

Revue de la littérature académique quantifiant les effets d'agglomération sur la productivité et l'emploi

Pierre-Philippe Combes¹ Miren Lafourcade²

31 janvier 2012

Rapport final, réalisé pour la Société du Grand Paris

Mission d'études des éléments de l'évaluation socio-économique
du réseau de transport du Grand-Paris

Lot 3

Calcul des effets économiques élargis de la réalisation du réseau de transport du Grand Paris,
appelés effets d'agglomération, en termes de productivité des entreprises et de PIB
supplémentaire

1. Aix-Marseille School of Economics, 2 Rue de la Charité, 13002 Marseille. Email : ppcombes@univmed.fr.

2. Université Paris-Sud et Paris School of Economics, 48 Bd Jourdan, 75014 Paris. Email : miren.lafourcade@ens.fr.

Table des matières

| | |
|--|-----------|
| Introduction | 2 |
| 1 Les stratégies empiriques d'estimation des effets d'agglomération | 8 |
| 1.1 Effets sur la productivité du travail et les salaires | 8 |
| 1.1.1 La stratégie pragmatique : les formes réduites | 9 |
| 1.1.2 L'approche théorique : les formes structurelles | 14 |
| 1.2 Effets sur la croissance, l'innovation et la localisation des firmes | 18 |
| 1.2.1 Croissance de l'emploi ou du PIB | 18 |
| 1.2.2 R&D, innovation et pôles technologiques | 20 |
| 1.2.3 Choix de localisation des entreprises et modèles logit | 20 |
| 1.3 Biais d'endogénéité dans l'estimation des effets d'agglomération | 23 |
| 1.3.1 Variables omises et causalité circulaire | 23 |
| 1.3.2 Techniques de panel et variables instrumentales | 24 |
| 2 L'ampleur des effets d'agglomération | 28 |
| 2.1 Gains de productivité et de salaires induits par la taille des marchés | 28 |
| 2.1.1 Économies d'agglomération liées à la densité | 29 |
| 2.1.2 Diffusion spatiale des économies d'agglomération liées à la densité | 32 |
| 2.1.3 Ampleur des externalités de localisation et d'urbanisation | 34 |
| 2.1.4 Économies d'agglomération liées au tri spatial des travailleurs | 35 |
| 2.2 Dynamiques de l'emploi et choix de localisation des entreprises | 38 |
| 2.2.1 Emploi total, spécialisation et diversité | 39 |
| 2.2.2 Modèles dynamiques et durée des chocs | 40 |
| 2.2.3 Choix de localisation, densité et accès aux marchés | 42 |
| 2.2.4 Choix de localisation et composition industrielle locale | 43 |
| 2.2.5 Choix de localisation, coûts du travail et entrepreneuriat | 45 |
| 2.3 Technopôles et diffusion spatiale des externalités technologiques | 46 |
| 2.3.1 Production des connaissances et spillovers informationnels aux États-Unis | 47 |
| 2.3.2 Spillovers informationnels et technologiques en Europe | 48 |
| 2.3.3 Proximité géographique, brevets et citations | 49 |
| Conclusion : Quelles leçons pour le réseau du Grand Paris ? | 51 |
| 2.3.4 Les infrastructures de transport : créations ou déplacements d'activités ? | 51 |
| 2.3.5 L'effet du Grand Paris : un scénario... parmi d'autres possibles | 52 |
| Bibliographie | 57 |

Introduction

L'accélération continue de l'urbanisation est un fait stylisé majeur des économies contemporaines, qui a contribué à changer le visage de toute la planète. Si, en 1950, on recensait 86 agglomérations de plus d'un million d'habitants dans le monde, ce chiffre est passé à 430 en 2005. En 1950, il n'existait que deux mégapoles de plus de 10 millions d'habitants (New York et Tokyo). On en dénombre plus de vingt depuis 2003 (Véron (2006)). Selon certaines prévisions, la part de la population urbaine devrait avoisiner 70% en 2050 (Nations Unies (2008)). Si de telles prévisions étaient avérées, le monde devrait alors compter plus de 6,4 milliards de citoyens. Selon le dernier recensement de la population en 2008, l'Ile-de-France, qui recouvre à peine 2% de la superficie de la France métropolitaine, regroupait 19% de sa population et produisait 29% de son Produit Intérieur Brut. La productivité du travail y était aussi plus élevée qu'ailleurs, puisque le salaire net mensuel moyen d'un francilien s'élevait alors à 2565€, soit 28,5% de plus que la moyenne de la Métropole. Ces chiffres ne laissent aucune place au doute : la métropolisation des économies constitue la manifestation la plus visible des avantages procurés par la concentration géographique des activités. Le concept d'"économies d'agglomération" traduit l'ensemble des bénéfices associés à la proximité géographique, au sens large.

Des travaux de recherche récents tentent de mesurer plus précisément les gains engendrés par l'agglomération des activités économiques. Ce rapport a pour objectif de fournir une revue minutieuse des études académiques ayant quantifié l'impact de la concentration des activités économiques sur la productivité des entreprises, les rémunérations perçues en retour par leurs salariés, les choix de localisation des entreprises, la croissance de l'emploi local et la diffusion des connaissances et innovations. En effet, le projet de Métro Grand Paris, du fait de son tracé en rocade, pourrait contribuer à densifier l'activité économique autour de la capitale, tout en fixant une limite naturelle à l'étalement urbain francilien. La revue de littérature proposée vise donc à situer l'importance des gains d'efficacité engendrés par la polarisation des activités économiques attendue du fait de cette future infrastructure.

On distingue généralement quatre familles d'économies d'agglomération, selon qu'elles sont liées à (i) à la diffusion des innovations et connaissances, (ii) à la taille des marchés locaux du travail et à la qualité des appariements salariés/entreprises, (iii) à l'accès des entreprises à leurs sous-traitants, ou (iv) à la taille du marché du bien final.

Les différents types d'économie d'agglomération

- (i) Le regroupement d'entreprises de haute technologie favorise l'émergence de nouvelles connaissances et le développement d'innovations radicales, du simple fait de la proximité de leurs personnels qualifiés, qui communiquent plus aisément que lorsqu'ils sont distants (Saxenian (1994)). De tels "spillovers technologiques" constituent le premier type de gain à la concentration spatiale, et la Silicon Valley en est un des exemples les plus emblématiques.

- (ii) Depuis Adam Smith, il est connu qu'une division plus fine du travail améliore la productivité et, depuis Marshall (1890), qu'un marché du travail plus dense facilite la division du travail. Ainsi, plus une région regroupe des entreprises hétérogènes, plus la demande de travail y est importante et diversifiée, ce qui attire des travailleurs disposant de multiples compétences, contribuant à favoriser un meilleur appariement entre l'offre et la demande de travail. La constitution d'un vaste bassin d'emplois, issu des demandes indépendantes de chaque entreprise, profite ainsi à l'ensemble des producteurs qui vont pouvoir puiser dans ce vivier de travailleurs les qualifications dont ils ont besoin et les savoir-faire indispensables au développement de leur activité. L'appariement entreprises-employés étant meilleurs, ces derniers gagnent également à ce rapprochement et en tire des revenus plus élevés. De plus, la proximité géographique des employeurs et des salariés réduit le risque de chômage et de pénurie de main d'œuvre auxquels ils sont peut être confrontés. Un marché du travail dense joue ainsi aussi un rôle d'assurance mutuelle.
- (iii) La présence, au sein d'une même localisation, d'un grand nombre de sous-traitants y augmente l'offre globale d'inputs, ce qui accroît, pour chaque firme, les chances de trouver à moindre coût (du fait de la concurrence entre sous-traitants et des économies d'échelle dont ceux-ci bénéficient) les facteurs de production dont elle a besoin, tout en réduisant ses coûts de transport ou de recherche.
- (iv) Enfin, les entreprises, qu'elles soient productrices de biens intermédiaires ou de biens finals, lorsqu'elles disposent d'une technologie à rendements croissants, et donc sont plus efficaces à mesure que leur échelle de production augmente, ont intérêt à s'implanter dans les zones les plus denses, dans lesquels les débouchés, plus importants, permettent de réduire leur coût unitaire de production en produisant à plus grande échelle.

Toutefois, la concentration géographique des activités économiques accroît aussi la concurrence sur le marché des produits et du travail, ce qui peut engendrer une hausse du coût des inputs les moins mobiles, travail ou espace, et donc des salaires et rentes foncières, et ainsi réduire le profit des entreprises. Par ailleurs, le regroupement géographique des entreprises d'un même secteur peut rendre la zone d'accueil tributaire des chocs affectant le secteur, comme ce fut le cas du textile et de la sidérurgie dans le Nord de la France. Les transports locaux peuvent également devenir saturés et la pollution peut détériorer la santé des employés. Ces déséconomies d'agglomération, si elles sont suffisamment fortes, peuvent l'emporter sur les économies d'agglomération. L'effet net théorique de ces différentes variables sur la rentabilité des entreprises et le niveau de vie des employés est donc a priori ambigu. Les modèles théoriques anticipent en fait un effet en cloche entre productivité et taille locale ou densité des activités économiques. Les économies d'agglomération commence par dominer, une plus forte densité accroissant la productivité jusqu'à un seuil critique au delà duquel ce sont les déséconomies d'agglomération qui se renforcent le plus et finissent par dominer les gains à la concentration, la productivité baissant alors avec la taille.

