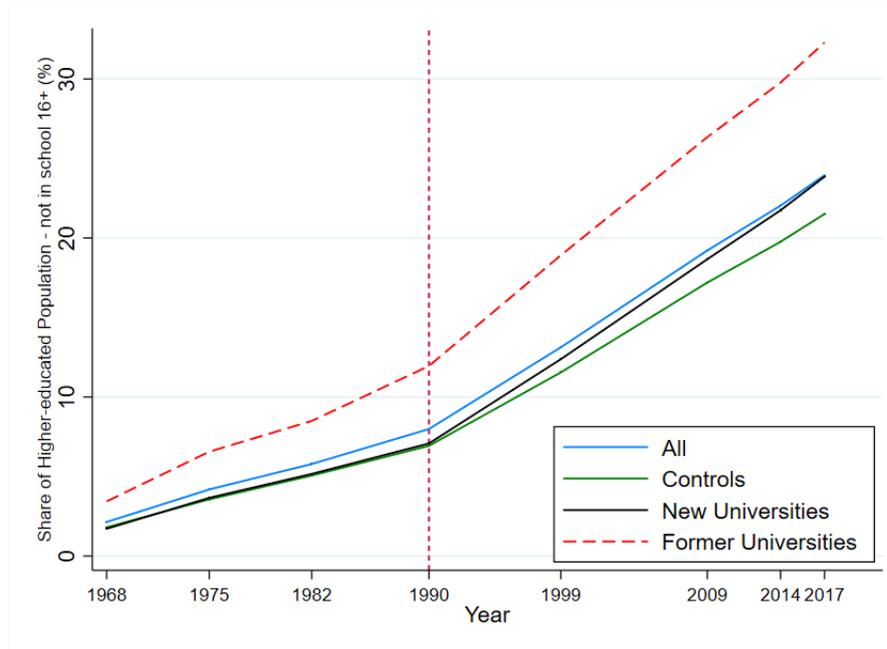
The identification strategy is to compare employment-zones where "new universities" were established (called the treated zones) with other employment-zones where no university-related effects have occurred (called the untreated-zones, which comprise a set of potential controls). Before any statistical analysis, Figure 2 shows average trends in the share of higher-educated population (not including those in school) in France for different categories of employment-zones over the period 1968-2014. If we compare the deviation of higher-educated population share in "new universities" zones (on average) from the national average (respectively the black line and the blue line in Figure 3.2), we find only a small convergence: the share increased from 7% in 1990 to 24% in 2017 where "new universities" were created, whereas the average in all employment-zones increased from 8% to 24%. However, this gives no real indication of the impact of the creation of "new universities," precisely because some other employment-zones had pre-existing universities, and experienced a large growth in their average share (dashed red line on Figure 3.2). This first comparison certainly underestimates the potential effect of the creation of "new universities." It is more appropriate to compare the outcome trajectory of "new universities" with that of a group that exclude zones with pre-existing universities and zones that experienced unusual population and human-capital flows over the period, such as the Parisian and adjacent areas (green line of Figure 3.2). This group is defined as our "donor pool": a set of potential controls used to construct the counterfactual. In this group, the average local share of higher-educated population increased from 7% to 21%. On average, this graph thus suggests a small positive impact of "new universities" on local population skills (around 3 p.p.).

**Methodology.** "New university" locations are not completely random. Moreover, the impacts of the creation of "new universities" could be heterogeneous according to the zones of their respective locations. To address these main empirical challenges, we implement the synthetic control method developed by [Abadie and Gardeazabal \(2003\)](#); [Abadie et al. \(2010, 2015\)](#) in the context of comparative case studies. This technique allows us to go further than the parallel

---

<sup>12</sup>We still discuss the effects in the Dunkerque case in Appendix.

Figure 3.2: Trends in Higher-educated Population Share in France, 1968-2017



trends assumption of the difference-in-difference method, using a matching approach to create a better counterfactual. The synthetic control method is based on the idea that using a weighted average of potential control zones is better than controlling with a specific zone alone, or an average of all the untreated zones. We construct a synthetic control zone for each treated zone by selecting potential controls in the donor pool and giving them a weight depending on a set of pre-intervention matching variables – including our outcome of interest. These matching variables are chosen based on usual predictors of local human capital and some controls for the factors that may underlie the designation of the location of "new universities." In our case, the synthetic control method is relevant for its capacity to account for time-varying unobserved effects, and to examine separately several affected areas, where the creation of "new universities" may have heterogeneous impacts.

The question of statistical inference is evaluated using placebo tests. Those tests aim to evaluate whether our estimated results are driven by chance. We re-run the synthetic method for fictive placebos to see if the magnitude of the results are indeed different from our real treated-zones results.

Finally, we run several robustness checks to test the sensitivity of our results and elimi-

nate some doubts on the SCM implementation. In particular, we run a standard Difference-in-Difference model to confirm SCM results and to take into account potential other contemporaneous shocks that may influence local human capital accumulation in the post-treatment period.<sup>13</sup>

**Main results** Using harmonized population-census data for the period 1968-2017, the synthetic control method makes it possible to conclude that the creation of "new universities" has a significant positive impact on local development, by increasing human-capital accumulation. This finding is further corroborated with the additional difference-in-differences event-study design.

On average, the creation of "new universities" increased the local higher-educated population share by 4 points within 27 years. This represents a 17% increase from the counterfactual level: the average share of higher-educated population (not including those in school) reached 24% in "new university" zones in 2017, while it would have been 20% without the U2000 Plan. This represents an increase of more than 7,300 higher-educated individuals on average.

We measure notable heterogeneity across zones. In six cases,<sup>14</sup> the local higher-educated population share is significantly larger as a result of the founding of the "new university," with an effect between 2 p.p. and 7 p.p. within 27 years (which represents between 8% and 47% of the respective counterfactuals). In the three other cases, we cannot confidently conclude that there exists a significant "new university" effect on local human-capital accumulation. Our results remain valid after additional robustness tests.

The analysis of potential implications for employment shows that human-capital gains co-occurred with positive effects on local employment density and the share of white-collar workers, on average, in the treated zones, suggesting that local human-capital accumulation actually resulted in more skilled employment at the local level. Using the same synthetic control method with the same data, we show that average employment density in the treated zones is 5%

---

<sup>13</sup>Keeping in mind that many local shocks may also come from university creation, we check for the influence of local labor demand shocks linked to urban tax-free zones creation (ZFU program), easier access to the area (through rail development), new preference for the proximity of the sea and potential spillovers coming from the proximity with other dense and dynamic local labor markets.

<sup>14</sup>The Lens, Béthune and Douai sites of the Artois University, the Calais site of the Littoral-Opal-Coast University, La Rochelle University, and the Lorient site of Bretagne-Sud University.

higher in 2017 compared to the counterfactual (although we found no convincing long-term impact on local unemployment on average, only a short-term deviation of 8% compared to the counterfactual level in 1999). In addition, "new university" creation increased the population share of white-collar workers by 1.7 points within 27 years, on average, in the treated zones. This represents a 17% increase from the counterfactual level.

The rest of the paper is organized as follows. Section 3.2 retraces the historical background of French universities and describes the U2000 Plan's implementation. Section 3.3 develops the empirical methodology and the data. Section 3.4 presents the estimation results. Section 3.5 presents robustness checks on alternative SCM and Diff-in-Diff method. Finally, we explore potential implications for local employment in Section 3.6.

## 3.2 Historical Background

### 3.2.1 The History of French Universities

The French higher-education system reflects centuries of development and reforms. After the creation of the earliest universities in the Middle Ages (first in Paris, Angers, Orléans, Montpellier, and Toulouse (Verger, 1986)), the French universities moved from ecclesiastical supervision to state supervision. Universities took the form of local guilds<sup>15</sup> attached to cities. Afterwards, the French Revolution led universities into a deep crisis, and indeed to their dissolution in 1793 – along with the entire guild system.

There followed more than 150 years without truly autonomous regional universities. Indeed, in 1806, the Napoleonic imperial university system established a highly centralized state-controlled organization. The state minister imposed vertical control on some faculties (theology, law, medicine, humanities and sciences), which were re-formed on the old medieval model, with a relatively small and unclear role (Musselin, 2012). These faculties were not really detached from secondary-education institutions (Karady, 1986). Although new higher-education centers progressively emerged in Aix, Bordeaux, Caen, Dijon, Douai, Grenoble, Lyon, Nancy, Poitiers, Rennes, and Strasbourg (Ver, 1986), the number of students remained very low during the 19th

---

<sup>15</sup> *Corporations* in French.

century, and few courses were offered (Antoine Prost 1968). Some reforms were implemented during the late 19th century, but this was negligible when compared to the massive transformation that took place in the mid-20th century.

The fundamental renewal of French universities followed in the wake of the events of May 1968, starting with the "Faure Law" of November 1968, which was rooted in three main socio-economic movements (Passeron, 1986): the first wave of the democratization of higher education; the social diversification of the demand for higher education; and the divergence between education and job opportunities. This law provided for considerable changes in the geography of French universities, with the creation of modern universities and the restoration of the autonomy of regional universities (in terms of organization, pedagogy, and finances). At the end of the 1970s, however, many people still lacked access to higher education, and regional inequalities and vertical state control remained relatively strong. The "Faure Law" was therefore reinforced with the "Savary Law" of 1984, which established new operating rules for French universities, and broadened their scope (Duport, 2008).

The map of French universities went through another major transformation in the late 20th century, with a period of higher-education decentralization in mid-sized and small cities. This decentralization movement was given impetus by a second wave of democratization and thus an increase in student numbers – partly due to pro-secondary-education national policies – and by the active involvement of local actors. Primarily on the initiative of local representatives, beginning in the 1980s, a number of small secondary universities were founded, scattered all around the country. In the context of a general movement towards decentralization, and a second phase of higher-education democratization,<sup>16</sup> these local establishments emerged most often without the consultation or participation of the central government (Filâtre, 2003; Ferréol, 2010).

In response to this surge of interest among local authorities, the national government introduced the University 2000 Plan in 1990 (also called "U2000"). This program aimed to consolidate and guide the movement towards the decentralization of universities, and achieved, inter alia,

---

<sup>16</sup>This second democratization period affected the generation born in 1960-1977, and was mostly due to the sharp increase in secondary education for these age groups. Several education reforms increased the rate of baccalaureate graduates by 30 percentage points within 15 years (Albouy and Tavan, 2007).

the creation of eight "new universities," which constitute an interesting natural experiment for investigating the impact of university decentralization on the local accumulation of human capital.

### 3.2.2 The University 2000 Plan

Announced in May 1990 by Lionel Jospin, then the Minister for Education, Youth and Sport, University 2000 was intended to address the territorial imbalance and the expected surge of new students by supporting and regulating universities' delocalization through a program of construction and renovation (Datar, 1998; Felouzis, 2001). U2000 also resolved various qualitative mismatches in terms of social inclusion and the diversification of instruction, but geographical realignment through financial assistance represented its most immediate and dominant thrust. Moreover, the key issues identified at the national conference for higher education in 1990<sup>17</sup> included a territorial focus, indicating that the university has a crucial role to play in local economic development (Poulain, 1997).

U2000's financial outlay represented more than 32 billion francs (7.2 billion euros in 2015), half of it paid by regional and local authorities, and involved 1.5 billion square meters of developed area from 1991 to 1995. (Datar, 1998). The ex-post financial results for the 23 *académies* outside *Ile-de-France* and *DOM-TOM* determined that around 21 billion francs were allocated (4.2 billion euros in 2015) – 41% of which was state-funded. Note, however, that the program did not affect only new delocalized establishments, and that a major portion (17 billion francs) of the budget was allocated for the strengthening of pre-existing universities.

University openings can be divided into two main categories: small secondary universities and "new universities." The first type is defined here based on the *"Atlas régional des effectifs d'étudiants,"* the main source of data on university establishments. "Secondary university" (or decentralized satellite university) refers to the relocation to a new site of one or more university programs leading to a national degree (excluding technical diplomas and engineering degrees). These small satellite universities are dependent on a larger one and their course-offerings and research activities are limited.<sup>18</sup> In contrast, the larger "new universities" (*"Universités nouvelles"*)

---

<sup>17</sup>"Assises nationales de l'enseignement supérieur," 26-29th of June 1990.

<sup>18</sup>As explained in the introduction, we will not analyze this type of university decentralization, but will rather

are full-service, autonomous, and multidisciplinary universities. From the opening, those institutions offer graduate degrees and show serious ambitions in terms of academic research. The creation of eight "new universities" was approved by U2000. Four of them were built in the Paris region, in order to ease the congestion of Parisian universities and to revitalize neighboring cities. Of the four others, two multipolar universities are located in the north of France, the University of Artois and the University of Littoral Opal Coast, and two in the west, the University of La Rochelle and the dual-site Bretagne-Sud University.

Regardless of the type of establishment, the archives of the central and local authorities and political reports on the U2000 Plan underline the enthusiasm and the high expectations of local actors regarding the founding of these universities, as they foresaw a significant impact on local economic development (see in particular Datar (1998), Duport (2008), Poulain (1997), and the reports of the *Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel*). Apart from the structural role of university facilities in an urban area – and the image they reflect beyond it – local actors expected a wide range of spillovers from universities, notably job-creation and the attraction and retention of a higher-educated population (Datar, 1998). Therefore, many local actors were heavily involved in the negotiations. The multiplication of stakeholders, at many levels, led to relatively complex negotiations and designation processes. Each founding was the result of specific partnership frameworks, involving multiple participants and organizations, which were either expressly created or pre-existing (local authorities, regional elected representatives, European Union actors, civic organizations, local persons of influence).<sup>19</sup> Several *ex-ante* studies were conducted in some cases, but they were mostly focused on town planning and architectural aspects, or on the living and economic conditions for future students. Prior analysis of local needs and potential economic and demographic impacts are very rare, which suggests that these considerations were not decisive in the designation process. The documentation that most often guided the selection of sites were

---

focus on the "new universities."

<sup>19</sup>The fact that many documents indicate that, in the designation process, the completion of a decentralization project often depended on few powerful local personalities, such as the city mayor or regional councilor, may pose some endogeneity questions. However, in the case of "new universities," which are relatively more documented, it appears that the political influence games were less decisive, and were subject to a more top-down, centralized designation process based on the need for re-balancing the map of higher-education in favor of locations where the demand for higher education exceeded the capacity of the existing facilities (Duport, 2008; Aust, 2007; Datar, 1998; CNE, 1996a, 1997, 1996b, 1999). See Section 3.5 for endogeneity issues discussion.

statistical projections, at the regional-level, of expected enrollments by 1995 and 2000 (Datar, 1998).

At the end, while each founding of a "new university" has involved many stakeholders, at many levels, the primary factor that affected the deliberation at the central level was based on the need for re-balancing the map of higher-education in France. This resulted in a relatively top-down, centralized designation process with a composite financing system, which from a statistical point of view, avoids potential endogeneity arising from the fact that "better" or "richer" localities may have greater chance to be selected for the program. In fact, the treated areas are located in regions with a GDP per capita below the national average in 1990 - around 80% of the national average. Moreover, their average budget per capita at the departmental-level does not exceed the national average - on average 2,905 francs per capita compared to 2,920 francs in 1990.

### 3.2.3 The "New Universities"

As presented in Table 3.1, the four "new universities" that we analyze were created in the 1990s with campuses in eleven employment-zones.

**La Rochelle University** Located in the employment-zone of the same name in western France, on the Atlantic coast, with a total population of almost 190,000 in 1990,<sup>20</sup> it was created by decree on January 20, 1993. Before its establishment, the nearest universities were Nantes University and Poitiers University, in the same "*Académie*," both about 120 km away from La Rochelle. The history of higher education in La Rochelle began with the creation of an University Institutes of Technology (IUT) in 1968, and the establishment of a local satellite of the Poitiers Faculty of Law in 1974 (comprising around 300 students) (CNE, 1997). But, it was not until the creation of the "new university" that La Rochelle became a real university town. The "new university" had around 2,400 students in 1993-94, its first academic year, and around 100 professors in 1995 (CNE, 1997). For the cohort 2011-2012, the completion rate for a *licence* (three-year undergraduate program) in three to four years were 42%.

---

<sup>20</sup>Population data comes from the INSEE census.

Table 3.1: The U2000 Plan's "New Universities"

		Nb of Students		Total Population		Higher-educated Share	
		First-year	2014	1990	2014	1990	2014
<b>Artois University</b>	1991						
Arras		2812	4156	229320	243453	8%	24%
Béthune		541	1156	290484	293015	6%	19%
Douai		345	1400	247176	246731	7%	20%
Lens		777	1797	376756	363016	5%	15%
<b>Bretagne-Sud University</b>	1995*						
Lorient		2144	3732	269360	295087	8%	24%
Vannes		1588	2987	254355	342335	9%	27%
<b>Littoral Opal Coast University</b>	1991						
Boulogne-sur-mer		1892	3122	158848	161400	7%	20%
Calais		1021	1963	153513	170055	6%	19%
Dunkerque		2392	4148	262715	253553	7%	20%
Saint-Omer**		0**	199	109510	118103	6%	19%
<b>La Rochelle University</b>	1993	2429	6340	189673	242601	9%	27%

*Note.* Numbers of university students exclude technical and engineering students. Students data come from annual Atlas regional de l'enseignement supérieur provided by the French Ministry of Higher Education, Research and Innovation. Population data are harmonized census data from INSEE.

\*Pre-independent phase during 1990-1995.

\*\*Saint-Omer did not receive university students before the 2000s (only Institute of Technology students).