Bien que les études empiriques mesurant les effets d'agglomération abondent ¹, les travaux permettant d'isoler séparément l'impact net sur la productivité, les salaires ou la croissance locale de ces quatre différentes sources d'économies d'agglomération sont extrêmement rares.

1. Voir la synthèse exhaustive fournie par Rosenthal et Strange (2004). Le foisonnement des études récentes sur la question des effets d'agglomération rend cependant cette synthèse quelque peu obsolète. C'est pourquoi ce rapport se concentre sur le renouvellement opéré par les travaux publiés sur les huit dernières années.

La revue de littérature que nous proposons a donc pour objectif de recenser les études empiriques qui se sont attachées à quantifier l'impact de variables économiques véhiculant un certain nombre des forces évoquées plus haut, comme la taille de l'économie locale, la présence de tel ou tel type d'entreprises dans la zone ou l'accès aux marchés externes (et donc implicitement les coûts de transport), sans identifier en général lequel des quatre vecteurs mentionnés propage cet impact. Nous présentons cependant dans une dernière section les travaux relatifs aux brevets et à l'innovation qui, eux, permettent de se concentrer sur un des vecteurs des gains à l'agglomération, celui relatif aux retombées technologiques. Sur la période récente, l'économie géographique empirique a de plus connu des avancées méthodologiques majeures, rendues possibles par deux principaux éléments.

Finesse des découpages géographiques, richesse des données individuelles

D'une part, les études mesurant l'impact des phénomènes de concentration spatiale sur la productivité des entreprises et la croissance locale s'appuyaient jusqu'alors sur des données relativement agrégées, à l'échelle par exemple des pays, des états fédéraux américains, ou des régions européennes de niveau NUTS1². Or depuis une dizaine d'années environ, de nouvelles bases de données ont été développées à des échelles géographiques beaucoup plus fines, dont les contours ne reposent plus sur de simples logiques administratives, mais reflètent davantage la réalité des choix de localisation des agents économiques.

Par exemple, certains pays européens ont défini des "marchés du travail locaux", à savoir des zones géographiques à l'intérieur desquelles la plupart des actifs résident et travaillent, et dans lesquelles les entreprises peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts. Ces zones constituent une partition du territoire particulièrement adaptée aux études locales sur le marché du travail, en particulier celles ayant trait à la productivité du travail. En minimisant les flux transfrontaliers de déplacement domicile-travail et les migrations alternantes des salariés, ces découpages reflètent la réalité de fonctionnement des marchés locaux de l'emploi. Ces derniers constituent l'échelle d'étude pertinente pour étudier les phénomènes d'agglomération, en particulier les gains de productivité liés à un meilleur appariement des entreprises et des salariés, à une meilleure diffusion des informations et des innovations technologiques, à l'usage des biens et services public locaux, voire aux relations entreprises/sous-traitants même si ces dernières peuvent également se produire à des niveaux légèrement plus agrégés.

En 1990, l'INSEE a par exemple défini 364 "zones d'emploi" en France, dont 341 sur le continent, nombre qui a été réduit à 322 en 2010. Ce découpage géographique est donc bien plus fin que le principal découpage administratif utilisé pour la France, à savoir les départements, au nombre de seulement 101 dont 94 seulement pour la Métropole continentale. En 1991, l'ISTAT italien a défini de manière similaire 784 "Systèmes locaux du travail" (Sistemi locali del lavoro), nombre réduit à 686 en 2001 (contre seulement 110 provinces de niveau NUTS3). L'office des statistiques britannique a quant à lui créé des zones de navettes domicile-travail (Travel-to-work areas), au nombre de 243 depuis 2007.

2. La Commission européenne a subdivisé les pays de l'espace économique européen en régions appelées "NUTS" (Nomenclature des Unités Territoriales Statistiques). Ces régions ne constituent pas forcément des unités administratives officielles, mais correspondent à des groupements d'unités administratives homogènes fondés sur la population résidente des pays européens. Les régions NUTS1 regroupent de 3 à 7 millions d'habitants. Le niveau NUTS2 correspond en France aux régions administratives, le niveau NUTS3 aux départements.

Ces marchés locaux du travail européens, au sein desquels résident en général environ 75% de la population active de la zone, constituent désormais l'entité spatiale élémentaire de référence pour les politiques de développement local. Les études empiriques doivent donc être conduites prioritairement à cette échelle. De plus, l'utilisation de ces échelles micro-spatiales n'exclut pas pour autant de considérer simultanément les effets d'agglomération opérant à des échelles géographiques plus larges, comme les gains de productivité associés au commerce inter-régional ou international, qu'il est possible de capturer à l'aide de variables telles que l'accès aux marchés, calculées à l'échelle des zones d'emploi.

Encore plus récemment, les bases de données nationales ont été enrichies de données individuelles sur les entreprises et leurs salariés, permettant de suivre sur longue période les effectifs de chaque établissement, les rémunérations horaires de leurs employés et leurs zones de résidence. Ces nouvelles données permettent d'enrichir les spécifications empiriques de variables capturant l'hétérogénéité individuelle des salariés ou des entreprises au sein des marchés locaux du travail. Suivre les salariés au cours du temps permet par exemple d'isoler les gains de productivité liés à la concentration géographique des activités économiques de ceux issus d'une certaine forme d'auto-sélection spatiale des travailleurs les plus qualifiés. En effet, ceux-ci sont plus enclins à résider dans les zones denses. Il est donc crucial de savoir si c'est l'agglomération des activités au sens large qui rend les travailleurs plus productifs, ou bien si ce sont les travailleurs intrinsèquement les plus productifs qui décident de se localiser dans les zones denses, ce qui les fait apparaître comme plus productives, même si ces travailleurs seraient tout aussi productifs s'ils résidaient ailleurs. Un tel tri spatial de la main d'œuvre (appelé aussi "sorting" dans la terminologie anglo-saxonne) est bien sûr potentiellement prévalent en Île-de-France et risque, s'il n'est pas bien pris en compte dans l'analyse empirique, de conduire à une surestimation des gains de productivité associés à la densification des activités économiques. Il est par exemple probable que les connaissances et les innovations se diffusent plus rapidement dans les zones où les travailleurs qualifiés sont surreprésentés. Cela peut se produire, d'une part, parce que ces travailleurs sont plus aptes à émettre et diffuser ces connaissances, mais aussi d'autre part parce qu'ils en sont des récepteurs plus efficaces. Compte tenu de l'extrême concentration des cadres supérieurs et autres travailleurs qualifiés en Île-de-France, évaluer l'impact potentiel de la composition de la main d'œuvre est absolument crucial dans l'optique du Grand Paris, ce que la revue de littérature que nous présentons s'attache à faire.

Sophistication croissante des méthodes d'évaluation

D'autre part, la possibilité de prendre en compte les caractéristiques individuelles des firmes et des salariés est allée de pair avec une sophistication croissante des méthodes économétriques utilisées pour mesurer les économies d'agglomération. Tout d'abord, la mesure des économies d'agglomération est exposée à différents biais d'estimation, appelés aussi "biais d'endogénéité", occasionnés d'une part, par l'existence d'une causalité inverse entre la variable expliquée et les variables explicatives et d'autre part, par l'omission de variables explicatives susceptibles d'affecter la variable expliquée. Par exemple, si la concentration géographique des activités économiques est susceptible d'accroître la productivité des travailleurs, il est également probable qu'une plus grande productivité induise des salaires plus élevés, ce qui attire encore plus fortement de nouveaux salariés dans les zones les plus denses. Ainsi, ce n'est plus seulement la densité des activités qui améliore la productivité mais aussi, à l'inverse, une plus grande productivité qui favorise la concentration géographique des activités. De même, les zones les plus denses peuvent bénéficier

de dotations spécifiques, comme par exemple de grandes infrastructures de transport ou de recherche. Si celles-ci sont omises dans la stratégie d'estimation, on risque d'attribuer leur impact sur la productivité ou les salaires à la densité, alors que celle-ci n'a pas d'effet direct.

La méthode dite des “variables instrumentales” permet de corriger ces biais d'endogénéité, qu'ils proviennent de la causalité inverse ou des variables omises. En termes techniques, les biais d'endogénéité proviennent de la présence d'une corrélation entre le terme d'erreur de la régression de y , la variable que l'on veut expliquer, sur x , la variable explicative, et x . Cette corrélation provient soit du fait que les variables omises, qui rentrent donc dans ce terme d'erreur, sont corrélées à x , soit du fait que y , qui inclut le terme d'erreur, a un aussi un effet inverse sur x . Lorsqu'on cherche à estimer l'impact causal de la variable x , suspectée d'endogénéité, sur la variable y , un “instrument” est une troisième variable z qui a un effet sur y transitant uniquement par l'effet qu'elle a sur x (autrement dit, elle n'est pas corrélée avec le terme d'erreur de la régression de y sur x).