**Bretagne-Sud University** This institution was created by decree on February 7, 1995 in the two neighboring employment-zones of Lorient and Vannes (containing around 260,000 inhabitants each in 1990) situated in the western region of Bretagne (Brittany). This "new university" differs from the others in that it had a pre-independence phase, beginning in 1990, during which the two locations were considered as separate decentralized satellites of nearby Brest and Rennes Universities, respectively, before joining forces to develop into a full-service dual-site university. The first secondary university satellites (for non-technical disciplines) were delocalized in Lorient in 1990 and in Vannes in 1969. Before Bretagne-Sud's creation, the nearest universities were Brest University (around 110 km and 150 km from Lorient and Vannes, respectively), Nantes University (150km and 100km) and Rennes University (130 km and 95 km). In 1996, the University of Bretagne-Sud had 2,965 students in Lorient and 2,679 in Vannes, and around 80 professors (excluding IUT)(CNE (1999)). Although the Lorient site is larger, it was decided not to designate either campus as the seat of the university; rather, the headquarters are shared between the two sites. For the cohort 2011-2012, the completion rate for a *licence* (three-year undergraduate program) in three to four years were 47%.

**University of Artois** This institution was created by decree on November 7, 1991 as a multipolar university located in the four neighboring employment-zones of Arras, Béthune, Douai, and Lens, which respectively had total populations of around 230,000, 290,000, 250,000, and 380,000 in 1990. Like its neighbor the University of Littoral, the University of Artois aimed to strengthen the higher-education supply in the dense northern region. These different sites already had some higher-education experience, with small secondary satellites having been created in Arras in 1988, and in Lens and Béthune in 1990. The campus of Artois University has the distinction of being located very near to the historical University of Lille (around 30-40 km), the more recent Valenciennes University (30-60 km), and to Picardie University in Amiens (60-80 km) – and is also not far from the sites of the "new university" of Littoral (less than 100 km). This proximity is mostly explained by the inability of the pre-existing higher-education network to handle the strong demographic growth in this particularly dense region. In 1993, in its second academic year, Artois University had around 120 professors and 4,500 students, around 2,800 in Arras, which is the seat of Artois University, 540 in Béthune, 780 in Lens, and 350 in Douai. For the cohort 2011-2012, the completion rate for a *licence* (three-year undergraduate program) in three to four years were 33%.

**University of the Littoral Opal Coast** It was created by decree on November 7, 1991. Like its Artois counterpart, the multipolar university of Littoral was constructed based on existing decentralized secondary satellites of Lille University, primarily to address congestion issues in the region. The University of Littoral is situated in the far north employment-zones of Calais (total population of around 153,000 in 1990), Dunkerque (around 263,000), and Boulogne-sur-mer (around 159,000).<sup>21</sup> The headquarters of the university are located in Dunkerque, but Calais was the first zone with a small secondary university satellite in 1976, followed by Dunkerque and Boulogne in 1985 and 1987. The nearest main university is the University of Lille, which is around 65 to 100 km from the different sites. In its second academic year, in 1993, the "new university" of Littoral had 1,892 university students in Boulogne, 1,021 in Calais, and 2,392 in Dunkerque. Recall that we will not consider this last site of Dunkerque in our analysis as it may

---

<sup>21</sup>The University of the Littoral Opal Coast includes also a site in Saint-Omer, but this satellite did not receive university students before 1999 (excluding technical students). Therefore, we do not consider the site of Saint-Omer as a real 1990's "new university," but rather as a decentralized secondary university satellite.

bias the results due to the border spillovers. There were around 140 professors in 1995 (CNE (1996b)). For the cohort 2011-2012, the completion rate for a *licence* (three-year undergraduate program) in three to four years were 41%.

### 3.3 Empirical Analysis

#### 3.3.1 The synthetic control method

To investigate the causal relationship between the creation of universities and local human-capital accumulation, we use the synthetic control method developed in Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2012) in the context of comparative case studies. This empirical approach is based on the computation of a synthetic control zone to reproduce the counterfactual, i.e. the situation of the zone where the new university is located if the creation had not occurred. The key principle of this method is that using a weighted average of potential control zones (those that did not receive the treatment) is better than controlling with a specific zone alone or an average of all the control zones. We use panel data to form a synthetic control zone by computing a specific weight (ranged between 0 to 1) for each of the zones in the set of potential controls. These weights are chosen to ensure that the new control will reflect as closely as possible the treated-zone according to several local factors and trends in the period before the university's creation (called the pre-intervention or pre-treatment period). Once the synthetic control zone is matched, the outcome variable is extrapolated to the post-intervention period with the parameter estimates and the effect of the university's creation is quantified as the difference between the treated and the synthetic zone values.

Following formal discussion in Abadie et al. (2010, 2015), suppose there is a sample of  $(J+1)$  employment-zones indexed by  $j$ , among which zone  $j = 1$  is the case of interest (i.e. is affected by the program) and the rest of zones ( $j = 2$  to  $j = J + 1$ ) constitute the potential controls<sup>22</sup> (or the donor pool, which is not subject to program shock).

Assume that the sample is a balanced panel including pre-intervention periods,  $T_0$ , and post-intervention periods,  $T_1$ , with  $T = T_0 + T_1$ , and  $T_0$  not necessarily equal to  $T_1$ . Then the zone

---

<sup>22</sup>If more than one zone is exposed to the program, the method can be applied successively to each treated zone. We describe the method only in the case of a unique treated-zone without a loss of generality.

$j = 1$  is exposed to the program (the location of a university site) during periods  $T_0 + 1, \dots, T$ , and the program has no impact during the period  $1, \dots, T_0$ . The program effect for the treated-zone at time  $t = T_0 + 1, \dots, T$  is given by

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N \quad (3.1)$$

with  $Y_{1t}^I$  and  $Y_{1t}^N$  the treated-zone's outcome with and without intervention, respectively.  $Y_{1t}^N$  is obviously not directly observable. We need an estimate of  $Y_{1t}^N$  to measure the  $\alpha_{1t}$ .

The synthetic control model assumes that the outcome  $Y_{jt}^N$  is given by :

$$Y_{jt}^N = \delta_t + \theta_t X_j + \lambda_t \mu_j + \epsilon_{jt} \quad (3.2)$$

where  $\delta_t$  is an unknown common factor with constant factor loadings across units;  $X_j$  is a vector of observed covariates (not affected by the intervention) and  $\theta_t$  the associated vector of parameters;  $\mu_j$  is a vector of unobserved factor loadings and  $\lambda_t$  the unknown common factors; and  $\epsilon_{jt}$  are unobserved transitory shocks with zero mean.

Considering now  $W = (w_2, \dots, w_{J+1})'$  a  $(J + 1)$  vector of weights, with  $0 \leq w_j \leq 1$  and  $w_2 + \dots + w_{J+1} = 1$ , each set of weights  $W$  defines a possible synthetic control. [Abadie et al. \(2010\)](#) show that, as long as there are  $(w_2^*, \dots, w_{J+1}^*)'$  such that

$$\sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0} \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* X_j = X_1 \quad (3.3)$$

the mean of the difference in outcomes between the treated and the synthetic control unit ( $Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ ) will be close to zero. Then

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \quad (3.4)$$

is an unbiased estimator of  $\alpha_{1t}$ . The synthetic control estimator of the impact of university location is given by the gap between the value of the outcome variable for the treated-zone and the value of the same outcome for the synthetic control at that period.

The conditions of equation 3.3 very rarely hold exactly in the data. In practice, the synthetic

control unit is selected so that these conditions hold approximately. Therefore, the value of  $W$  is chosen such that the characteristics of the synthetic control best reproduce the characteristics of the treated-zone over the period. With  $X_1$  a  $(k \times 1)$  vector containing the values of pre-intervention characteristics of the treated-zone – that we aim to match – and  $X_0$  a  $(k \times J)$  matrix collecting the values of these variables for the potential control zones.<sup>23</sup> The optimal weights  $W^*$  are chosen to minimize the magnitude of the difference between pre-intervention characteristics of the treated-zone and synthetic control ( $X_1 - X_0W$ ). Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2012) choose the weights that minimize

$$\sum_{m=1}^k v_m (X_{1m} - X_{0m}W)^2 \quad (3.5)$$

where  $v_m$  is a weight that reflects the relative importance of the  $m$ -th matching variable regarding their predictive power of the outcome variable.

Furthermore, rather than matching the synthetic control on pre-intervention averaged outcome (and predictors), we fit synthetic control based on trends in the outcome variable.<sup>24</sup> When forcing the construction of synthetic control to match trends in the outcome variable,<sup>25</sup> we are able to measure the growth deviation caused by the treatment – compared to the growth rate that would have occurred in the absence of the treatment. This strategy aims to take into account the trend towards mass higher-education over our analysis period and to avoid comparing zones with similar average level of human capital during the 1968-1990 period, but with very different dynamics. We will keep the level of the share of higher-educated people (not including those in school) in the last pre-treatment period (1990) in our set of matching variables. Our confidence in the validity of the counterfactual is improved by exploiting these pre-intervention trends to compute the synthetic control.

We can evaluate the "quality" of the synthetic control zone by examining its degree of "fit" and "balance". The fit corresponds to the closeness of the treated-zone and its synthetic control zone in the pre-intervention period. We can appreciate this fit on the graphical representations that plot values of the outcome variable in each census for the treated-zone and its synthetic

<sup>23</sup>The set of pre-intervention variables may include the pre-intervention values of the outcome variable.

<sup>24</sup>See Galiani and Quistorff (2017) for methodology and Stata implementation.

<sup>25</sup>Each zone's outcome is scaled so that it takes the value 1 in the last pre-treatment period (1990).

control (a superposition of the two respective lines suggest a perfect match). More precisely, the pre-intervention fit is evaluated by the root mean square prediction error (RMSPE) of the synthetic control estimator. It measures the lack of fit, i.e. how far on average the error is from 0 over the pre-treatment period (low RMSPE indicates a better control), and is defined as

$$RMSPE = \left( \frac{1}{T_0} \sum_{t=1}^{T_0} \left( Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{1/2} \quad (3.6)$$

The degree of balance on predictors is given by the comparison between values of pre-treatment characteristics for each treated-zone and the computed values of these characteristics for its synthetic control. Small gaps between the respective covariates suggest a better match.

Cavallo et al. (2013) and Galiani and Quistorff (2017) extend the synthetic control method to allow multiple events, i.e. for more than one unit to experience treatment. With their development, it is possible to run new estimates that aggregate the zone-specific effects into an average effect.

Let us now consider a sample of  $G$  treated-zones indexed by  $g$ , and  $J$  potential control zones where the treatment never occurred. As explained above we can estimate the effects  $\hat{\alpha}_{gt}$  for each of the treated-zones.<sup>26</sup> The estimated average effect over all the  $G$  treated-zones is given by

$$\bar{\alpha}_t = \frac{1}{G} \sum_{g=1}^G \hat{\alpha}_{gt} \quad (3.7)$$

### 3.3.2 Inference

**In-space Placebos** Following Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015), the statistical significance of synthetic control method results is determined by running placebo tests. That means iteratively re-estimating the model using each zone of the donor pool as a false treated-unit ("in space-placebo" implementation). This provides a distribution of placebo effects, i.e. the set of estimated gaps for all the placebo zones. Comparing this distribution with the estimated effect of the initial treated-zone allows us to appreciate whether or not the results are driven entirely by chance. In our case, that is to say to study whether the estimated effect

---

<sup>26</sup>Unlike Galiani and Quistorff (2017), we keep here the  $t$  subscript as the treatment period is the same for all the treated zones in our data-base.

of a "new university" creation is important relative to the distribution of the effects estimated for the zones not exposed to this creation. If we find many placebo effects as large as the effect of the initial treated-zone, then it is quite plausible that this initial effect was observed by chance. Therefore, estimates do not provide significant evidence of "new university" impact on the outcome variable. In space-placebo tests are graphically represented by plotting the effect gaps for the treated-zone together with the effect gaps for each of the placebo runs.

It is therefore possible to compute the corresponding *p-values* (i.e. significance levels) of these placebo tests. The *p-values* directly derive from the quantitative comparison of the distribution of placebo effects and the initial estimates. They are the proportions of placebo effects that are at least as large as the main effect for each post-treatment period. Let  $\hat{\alpha}_{j,t}^{PL}$  be the estimated effect for the post-treatment period  $t$  when control zone  $j$  is assigned to a placebo treatment at the same time as the "real" treated-zone. This effect is estimated using the exact same procedure as for  $\hat{\alpha}_{1t}$ . Then *p-values* are computed as

$$p - value_t = Pr(|\hat{\alpha}_{j,t}^{PL}| \geq |\hat{\alpha}_{1t}|) = \frac{1}{J} \sum_{j=2}^{J+1} I(|\hat{\alpha}_{j,t}^{PL}| \geq |\hat{\alpha}_{1t}|) \quad (3.8)$$

A *p-value* of 0.10 for a specific year indicates that, when the treatment is fictively reassigned at random among the donor pool, the probability of obtaining an estimate at least as large as the one obtained for the treated-zone is 10%. More briefly, there is a probability of 10% that the measured effect would happen by chance.

However, in the same way that the treated-zone estimates can be artificial if the synthetic control method does not sufficiently fit the pre-intervention characteristics, lack of fit in placebo runs could bias the inference tests. A solution is to drop the inaccurate placebos in order to better appreciate the exceptional nature of the treated-zone results. In our placebo test, we choose to exclude the placebo runs that show a pre-intervention RMSPE at least four times as large as the pre-intervention RMSPE of the real treated-unit.

In the case of multiple treated-zones, as explained in Cavallo et al. (2013), the average smooths out some noise in the estimate, and we need to account for this in the inference verification. They proposed constructing a distribution of average placebos  $\bar{\alpha}_t^{PL}$  constructed from all possible averages where a single placebo estimate is picked from each treated-zone's placebo test

$(\hat{\alpha}_{gt}^P L)$ , and then averaging over the picked placebos (ten in our case). In total, there is a set of  $N_{\overline{PL}} = \prod_{g=1}^G J_g$  possible placebo averages. We restrict again the pool to the placebos that show sufficiently good match quality (i.e. at most twice the fit of the treated-zones match). Formally we have

$$p - value_t = Pr(|\bar{\alpha}_t^{PL}| \geq |\bar{\alpha}_t|) = \frac{1}{N_{\overline{PL}}} \sum_{i=1}^{N_{\overline{PL}}} I(|\bar{\alpha}_t^{PL(i)}| \geq |\bar{\alpha}_t|) \quad (3.9)$$

where  $i$  indexes a possible placebo average.

**In-time Placebos** Abadie et al. (2015) also propose a second in-time placebo test. In the same way as for the previous in-space placebo tests, in-time placebo tests are fictive estimate runs where the treatment is assigned in a period that differs from the real one. The validity of the initial results is questioned if estimates also show significant effects for dates when the treatment did not occur.

**Robustness** Finally, our identification methodology and the placebo tests are valid under the assumption of no similar university-effects in the potential control units. These university-effects could appear in the untreated-zones for two main reasons. First, because of spillovers effects. The "new university" creation could affect the outcome not only in the treated-zone but also in the neighboring employment-zones. We therefore conduct a robustness check by running the same synthetic control method but with a new donor pool that does not include the zones neighboring the "new universities" zones. The second reason is because U2000 also led to the decentralization of small satellite universities in many commuting-zones of our donor pool. They may have similar positive effects in their areas. Therefore, the magnitude of our effects could be underestimated. Another robustness check is based on an alternative donor pool that excludes the zones with this type of decentralized secondary university.

### 3.3.3 Data

Data on universities' localization, creation dates, and student population at the university level are drawn from the annual *Regional Atlas of student population*, published by the Ministry of Higher Education and Research between 1993 and 2017, and the multiple university evaluation

reports produced by the National Evaluation Committee since 1984. Employment-zone-level data on population, skills, employment, activity, and other local attributes are drawn from harmonized population census data for the years 1968, 1975, 1982, 1990, 1999, 2009, 2014 and 2017. The employment-zones are defined by the geography applicable in 2016.<sup>27</sup> We therefore have four pre-intervention years, and four years of post-intervention data for our analysis of "new universities," all of which were established in the 1990s.

The sample is restricted to employment-zones in metropolitan France and we choose to exclude three types of potential controls from our donor pool: zones with pre-existing universities; Parisian zones; and border zones.<sup>28</sup>

Because the synthetic control aims to reproduce the situation that would have occurred without the universities' creation, we obviously discard from our sample the 39 zones with at least one "old" university (those that existed before the 1990s).

The Parisian zones (specifically, zones in the region of Ile-de-France) are relatively complex in design and structure because of the proximity to Paris, and may bias the synthetic control method. These 20 zones are indeed characterized by larger inter-zone population flows, especially higher-educated worker flows, and many important historical universities are located in the region, with potential spillover effects between areas.<sup>29</sup>

Similarly, because of international population flows (in particular higher-educated population), the comparisons with border zones are delicate (and the progressive establishment of the Schengen area during our sample period may exacerbate the issue). Moreover, these zones may benefit from spillovers from zones across the border, potentially with universities of their own. For these reasons, we also discard from our donor pool the 45 border zones.

Finally, our dataset is a balanced panel of 209 employment-zones from 1968 to 2017. It includes 9 zones with "new universities" (our treated-zones), 98 zones with at least one de-

---

<sup>27</sup>INSEE redesigned the boundaries of employment-zones in 2010. We adjusted the geography for former periods using the 2010 table of correspondence, which provides a consistent basis for comparing local outcomes over time.

<sup>28</sup>Therefore, we also exclude from our treated group the Parisian "new-universities" and the border zone of Dunkerque where University of Littoral have a campus.