Supposons par exemple que l'on cherche à estimer l'effet causal de la cigarette sur la santé, une corrélation négative entre l'état de santé et le nombre de cigarettes fumées n'implique pas nécessairement que fumer nuit à la santé car d'autres variables potentiellement omises, comme le revenu ou le statut socioprofessionnel, peuvent affecter la santé et la consommation de cigarettes simultanément. La méthode des variables instrumentales consiste à trouver une variable ou “instrument” qui n'affecte la santé qu'à travers la consommation de cigarettes, comme les droits d'accise sur le tabac. Si de telles taxes n'ont un effet sur la santé que parce qu'elles réduisent la consommation de cigarettes (toutes choses égales par ailleurs), une corrélation positive entre l'effet des taxes sur le tabac et l'état de santé signifie que fumer nuit effectivement à la santé et que la relation entre les deux variables est donc bien causale. La littérature académique relative à l'estimation des économies d'agglomération a récemment proposé un certain nombre de stratégies fondées sur l'instrumentation afin de limiter les biais d'endogénéité liés soit à des phénomènes de causalité inverse dus aux choix de localisation endogènes des agents, soit à l'omission de variables caractérisant les zones denses et affectant la productivité.

Les techniques économétriques de panel, rendues possibles par l'utilisation de données individuelles disponibles à diverses dates, permettent de réduire encore plus substantiellement les biais d'endogénéité liés à l'existence de variables omises. L'introduction d'effets fixes individuels permet en effet de capturer l'impact sur la productivité des caractéristiques individuelles des salariés, qu'elles soient observées (expérience, diplôme ou catégorie socioprofessionnelle) ou inobservées (habileté ou talent particulier). L'effet de la densité est alors estimé net de ces effets individuels. Notre revue de littérature permet d'évaluer dans quelle mesure ces biais d'endogénéité peuvent conduire à des évaluations potentiellement erronées des gains de productivité induits par l'agglomération spatiale des activités.

Ensuite, les équations estimées ont pris une forme de plus en plus “structurelle”, i.e. plus strictement conforme aux spécifications théoriques prédites par exemple, par les modèles d'économie géographique. Cela permet tout à la fois de fournir une interprétation économique des relations causales estimées, et d'atténuer les biais de variables omises en incorporant l'ensemble des variables dictées par la théorie. Cette tendance a conduit à une dichotomie très nette des études empiriques cherchant à mesurer les effets d'agglomération.

D'un côté, la majorité des études estiment des “formes réduites”, i.e. incluant toutes les variables potentiellement pertinentes pour expliquer les gains de productivité, sans qu'il soit pour autant possible d'identifier précisément tous les mécanismes théoriques sous-jacents. Ces études sont sans conteste les

plus utilisées par les pouvoirs publics, car elles ont l'avantage d'inclure dans une même spécification toutes les variables d'intérêt. Toutefois, de telles formes réduites peuvent soulever de graves problèmes d'interprétation. Si certaines variables explicatives sont interdépendantes, une forme réduite ne permet pas d'évaluer correctement l'effet de chacune des variables prise séparément, dont l'impact peut être phagocyté par les variables qui lui sont corrélées. De plus, les formes réduites incluent généralement un grand nombre de variables explicatives, potentiellement endogènes, pour lesquelles il risque d'être difficile de trouver les instruments adéquats.

De l'autre côté, un petit nombre d'études suivent une toute autre route. Elles commencent par développer un modèle théorique permettant d'expliquer le niveau de la productivité ou des salaires, et déduisent ensuite les spécifications adéquates susceptibles d'être estimées. Il est dans ce cadre possible de savoir quelles sont les variables exogènes et endogènes, et d'identifier précisément les mécanismes économiques les reliant à la variable d'intérêt. En contrepartie, ces formes structurelles sont cependant plus opaques pour les pouvoirs publics car elles incorporent un plus petit nombre de variables explicatives, dont l'interprétation n'est pas toujours immédiate ou intuitive, et pas nécessairement la variable d'intérêt dont ils aimeraient mesurer l'impact (qui est souvent interagie avec d'autres variables), et pour laquelle ils ne peuvent disposer que d'une mesure partielle.

Plan du rapport

La revue de littérature que nous proposons se veut la plus exhaustive possible pour ce qui est des études récentes et du cadre européen. Le Chapitre 1 présente les différentes stratégies d'estimation utilisées pour mesurer les effets d'agglomération induits par la polarisation des activités économiques et nous invite à réfléchir aux problèmes théoriques et méthodologiques associés à l'interprétation de ces effets. Le Chapitre 2 se focalise sur le résultat des estimations réalisées à partir des stratégies présentées dans le Chapitre 1. Le rapport conclut par des recommandations de politique économique destinées à éclairer le débat public sur les effets économiques attendus du réseau de transport du Grand Paris, en matière d'économies d'agglomération.

Chapitre 1

Les stratégies empiriques d'estimation des effets d'agglomération

La relation ayant suscité le plus d'intérêt de la part de la recherche académique est celle reliant la productivité des entreprises ou le salaire de leurs employés à différentes mesures de la concentration géographique des activités. La Section 1.1 de ce chapitre se concentre donc tout d'abord sur cette question, en détaillant successivement les formes réduites et structurelles qui ont été estimées. La Section 1.2 répertorie les stratégies qui ont été implémentées pour estimer les dynamiques locales associées à ces économies d'agglomération, en distinguant les effets en termes d'emplois, d'innovation technologique et d'attractivité pour les entreprises. Quant à la Section 1.3, elle s'attarde sur les biais d'endogénéité potentiels associés à la mesure des effets d'agglomération, et sur les techniques économétriques permettant de les corriger.

1.1 Effets sur la productivité du travail et les salaires

Les effets d'agglomération, qu'ils soient mesurés en termes de productivité des entreprises ou de salaires, peuvent l'être de deux façons. La majorité des études s'appuie sur l'estimation de formes réduites, qui incluent tous les variables susceptibles d'influencer la productivité ou les salaires, en ignorant souvent la possibilité que ces variables puissent être multicolinéaires ou endogènes. Une poignée d'études, pour la plupart récentes, adopte cependant une approche moins pragmatique et plus structurelle. Ces études ne considèrent que des variables explicitement requises par le modèle théorique sous-jacent, dont l'interprétation est rendue plus complexe par l'imbrication non linéaire des relations d'équilibre du modèle. Dans le cadre d'une approche structurelle, le modèle choisi est bien évidemment supposé pertinent pour expliquer le phénomène étudié. Or certaines hypothèses théoriques peuvent parfois apparaître excessives, voire héroïques pour ceux qui privilégient les considérations normatives. Cela explique sans doute pourquoi les approches structurelles restent encore rares et, lorsqu'elles existent, rencontrent moins de succès que les formes réduites auprès des décideurs publics. Nous détaillons successivement ces deux approches.

1.1.1 La stratégie pragmatique : les formes réduites

La stratégie d'estimation la plus pragmatique adoptée pour quantifier les gains de productivité ou de salaire associés à une plus grande concentration géographique des activités est la forme réduite. Si les méthodes d'estimation diffèrent peu d'une étude à l'autre, les spécifications estimées n'en sont pas moins très diverses. Les plus "économiques" se concentrent sur les effets d'agglomération véhiculés par la taille des marchés, alors que les plus "prolixes" tentent de dissocier l'impact de la taille et les effets de composition associés à la présence de différents secteurs ou différentes catégories de travailleurs.

Densité locale d'activités et externalités d'agglomération

C'est la densité des activités économiques qui s'est rapidement imposée comme déterminant de la productivité et des salaires, plutôt que la taille globale du marché. Cette dernière est en effet bien plus sensible au découpage spatial utilisé, contrairement à la densité, qui est une variable plus continue. L'Île-de-France en est d'ailleurs un bel exemple. La population active occupée varie très fortement d'une commune à l'autre, la région francilienne étant constituée d'un patchwork de petites et de grandes communes juxtaposées. À l'inverse, la densité des emplois francilien présente une évolution spatiale plus continue, ce qui rend l'analyse plus pertinente, tant du point de vue économique qu'économétrique.

Le premier article à s'être intéressé à la question de l'impact de la densité sur la productivité des entreprises est celui de Ciccone et Hall (1996), dans le cadre des comtés américains. À leur suite, de nombreux travaux ont estimé une relation log-linéaire du type :

$$\ln prod_z = \alpha + \beta \ln dens_z + \varepsilon_z, \quad (1.1)$$

où $\ln prod_z$ est le logarithme de la productivité du travail moyenne dans la zone z , $dens_z$ une mesure de la densité des activités au sein de cette zone, généralement le nombre d'actifs occupés par km^2 de surface et ε_z un terme d'erreur représentant tout ce que la densité ne parvient pas à expliquer. L'expression en logarithme permet d'estimer directement l'élasticité de la variable dépendante à la variable explicative, notée β , c'est-à-dire la variation en % de la productivité engendrée par une hausse de 1% de la densité¹. La spécification (1.1) offre donc la possibilité d'évaluer la part des écarts de productivité imputable aux écarts de densité entre les zones, ou d'estimer l'impact potentiel sur la productivité d'une hausse ou d'une baisse de la densité au sein d'une zone.

L'équation (1.1) inclut aussi parfois une seconde variable de taille, permettant d'apprécier le fait qu'une "grande" zone peut être tout simplement plus étendue, et pas nécessairement plus dense. Les deux variables, superficie et densité, peuvent influencer simultanément la productivité, mais pour le décideur politique, modifier la première - via un nouveau plan d'occupation des sols, par exemple - ou la seconde - via la création d'une zone franche urbaine, par exemple - relèvent de deux stratégies très différentes. Il est donc important d'identifier séparément le rôle joué par ces deux déterminants. Certains auteurs préfèrent néanmoins considérer l'emploi total (dans une spécification log-linéaire, cela revient au même que d'inclure simultanément la superficie et la densité), plutôt que les deux variables séparées. Il est alors

1. Une autre manière d'interpréter le coefficient β consiste à dire qu'un doublement de la densité multiplie la productivité par 2^β , et donc engendre une hausse de la productivité de $(2^\beta - 1) \times 100\%$. En effet, si on compare deux zones i et j telles que $dens_i = 2 \times dens_j$, la différence de productivité associée est telle que $\log(prod_i/prod_j) = \beta \log(dens_i/dens_j) = \beta \log 2$, ce qui implique bien que $prod_i/prod_j = 2^\beta$.

plus difficile d'interpréter l'effet de cette variable sur la productivité, car elle capture à la fois l'épaisseur et l'étendue de la zone étudiée.

Malgré sa nature très réduite, la spécification (1.1) n'en repose pas moins sur des fondements théoriques simples, qui ont été détaillés par Combes *et al.* (2008a). Supposons que, quels que soient la région et le secteur considérés, les entreprises produisent à coûts marginaux constants², lesquels coûts intègrent les salaires et le coût des inputs (primaires ou intermédiaires) autres que la main d'œuvre. Supposons également que la technologie et l'efficacité du travail soient spécifiques à chaque entreprise. Le profit de l'entreprise est la somme des ventes qu'elles réalise sur les marchés, nette du coût des facteurs de production et des coûts aux échanges (dont le transport) occasionnés par l'approvisionnement des marchés extérieurs. Les ventes de l'entreprise correspondent au produit de son volume de production et du bénéfice de la vente par unité produite, qui peut être appelé, un peu abusivement, le "prix" du produit (pour l'entreprise). Ce prix représente en fait la moyenne du bénéfice dégagé par l'entreprise sur chaque unité vendue sur les marchés de destination, net des coûts de transport et des consommations intermédiaires. Ce "prix" est plus élevé quand l'entreprise est sur (ou proche d') un grand marché car cette proximité réduit ses coûts d'approvisionnement. À l'inverse, ce "prix" est plus faible lorsque la concurrence est forte, comme sur les grands marchés. Cet effet contrecarre le premier, et agit comme un facteur de dispersion. Même en faisant l'hypothèse d'une concurrence parfaite sur les marchés des biens et des inputs, il est possible de montrer que la condition de premier ordre de l'utilisation optimale du travail par les entreprises, à savoir l'égalisation du salaire nominal de chaque travailleur à sa productivité marginale, conduit à une équation de salaire permettant d'estimer les effets d'agglomération. Le salaire nominal y est directement proportionnel aux compétences individuelles du travailleur, il croît avec le niveau technologique de l'entreprise et avec son "prix" - tel que nous venons de le définir - et décroît avec le coût des inputs autres que le travail.

La densité d'une zone peut donc accroître la productivité des entreprises qui s'y trouvent, soit parce qu'elle améliore leur technologie ou l'efficacité de leur main d'œuvre, soit parce qu'elle accroît le prix de leurs biens, soit encore parce qu'elle réduit le prix de leurs inputs, du fait de l'économie de coûts de transport engendrée par la plus grande proximité des acheteurs (consommateurs ou sous-traitants). À l'inverse, la densité peut aussi décroître la productivité des entreprises en créant les conditions propices à l'apparition de phénomènes de congestion sur le marché de certains inputs, qui en accroissent les coûts. On peut illustrer ce dernier phénomène par la saturation des réseaux de transport en commun à Paris, qui tend à accroître la durée des navettes domicile-travail de la main d'œuvre, ou par la hausse des prix de l'immobilier en Île-de-France, dans un contexte où l'offre foncière locale présente une grande rigidité. Le coefficient β de l'équation (1.1) ne représente que le gain de productivité "net" associé à la densité des activités économiques, puisque les effets positifs de l'agglomération sont en partie compensés par ses effets négatifs.

Il est important de noter que les travaux de Moretti (2004a) et de Combes *et al.* (2008a) insistent sur le fait que, lorsqu'on cherche à évaluer les effets d'agglomération en termes de salaires, et plus seulement de productivité, c'est le salaire *nominal*, et non le salaire réel, qui doit constituer la variable dépendante de l'équation (1.1). Il n'y a aucune raison de déduire des salaires nominaux le coût de la vie locale (en particulier les coûts du foncier ou du logement), ni même d'introduire ces éléments dans la spécification. Ce qui doit être estimé est une équation reliant la productivité du travail, donc son rendement nominal,

2. Cette hypothèse représente une simplification considérable en comparaison des modèles d'économie géographique "standard", qui s'appuie le plus souvent sur la concurrence imparfaite, comme nous le verrons plus loin.

et non une équation dans laquelle les migrations des travailleurs seraient conditionnées par les différences de revenu réel entre les zones.

Si l'estimation de l'impact net global de la densité sur la productivité du travail et les salaires est certainement très utile pour la définition des politiques économiques locales, il faut toutefois rester prudent dans les interprétations, et ne pas fonder les décisions de politique économique sur l'estimation de cette seule forme réduite. Il existe de sérieuses limites à l'utilisation d'une spécification aussi simple, et des précautions méthodologiques additionnelles s'avèrent nécessaires, comme nous le montrerons plus loin.

Coûts de transport et accès aux marchés

Les gains d'efficacité ne proviennent cependant pas uniquement de la concentration *locale* des activités. Les marchés externes peuvent aussi avoir un impact positif sur la productivité du travail, comme nous venons de le voir, mais la magnitude de cet impact est moindre du fait de leur éloignement géographique. Formellement, cette notion de "centralité" peut être appréhendée à l'aide d'une variable que le géographe Harris (1954) appelle le "potentiel marchand", et que l'on peut calculer comme la somme des activités présentes sur les marchés, escomptées du coût de transport permettant d'y accéder. C'est ici la mobilité pendulaire des salariés entre les zones qui véhicule les effets d'agglomération. L'intégration du potentiel marchand dans la spécification (1.1) permet de mettre en évidence le fait que, lorsque les biens sont échangeables, ce n'est pas seulement la taille de l'économie locale qui importe, mais également la taille des marchés externes, qui peuvent être proches, auquel cas ils ont un poids plus important, ou bien éloignés. En d'autres termes, dans un cadre où les biens et/ou les personnes sont mobiles entre les zones, il n'y a pas de raison de penser que les effets d'agglomération véhiculés par la densité ou la superficie ne débordent pas des limites spatiales de la zone.

Le potentiel marchand le plus souvent considéré est le potentiel marchand *externe* à la zone étudiée, car la densité et la superficie locales capturent déjà les économies d'agglomération sur le marché local. La spécification estimée devient donc :

$$\ln prod_z = \alpha + \beta \ln dens_z + \delta \ln sup_z + \eta \ln PM_z + \varepsilon_z, \quad (1.2)$$

où sup_z est la superficie de la zone et $PM_z = \sum_{j \neq z} \frac{dens_j}{dist_{jz}}$ son potentiel marchand, calculé ici comme la densité associée aux marchés externes, divisée par la distance permettant d'y accéder, qui constitue une bonne approximation en coupe des coûts de transport (Combes et Lafourcade (2005)).

Le potentiel marchand élargit l'éventail des instruments de politique économique à la disposition des décideurs publics, puisqu'il permet de capturer et d'estimer séparément les effets d'agglomération véhiculés par les marchés externes (nationaux ou étrangers) de ceux véhiculés par le marché local. Certains auteurs vont même plus loin et évaluent l'intensité avec laquelle ces effets de débordement se propagent dans l'espace, en calculant des potentiels marchands associés à différents seuils de distance de la zone. Ils intègrent par exemple dans la spécification (1.2), la densité de la zone, la densité des zones situées dans un rayon inférieur à 50 kilomètres, la densité des zones situées dans un rayon de 50 à 100 kilomètres, et ainsi de suite.

La productivité moyenne d'une zone masque cependant souvent une grande hétérogénéité entre les secteurs. En conséquence, certains auteurs utilisent comme variable dépendante des équations (1.1) et (1.2) la productivité moyenne d'un secteur donné, et ils empilent ensuite les secteurs pour réaliser l'estimation.

Deux types de variables explicatives additionnelles peuvent alors être intégrées dans la régression : des variables mesurant la composition sectorielle de la zone, mais ne dépendant pas du secteur, comme la diversité de son tissu industriel, mais aussi des caractéristiques locales du secteur d'activité étudié, comme son degré de concentration dans la zone, ou la taille moyenne de ses entreprises, qui constitue une mesure indirecte des économies d'échelle et une mesure directe du degré de concurrence entre les firmes.