<sup>29</sup>According to Aliaga (2015), the region of Ile-de-France is characterized by a very strong polarization of employment and important flows between the municipalities in the Parisian area of influence. Consequently, using the same parameters as for other regions would lead to construct a single big employment-zone (equivalent to Paris' area of influence). A specific method is instead used in the case of Ile-de-France, with further work on home-to-work flows, and a limitation of the distance effect (many long commute distances), that provides for smaller employment-zones based on local employment hubs.

centralized secondary satellite dependent on a proper university,<sup>30</sup> and 138 zones without any university establishment, as defined by the annual regional atlas of higher education, i.e. a university site that provides higher education leading to a national degree (excluding technical diplomas (DUT) and engineering degrees).<sup>31</sup> We choose to keep the employment-zones with secondary satellites in our donor pool in order to keep much more potential control in our donor pool. We therefore expect more appropriate synthetic controls. Due to their small size compared to "new universities" (in terms of students, course-offerings, grade level, and research activities), and their non-autonomous status, we expect very smaller local effects. Consequently, we assume that zones with small secondary universities are, by comparison, similar to "empty" zones.<sup>32</sup>

Our main outcome variable of interest is the share of population (not in school, and aged 16 and over) with higher education diplomas (a minimum of two years of study after the *baccalauréat*). As introduced in subsection 3.3.1, given the strong upward trend in higher education during our analysis period, we normalized our outcome variable. We force the synthetic control to match trends in the higher-educated population share by scaling each zone's outcome variable so that it takes the value 1 in the last pre-intervention period (1990).<sup>33</sup> We therefore focus on the variations in local human-capital accumulation.

For each treated zone, recall that donor pool weights are computed in order to obtain a synthetic zone that best reflects the values of a set of matching variables in the treated-zone before the creation of the "new university" (over the 1968-1990 period).<sup>34</sup> Given the available data in the INSEE population census, the set of matching variables includes usual predictors of local human capital and some controls for the local factors that may underlie the "new universities" designation: scaled share of higher-educated people - aged 16 and more and not including those in school (our outcome variable); this share of higher-educated population in 1990, the last pre-treatment period ; share of managers and professionals in the employed population (aged 25-54); employment rate (in population aged 25-54); employment and population density;

---

<sup>30</sup>Including Saint-Omer.

<sup>31</sup>A zone with only small University Institutes of Technology (IUT) is not considered as a university area.

<sup>32</sup>A robustness test is based on an alternative donor pool without these type of zones.

<sup>33</sup>We still keep the higher-educated population share level in 1990 as a matching variable.

<sup>34</sup>Because there was no significant announcement effect before 1990, we assume the absence of anticipation effects before treatment, such that the pre-treatment indicators are exogenous and not affected by the (future) university opening.

unemployment rate; participation rate (25-54); four sectoral shares of employment (Agriculture/Manufacturing/Construction/Services); shares of population by age group (0-14; 15-24; 25-39; 40-64; 65 and more); and distance from the nearest pre-existing university. These predictors are averaged over the pre-treatment period.

## 3.4 Results

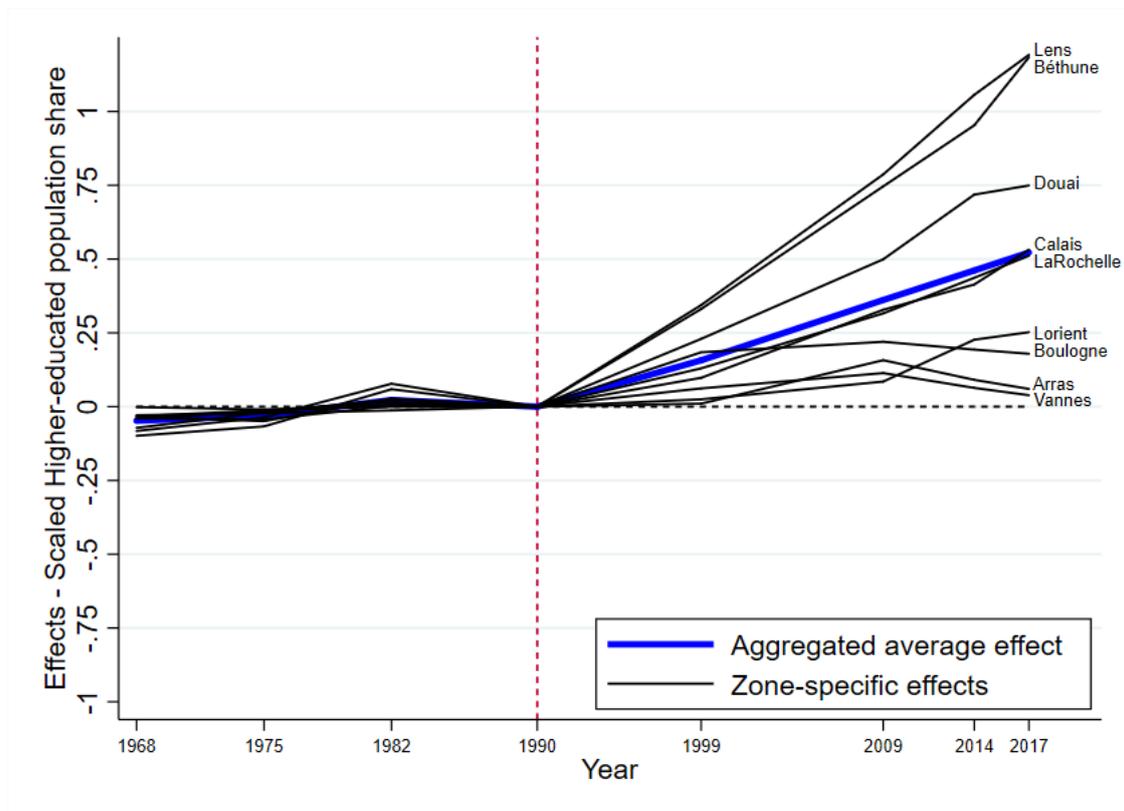
### 3.4.1 Computing Synthetic Controls

Although the general upward trend in outcome variable is observed in all cases, the pre-intervention trajectory in treated-zones (black lines on Figure 12 graphs) responds to slightly different patterns than the other zones' average (i.e. zones without "new universities," the blue line on Figure 12 graphs) or the donor pool average (the dashed line on Figure 12 graphs). As stated previously, these groups, taken as a whole, may not provide suitable comparisons for each treated-zone. The pre-treatment trajectories of the higher-educated population share (index 1 in 1990) in the synthetic control zones displayed on Figure 13 confirm that this method provides a better counterfactual. In contrast with Figure 12, lines of a treated-zone and its synthetic control are nearly always superposed over the pre-intervention period. These results are confirmed by the low values of the RMSPE and the balance of predictors presented in Table 5. Pre-treatment characteristics of each treated-zone are more closely aligned with the respective synthetic controls than the donor pool's average. For each synthetic control, Tables 6 to 15 in the appendix display the computed weights for each zone of the donor pool.

### 3.4.2 Impacts of "new university" creation

Figure 3.3 shows the effects of "new university" creations on local human-capital accumulation. The blue thick line illustrates the aggregate average effect, whereas dark lines represent each treated-zone's specific effect. The synthetic control estimator of university impact is provided by the gap between the solid lines and the horizontal zero-effect dashed line. Beginning after the 1990 census, the diverging movement of the treated-zones' average line confirms that, following the creation of "new universities," the local share of higher-educated population grew faster in

Figure 3.3: Local Human-Capital Effects in "New Universities" Zones



those areas compared to the synthetic control. On average, the creation of "new universities" increased the local human-capital accumulation by 16 p.p. within around 7 years, by 36 p.p. within 15 years, by 46 p.p. within 20 years, and by 52 p.p. within 25 years. At a mean of 7% in 1990, these estimates indicate that the share of higher-educated population is on average 1 point greater in those zones in 1999, 3 points greater in 2009 and 2014, and 4 points greater in 2017, as a result of the "new universities" creation.<sup>35</sup>

Case-by-case lines show important heterogeneity in the effects. Our estimates indicate substantial positive effects in half of the cases: Lens, Béthune Douai, La Rochelle, and Calais. While the effects remain positive (respective lines above the zero-line), the group of Lorient and Boulogne zones show weaker effects. Finally, in the case of Arras and Vannes, the estimated effects fail to deviate from the zero-line after 25 years. The zone-specific gains related to "new university" creation are displayed in Table 3.2.

<sup>35</sup>Higher-educated population share level in 1990 times the estimated growth difference between treated-zones and synthetic control.

Table 3.2: Higher-educated Population Share : 1990 vs. 2017

	1990	2017			
		Counterfactual		Estimated Gains	Observed
Average	7%	20%	+	<b>4 p.p.</b> (0,00)	= 24%
Béthune	6%	15%	+	<b>7 p.p.</b> (0,00)	= 22%
Lens	5%	12%	+	<b>5 p.p.</b> (0,00)	= 17%
Douai	7%	17%	+	<b>5 p.p.</b> (0,01)	= 22%
La Rochelle	9%	25%	+	<b>5 p.p.</b> (0,06)	= 30%
Calais	6%	18%	+	<b>3 p.p.</b> (0,06)	= 21%
Lorient	8%	24%	+	<b>2 p.p.</b> (0,18)	= 26%
Boulogne	7%	21%	+	<b>1 p.p.</b> (0,23)	= 22%
Arras	8%	25%	+	<b>1 p.p.</b> (0,37)	= 26%
Vannes	9%	29%	+	<b>0 p.p.</b> (0,41)	= 29%

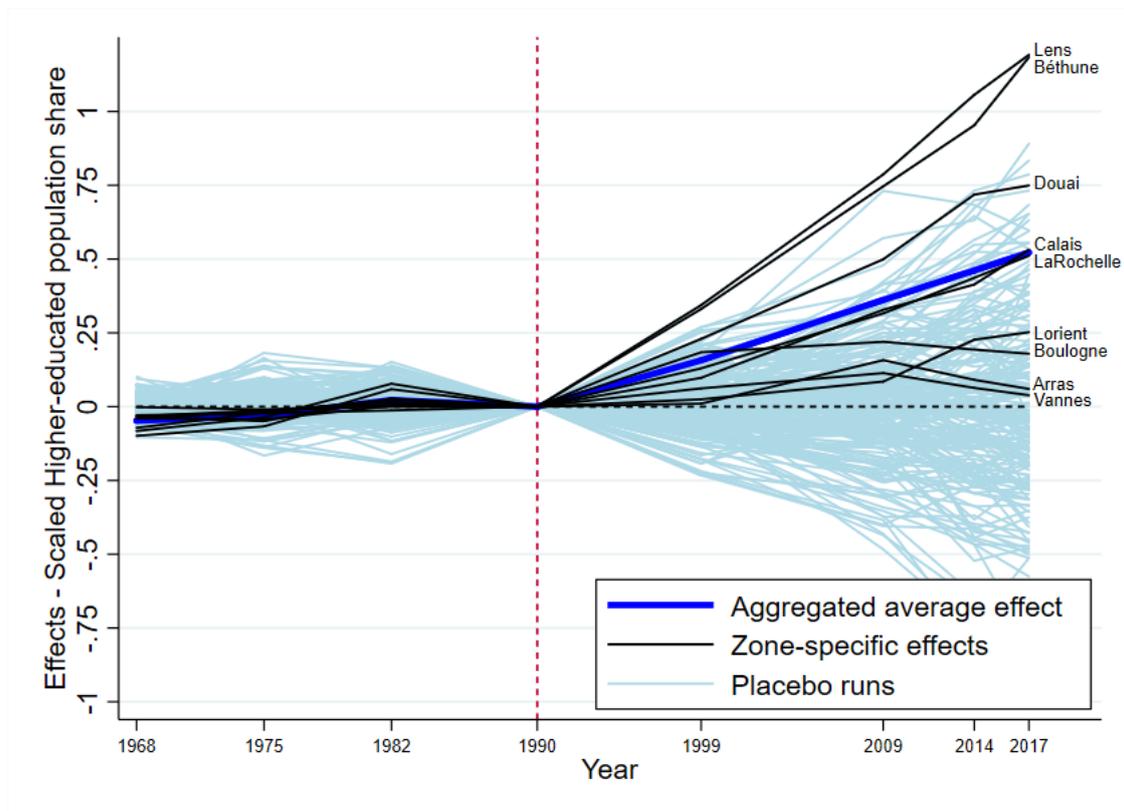
*Notes.* The magnitude of the "new university" impact is measured as the outcome deviation of the "new university" zone in comparison to the counterfactual. For example, as a result of the creation of La Rochelle University in 1993, the local share of higher-educated population (aged 16 and more, not including students) is greater by 5 points in 2017 (30% compared to 25% for the synthetic control). The p-values in brackets indicate the probability that the related result would happen purely by chance. For a distribution of 200 placebos, a 10% p-value means that only 20 placebo-zones show an estimated effect greater than the given treated-zone.

### 3.4.3 Placebo Tests

Figure 3.4 displays the synthetic control results in comparison with placebo effects distribution. Light-blue lines represent all the fictive placebo runs.<sup>36</sup> This first graphical representation confirms the credibility of previous comments on substantial positive impacts in the five cases of Lens, Béthune Douai, Calais, and La Rochelle: only a few placebo-lines (around ten at the most) show a larger effect than their respective lines. With a placebo distribution of 200 zones, the results in those cases have a very low probability of being entirely due to chance. For the other zones, we cannot confidently draw conclusions about the significance of the impacts with a simple graphical observation: the dark lines are not sufficiently unusual compared to the placebo effects distribution.

<sup>36</sup>Along with the treatment effects, placebo effects are the gaps between values of outcome variable for each placebo-zone versus its synthetic control.

Figure 3.4: Local Human Capital Effects in "New Universities" Zones vs. Placebo Tests



However, Figure 3.4 plots all possible placebo effects, for all the donor pool, regardless of the fit quality in the pre-intervention period. As explained in part 3.3.2, lack of fit in placebo runs could bias the inference tests. Figure 14 displays the zone-specific placebo test graphics for each treated-zone, excluding the placebo runs that show a pre-intervention RMSPE at least four times larger than the pre-intervention RMSPE of the corresponding treated-zone. The associated year-specific *p-values* are reported in Table 6.

The probability that the estimation of the average aggregated impact has arisen purely by chance is nil. However, for the case-by-case analysis, logically, the smaller the estimated impact, the bigger the probability that estimation has happened by chance. Overall, we can credibly conclude that there exists a positive impact of "new university" creation on local human-capital accumulation in six of the U2000 Plan candidates: the Lens, Béthune and Douai sites of the Artois University, the Calais site of the Littoral Opal Coast University, La Rochelle University,

and, to a lesser degree, the Lorient site of Bretagne-Sud University.<sup>37</sup>

As explained in subsection 3.3.2, we also run time-placebo tests. We run our synthetic control model with a fictive date of intervention. In order to have enough matching periods, we reassign the "new university" creations as if they had occurred in 1982. Figure 15 shows no significant effect from the hypothetical "new university" creations between 1982 and 1990 (despite the good pre-intervention fit). This enhances the credibility of our previous results.

### 3.5 Robustness

**Endogeneity issue** Placebo tests confirm that our results are not a mere statistical construction. However, inference analysis may not fully eliminate the uncertainty that those results do not arise from the potential endogenous decision of where to locate "new universities." In particular, a plausible alternative hypothesis is that powerful local politicians – with strong networks within the central government – were elected around the time of U2000 Plan and managed to obtain national funds to create "new universities" in their towns. If this change in local government also strongly affected the local economic dynamics afterwards, our estimated impacts could be the result of this new local leadership (which also made possible the university's creation) rather than of the "new university" creation itself.

As mentioned in Section 3.2.2, several official documents note that in the cases of "new universities," which represent relatively large public investments compared to smaller secondary university satellites, local politicians did not play a critical role in the localization decisions, which from the central government's perspective were driven by regional planning issues above all. In addition, a simple glance at the national and municipal election history in our treated zones is enlightening regarding the heterogeneity and complexity of potential political connections and affiliations between the local and national levels during this period. Over 1986-2000, around the time of the U2000 Plan, France had seven prime ministers, nine governments,

---

<sup>37</sup>In the case of Lorient, we tolerate a *p-value* of 20% because our test suffers from a low number of fake placebos, that show a pretreatment RMSPE no more than two times the RMSPE of Lorient case. In fact, only 2 placebo-zones among 10 show an estimated effect greater than the Lorient zone. If we choose to keep all the placebo tests that show a pretreatment RMSPE no more than three times the RMSPE of Lorient SCM, the *p-value* drops to 7%, with 2 placebos among 29 that show an estimated effect larger than the effect of the treated zone.

and three periods of cohabitation at the national level (Mitterrand–Chirac 1986–1988, Mitterrand–Balladur 1993–1995, and Chirac–Jospin 1997–2002).<sup>38</sup> These many political changes may have encumbered the policy influence-games between local and central authorities over a medium or long-term period.

Moreover, only one of our treated zones, Boulogne, had elected a new mayor around the period of the U2000 decisions, specifically in the municipal elections of 1989 (Jean Muselet, right wing). However, he has not remained in place throughout the post-treatment period in the zone of Boulogne-sur-mer, which does not benefit in a significant way from the university's creation, in terms of human-capital accumulation. This can therefore rule out the hypothesis that the arrival of an influential mayor at the moment of U2000 negotiations led to both a university's establishment and the accumulation of human capital afterwards.

One last concern is that the election of a new mayor in a "new university" zone between the last pre-treatment date and the first post-treatment date, though unrelated to the university's creation, could entail a radical change in political governance, and thus act as a shock affecting local human-capital accumulation. In this case, our estimated impacts could in fact be the result of this coincidental event rather than of the creation of the "new university."<sup>39</sup> This is the case in four of our treated zones: a new mayor was elected in Arras in 1995, in Lens and Lorient in 1998, and in La Rochelle in 1999. We therefore check whether those zones provide different estimated results than the other treated zones.