Diversité sectorielle et externalités d'urbanisation

La densité locale des emplois, la superficie et les potentiels marchands ne dépendent pas du secteur et s'inscrivent dans un ensemble plus large de caractéristiques véhiculant ce que l'on peut appeler des externalités "inter-sectorielles" ou d'"urbanisation". Il est également courant d'intégrer dans la spécification (1.2) une variable permettant d'appréhender la diversité sectorielle de la zone. Cette diversité est généralement mesurée à l'aide d'indicateurs de concentration sectorielle, comme les indices de Gini ou d'Herfindahl, calculés sur la base du pourcentage représenté par chaque secteur dans la population active occupée de la zone.

La diversité englobe deux types très différents de forces d'agglomération. L'urbaniste Jacobs (1969) a été l'une des premières à défendre l'idée qu'une fertilisation croisée des idées entre les secteurs d'activité pouvait jouer un rôle moteur dans les processus d'innovation : des innovations radicales développées dans certains secteurs peuvent être empruntées par d'autres et en améliorer sensiblement la productivité. Les zones plus diverses en termes de tissu industriel bénéficieraient ainsi d'une plus forte croissance. De plus, la diversité sectorielle permet aux marchés de mieux résister aux chocs négatifs frappant des activités spécifiques, en permettant aux emplois de se déverser vers d'autres secteurs. Il est par ailleurs bien connu que la préférence pour la diversité, modélisée en concurrence monopolistique sous la forme d'un grand nombre de variété différenciées intégrées dans les fonctions d'utilité ou de production, accroît les incitations des agents économiques à se concentrer dans l'espace géographique, comme l'a montré par exemple le modèle séminal de Krugman (1991).

Toutefois, il ne faut pas non plus oublier qu'une plus forte urbanisation est aussi propice à l'accroissement de la concurrence pour l'usage du sol et donc, à la hausse de la rente foncière, qui provoque l'étalement urbain et l'intensification des navettes domicile-travail, avec ses corollaires que sont la congestion des réseaux de transport et la croissance des émissions de gaz à effet de serre. L'impact net attendu des externalités d'urbanisation, qui engendrent aussi des forces de dispersion, est donc en théorie ambigu.

Clusters d'activités et externalités de spécialisation

Lorsque des entreprises d'un même secteur se regroupent géographiquement, la diffusion des connaissances spécifiques à ce secteur s'en trouve renforcée, et les innovations du secteur sont susceptibles d'être plus rapidement imitées et adoptées. Ce n'est alors plus la densité de l'économie locale qui induit des gains de productivité, mais plus spécifiquement celle du secteur considéré. Ces externalités dites de "localisation" ou de "spécialisation" sont aussi plus communément appelées effets MAR (Marshall-Arrow-Romer), en hommage aux premiers économistes ayant modélisé ou souligné leur existence.

La variable la plus souvent utilisée pour capturer ces externalités "intra-sectorielles" est le degré de spécialisation de la zone dans le secteur d'activité considéré, que l'on peut mesurer par la part du secteur dans l'économie locale, comme l'ont fait par exemple Glaeser *et al.* (1992). La taille moyenne des

entreprises locales, qui peut refléter l'importance des économies d'échelle internes à l'entreprise, est aussi parfois utilisée. Une autre variable possible est la part des professions dites "qualifiées" dans l'emploi local du secteur, qui permet d'évaluer le rôle positif indirect joué par ces personnes sur le niveau technologique local. Un examen relativement complet de ces variables intra-sectorielles est proposé dans Rosenthal et Strange (2004).

Ici encore, il fait bien garder à l'esprit que la concentration géographique d'un secteur d'activité peut tout aussi bien intensifier la concurrence entre les entreprises du secteur. Par ailleurs, les clusters peuvent aussi avoir des effets négatifs lorsqu'on prend en compte le risque que fait courir une trop grande spécialisation face à un choc sectoriel, en particulier lorsque les travailleurs sont peu mobiles, comme c'est le cas en France. Ces déséconomies d'agglomération, si elles sont suffisamment fortes, peuvent l'emporter sur les économies d'agglomération. L'effet net théorique est donc ici encore ambigu.

Tri spatial de la main d'œuvre

Les connaissances et les innovations se diffusent plus rapidement dans les zones où les travailleurs qualifiés sont surreprésentés. Or les salariés les plus qualifiés (diplômés, travailleurs expérimentés ou disposant de compétences spécifiques), qui sont aussi les plus productifs, sont plus enclins à résider dans les zones urbaines les plus denses. Une partie des gains de productivité liés à la densité provient donc d'une certaine forme d'auto-sélection spatiale des travailleurs, comme l'avait d'ailleurs souligné dès le départ l'étude originale de Ciccone et Hall (1996).

Dans la productivité locale, il est donc important d'essayer de dissocier l'effet des externalités locales pures, des gains de productivité engendrés par le rendement intrinsèquement plus élevé de certaines professions. S'il existe d'importantes disparités spatiales dans la répartition de ces professions, et que cette disparité n'est pas prise en compte dans l'analyse économétrique, certaines zones peuvent apparaître comme plus productives, alors qu'elles ne bénéficient en réalité d'aucune externalité locale, parce qu'elles accueillent une plus grande part de travailleurs qualifiés, et qui le seraient tout autant dans les autres zones.

S'il est intéressant de comprendre pourquoi la répartition spatiale des compétences peut être inégale, une question sur laquelle nous reviendrons plus loin, les études disponibles proposent des stratégies permettant d'identifier séparément cet effet de composition locale, des économies d'agglomération induites par les autres déterminants de la productivité du travail. Comme l'indiquent Hellerstein *et al.* (1999), des mesures de la distribution des niveaux d'éducation et des métiers recensés au sein de la zone doivent être intégrées dans les spécifications permettant d'estimer les effets d'agglomération. L'équation (1.2) devient alors :

$$\ln prod_{zs} = \alpha + \beta \ln dens_z + \delta \ln sup_z + \eta \ln PM_z + \lambda \ln div_z + \theta \ln spe_{zs} + \mu HC_z + \varepsilon_{zs}, \quad (1.3)$$

où $spe_{zs} = emp_{zs}/emp_z$ est la part du secteur s dans l'activité totale de la zone, $div_z = 1/\sum_s (spe_{zs})^2$, une mesure de la diversité de son tissu industriel et HC_z une mesure de son niveau moyen d'éducation.

Lorsque des données individuelles sont disponibles, on peut même espérer distinguer l'effet des compétences de l'individu et de l'externalité globale qu'il exerce sur la productivité des autres travailleurs de la zone.

La spécification (1.3) devient alors :

$$\ln prod_{it} = \alpha + \beta \ln dens_{z(i)t} + \delta \ln sup_{z(i)} + \eta \ln PM_{z(i)t} + \lambda \ln div_{z(i)t} + \theta \ln spe_{z(i)s(i)t} + \mu HC_{z(i)} + C_{it}\theta + \gamma_i + \varphi_t + \varepsilon_{it}, \quad (1.4)$$

où $prod_{it}$ est la productivité de l'individu i salarié dans la zone z et le secteur s à la date t , γ_i un effet de compétences propre à l'individu i invariant au cours du temps (comme le sexe), C_{it} un vecteur de caractéristiques individuelles variables au cours du temps (comme l'âge ou l'expérience), et φ_t une indicatrice temporelle permettant d'appréhender la part de la productivité affectée par des chocs conjoncturels exogènes à l'individu.

Si l'on travaille sur données agrégées, seul le résultat combiné des compétences individuelles et de l'externalité de capital humain peut être identifié, via le coefficient μ associé à l'équation (1.3).

Dotations factorielles locales

Enfin, les études insistent aussi sur le rôle joué par les dotations factorielles. Les dotations en biens publics qui relèvent de la compétence des décideurs politiques (centres de recherche, universités, infrastructures de transport), les avantages ou désavantages de "première nature" (accès à la mer, climat, enclavement, ressources naturelles), la qualité des institutions locales ou le niveau des technologies disponibles, ont aussi une incidence sur la productivité des entreprises et des travailleurs. La gouvernance et le progrès technique sont certainement plus importants dans le contexte des pays émergents (comme la Chine), des pays marqués par leur histoire coloniale (comme la Côte d'Ivoire ou l'Inde), ou ayant du effectuer une difficile transition vers l'économie de marché (comme les pays d'Europe Centrale ou de l'Est). Elles sont probablement moins pertinentes pour d'autres régions du monde, comme l'Europe occidentale, dont les régions sont plus homogènes de ce point de vue. En revanche, pour ces dernières, les dotations en biens publics locaux sont souvent très disparates.

Comme nous le verrons plus loin, si l'effet de chacune de ces variables de dotations reste très difficile à identifier, les effets d'agglomération peuvent être purgés de leur impact global. Dans ce cas, ce qui est estimé n'est pas le fait que les zones les plus denses soient en moyenne mieux dotées en infrastructures publiques, mais bien l'impact des variables d'agglomération (den_{zt} , PM_{zt} , div_{zt} , spe_{zst} et HC_{zt}) sur la variable d'intérêt, à *répartition spatiale donnée des biens publics locaux*.