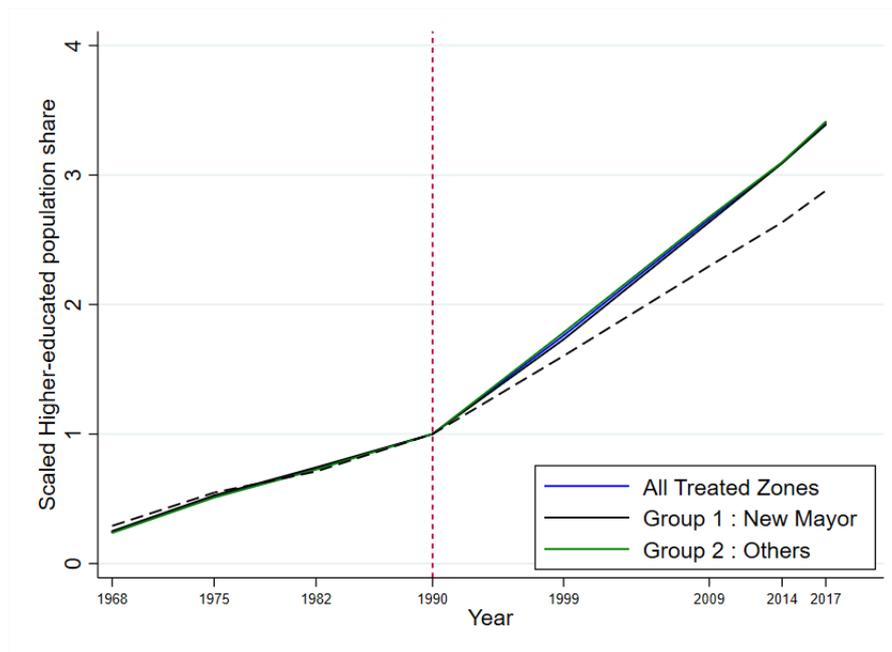
Figure 3.5 shows that when we aggregate the SCM results according to the zone's mayoral change, we do not see any difference in the estimates. This strengthens our conviction that our results are not driven by any local political changes.

---

<sup>38</sup>Cohabitation in France's dual-executive governmental system occurs when the president represents a different political tendency than the majority party in the National Assembly. The president then appoints a prime minister in the opposition to form a government.

<sup>39</sup>A radical change in local authorities seems to be the only shock that would be sufficiently important to muddy the predicted effects and not be captured by the SCM. In particular, other changes in the local economies, such as the entries and exits of firms, do not contradict our results as long as i) they are correlated with local socioeconomic dynamics which are controlled for in the SCM, ii) they may be indirect effects of the universities' creation. Nonetheless, when observing the historical economic forces in the treated zones, especially in the "successful" areas in terms of human-capital accumulation gains, no common phenomena, such as the building of major infrastructure or the founding of large companies, took place in our treated zones in the 1990s. There remain questions, however, regarding the opening of the Eurotunnel in 1994, which may have affected the human-capital accumulation in the zone of Calais (its French terminal), and the neighboring zone of Boulogne. On average, in these two zones, the effect of university creation is lower compared to the other zones' average, suggesting that the potential disruption represented by the Eurotunnel's opening do not drive the results.

Figure 3.5: Local Human-Capital Effects in Treated Zones: Group 1 vs. Group 2



**Spatial spillovers bias** As explained in Section 3.3.2, to further gauge the sensitivity of our results, we conduct additional analyses with alternative donor pools.

First, the creation of "new universities" may affect the outcome not only in the treated-zone but also in other control zones, especially in the neighboring employment-zones. For example, if the creation of La Rochelle University led to an increase of qualified population in the neighboring zones, the inclusion of these zones in the donor pool may underestimate the effect for the La Rochelle zone. Conversely, if the university's creation attracts higher-educated population from the neighboring zones, our estimates may be overstated. Even if our choice of employment-zones as the spatial unit for our analysis reduces the necessity for this caveat,<sup>40</sup> we still have concerns about this no-spillover assumption. We therefore run a similar synthetic control method but with a second donor pool (DP2), which excludes the neighboring zones of "new universities" zones.

Secondly, we conduct a third set of specifications based on a third donor pool (DP3), which excludes zones with decentralized secondary university sites. Indeed, if these small university satellites comparably affect the local higher-educated population share of their zones, we may

<sup>40</sup>Recall that employment-zones are based on the assumed flows of movement from residence to work, so this spatial designation already takes into account some spillovers effects, as compared to a city-level designation.

underestimate the impact of "new universities" by keeping these zones in the potential controls. Additionally, we run a last synthetic control method with a fourth donor pool (DP4), which excludes zones with decentralized secondary university sites, but keep the neighboring zones of "new universities" zones.

Table 3.3 summarizes the effects of the "new universities" for the four treatment periods (with associated *p-values*), depending on our four different donor pools. The results do not differ from the previous estimates with Donor Pool 1.

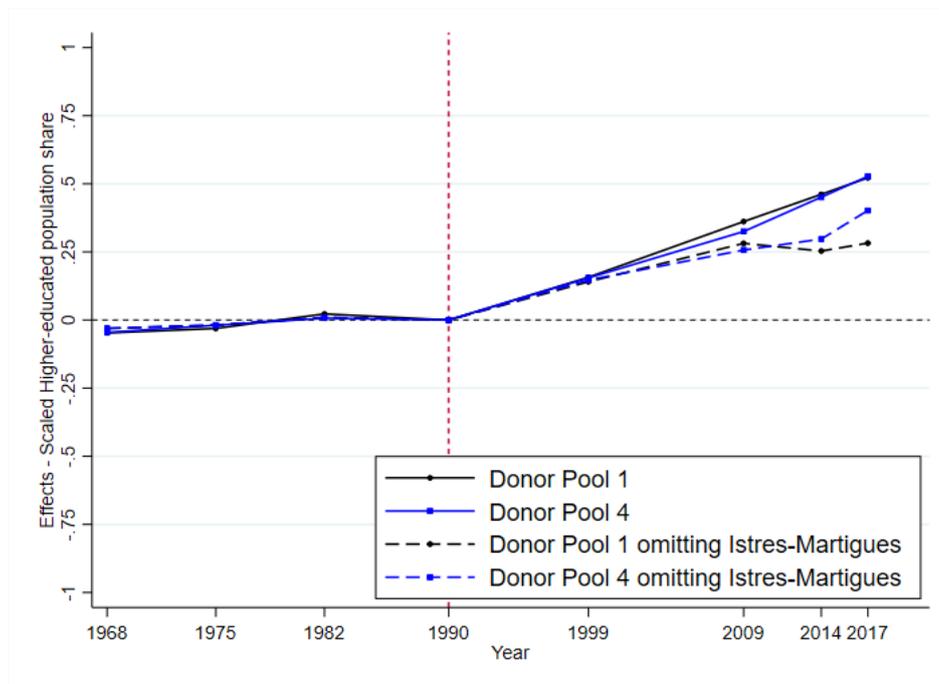
Table 3.3: Average Human Capital Gains (percentage points) - Donor Pool Comparisons

	1999	2009	2014	2017
<b>Donor Pool 1 (Benchmark)</b>	1.1 (0.000)	2.6 (0.000)	3.3 (0.000)	3.7 (0.000)
<b>Donor Pool 2</b>	1.1 (0.000)	2.5 (0.000)	3.2 (0.000)	3.6 (0.000)
<b>Donor Pool 3</b>	1.1 (0.000)	2.2 (0.000)	3.1 (0.000)	3.6 (0.000)
<b>Donor Pool 4</b>	1.1 (0.000)	2.3 (0.000)	3.2 (0.000)	3.7 (0.000)

*Notes.* We provide the synthetic control results for our four different donor pools for the four census. Results are computed the same way as in Table 3.2. The magnitude of the "new university" impact is measured as the outcome deviation (in percentage points) of the "new university" zone in comparison to the counterfactual. The *p-values* in brackets indicate the probability that the related result would happen purely by chance. Donor Pool 1 includes 200 placebos, Donor Pool 2 includes 179 placebos (= Donor Pool 1 without "new universities" neighboring zones), Donor Pool 3 includes 107 placebos (= Donor Pool 2 without zones with small secondary university satellites), and Donor Pool 4 includes 119 placebos (= Donor Pool 1 without zones with small secondary university satellites).

**Donor Pool weights** Another concern comes from the sensitivity of our main results to changes in the donor zones weights. In particular, we see in the tables of Appendix .6 that the synthetic controls for the cases of Lens, Béthune and Douai are principally computed on the basis of a unique control zone: Istres - Martigues. We run additional robustness checks by excluding this zone of our donor pool. By doing so, we accept a sort of breach in the synthetic control method by two aspects: we interfere in the computation of the best synthetic units, potentially sacrifice some goodness of fit and balance, and we create specification-searching problem. When we rerun the SCM using Donor Pool 1, but excluding the zone of Istres - Martigues, for the interest cases of Lens, Béthune and Douai, weights are now positive for two control zones: Le Creusot (around 60% of the synthetic control) and Sète (about 40%). We still

Figure 3.6: Sensitivity of Local Human-Capital Effects: omitting Istres - Martigues



estimate a strong positive effect of universities creation but that substantially decreases in value for the last two periods. However, the two control areas of Le Creusot and Sète host secondary university satellites which also benefit from U2000 program. Figure 3.6 shows that when using Donor Pool 4 (or Donor Pool 3), which excludes these potential "bad donors", but omitting the zone of Istres - Martigues, estimates are much more similar to our main effects.

**Alternative Diff-in-Diff method - Control for other contemporaneous shocks** Although the SCM addresses omitted variable bias by controlling for time-variant unobservables, one concern remains in the identification if other contemporaneous shocks influence spatial sorting of highly educated individuals in the post-treatment period. To mitigate this concern, we present in Table 3.4 the results of standard Difference-in-Difference (DD) model including covariates to control for post-treatment shocks. In particular, we check for the influence of local labor demand shocks linked to urban tax-free zones creation (ZFU program<sup>41</sup>), easier access

<sup>41</sup>The *Zones Franches Urbaines* (ZFU) program announced in 1996 have led to the creation of 38 urban tax-free zones in 1997, 41 in 2004 and 15 in 2006. Four of our treated zones were concerned by the program: Calais (1997), Lens (2004), Douai (2004) and La Rochelle (2004). This policy was designed to revitalize economic activity and promote social and professional integration in some vulnerable suburban areas. Tax credits were granted to firms settling in these ZFUs, resulting in local employment improvement (Charnoz, 2014; Briant et al., 2015).

to the area (through rail development<sup>42</sup>), new taste for local amenities such as the proximity of the sea<sup>43</sup> that might attract higher-educated workers in the new university areas, and potential spillovers coming from the proximity with other dense and dynamic local labor markets (measured by market potential<sup>44</sup>).

The difference-in-differences design uses the same initial donor pool than SCM (9 treated zones and 200 untreated) and estimates the average effect of new universities creation based on the following equation:

$$Y_{zt} = \alpha + \gamma Post_{zt} \times NewUni_z + \theta X_{z,t} + \eta_z + \delta_t + \epsilon_{zt} \quad (3.10)$$

where  $Y_{zt}$  is our outcome of interest (scaled share of higher-educated people - aged 16 and more and not including those in school) for the zone  $z$  in year  $t$ . We are interested in the interaction term between the dummy variable for post-treatment observations ( $Post_{zt}$ ) and a variable taking the value one if the zone  $z$  is a new-university-zone, zero otherwise ( $NewUni_z$ ).  $X_{z,t}$  is a set of covariates including our controls for other contemporaneous shocks.  $\eta_z$  and  $\delta_t$  are zone fixed effects and year fixed effects.  $\epsilon_{zt}$  are the errors terms.

Results reported in columns (1) to (6) of Table 3.4 show that DD method similarly concludes on a strongly significant positive effect of new universities creation on the local accumulation of higher-educated people.

For comparison and robustness purposes, we first present in column (1) of Table 3.4 a DD model where  $X_{z,t}$  contains all the matching variables used in the SCM model: share of managers and professionals in the employed population (aged 25-54); employment rate (in population aged 25-54); employment and population density; unemployment rate; participation rate (25-54); sectoral shares; age group shares; and distance from the nearest pre-existing university. We confirm a positive and very significant impact of new universities creation on the local

<sup>42</sup>Due to a lack of available local data on the chronological development of the Train Express Regional (TER), we focus on high-speed train TGV network. Since its introduction between Paris and Lyon in 1981, the French high-speed rail network, initially centered on Paris, has gradually expanded to connect many French cities with the construction of new lines to the south, west, north and east of France.

<sup>43</sup>Most of the treated zones are located near the sea (see map 3.1).

<sup>44</sup>Following Harris (1954), the market potential variable for a zone  $z$  is computed as the sum of each other local labor market's density ( $Dens_{i \neq z}$ ) divided by the distance separating this other zone  $i$  and  $z$ :  $PM_z = \sum_{i \neq z}^n \frac{Dens_i}{distance_{i-z}}$ .

Table 3.4: DiD estimations of new university impact on scaled higher-educated population share

Scaled Higher-educated population share						
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Impact of the university creation</b>	0.277*** (0.0444)	0.173*** (0.0454)	0.173*** (0.0383)	0.227*** (0.0390)	0.200*** (0.0396)	0.177*** (0.0396)
Tax-free Zones	-	-	-	-0.178*** (0.0270)	-0.183*** (0.0268)	-0.192*** (0.0268)
Distance to TGV station	-	-	-	0.004 n.s. (0.0010)	0.009 n.s. (0.0107)	0.010 n.s. (0.0107)
Distance to the sea						
x1975	-	-	-	-	-0.005 n.s. (0.0124)	-0.003 n.s. (0.0123)
x1982	-	-	-	-	0.005 n.s. (0.0128)	0.006 n.s. (0.0128)
x1990	-	-	-	-	-0.010 n.s. (0.0124)	-0.010 n.s. (0.0123)
x1999	-	-	-	-	-0.012 n.s. (0.0124)	-0.013 n.s. (0.0124)
x2009	-	-	-	-	-0.029*** (0.0124)	-0.033*** (0.0124)
x2014	-	-	-	-	-0.040*** (0.0124)	-0.044*** (0.0124)
x2017	-	-	-	-	-0.049*** (0.0124)	-0.053*** (0.0124)
ln Market Potential	-	-	-	-	-	-0.968*** (0.2902)
All SCM covariates	Yes	No	No	No	No	No
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Zone fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	1,672	1,672	1,672	1,672	1,672	1,672
R <sup>2</sup>	0.97	0.97	0.98	0.98	0.98	0.98

Notes. This Table presents Diff-in-Diff estimation results based on equation 3.10. The data sample contains 1,672 observation (9 treated + 200 untreated zones  $\times$  8 periods). Each column is a separate regression. Column (1) reports the regression including all the covariates used in the SCM: share of managers and professionals in the employed population (aged 25-54); employment rate (in population aged 25-54); employment and population density; unemployment rate; participation rate (25-54); four sectoral shares of employment (Agriculture/Manufacturing/Construction/Services); shares of population by age group (0-14; 15-24; 25-39; 40-64; 65 and more). As these covariates are potential bad controls, the rest of the columns exclude these variables. All specifications include a constant term.

Variables of distance are divided by 100.

Coefficients are reported with robust standard errors in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

accumulation of human capital. However, the problem with this specification is that many (if not all) of the control variables are very likely to be themselves outcomes of the treatment (see next section 3.6). They thus are bad controls since they represent indirect effects of new universities on the local concentration of graduates. We exclude this set of controls in the other robustness specifications (columns (2) to (6)).

Our preferred specification, which includes three controls for local post-treatment shocks (tax-free zone indicator, distance to the nearest TGV station and the interaction of the distance to the sea with year indicators) plus year and zone fixed effects, is presented in column (5) of Table 3.4. We estimate that, on average, the creation of a new university increased the local human-capital accumulation by 20 percentage points. Note that the value of this impact decreases to 18 percentage points when we add our additional control of market potential (see column(6)). However, because the U2000 program was national in scope, and in particular it financed the modernization of major historical universities in big French cities, it had a direct influence on this market potential variable, which may also represents a bad control in the DD specification.

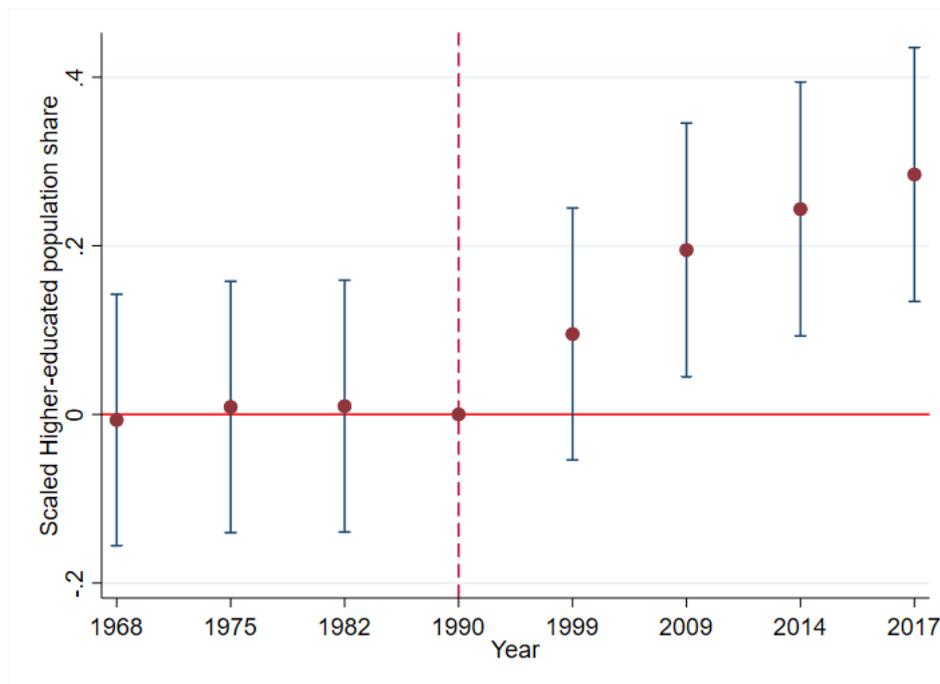
We finally run a difference-in-differences event-study design to evaluate treatment effects of the different pre- and post-treatment periods. We include the same controls as in column (5) of Table 3.4. Estimation results are reported in Table 16 in annex. Compared to SCM, event study results plot on Figure 3.7 show very similar progressive increase of the outcome variable after the creation of new universities in the 1990s. This reinforces our confidence that the main SCM results are not driven by post-treatment specific shocks affecting the treated zones.