1.1.2 L'approche théorique : les formes structurelles

Nous avons souligné que les formes réduites ne permettaient pas d'exploiter toute la richesse des modèles d'économie géographique. Par exemple, elles ne rendent compte que très partiellement du rôle, pourtant très important, joué par les barrières aux échanges. La seule variable permettant de les appréhender est en effet le potentiel marchand, mais les coûts aux échanges y sont supposés inversement proportionnels à la distance, dans le cadre d'une fonction complètement *ad hoc*. Or les barrières aux échanges intègrent de multiples composantes qui ne sont pas proportionnelles à la distance. Ces éléments qui peuvent être catégorisés en quatre types de coûts, dénommés par Spulber (2007) les "4T", englobent :

- (i) les coûts de Transaction, incluant l'ensemble des frais associés à la préparation et à la réalisation d'un échange marchand (dépenses visant à compenser un handicap informationnel, comme la méconnaissance de la langue, des pratiques culturelles ou managériales, des cadres législatif et réglementaire du pays

avec lequel on réalise l'échange), mais aussi les frais de publicité, de marketing ou les coûts du change,

- (ii) les barrières Tarifaires et non Tarifaires (droits de douane, quotas, restrictions volontaires aux échanges et autres obstacles techniques au commerce, comme les normes sanitaires),
- (iii) les coûts de Transport, qui recouvrent les dépenses de fret liées à l'acheminement des marchandises, les coûts logistiques (gestion des flux, des stocks et des entrepôts) ou les dépenses associées au déplacement des personnes,
- (iv) la valeur du Temps, incluant les délais de livraison, les procédures de dédouanement aux frontières ou l'obsolescence de plus en plus rapide des produits.

Une des contributions les plus importantes des modèles d'économie géographique est de montrer qu'en équilibre général, les coûts aux échanges affectent la plupart des variables endogènes : le rendement des facteurs de production, mais aussi les quantités produites ou achetées, et le prix des biens (finaux et intermédiaires). Par conséquent, dans les exercices empiriques, presque toutes les variables explicatives dépendent des barrières aux échanges. Il est donc difficile de démêler l'écheveau de tous ces effets directs et indirects sans l'appui d'un modèle théorique.

Pour cette raison, les formes dites "structurelles" préfèrent s'appuyer sur les relations d'équilibre prédites par des modèles théoriques parfois complexes, plutôt que de juxtaposer des variables de bon sens, certes plus faciles à interpréter, mais dont il est difficile d'identifier l'effet réel. Si l'approche structurelle permet d'ouvrir l'éventail des effets d'agglomération induits par telle ou telle politique économique, elle le fait au prix d'une réduction drastique du nombre de politiques pouvant être étudiées simultanément. Il convient ici pour les décideurs publics d'arbitrer entre la compréhension globale, mais très approximative, des mécanismes d'ensemble reliant la productivité à la distribution spatiale des activités, et la compréhension fine de l'effet induit par la modification d'un paramètre, malheureusement très spécifique, du modèle.

Les formes structurelles permettant de lier la productivité, ou plus exactement les salaires, à la concentration géographique des activités, sont le plus souvent issues des modèles de concurrence monopolistique à la Dixit-Stiglitz (Dixit et Stiglitz (1977)), comme par exemple les modèles fondateurs de Krugman (1991) et de Krugman et Venables (1995). Comme ces modèles ne considèrent généralement qu'un petit nombre de secteurs et de régions (en général deux), la première étape pour les économètres consiste à les étendre à un cadre multirégional et multisectoriel, afin d'en accroître le réalisme. Dans les modèles de concurrence monopolistique, les biens sont différenciés en un grand nombre de variétés, qui sont toutes demandées par les consommateurs. En ayant accès à toute la gamme des variétés offertes sur les marchés, les acheteurs ont en effet la possibilité de se rapprocher de leur variété "idéale", ce qui contribue à accroître leur satisfaction. Ils ont donc une préférence pour la diversité. Les entreprises de certains secteurs, comme l'industrie manufacturière, disposent d'une technologie à rendements croissants : le lancement de la production requiert un investissement initial ou le paiement d'un coût fixe, si bien que l'augmentation des quantités produites permet de réduire le coût unitaire de la production. Ces économies d'échelle donnent un avantage concurrentiel aux entreprises implantées sur les grands marchés. Le pouvoir de marché des entreprises est lié au fait que chaque variété n'est produite que par une seule entreprise (c'est le côté "monopolistique" de la concurrence). Il permet de couvrir les coûts fixes de production, mais ne constitue pas une pression concurrentielle en soi. Chaque entreprise ne peut en effet influencer le prix de ses

concurrentes : autrement dit, il n’y a pas d’interactions stratégiques entre les firmes. Dans ce contexte, l’entrée d’une nouvelle entreprise sur le marché ne se traduit pas par la compression des marges extraites par ses concurrentes, mais par la réduction mécanique de leur part de marché : les consommateurs vont, du fait de leur goût pour la diversité, acheter la nouvelle variété et, à budget constant, réduire le montant des ressources consacrées à l’achat des anciennes variétés. Tant que les profits sont positifs, de nouvelles entreprises entrent sur le marché. Le processus s’arrête lorsque l’entrée d’une firme supplémentaire annule le profit de toutes ses concurrentes (c’est le côté “concurrentiel” de la concurrence monopolistique). En sus de la main d’œuvre, la fonction de production des entreprises peut inclure d’autres inputs, comme des biens intermédiaires composites produits à partir des différentes variétés fabriquées sur les marchés. L’échange des variétés entre les régions ou les exportations sont soumises au paiement d’un coût dont le montant représente généralement une certaine fraction du prix sortie usine de chaque variété, hypothèse dite de l’“iceberg” (Samuelson (1954)).

Dans ce type de modèle³, les inégalités de salaires sont le résultat d’un arbitrage fondamental effectué par les entreprises disposant de la technologie à rendements croissants. Parce qu’elles supportent d’importants coûts fixes, ces firmes sont incitées à concentrer leur production dans un petit nombre d’établissements pour tirer partie de leurs économies d’échelle. Cette concentration a cependant un coût en contrepartie : les entreprises doivent payer l’acheminement des marchandises vers les acheteurs finaux, qui peuvent être très éloignés des sites de production. Lorsque les barrières aux échanges sont élevées, les entreprises préfèrent multiplier les sites de production, en dépit des déséconomies d’échelle engendrées par cette fragmentation spatiale, pour se rapprocher des acheteurs. En revanche, si les barrières aux échanges sont faibles, les entreprises arbitrent en faveur de la concentration, et choisissent de s’implanter sur les grands marchés - par exemple les grandes capitales régionales - pour avoir accès à un plus grand nombre d’acheteurs. Les petits marchés, qui peuvent être facilement approvisionnés en raison du faible coût des exportations, perdent donc une partie de leur industrie.

Si les consommateurs sont mobiles, en particulier la main d’œuvre qualifiée, la concentration géographique des secteurs à rendements croissants va de pair avec celle de l’emploi et des services afférents. Cette polarisation des emplois accroît en retour la demande sur les grands marchés, ce qui les rend encore plus attractifs pour de nouvelles entreprises. S’enclenche ainsi un processus auto-entretenu d’agglomération des activités à rendements croissants, de leurs sous-traitants et de la population. Les secteurs à rendements décroissants, comme l’agriculture, restent en revanche dispersés, avec une localisation conditionnée par l’accès aux matières premières et aux ressources naturelles. Toutefois, si les barrières aux échanges continuent à diminuer, le processus peut s’inverser. En effet, comme nous l’avons déjà précisé, l’agglomération des activités finit par engendrer des phénomènes de congestion sur les marchés : les salaires nominaux des grandes régions augmentent pour y attirer la main d’œuvre nécessaire, de même que les rentes foncières, ce qui finit par grever le profit des entreprises et rend ces marchés moins attractifs. On peut donc assister *in fine* à une re-dispersion géographique partielle de certaines activités au profit des petites régions initialement désavantagées.

Deux principales stratégies empiriques permettent d’estimer les prédictions associées à ces modèles. La première, proposée par Hanson (2005), est pertinente pour de petites échelles géographiques ou intra-

3. Le lecteur trouvera dans Crozet et Lafourcade (2009) une introduction à ces modèles et leurs implications de politique économique, et dans Combes *et al.* (2008b) une présentation exhaustive de leur fonctionnement et des tests empiriques associés.

nationales (l'auteur étudie les comtés américains). Les mécanismes d'agglomération reposent alors sur une parfaite mobilité interrégionale de la main d'œuvre et des phénomènes de migrations vers les zones offrant les salaires réels les plus élevés. L'approche de Redding et Venables (2004) est plus adaptée à des situations dans lesquelles la mobilité spatiale de la main d'œuvre est plus faible (le modèle suppose une absence totale de mobilité interrégionale, mais autorise la mobilité intersectorielle de la main d'œuvre), mais où la consommation d'inputs intermédiaires joue un rôle crucial. Il est donc plus conforme à une échelle internationale (les auteurs étudient les disparités de PIB par tête et/ou de salaires de près d'une centaine de pays développés et en développement).

Dans les deux cas, il est possible d'exprimer les salaires ou la productivité de la main d'œuvre comme une fonction des potentiels marchands des zones étudiées. Ces potentiels, qui sont proches conceptuellement de ceux construits par Harris (1954), englobent cependant des effets beaucoup plus sophistiqués. Nous avons vu que le potentiel marchand reposait sur l'intuition que les entreprises localisées près d'un grand marché disposaient d'un avantage concurrentiel. Les modèles théoriques confortent cette intuition : sur les marchés lointains, les barrières aux échanges rendent les entreprises moins efficaces que les entreprises locales. Les entreprises gagnent davantage de parts de marché sur les régions proches que sur les régions éloignées : elles y vendent donc plus de biens. En d'autres termes, le profit des entreprises est une somme spatialement escomptée de la taille des marchés (en termes de population, d'emploi ou encore mieux, de revenu). C'est précisément ce à quoi correspond le potentiel marchand de Harris (1954). Cependant, dans un modèle d'équilibre général en situation de concurrence imparfaite, il est facile de voir que les pertes de parts de marché imputables à l'éloignement géographique dépendent du degré d'asymétrie associé au nombre d'entreprises et à la demande des marchés. Ces deux variables sont naturellement endogènes dans les modèles d'économie géographique avec choix de localisation optimal des ménages et des entreprises. Ainsi, à l'équilibre, les exportations de chaque entreprise dépendent non seulement de la taille des marchés, mais aussi des effets prix, et donc de l'intensité relative de la concurrence sur chaque marché.

Dans ce contexte, les salaires sont une fonction de ce que les chercheurs appellent le potentiel marchand *réel* qui s'apparente à celui de Harris (1954), mais qui est corrigé de ces effets prix. Ce potentiel marchand réel dépend des prix "moyens" (ou indices de prix) pratiqués dans toutes les localités, qui sont eux-mêmes une fonction des prix sortie-usine de l'ensemble des variétés produites sur les marchés. Le principal problème pour l'économètre consiste à mesurer ces indices de prix. Or malgré le développement récent et spectaculaire des bases de données, il n'existe à ce jour aucun indice régional permettant d'appréhender les disparités spatiales de prix à l'intérieur des pays. De plus, les indices de prix, en sus d'être endogènes dans les modèles, ne sont pas des fonctions linéaires des paramètres inconnus du modèle, et ils ne peuvent donc être estimés par des méthodes économétriques simples, comme les Moindres Carrés Ordinaires.

Hanson (2005) s'attaque au problème en considérant l'extension du modèle de Krugman (1991) proposé par Helpman (1998). Celle-ci confère un rôle très important aux marchés locaux du logement, secteur qui remplace l'agriculture dans le modèle originel. Les décisions de migration de la main d'œuvre y dépendent des revenus nominaux et indices de prix des biens différenciés, mais aussi du prix des logements. En utilisant la condition de mobilité parfaite des travailleurs et les équations d'équilibre du marché du logement en chaque lieu, Hanson (2005) montre que les indices de prix peuvent s'écrire comme une fonction des salaires, des PIB régionaux et de l'offre locale de logements qui sont, à la différence des indices de prix, observables par l'économètre. Le modèle structurel de Hanson (2005) montre donc que les effets

prix, lorsque le modèle est bien spécifié, modifient la définition des potentiels marchands par rapport aux formes réduites. Surtout, ces potentiels marchands réels dépendent des autres variables observables incluses dans la spécification, et d'un certain nombre de paramètres devant être estimés, comme l'illustre la spécification suivante :

$$\ln w_z = \alpha_0 + \alpha_1 \ln \underbrace{\left(\sum_j e^{-\alpha_2 \text{dist}_{jz}} (\lambda R_j)^{\alpha_3} w_j^{\alpha_4} L_j^{\alpha_5} \right)}_{PM_z^r} + \varepsilon_z, \quad (1.5)$$

où w_z est le salaire nominal de la zone z , R_z son revenu régional, L_z , une mesure de son stock de logements disponibles, et PM_z^r , son potentiel marchand réel.

L'approche de Hanson (2005) est structurelle pour deux raisons : la spécification estimée est directement issue du modèle théorique, et les paramètres à estimer correspondent à ceux du modèle théorique. Les valeurs estimées pour ces paramètres constituent un test de validité du modèle et plus seulement une mesure des effets d'agglomération⁴.

L'approche proposée par Redding et Venables (2004) est également structurelle, mais elle s'intéresse à des effets d'agglomération opérant à des échelles géographiques plus agrégées. De ce fait, le modèle ne confère plus aucun rôle au marché du logement, mais se concentre sur les effets d'agglomération véhiculés par les relations de sous-traitance et la consommation d'inputs intermédiaires. L'équation de salaire obtenue est cependant très similaire à la précédente. Elle inclut néanmoins deux variables de potentiel marchand réel : la première se rapporte au marché des biens finaux, la seconde au marché des biens intermédiaires. Toutefois, ces potentiels marchands ne dépendent plus seulement des variables observables du modèle, comme dans l'étude de Hanson (2005). Afin de contourner ce problème, les auteurs adoptent une stratégie astucieuse. Grâce aux différentes spécifications prédites par la théorie, qui permettent également de faire le lien entre les volumes commerciaux échangés et les potentiels marchands réels, ils estiment ces derniers à l'aide d'une équation de gravité, pour laquelle toutes les variables sont observables empiriquement. Ils réinjectent ensuite les estimateurs obtenus dans l'équation de salaire, afin d'estimer l'impact des effets d'agglomération véhiculés par l'accès aux marchés internationaux.

1.2 Effets sur la croissance, l'innovation et la localisation des firmes

Si la productivité du travail et les salaires constituent les variables d'intérêt souvent privilégiées par l'analyse empirique, certaines études évaluent également l'impact des effets d'agglomération sur d'autres variables d'intérêt : croissance de l'emploi local, innovation et progrès technique, choix de localisation des entreprises... Cette section présente les différentes stratégies empiriques qui ont été utilisées à cette fin.

1.2.1 Croissance de l'emploi ou du PIB

En dehors de Ciccone et Hall (1996), deux autres études ont contribué à renouveler en profondeur l'analyse empirique des économies d'agglomération. Il s'agit de Glaeser *et al.* (1992) et de Henderson

4. Les modèles de concurrence monopolistique font par exemple l'hypothèse que l'élasticité de substitution entre les variétés est strictement supérieure à 1. L'estimation structurelle de l'équation de salaire doit donc donner un coefficient estimé compatible avec cette valeur théorique.

et al. (1995). La principale différence avec l'étude séminale de Ciccone et Hall (1996) concerne la variable dépendante expliquée, qui est maintenant la croissance de l'emploi local. Cette variable est très importante pour les décideurs politiques, en particulier dans un contexte où les disparités interrégionales de taux de chômage sont très importantes, comme elles peuvent l'être en France ou en Europe. De plus, comme pour la productivité ou les salaires, les modèles d'économie géographique apportent des fondements théoriques solides aux spécifications, même s'il faut bien préciser que ni Glaeser *et al.* (1992) ni Henderson *et al.* (1995) n'avaient pour objectif de tester la validité de ces modèles. Ils se plaçaient plutôt dans le cadre des modèles de croissance régionale endogène, plus en vogue dans les années 1990.

La condition d'équilibre qui motive les équations de salaire décrites plus haut, peut être inversée de manière à obtenir une relation théorique entre d'une part l'emploi, et d'autre part, les salaires, le prix des produits, le niveau technologique, les compétences de la main d'œuvre et le coût des autres inputs. Par conséquent, si on utilise les mêmes hypothèses sur la façon dont les externalités inter- et intra-sectorielles affectent ce groupe de variables explicatives, il est possible d'identifier les mêmes effets d'agglomération pour la croissance de l'emploi local que pour la productivité et les salaires (et transitant par la densité, les potentiels marchands, la diversité, la spécialisation, ou les qualifications)⁵.

Le modèle proposé par Viladecans-Marsal (2004) insiste sur la nécessité d'expliquer la croissance locale par le niveau de la production, et pas seulement le niveau de salaire. Cette inclusion soulève de graves problèmes d'endogénéité, comme nous le verrons plus loin. Certains auteurs préfèrent donc s'en dispenser, comme par exemple Combes *et al.* (2004). Mais cela modifie l'interprétation des effets d'agglomération. En effet, un choc positif de la densité sur la productivité ne se répercute sur l'emploi que pour certaines valeurs de l'élasticité de la demande par rapport à l'offre de main d'œuvre. Si la demande n'est pas très élastique, une externalité locale positive qui augmente la productivité sans augmenter la production provoque une contraction de l'emploi. Si l'on estime un effet négatif de la spécialisation sur l'emploi, le décideur ne sait donc pas si la spécialisation a un impact négatif sur la productivité (et donc sur l'emploi) parce que les deux variables sont positivement corrélées, ou si la spécialisation a un impact positif sur la productivité qui réduit ensuite le volume d'emploi. Ce type de considérations permet de comprendre pourquoi il est peu informatif d'estimer l'effet de la spécialisation sur la croissance locale.