### **3.6 Exploring the Employment Implications of the Creation of "New Universities"**

Although the principal analysis is designed to explore the effects of university creations on local accumulation of human capital, our results raise questions about other aspects of universities' local impact, in particular about their effect on employment levels. As mentioned above, local authorities hoped that universities would produce positive direct and indirect effects on local employment. In particular, the creation and attraction of a skilled population was expected to translate into more skilled jobs on the labor market, with spillovers effects on general employment and economic activity.

Therefore, having shown the positive effects of "new university" creation on local human-capital accumulation, the question that naturally arises is whether the founding of these universities and the associated human-capital accumulation in the local labor pool have led to more

Figure 3.7: Event study estimates of the impact of new universities creation



Notes. This graph plots the coefficients for interaction terms between year dummies and treatment status obtained from an event-study difference-in-differences design. The graph is based on our preferred model, which includes a dummy for tax-free zone, controls for distance to the nearest TGV station, an interaction between year and distance to the sea, zone fixed effects, and year fixed effects. Confidence intervals are represented at the 95% level.

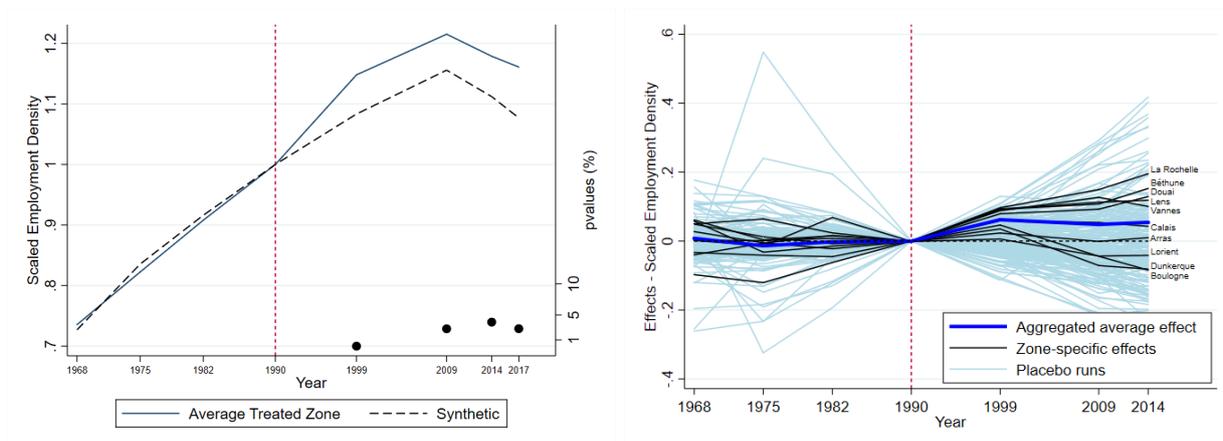
employment, less unemployment, and more skilled jobs at the local level. Using precisely the same synthetic control method, we also explore the U2000 program's impact on employment density, unemployment, and the share of white-collar occupations in the local labor markets.

Figures 3.8 to 3.10 depict the estimated results for the average of the treated zones. While the synthetic control method shows a good match in the pre-treatment period with the actual trends, when averaging over all the treated zones, it actually smooths out some of the heterogeneity in the match for each individual treated zone. In contrast with the synthetic controls in the principal case of higher-educated population share, which showed a reasonably good match for all the treated zones, for these new outcomes, case-by-case results should be interpreted with some caution, keeping an eye on the pre-RMSPE value. For this reason, we choose to comment on only the average effects – case-by-case results and robustness tests are presented in the Annex.

**Employment Density** On average, the creation of "new universities" increased the growth of local employment density by 6 p.p. within around 7 years, and this gap remains stable within 15 years (6 p.p.), and 20 years (7 p.p.), and reaches 8 p.p within 25 years. For a mean of 85 jobs per km<sup>2</sup> in 1990, these estimates indicate that the employment density is on average 6 jobs greater in those zones in 1999, 5 jobs greater in 2009, 6 jobs greater in 2014, and 7 jobs greater in 2017 as a result of "new universities" creation,<sup>45</sup> which represents an increase of 8% compared to the counterfactual in 2017. Testing statistical inference, we see that the probability that the estimation of the average aggregated impact has arisen purely by chance is nil for the first period, but that the *p-values* are around 3% for the last periods.

Those results therefore indicate that university creations rather had a short-term impact on local density, presumably due to the immediate direct effects from university-building, and new demand from students and university workers. The estimated gap with the counterfactual slowly increased in the second decade of the universities' existence, suggesting no substantial cumulative long-term spillover effects on local employment density.

Figure 3.8: Effects of "New Universities" Creation on Local Employment Density



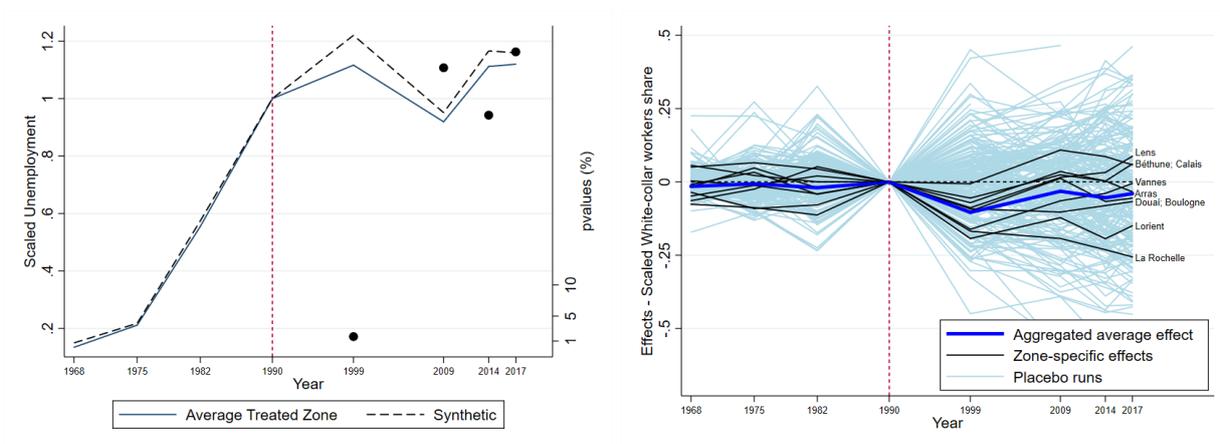
**Unemployment** On average, the creation of "new universities" reduced the growth of local unemployment by 10 p.p. by 1999. This effect is no longer statistically credible, however, after 15, 20 and 25 years (based on a threshold of 10%). At a mean of 12.6% in 1990, the first period estimate indicates that the unemployment rate in those zones is on average lower by 1.3 p.p.

<sup>45</sup>Employment density levels in 1990 times the estimated growth difference between treated-zones and synthetic control zones.

by 1999, as a result of the "new universities" creation,<sup>46</sup> which represents a decrease of 10% compared to the counterfactual.

Those results indicate that university creation seems to have had no long-term effect on local unemployment. Although we found a significant reduction in unemployment by 1999, this immediate impact disappears in the subsequent periods, suggesting a temporary shock in "new university" zones that subsequently subsides in the next period, to return to the counterfactual trend.

Figure 3.9: Effects of "New Universities" Creation on Local Unemployment



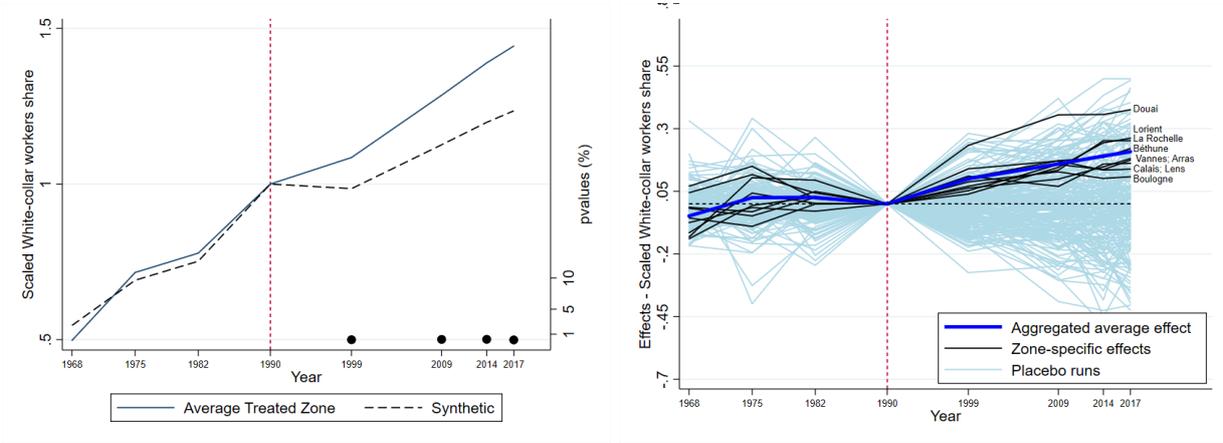
**White-collar workers share** On average, "new universities" creation increased the growth of the local white-collar employment share by 10 p.p. within around 7 years, 16 p.p. within 15 years, 10 p.p. within 20 years, and 21 p.p. within 25 years. At a mean level of 8.4% in 1990, these estimates indicate that the share of white-collar workers in those zones is progressively larger on average, rising to a difference of 1.7 p.p. in 2017, as a result of the "new universities" creation,<sup>47</sup> which represents an increase of 17% compared to the counterfactual.

Those results indicate that the impact of the creation of "new universities" on the share of local white-collar workers is analogous to the impact on local human-capital accumulation. This is a first element in answering the question of whether human-capital gains have actually translated into skilled employment at the local level.

<sup>46</sup>The unemployment level in 1990 times the estimated growth difference between treated-zones and synthetic control.

<sup>47</sup>White-collar workers' share in 1990 times the estimated growth difference between treated-zones and synthetic control.

Figure 3.10: Effects of "New Universities" Creation on Local White-Collar Workers Share



### 3.7 Conclusion

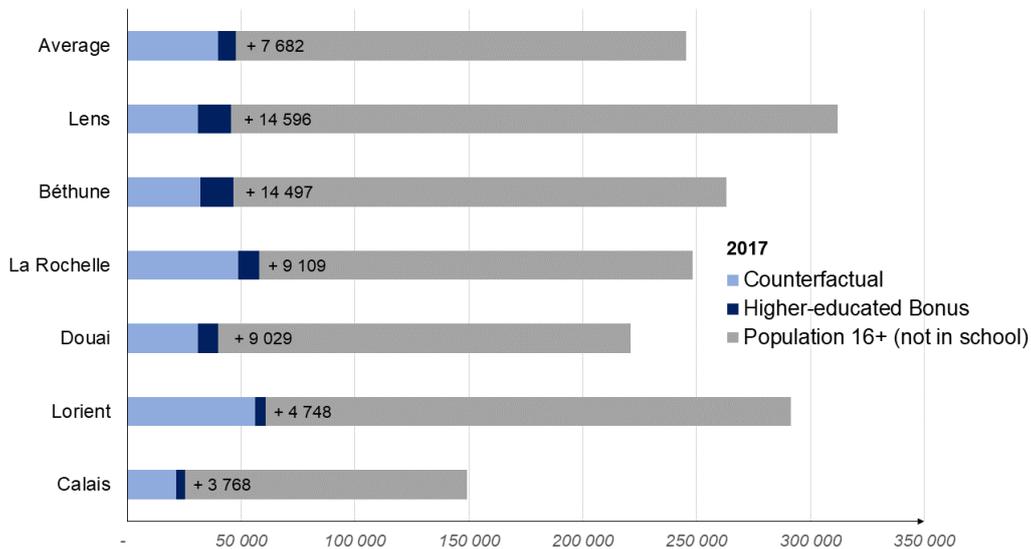
Local and regional authorities' initiatives for developing higher education in the late 20th century were supported by the national program "University 2000," using a new decentralizing approach. The engagement of local actors reveals a vision of the university as an important economic development lever, in particular because it is expected to boost human-capital accumulation in the locality. We examined the causal effects of "new universities" creation on local share of higher-educated population (aged 16 and more and not including students) by implementing the synthetic control method, using census data over the 1968-20147 period.

Our estimates provide the first causal evidence regarding the positive effects of French "new university" creation on local human-capital accumulation. Taking into account the upward trends in higher-education in the late 20th century, the results show that, on average, the creation of "new universities" increased the higher-educated population share by 4 points within 27 years. This represents a 18% increase from the counterfactual level: the average share of higher-educated population reached 24% in "new university" zones in 2017, though it would have been 20% without the U2000 Plan. This represents about 8,000 higher-educated individuals on average (see Figure 3.11 for effects in terms of population level).

We measure notable heterogeneity across zones. In six cases,<sup>48</sup> the local higher-educated population share is significantly larger as a result of the "new university" creation, with an

<sup>48</sup>The Lens, Béthune and Douai sites of the Artois University, the Calais site of the Littoral Opal Coast University, La Rochelle University, and the Lorient site of Bretagne-Sud University.

Figure 3.11: Higher-educated Population Share in Total Population in 2017 (not in school 16+)



effect between 2 p.p. and 7 p.p. within 27 years (which represents between 8% and 45% of the respective counterfactuals). In the three other cases, we cannot confidently conclude that there exists a significant "new university" effect on local human-capital accumulation.

Our study confirms the validity of local actors' expectations, regarding university-creation as a vector of local development. In addition to its well-known role in reducing on social and geographical inequalities for individuals, university decentralization in middle-sized areas has a positive impact on local development by increasing human-capital accumulation. Moreover, exploring the employment implications of "new universities" creation, we found credible evidence that, on average, human-capital gains co-occurred with gains in skilled employment, providing some arguments against the idea that there is a mismatch between higher-education demand and supply in medium-sized cities containing universities. In terms of general employment, using the same synthetic control method, we show that average employment density in the treated zones was 8% higher in 2017 compared to the counterfactual (although we found no convincing impact on local unemployment on average). More specifically, we show that, on average, "new university" creation increased the white-collar workers population share by 1.7 points within 27 years. This represents a 17% increase from the counterfactual level.

The synthetic control method helps us to identify the causal effects of a university's creation,

but it does not provide information about the mechanisms underlying local human-capital accumulation – nor about the implications for skilled employment. In particular, we cannot distinguish between direct and indirect causal effect. We do not know if the "new universities" create and/or attract and/or maintain higher-educated population in their zones. The underlying demographic causes could include the fact that newly higher-educated individuals decide to work and stay in the zone where they went to university, but also the fact that the existence of a university could be a positive signal that attracts new companies and higher-educated workers.

# .1 Trends in Higher-educated Population Share 1968-2014

Figure 12: Trends in Higher-educated population share 1968-2014

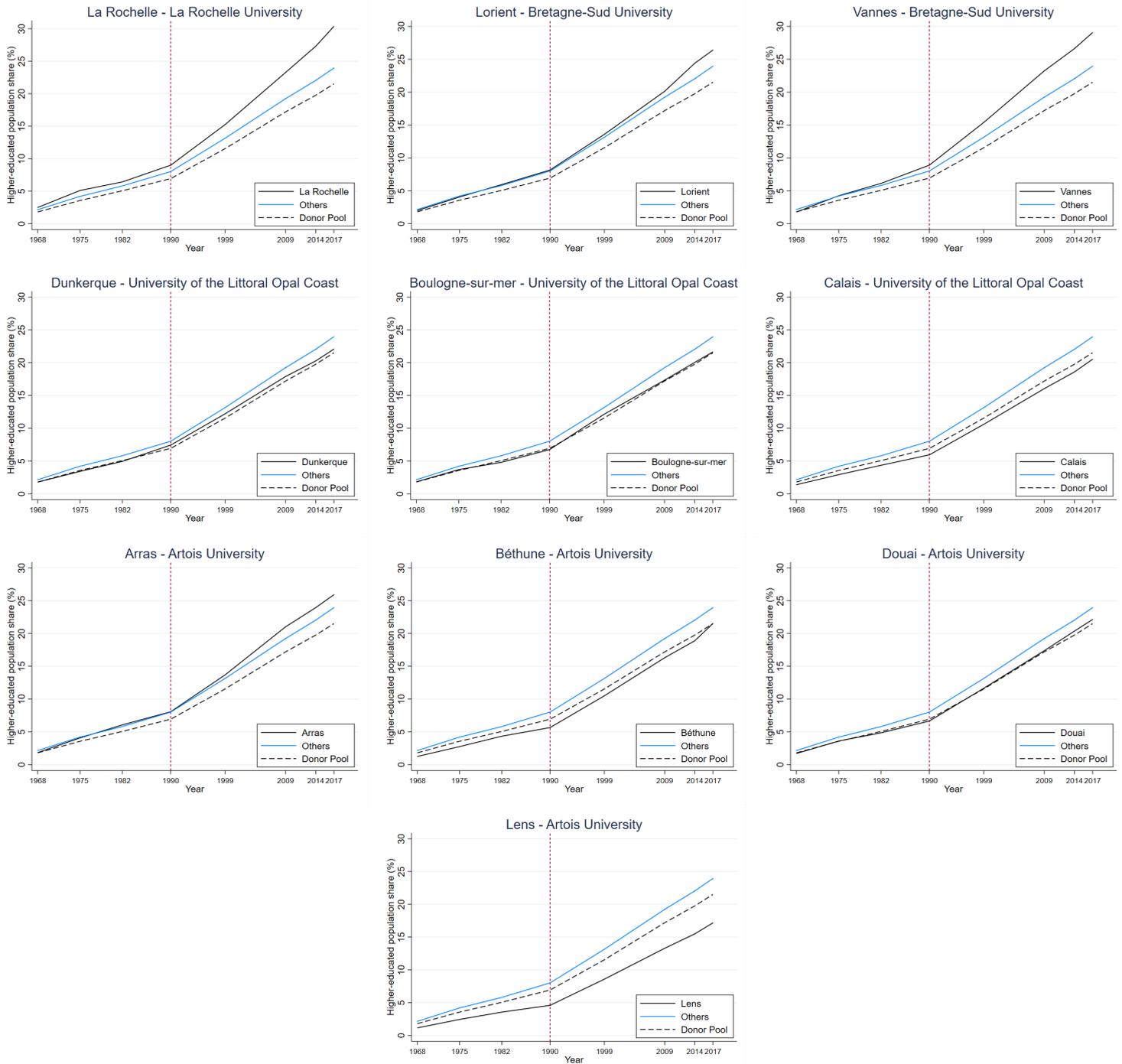
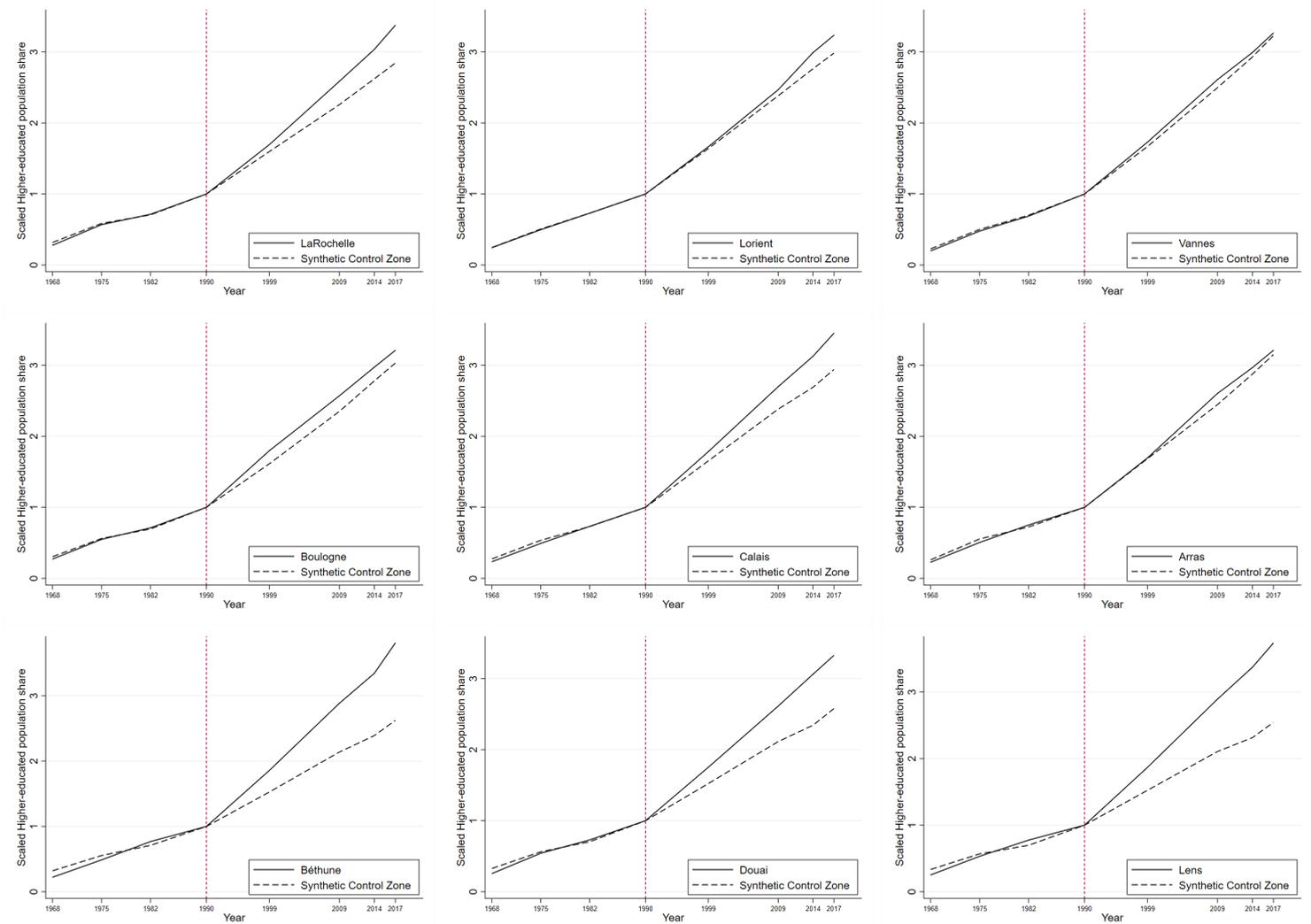


Figure 13: Trends in Higher-educated population share 1968-2014 (Scaled in 1990) - Synthetic Control Comparisons (Scaled in 1990)



## .2 Fit &amp; Balance

Table 5: Fit &amp; Balance of synthetic controls

	Donor Pool	Arras		Béthune		Boulogne		Calais		Douai	
		<i>obs</i>	<i>SC</i>								
RMSPE (pretreatment)		<b>0,033</b>		<b>0,067</b>		<b>0,019</b>		<b>0,029</b>		<b>0,041</b>	
Scaled Higher-educated population share (outcome)	0,63	0,62	0,63	0,62	0,65	0,63	0,64	0,61	0,64	0,63	0,65
Higher-educated population share in 1990	7%	8%	7%	6%	9%	7%	8%	6%	8%	7%	9%
Professionals employment share	6%	7%	7%	5%	7%	7%	7%	6%	7%	6%	7%
Population density (log)	3,5	4,1	4,1	5,5	4,9	4,9	4,9	4,7	4,7	5,7	5,0
Jobs density (log)	2,7	3,2	3,2	4,5	4,1	4,0	4,0	3,9	3,9	4,8	4,2
Unemployment	5%	4%	5%	5%	7%	7%	7%	7%	7%	6%	7%
Employment rate	73%	70%	70%	59%	64%	65%	65%	65%	66%	61%	65%
Participation rate	77%	74%	74%	63%	70%	70%	70%	70%	71%	66%	70%
Services	46%	53%	53%	45%	50%	56%	56%	47%	48%	48%	50%
Agriculture	19%	16%	16%	6%	3%	9%	9%	8%	8%	3%	3%
Manufacturing	25%	23%	23%	40%	36%	26%	26%	35%	35%	42%	35%
Construction	10%	8%	9%	9%	12%	8%	9%	10%	10%	7%	12%
0-14	23%	25%	24%	25%	26%	27%	24%	28%	26%	26%	26%
15-24	15%	16%	16%	16%	16%	17%	15%	17%	16%	17%	16%
25-39	19%	20%	20%	19%	23%	20%	20%	20%	21%	20%	24%
40-64	28%	26%	27%	27%	27%	25%	28%	24%	26%	26%	26%
65 and more	20%	17%	17%	17%	12%	15%	17%	16%	14%	15%	11%
Distance	62	44	51	33	35	103	34	94	41	30	31

	Donor Pool	Dunkerque		La Rochelle		Lens		Lorient		Vannes	
		<i>obs</i>	<i>SC</i>								
RMSPE (pretreatment)		<b>0,065</b>		<b>0,022</b>		<b>0,059</b>		<b>0,005</b>		<b>0,020</b>	
Scaled Higher-educated population share (outcome)	0,63	0,59	0,65	0,64	0,65	0,64	0,65	0,62	0,62	0,59	0,61
Higher-educated population share in 1990	7%	7%	9%	9%	9%	5%	9%	8%	8%	9%	7%
Professionals employment share	6%	7%	7%	8%	8%	5%	7%	7%	6%	6%	6%
Population density (log)	3,5	5,2	5,0	4,4	4,4	6,5	5,0	4,4	4,4	3,9	3,9
Jobs density (log)	2,7	4,4	4,2	3,6	3,6	5,5	4,3	3,6	3,6	3,1	3,1
Unemployment	5%	6%	7%	6%	6%	7%	8%	6%	6%	5%	5%
Employment rate	73%	65%	65%	70%	70%	57%	65%	70%	70%	73%	73%
Participation rate	77%	70%	70%	75%	74%	61%	70%	74%	74%	77%	77%
Services	46%	53%	52%	61%	61%	44%	51%	52%	52%	50%	50%
Agriculture	19%	6%	5%	9%	9%	1%	2%	15%	15%	24%	24%
Manufacturing	25%	33%	32%	20%	19%	45%	35%	22%	22%	14%	14%
Construction	10%	9%	11%	10%	10%	10%	12%	11%	11%	12%	12%
0-14	23%	29%	26%	24%	22%	27%	26%	24%	24%	25%	23%
15-24	15%	17%	16%	16%	15%	17%	16%	15%	15%	16%	15%
25-39	19%	22%	23%	20%	20%	19%	24%	19%	20%	19%	20%
40-64	28%	24%	26%	27%	28%	27%	26%	28%	28%	27%	28%
65 and more	20%	12%	13%	18%	19%	14%	10%	17%	17%	18%	18%
Distance	62	68	32	121	32	28	28	110	49	92	61

Notes. In each case, the 'quality' of the synthetic control is given by the comparison between average values of pre-treatment observed characteristics for the treated-zone (*obs.*) and the estimated values of these characteristics for its synthetic control (*SC*) (i.e. the balance on predictors) and the root mean square prediction error (RMSPE) of the synthetic control estimator (i.e. the fit of the outcome variable for the pre-intervention period). The set of pre-intervention characteristics include: scaled higher-educated population share, index 1 in 1990 (the outcome variable); higher-educated population share in 1990; share of managers and professionals in employed population (aged 25-54); employment rate (in population aged 25-54); employment and population density; unemployment rate; participation rate (25-54); four sectoral shares of employment (Agriculture/Manufacturing/Construction/Services); shares of population by age group (0-14; 15-24; 25-39; 40-64; 65 and more); and distance from the nearest pre-existing university.

### .3 Estimated Effects by Zone with *p-values*

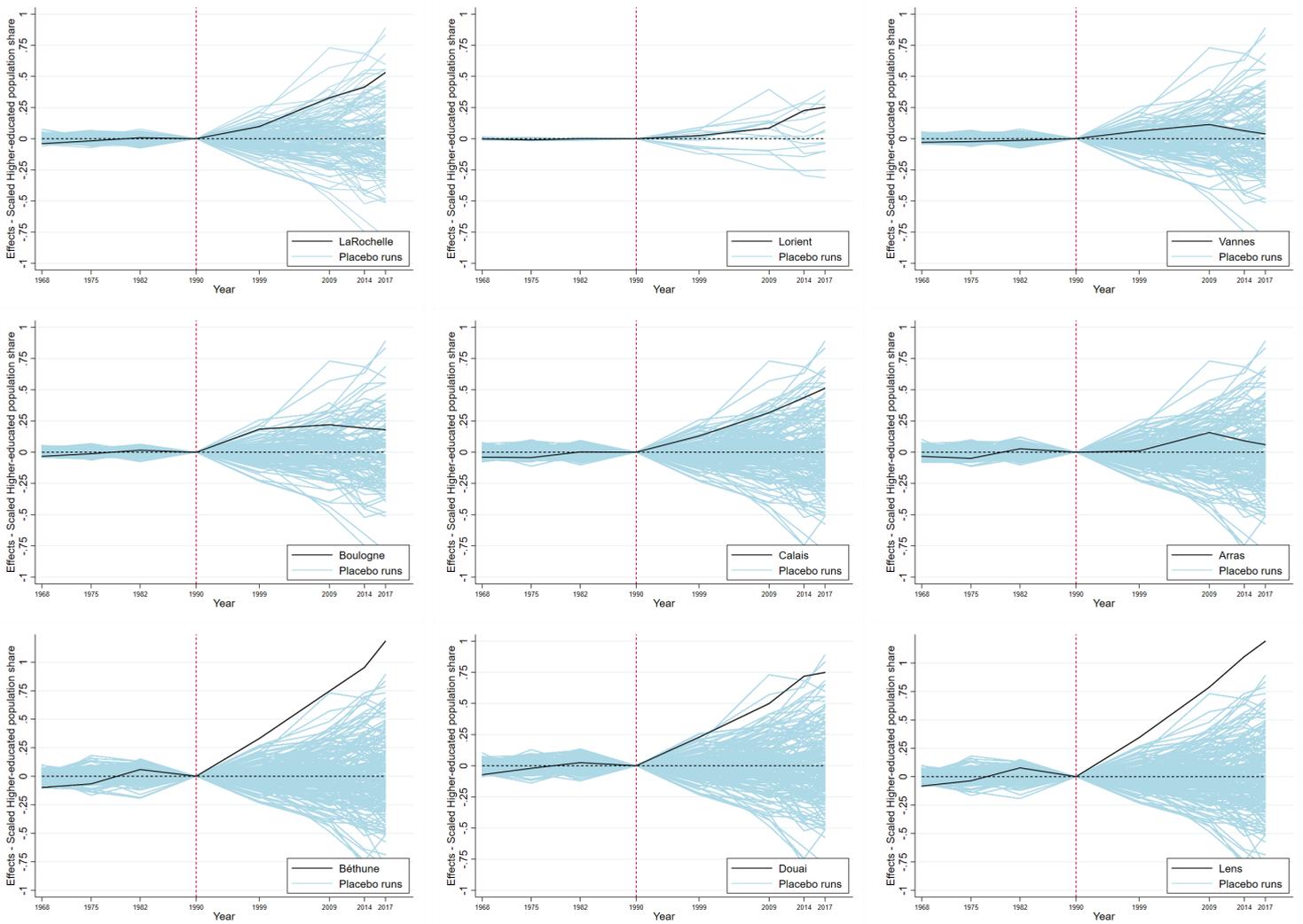
Table 6: Estimated gains (percentage points) with corresponding *p-values*

	1999	2009	2014	2017
<b>Average</b>	1.1 (0.000)	2.6 (0.000)	3.3 (0.000)	3.7 (0.000)
<b>Béthune</b>	1.9 (0.000)	4.2 (0.000)	5.4 (0.000)	6.7 (0.000)
<b>Lens</b>	1.6 (0.000)	3.6 (0.000)	4.9 (0.000)	5.5 (0.000)
<b>Douai</b>	1.5 (0.02)	3.3 (0.01)	4.8 (0.005)	5.0 (0.015)
<b>La Rochelle</b>	0.9 (0.14)	3.0 (0.062)	3.7 (0.073)	4.8 (0.057)
<b>Calais</b>	0.8 (0.112)	1.9 (0.086)	2.6 (0.066)	3.0 (0.061)
<b>Lorient</b>	0.2 (0.313)	0.7 (0.333)	1.9 (0.146)	2.1 (0.188)
<b>Boulogne</b>	1.2 (0.038)	1.5 (0.13)	1.3 (0.211)	1.2 (0.227)
<b>Arras</b>	0.1 (0.425)	1.3 (0.205)	0.7 (0.31)	0.5 (0.365)
<b>Vannes</b>	0.6 (0.226)	1.0 (0.247)	0.6 (0.355)	0.3 (0.409)

*Notes.* We summarize our results for each treated-zone and for the aggregate average. We use synthetic control estimates and higher-educated population share in 1990 at the local level to report specific impact in terms of human-capital gains after U2000 Plan. The magnitude of the "new university" impact is measured as the outcome deviation of the "new university" zone in comparison to the counterfactual. For example, as a result of the creation of La Rochelle University in 1993, the local share of higher-educated population is greater by 4,8 points in 2017 (30,4% compared to 25,6% for the synthetic control). The *p-values* in brackets indicate the probability that the related result would happen purely by chance. For a distribution of 200 placebos, a 10% *p-value* means that only 20 placebo-zones show an estimated effect greater than the given treated-zone.

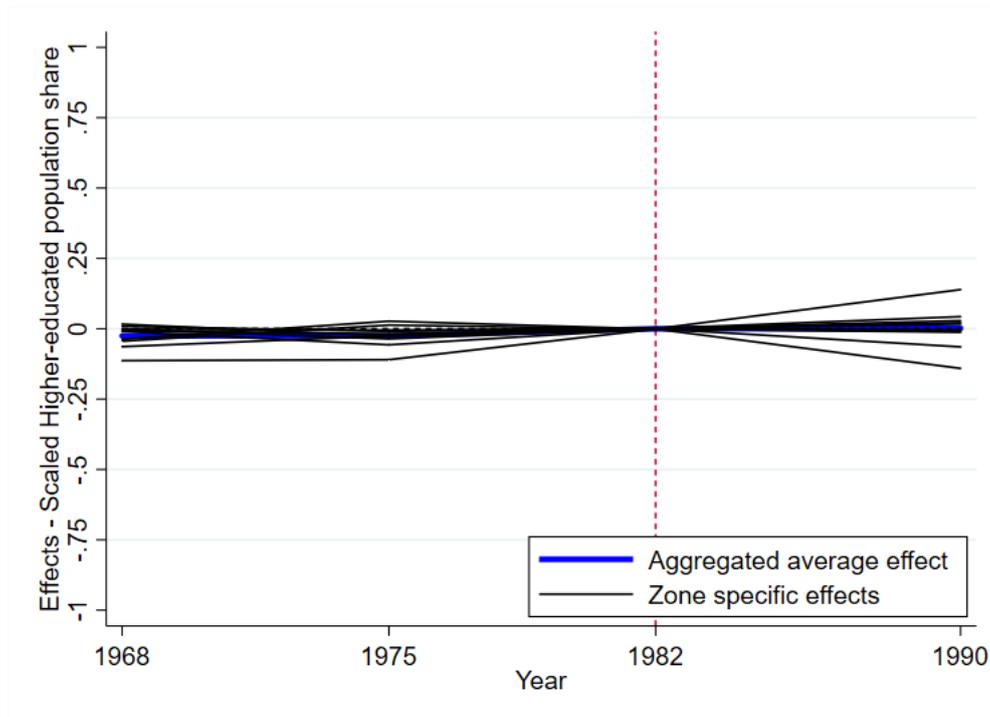
## .4 In-Space Placebo Tests by Zone

Figure 14: Gaps in scaled higher-educated population share - treated-zone vs. placebo tests (<math> < 2 \times \text{RMSPE}</math>)



## .5 In-Time Placebo Tests

Figure 15: Local Human Capital Effects in "New Universities" Zones: Placebo 1982



**.6 Synthetic Weights by Zone**

Table 7: La Rochelle University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Châlons-en-Champagne	0,245	-	-
Tergnier	0	0,006	0,009
Honfleur	0	0,506	0,557
Lunéville	0,115	-	-
Dinan	0	0,287	0,240
Bourg-en-Bresse	0,146	-	-
Sète	0,267	-	-
Cannes - Antibes	0,227	-	-
Istres - Martigues	0	0,201	0,194

Table 8: Arras - Artois University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Châlons-en-Champagne	0,244	-	-
Tergnier	0	0,313	0,371
Avranches	0	0,003	0
Bayeux	0	0,218	0,299
Saint-Lô	0,146	-	-
Saint-Omer	0,254	-	-
Berk-Montreuil	0	0,308	-
Dinan	0	0,126	0,223
Rodez	0,088	-	-
Sète	0,241	-	-
Ajaccio	0	0	0,04
Ghisonaccia - Aléria	0,018	0,032	0,067

Table 9: Lorient - Bretagne Sud University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Tergnier	0	0,063	0,088
Avranches	0	0,057	0,081
Granville	0	0,151	0,102
Saint-Lô	0,038	-	-
Berck - Montreuil	0,233	0,047	-
Châteaubriant	0	0,057	0
Dinan	0	0,314	0,346
Guingamp	0	0	0,039
Carhaix-Plouguer	0	0,027	-
Quimper	0,266	-	-
Fougères	0,064	-	-
Saint-Malo	0,011	-	-
Nîmes	0,014	-	-
Agde - Pézenas	0	0,028	0,043
Sète	0,114	-	-
Istres - Martigues	0,242	0,314	0,294
Draguignan	0,004	-	-
Fréjus - Saint-Raphaël	0,018	0	0
Ajaccio	0	0	0,006

Table 10: Vannes - Bretagne Sud University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Bayeux	0	0,014	0,07
Coutances	0,022	0,184	0,116
Saint-Lô	0,307	-	-
Berck - Montreuil	0	0,129	-
Dinan	0	0,355	0,437
Saint-Malo	0,36	-	-
Vitré	0,081	0,194	0,231
Saint-Flour	0,069	0	0
Fréjus - Saint-Raphaël	0,082	0,075	0,049
Porto-Vecchio	0,058	0	0
Bastia	0	0,026	0,071
Ghisonaccia - Aléria	0,023	0,022	0,029

Table 11: Lens - Artois University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Alès	0	0	0,05
Istres - Martigues	1	1	0,95

Table 12: Boulogne - Littoral University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Tergnier	0,199	0	0,014
Saint-Omer	0,104	-	-
Dinan	0	0,178	0,163
Alès	0	0,009	0,005
Bagnols-sur-Cèze	0,072	0	0
Agde - Pézenas	0	0,153	0,155
Sète	0,451	-	-
Istres - Martigues	0,125	0,657	0,647
Salon-de-Provence	0,051	-	-
Ajaccio	0	0,003	0,015

Table 13: Calais - Littoral University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
La Vallée de la Bresse	0	0,149	0,139
Tergnier	0	0,158	0,145
Cambrai	0,088	-	-
SaintOmer	0,299	-	-
Les Herbiers	0	0,001	0,009
Alès	0,108	0,045	0,066
Agde - Pézenas	0,033	0,108	0,102
Sète	0,021	-	-
Istres - Martigues	0,458	0,539	0,539

Table 14: Béthune - Artois University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Le Creusot - Montcea	0,100	-	-
Alès	0,068	0,339	0,409
Istres - Martigues	0,832	0,661	0,591

Table 15: Douai - Artois University

<b>Employment-Zones</b>	<b>Donor Pool 1</b>	<b>Donor Pool 2</b>	<b>Donor Pool 3</b>
Le Creusot - Montceau	0,075	-	-
Alès	0	0,084	0,158
Istres - Martigues	0,925	0,916	0,842

## .7 Diff-in-Diff Event Study Design

Table 16: Dynamic DD estimations of new university impact on scaled higher-educated population share

	(1)
Pre new-universities creation	
t=1968	-0.007 n.s. (0.0760)
t=1975	0.009 n.s. (0.0760)
t=1982	0.010 n.s. (0.0761)
Post new-universities creation	
t=1999	0.095 n.s. (0.0762)
t=2009	0.195*** (0.0767)
t=2014	0.244*** (0.0768)
t=2017	0.285*** (0.0768)
Tax-free Zones	-0.191*** (0.0269)
Distance to TGV station	0.009 n.s. (0.0107)
Distance to the sea	
x1975	-0.004 n.s. (0.0126)
x1982	0.006 n.s. (0.0130)
x1990	-0.010 n.s. (0.0126)
x1999	-0.015 n.s. (0.0126)
x2009	-0.029** (0.0126)
x2014	-0.039*** (0.0126)
x2017	-0.047*** (0.0127)
Year fixed effects	Yes
Zone fixed effects	Yes
Observations	1,672
R <sup>2</sup>	0.97

Notes. This Table presents Diff-in-Diff Event Study estimation results based on equation described in Section 3.5. The data sample contains 1,672 observation (9 treated + 200 untreated zones  $\times$  8 periods). Variables of distance are divided by 100.

Coefficients are reported with robust standard errors in parenthesis. \*\*\*, \*\*, and \* indicate significance at the 0.01, 0.05, and 0.1 levels.

**.8 Employment Outcomes Effects by Zone**

Table 17: Employment Outcomes Average Effects - Donor Pools comparisons

	Employment Density				White-collar workers share				Unemployment			
	RMSPE	1999	2009	2014	RMSPE	1999	2009	2014	RMSPE	1999	2009	2014
<b>Donor Pool 1 (benchmark)</b>	6,5 (0,000)	5,9 (0,028)	6,7 (0,039)	8,4 (0,028)	10,0 (0,001)	15,9 (0,002)	19,1 (0,002)	20,8 (0,001)	-10,4 (0,017)	-3,2 (0,448)	-5,4 (0,372)	-4,0 (0,473)
<b>Donor Pool 2</b>	7,2 (0,000)	6,9 (0,011)	8,0 (0,015)	9,8 (0,011)	9,9 (0,002)	15,1 (0,003)	18,2 (0,003)	19,8 (0,001)	-10,3 (0,019)	-2,6 (0,47)	-6,5 (0,302)	-4,8 (0,417)
<b>Donor Pool 3</b>	5,9 (0,000)	4,0 (0,000)	4,4 (0,000)	5,9 (0,000)	10,8 (0,002)	13,8 (0,011)	17,2 (0,011)	17,9 (0,006)	-4,3 (0,206)	5,6 (0,113)	4,6 (0,198)	7,4 (0,107)
<b>Donor Pool 4</b>	5,1 (0,004)	3,1 (0,164)	3,6 (0,165)	5,1 (0,119)	10,9 (0,001)	14,2 (0,006)	17,1 (0,007)	18,0 (0,004)	-3,9 (0,216)	5,4 (0,124)	4,9 (0,2)	7,3 (0,113)

Notes. We present the synthetic control results for our three different donor pools for the different census. Results are computed the same way as Table??. The magnitude of the 'new university' impact is measured as the outcome's trend deviation in the 'new university' zone compared to the counterfactual's trend. The *p-values* in brackets indicate the probability that the related result would happen purely by chance. Donor Pool 1 includes 200 placebos, Donor Pool 2 includes 179 placebos (= Donor Pool 1 without 'new universities' neighboring zones), Donor Pool 3 includes 107 placebos (= Donor Pool 2 without zones with small secondary university satellites), and Donor Pool 4 includes 119 placebos (= Donor Pool 1 without zones with small secondary university satellites).

## .9 Employment Outcomes In-Time Placebo Tests

Figure 16: In-Time Placebo Employment Density Outcome

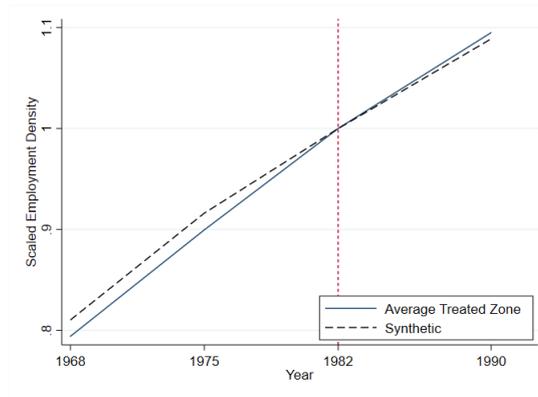


Figure 17: In-Time Placebo Unemployment Outcome

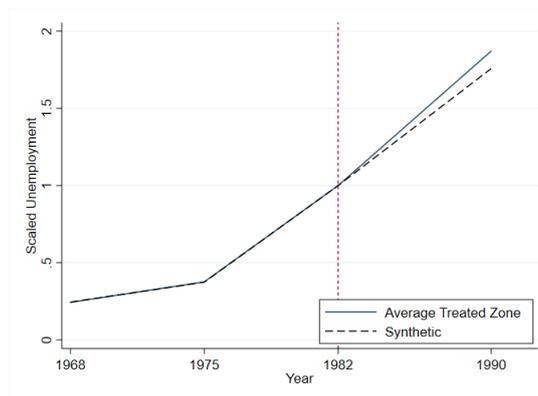
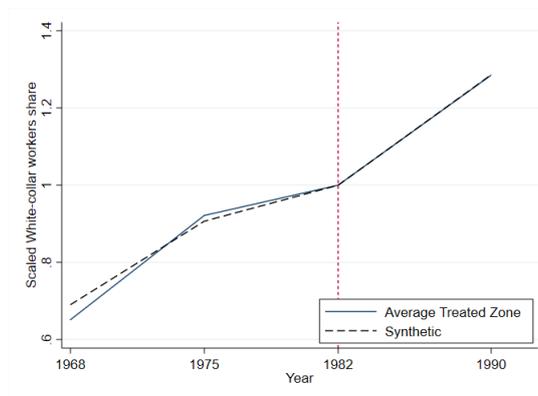


Figure 18: In-Time Placebo White-collar Workers Share Outcome



# Bibliography

- (1986). *Histoire des universités en France*. Bibliothèque historique Privat.
- (1996a). *L'Université d'Artois*. Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel.
- (1996b). *L'Université du Littoral*. Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel.
- (1997). *L'Université de La Rochelle*. Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel.
- (1999). *L'Université de Bretagne-Sud*. Comité National d'Evaluation des établissements publics à caractère scientifique, culturel et professionnel.
- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.
- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.
- Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *The American Economic Review*, 93(1):113–132.
- Abel, J. R. and Deitz, R. (2012). Do colleges and universities increase their region's human capital?. *Journal of Economic Geography*, 12(3):667 – 691.
- Abel, J. R., Dey, I., and Gabe, T. M. (2012). Productivity And The Density Of Human Capital. *Journal of Regional Science*, 52(4):562–586.
- Abel, J. R., Florida, R., and Gabe, T. M. (2018). Can low-wage workers find better jobs? Staff Reports 846, Federal Reserve Bank of New York.
- Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.
- Acemoglu, D. and Angrist, J. (2001). How large are human-capital externalities? evidence from compulsory-schooling laws. In *NBER Macroeconomics Annual 2000, Volume 15*, pages 9–74. National Bureau of Economic Research, Inc.
- Aghion, P., Bergeaud, A., Blundell, R., and Griffith, R. (2019). The innovation premium to soft skills in low-skilled occupations. CEP Discussion Papers dp1665, Centre for Economic Performance, LSE.

- Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.
- Albouy, V. and Tavan, C. (2007). Massification et démocratisation de l’enseignement supérieur en France. *Série des documents de travail de la Direction des Études et Synthèses Économiques*.
- Algan, Y., Malgouyres, C., and Senik, C. (2020). Territoires, bien-être et politiques publiques. Les notes du CAE 55, Conseil d’analyse économique.
- Aliaga, C. (2015). Les zonages d’étude de l’insee : une histoire des zonages supracommunaux définis à des fins statistiques. insee méthode n 129, chapitre 4 - les zones d’emploi. Technical report, INSEE.
- Amarante, P., Silva, M., and Monte, P. (2019). Does the spatial density of employment stimulate inter-firm worker mobility? an analysis of Brazilian municipalities. *The Annals of Regional Science*, 63.
- Andersson, F., Burgess, S., and Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1):112–128.
- Andersson, M. and Thulin, P. (2013). Does spatial employment density spur inter-firm job switching? *The Annals of Regional Science*, 51(1):245–272.
- Andersson, R., Quigley, J. M., and Wilhelmson, M. (2004). University decentralization as regional policy: the Swedish experiment. *Journal of Economic Geography*, 4(4):371–388.
- Andersson, R., Quigley, J. M., and Wilhelmson, M. (2009). Urbanization, productivity, and innovation: Evidence from investment in higher education. *Journal of Urban Economics*, 66(1):2 – 15.
- Andini, M., de Blasio, G., Duranton, G., and Strange, W. C. (2013). Marshallian labour market pooling: Evidence from Italy. *Regional Science and Urban Economics*, 43(6):1008 – 1022.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Askenazy, P., Behaghel, L., Laouenan, M., and Meurs, D. (2019). Quarante ans d’analyse du travail et de l’emploi : points de vue de quatre économistes / forty years of labour studies : the perspective of four economists. *Travail et emploi*, pages 69–94.
- Aust, J. (2007). Le sacre des présidents d’université. une analyse de l’application des plans université 2000 et université du troisième millénaire en Rhône-Alpes. *Sociologie du Travail*, 49(2):220 – 236.
- Baker, G., Gibbs, M., and Holmstrom, B. (1994). The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics*, 109(4):881–919.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2):407–443.
- Bayer, C. and Kuhn, M. (2019). Which Ladder to Climb? Decomposing Life Cycle Wage Dynamics. IZA Discussion Papers 12473, Institute of Labor Economics (IZA).

- Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy*, 81(4):813–846.
- Becker, S. O., Heblich, S., and Sturm, D. M. (2021). The impact of public employment: Evidence from bonn. *Journal of Urban Economics*, 122:103291.
- Bentsen, K. H., Munch, J. R., and Schaur, G. (2019). Education spillovers within the workplace. *Economics Letters*, 175:57–59.
- Blackwell, M., Cobb, S., and Weinberg, D. (2002). The economic impact of educational institutions: Issues and methodology. *Economic Development Quarterly*, 16(1):88–95.
- Bleakley, H. and Lin, J. (2012). Thick-market effects and churning in the labor market: Evidence from us cities. *Journal of Urban Economics*, 72(2):87 – 103.
- Bonander, C., Jakobsson, N., Podestà, F., and Svensson, M. (2016). Universities as engines for regional growth? using the synthetic control method to analyze the effects of research universities. *Regional Science and Urban Economics*, 60:198 – 207.
- Bosquet, C. and Overman, H. G. (2019). Why does birthplace matter so much? *Journal of Urban Economics*, 110:26–34.
- Briant, A., Lafourcade, M., and Schmutz, B. (2015). Can tax breaks beat geography? lessons from the french enterprise zone experience. *American Economic Journal: Economic Policy*, 7(2):88–124.
- Brunner, A. and Maurin, L. (2020). Rapport sur la pauvreté en france, édition 2020-2021. Technical report, Observatoire des inégalités.
- Caliendo, L., Monte, F., and Rossi-Hansberg, E. (2015). The Anatomy of French Production Hierarchies. *Journal of Political Economy*, 123(4):809–852.
- Caliendo, L. and Rossi-Hansberg, E. (2012). The impact of trade on organization and productivity. *The Quarterly Journal of Economics*, 127(3):1393–1467.
- Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *The Review of Economics and Statistics*, 95(5):1549–1561.
- Charnoz, P. (2014). Do place-based policies help people ? Evidence from French enterprise zones. Working Papers 2014-44, Center for Research in Economics and Statistics.
- Charnoz, P., Lelarge, C., and Trevien, C. (2018). Communication Costs and the Internal Organisation of Multi-plant Businesses: Evidence from the Impact of the French High-speed Rail. *Economic Journal*, 128(610):949–994.
- Charruau, P. and Epaulard, A. (2017). D’où viennent les écarts de salaire entre les territoires ? La note d’analyse 55, France Stratégie.
- Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98(C):17–49.
- Chetty, R. and Hendren, N. (2018). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates\*. *The Quarterly Journal of Economics*, 133(3):1163–1228.

- Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623.
- Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review*, 86(1):54–70.
- Ciccone, A. and Peri, G. (2006). Identifying Human-Capital Externalities: Theory with Applications. *The Review of Economic Studies*, 73(2):381–412.
- Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.
- Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). Estimating Agglomeration Economies with History, Geology, and Worker Effects. In *Agglomeration Economics*, NBER Chapters, pages 15–66. National Bureau of Economic Research, Inc.
- Combes, P.-P. and Gobillon, L. (2015). Chapter 5 - the empirics of agglomeration economies. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 247–348. Elsevier.
- Combes, P.-P., Gobillon, L., and Lafourcade, M. (2015). Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du Grand Paris. CEPREMAP Working Papers (Docweb) 1504, CEPREMAP.
- Combes, P.-P., Gobillon, L., and Lafourcade, M. (2016). Gains de productivité statiques et d'apprentissage induits par les phénomènes d'agglomération au sein du grand paris. CEPREMAP Working Papers (Docweb) 1602, CEPREMAP.
- Combes, P.-P., Lafourcade, M., Thisse, J.-F., and Toutain, J.-C. (2011). The rise and fall of spatial inequalities in France: A long-run perspective. *Explorations in Economic History*, 48(2):243–271.
- Coquard, B. (2019). *Ceux qui restent : faire sa vie dans les campagnes en déclin*. La Découverte.
- Courlet, C. and Pecqueur, B. (2013). *L'économie territoriale*. Presses universitaires de Grenoble.
- Datar (1998). Développement universitaire et développement territorial. l'impact du plan u 2000 : 1990-1995. Technical report, Délégation à l'aménagement du territoire et à l'action régionale (DATAR).
- Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2022). Matching in Cities. *Journal of the European Economic Association*.
- Davezies, L. (2012). *La crise qui vient : la nouvelle fracture territoriale*. Seuil/La République des idées.
- D'Costa, S. and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, 48:168–179.
- De la Roca, J. and Puga, D. (2017). Learning by working in big cities. *Review of Economic Studies*, 84(1):106–142.

- Delpirou, A. (2018). La couleur des gilets jaunes. *La Vie des idées*.
- Dherbécourt, C. (2015). La géographie de l'ascension sociale français. Document de travail 2015-06, France Stratégie.
- Dherbécourt, C. and Deshard, F. (2019). La répartition territoriale des emplois publics. Document de travail 2019-03, France Stratégie.
- Doeringer, P. and Piore, M. (1971). *Internal Labor Markets and Manpower Analysis*. M. E. Sharpe Incorporated, reprinted 1985.
- Drucker, J. and Goldstein, H. (2007). Assessing the regional economic development impacts of universities: A review of current approaches. *International Regional Science Review*, 30(1):20–46.
- Duport, J.-P. (2008). Aménagement du territoire, enseignement supérieur et recherche : entre proximité et excellence. Technical report, Conseil Economique.
- Duranton, G. and Jayet, H. (2011). Is the division of labour limited by the extent of the market? evidence from french cities. *Journal of Urban Economics*, 69(1):56–71.
- Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 48, pages 2063–2117. Elsevier, 1 edition.
- Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.
- Duranton, G. and Rodríguez-Pose, A. (2005). When economists and geographers collide, or the tale of the lions and the butterflies. *Environment and Planning A*, 37:1695–1705.
- Durieux, S., Eusebio, P., and Lévy, D. (2015). Un million d'emplois liés à la présence de touristes. Insee Première 1555, INSEE.
- Eribon, D. (2009). *Retour à Reims*. Fayard.
- Ernaux, A. (2008). *Les Années*. Gallimard.
- Faggio, G. and Overman, H. (2014). The effect of public sector employment on local labour markets. *Journal of Urban Economics*, 79:91–107. Spatial Dimensions of Labor Markets.
- Felouzis, G. (2001). De l'université à la diversité : l'émergence d'une « université moyenne. In Georges Felouzis, *La condition étudiante. Sociologie des étudiants et de l'Université*, pages 263–274. Presses Universitaires de France.
- Ferréol, G. (2010). Développement universitaire et logiques territoriales : entre cadrage national et arrangements locaux. *Éducation et sociétés*, 25(1):171–189.
- Fielding, A. J. (1989). Inter-regional migration and social change: A study of south east england based upon data from the longitudinal study. *Transactions of the Institute of British Geographers*, 14(1):24–36.
- Fielding, T. (1993). Migration and the metropolis: An empirical and theoretical analysis of inter-regional migration to and from south east england. *Progress in Planning*, 39:75–166.

- Filâtre, D. (2003). Les universités et le territoire : nouveau contexte, nouveaux enjeux. In *Georges Felouzis, Les mutations actuelles de l'Université*, pages 19–45. Presses Universitaires de France.
- Filâtre, D. and Soldano, C. (2012). *Les systèmes régionaux de l'enseignement supérieur en France : disparités et inégalités territoriales*, pages 153–167. Louvain-la-Neuve: De Boeck Supérieur.
- Findlay, A., Mason, C., Houston, D., Mccollum, D., and Harrison, R. (2009). Escalators, elevators and travelers: The occupational mobility of migrants to south-east england. *Journal of Ethnic and Migration Studies*, 35:861–879.
- Finney, M. M. and Kohlhase, J. E. (2008). The effect of urbanization on labor turnover. *Journal of Regional Science*, 48(2):311–328.
- Florida, R., Mellander, C., and Stolarick, K. (2008). Inside the black box of regional development—human capital, the creative class and tolerance. *Journal of Economic Geography*, 8(5):615 – 649.
- Frenette, M. (2009). Do universities benefit local youth? evidence from the creation of new universities. *Economics of Education Review*, 28(3):318–328.
- Galiani, S. and Quistorff, B. (2017). The synth\_runner package: Utilities to automate synthetic control estimation using synth. *Stata Journal*, 17(4):834–849(16).
- Glaeser, E., Kallal, H. D., Scheinkman, J., and Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6):1126–52.
- Glaeser, E. L. and Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–342.
- Glaeser, E. L. and Resseger, M. G. (2010). The complementarity between cities and skills\*. *Journal of Regional Science*, 50(1):221–244.
- Glaeser, E. L., Saiz, A., Burtless, G., and Strange, W. C. (2004). The rise of the skilled city [with comments]. *Brookings-Wharton Papers on Urban Affairs*, pages 47–105.
- Glaeser, E. L., Scheinkman, J., and Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of Monetary Economics*, 36(1):117 – 143.
- Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.
- Gordon, I., Champion, T., and Coombes, M. (2015). Urban escalators and interregional elevators: The difference that location, mobility, and sectoral specialisation make to occupational progression. *Environment and Planning A: Economy and Space*, 47(3):588–606.
- Güvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). What do data on millions of u.s. workers reveal about lifecycle earnings dynamics? *Econometrica*, 89(5):2303–2339.
- Harris, C. D. (1954). The market as a factor in the localization of industry in the united states. *Annals of the Association of American Geographers*, 44(4):315–348.

- Harter, J. and Arora, R. (2010). *The Impact of Time Spent Working and Job Fit on Well-Being Around the World*, pages 398–435.
- Huffman, D. and Quigley, J. M. (2002). The role of the university in attracting high tech entrepreneurship: A silicon valley tale. *The Annals of Regional Science*, 36(3):403–419.
- Huitfeldt, I., Kostøl, A. R., Nimczik, J. S., and Weber, A. (2021). Internal Labor Markets: A Worker Flow Approach. CEPR Discussion Papers 16434, C.E.P.R. Discussion Papers.
- Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–598.
- Jofre-Monseny, J., Silva, J. I., and Vázquez-Grenno, J. (2020). Local labor market effects of public employment. *Regional Science and Urban Economics*, 82:103406. Local public policy evaluation.
- Karady, V. (1986). De napoléon à duruy : les origines et la naissance de l'université contemporaine. In *Jacques Verger, Histoire des Universités en France*, pages 261–322. Bibliothèque historique Privat.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.
- Krugman, P. (1992). *Geography and Trade*, volume 1 of *MIT Press Books*. The MIT Press.
- Lazear, E. and Oyer, P. (2004). The structure of wages and internal mobility. *American Economic Review*, 94(2):212–216.
- Le Moigne, M. (2021). Exploring the "fissured workplace": Fragmentation of internal job ladders and workers' careers. Working paper.
- Lee, J. (2019). The local economic impact of a large research university: Evidence from uc merced. *Economic Inquiry*, 57(1):316–332.
- Lee, J. (2021). The role of a university in cluster formation: Evidence from a national institute of science and technology in korea. *Regional Science and Urban Economics*, 86:103617.
- Liu, S. (2015). Spillovers from universities: Evidence from the land-grant program. *Journal of Urban Economics*, 87:25 – 41.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.
- Manning, A. (2004). We can work it out: The impact of technological change on the demand for low-skill workers. *Scottish Journal of Political Economy*, 51(5):581–608.
- Marshall, A. (1890). *The Principles of Economics*. McMaster University Archive for the History of Economic Thought.
- Marshall, A. (2013). *Industrial Organization, Continued. The Concentration of Specialized Industries in Particular Localities*, pages 222–231. Palgrave Macmillan UK, London.

- Martin, R. (1999). The new 'geographical turn' in economics: some critical reflections. *Cambridge Journal of Economics*, 23(1):65–91.
- Mazzolari, F. and Ragusa, G. (2013). Spillovers from High-Skill Consumption to Low-Skill Labor Markets. *The Review of Economics and Statistics*, 95(1):74–86.
- Melo, P. C., Graham, D. J., and Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342.
- Monso, O. (2006). Changer de groupe social en cours de carrière. davantage de mobilité depuis les années quatre-vingt. Insee Première 1112, INSEE.
- Moretti, E. (2004a). Chapter 51 - human capital externalities in cities. In Henderson, J. V. and Thisse, J.-F., editors, *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2243 – 2291. Elsevier.
- Moretti, E. (2004b). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1):175 – 212.
- Moretti, E. (2004c). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *The American Economic Review*, 94(3):656–690.
- Moretti, E. (2011). Chapter 14 - local labor markets. volume 4 of *Handbook of Labor Economics*, pages 1237–1313. Elsevier.
- Musselin, C. (2012). Brève histoire des universités. In *Frédéric Forest, Les universités en France. Fonctionnement et enjeux*, pages 13–25. Publications des Universités de Rouen et du Havre.
- Newbold, K. B. and Brown, W. M. (2012). Testing and extending the escalator hypothesis: Does the pattern of post-migration income gains in toronto suggest productivity and/or learning effects? *Urban Studies*, 49(15):3447–3465.
- OCDE (2013). *International Migration Outlook 2013*.
- Passeron, J. C. (1986). 1950-1980 : l'université mise à la question : changement de décor ou changement de cap ? In *Jacques Verger, Histoire des Universités en France*, pages 367–418. Bibliothèque historique Privat.
- Poulain, M. (1997). Universités 2000 : quelle université pour demain ? : assises nationales de l'enseignement supérieur, sorbonne, 26-29 juin 1990, paris, ministère de l'Éducation nationale, direction de la programmation et du développement universitaire, paris, la documentation française, 1991.
- Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics*, 34(3):380 – 400.
- Razafindranovona, T. (2017). Malgré la progression de l'emploi qualifié, un quart des personnes se sentent socialement déclassées par rapport à leur père. Insee Références 1659, INSEE.
- Rodriguez-Pose, A. and Tselios, V. (2012). Individual Earnings and Educational Externalities in the European Union. *Regional Studies*, 46(1):39–57.
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5):1002–1037.

- Rosenthal, S. and Strange, W. (2004). Evidence on the nature and sources of agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4, chapter 49, pages 2119–2171. Elsevier, 1 edition.
- Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.
- Senik, C. (2008). Is man doomed to progress? *Journal of Economic Behavior and Organization*, 68(1):140–152.
- Senik, C. (2020). *Bien-être au travail - Ce qui compte*. Presses de Sciences Po.
- Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.
- Shimer, R. and Smith, L. (2000). Assortative matching and search. *Econometrica*, 68(2):343–369.
- Spanos, G. (2019). Firm organization and productivity across locations. *Journal of Urban Economics*, 112:152–168.
- Spiess, C. K. and Wrohlich, K. (2010). Does distance determine who attends a university in germany? *Economics of Education Review*, 29(3):470–479.
- Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.
- Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68(C):53–67.
- Verger, J. et Vulliez, C. (1986). Naissance de l’université. In [Ver \(1986\)](#).
- Walther, O. (2011). Paul krugman : (presque) un nobel de géographie. *Articulo - Journal of Urban Research*.
- Wheeler, C. H. (2006). Cities and the growth of wages among young workers: Evidence from the nlsy. *Journal of Urban Economics*, 60(2):162–184.
- Wheeler, C. H. (2008). Local market scale and the pattern of job changes among young men. *Regional Science and Urban Economics*, 38(2):101 – 118.
- Winters, J. V. (2011). Why are smart cities growing? who moves and who stays\*. *Journal of Regional Science*, 51(2):253–270.

## RÉSUMÉ

---

Cette thèse appréhende les inégalités géographiques de salaire et de promotion sociale entre les différentes zones d'emploi françaises. Elle met l'accent sur la concentration des diplômés du supérieur comme source majeure d'économies d'agglomération dans les marchés du travail locaux.

Le premier chapitre reconsidère l'importance respective de la densité d'emploi et du capital humain local dans l'explication des inégalités géographiques de salaires en France sur la période 2009-2015. L'exploitation de données individuelles de panel fait ressortir des gains d'agglomération reliés davantage aux externalités locales de capital humain, plutôt qu'à des effets de densité pure. Elle montre également que l'expérience professionnelle accumulée par les salariés dans des zones très riches en capital humain leur procure une prime salariale substantielle à moyen terme. Enfin, les inégalités géographiques de salaires semblent davantage provenir de différences de capital humain dans le privé plutôt que dans le secteur public, lequel est mieux réparti sur le territoire.

Le second chapitre porte sur l'influence des marchés du travail locaux sur la mobilité sociale des individus en cours de carrière. En France, sur une période de six ans, les salariés travaillant dans des zones denses ou grandes, et dans des zones riches en capital humain, ont plus de chances d'être promus vers une catégorie socioprofessionnelle supérieure. De nouveau, les estimations suggèrent que les effets d'agglomération sur la mobilité sociale proviennent principalement d'externalités local de capital humain (et de la proximité d'autres marchés denses), plutôt que d'effets exclusifs à la taille ou la densité des zones d'emploi. Par ailleurs, la densité locale augmente les promotions intra et inter-entreprises, l'effet étant encore plus important sur la promotion à l'extérieur de la firme. Enfin, l'expérience accumulée dans les zones les plus denses ou riches en capital humain est transférable et augmente les chances de promotion après une mobilité vers des zones moins denses ou moins riches en capital humain.

Le troisième chapitre se pose la question de la création et du maintien de populations diplômées au niveau local. Il évalue l'impact sur l'accumulation locale de capital humain, de la création des Universités nouvelles dans le cadre du plan U2000, mis en œuvre en France au début des années 1990. En utilisant la méthode des contrôles synthétiques, l'analyse montre que ces créations universitaires ont conduit à une augmentation significative de la concentration locale de diplômés du supérieur long (non scolarisés). En moyenne, l'implantation d'un site universitaire a augmenté la part des diplômés d'environ trois points en 25 ans – ce qui représente 17% du contrefactuel –, avec des effets hétérogènes selon les cas. En outre, l'examen des potentielles implications en matière d'emploi local suggère que, en moyenne, les gains en capital humain se sont accompagnés d'une augmentation de l'emploi qualifié (cadres et professions intellectuelles supérieures).

## MOTS CLÉS

---

Economie géographique, Agglomération, Capital humain, Mobilité sociale, Décentralisation universitaire

## ABSTRACT

---

This thesis examines the geographical inequalities in wages and social promotion across French employment zones. It focuses on the concentration of higher-educated people as a major cause of agglomeration economies in local labor markets.

The first chapter reconsiders the matter of density and human capital in the case of French local labor markets over the recent period 2009-2015. It has three main contributions. The first contribution is to provide a new exploration of the role of local human capital in agglomeration gains in France with respect to the role of pure density. Using a panel of individual administrative data and taking into account potential sorting bias, it shows that agglomeration gains in French local labor markets are much more linked to human capital externalities than pure local density effects. The second contribution is to explore the dynamic effects related to local human capital externalities. It finds that there are substantial learning advantages in zones where the human capital is abundant, which leads to a substantial medium-term wage premium. Finally, a third contribution of this paper is to investigate the role of local human capital externalities according to whether they come from private or public higher-educated workers. It finds that geographic wage inequalities appear to stem more from differences in human capital in the private sector than in the public sector. This is probably more a consequence of public policies aimed at distributing public employment equitably across territories, rather than the "proof" that higher educated workers in the public sector do not generate human capital externalities.

The second chapter examines the influence of local labor markets on individuals' social mobility during their working life. In France, over a period of six years, individuals working in big or dense areas and areas with abundant human capital are more likely to be promoted to a higher socioprofessional status – a promotion associated with a substantial wage increase and/or better working conditions. Applying an empirical strategy to remove spatial sorting bias in a sample of about 350,000 workers in metropolitan France over the period 2009-2015, it finds that density, local human capital, and labor market size significantly increase the likelihood of being promoted, but have no impact on demotions. Again, the results suggest that the agglomeration effects on upward mobility come primarily from human capital externalities (and proximity to other dense markets), rather than pure urbanization and scaled effects. In addition, it shows that local density increases both internal (within the same firm) and external (in another firm) promotions, with the impact of density being even larger on external promotion. Finally, experience accumulated in the densest areas is portable, and increases the chances of promotion after relocation to less dense areas.

The third chapter measures the effects on human-capital accumulation at the local level of the "new universities" created as part of the U2000 Plan implemented in France in the early 1990s. Established in 1990, this national program resulted in the creation of eight universities (spread over 15 sites), including four outside the Paris region (over 10 sites). Using the synthetic control method, it shows that the opening of "new universities" has led to a significant increase in the local share of higher-educated people (not including those in school). On average, the creation of "new universities" increased this share by 4 p.p. within 25 years (about 17% of the counterfactual situation), though the effect differs across cases. Moreover, exploring the employment implications of "new universities" creation, it finds credible evidence that, on average, human-capital gains co-occurred with gains in skilled jobs.

## KEYWORDS

---

Economic geography, Agglomeration, Human capital, Social mobility, Higher-education policy