Combes *et al.* (2004) proposent une seconde extension permettant d'identifier plus précisément les sources de la croissance de l'emploi local, qu'ils décomposent en deux termes : la croissance de l'emploi moyen des entreprises existantes et la croissance du flux des nouvelles entreprises. Ils peuvent ainsi évaluer l'impact de la densité et des autres variables capturant les effets d'agglomération sur les marges intensives et extensives de l'emploi. Cette décomposition est utile pour les décideurs politiques, qui peuvent ainsi distinguer la dynamique de l'emploi local créée par l'arrivée de nouvelles entreprises, de celle bénéficiant aux entreprises en place. Si ces deux dynamiques sont contradictoires, ils peuvent décider de préférer stimuler l'une ou l'autre. Certains auteurs, comme Rosenthal et Strange (2003), se concentrent uniquement sur les déterminants de la création des entreprises américaines.

Quelques rares études associent les deux variables d'intérêt que sont la productivité et la croissance. Elles utilisent des données de production mais estiment les déterminants de sa croissance, dans l'esprit de Glaeser *et al.* (1992) et de Henderson *et al.* (1995). Cette stratégie a l'avantage de réduire les problèmes d'endogénéité et de clarifier certaines interprétations. Des auteurs comme Henderson (2003) et Cingano

5. Combes (2000) identifie cependant un certain nombre de préoccupations associées à la manière dont les variables d'agglomération sont spécifiées.

et Schivardi (2004) estiment des modèles dynamiques étendus permettant d'étudier la croissance de la production, tandis que Henderson (1997) et Combes *et al.* (2004) font de même pour la croissance de l'emploi. Dans ce cadre, la production ou l'emploi sectoriels sont expliqués par leurs niveaux antérieurs, et par les niveaux actuels des variables capturant les externalités intra- et intersectorielles. Ces approches présentent des avantages à la fois sur le plan économétrique et pour la conception de politiques publiques locales. Du point de vue économétrique, les méthodes de panel dynamique utilisées, plus communément appelées Méthode des Moments Généralisée (ou GMM) depuis les travaux de Arellano et Bond (1991), permettent de neutraliser les biais d'endogénéité associés aux variables d'agglomération comme la densité. Ce qui est intéressant pour le décideur politique, c'est qu'une évaluation de la durée ou de l'inertie des effets d'agglomération peut être effectuée. Il est par exemple possible de savoir pendant combien d'années une augmentation locale de la densité influence la productivité ou la dynamique de l'emploi, et la façon dont cet impact décroît au cours du temps.

1.2.2 R&D, innovation et pôles technologiques

S'il est désormais bien établi que la population, l'emploi et la production tendent à se concentrer dans l'espace géographique, certaines activités présentent des degrés de polarisation encore plus fortes. C'est le cas de la R&D et de ses résultats, les innovations, ce qu'illustrent de nombreuses études de cas décrites par Porter (1990) et Saxenian (1994). Cette concentration spatiale, qui est capturée par l'évolution d'indices sophistiqués dans Feldman (1994) et Audretsch et Feldman (1996) pour les États américains, est souvent considérée en soi comme une preuve de l'existence d'externalités technologiques locales. La distance limite en effet la diffusion spatiale des connaissances, ce qui encourage les entreprises innovantes à s'implanter au même endroit, afin d'y bénéficier des connaissances produites par les autres. Marshall (1890) évoquait déjà cette idée que l'"atmosphère" industrielle pouvait contribuer à sédimenter les compétences et les savoir-faire et à encourager l'innovation.

Des études empiriques spécifiques s'efforcent de tester l'existence de ces externalités technologiques. Elles permettent d'en identifier les mécanismes économiques sous-jacents, contrairement à la grande majorité des travaux étudiant l'impact de la densité sur la productivité, les salaires ou la croissance locale. Plus précisément, ces études estiment ce que Pakes et Griliches (1984) appellent une fonction de production de connaissances. Le nombre de brevets produits par une entreprise ou une région est une fonction du nombre d'entreprises présentes dans la région, des inputs régionaux, notamment de la R&D, ainsi que de certaines externalités émanant de la R&D menée dans les universités ou les entreprises proches. L'impact des variables de densité, de spécialisation ou de diversité est également étudiée. Ces travaux examinent de plus le rôle joué par la distance sur la diffusion des connaissances, en comparant l'emplacement des dépôts de brevets et la localisation des utilisateurs de ces brevets et leurs citations.

1.2.3 Choix de localisation des entreprises et modèles logit

Les caractéristiques économiques locales comme la densité ou les potentiels marchands n'affectent pas seulement la productivité, les salaires, la croissance de l'emploi ou l'innovation. De très nombreuses études ont par exemple cherché à évaluer leur impact sur les choix de localisation des entreprises. Carlton (1983) a été l'un des premiers à proposer d'utiliser à cette fin les modèles de choix discrets développés notamment par McFadden (1974). Du point de vue théorique, l'économie géographique prédit la façon dont

les entreprises se répartissent dans l'espace selon le degré d'accessibilité relatif des régions, une fois prises en compte les différences régionales de coûts de production. Toutes les variables explicatives que nous avons déjà mentionnées déterminent donc les stratégies d'implantation des entreprises. En effet, comme elles influencent leur productivité, elles ont nécessairement un impact sur leurs choix de localisation. Comme pour les équations de productivité ou de salaire, deux stratégies d'estimation peuvent être envisagées, avec les mêmes avantages et inconvénients : les formes réduites, qui ont été proposées initialement par Carlton (1983), ou les formes structurelles, comme celle proposée par Head et Mayer (2004).

Ces travaux se sont plus particulièrement intéressés aux entreprises multinationales. Les caractéristiques économiques locales déterminant les décisions d'implantation sont en effet plus facilement identifiables pour des firmes qui sont libres des contingences historiques ou territoriales pesant sur les entreprises domestiques. Or une multinationale implante ses filiales au sein des régions dans lesquelles elle anticipe que ses profits seront les plus élevés. Si les régions étaient classées de la même manière par toutes les entreprises, et si ces entreprises choisissaient leur localisation toutes au même moment, les filiales devraient se répartir équitablement entre les régions, ou bien se concentrer dans une seule et même région. C'est ce que prédisent justement de nombreux modèles d'économie géographique. Mais ce n'est évidemment pas ce qui se passe en réalité.

C'est à la contribution fondamentale de McFadden (1980) que l'on doit de pouvoir échapper à ce paradoxe. Son modèle de choix discrets intègre deux éléments dans les préférences des individus ou des entreprises : une composante déterministe, que l'on retrouve chez tous les agents à l'identique, et une composante aléatoire, spécifique à chaque agent. Si on applique ce modèle aux choix de localisation, cela signifie que les profits dégagés par une entreprise dans une région sont composés du profit réalisé par toutes les entreprises qui se localiseraient dans cette région, et d'un terme aléatoire spécifique à l'entreprise⁶. Chaque entreprise tire de façon aléatoire la composante associée à une région donnée, indépendamment des autres entreprises et des composantes associées aux autres régions. En conséquence, le classement de ces éléments aléatoires diffère d'une région à l'autre pour chaque entreprise, de même que les profits. Les entreprises ne choisissent donc pas toutes la même région *in fine*. De toute évidence, plus la variance de la composante aléatoire par rapport à la moyenne de la composante déterministe est importante, plus la distribution spatiale des entreprises est équilibrée entre les régions. Symétriquement, plus la variance est faible, plus la distribution spatiale des entreprises est inégale.

La seconde contribution de McFadden (1974) a été de montrer que, sous certaines hypothèses relatives à la distribution du terme aléatoire, l'impact des variables déterminant le choix discret peut être estimé à l'aide d'un modèle logit. Comme il est facile de montrer que les déterminants de la productivité des entreprises influencent également leurs profits, les travaux empiriques dans ce domaine estiment des modèles multinomiaux conditionnels, dans lesquels la probabilité de localisation, évaluée par la part des entreprises présentes dans la région, dépend des mêmes déterminants que la productivité ou les salaires (cf. équation (1.3)).

Les modèles conditionnels logit font l'hypothèse que les termes aléatoires sont identiquement et indépendamment distribués entre les régions. C'est une hypothèse forte, car certaines régions peuvent être plus centrales que d'autres, ou appartenir à un même pays au sein d'un ensemble plus grand,

6. Il est facile de justifier la présence de ce terme aléatoire. Par exemple, les entreprises japonaises peuvent être plus efficaces dans une région donnée parce qu'elles y sont proches d'autres entreprises japonaises, avec qui elles peuvent mutualiser un certain nombre de coûts.

comme l'Europe. Une première solution pour échapper à ce problème consiste à inclure des variables de contrôle spécifiques aux pays (institutions, taux d'imposition, niveau de développement technologique, par exemple), ou à l'espace économique considéré (appartenance à l'UE par exemple). Une stratégie alternative, appelée modèle logit emboîté, fait l'hypothèse que les choix de localisation sont effectués en deux étapes. Dans un premier temps, les entreprises choisissent le pays dans lequel elles souhaitent s'implanter puis, conditionnellement à ce choix, elles optent pour telle ou telle région. Deux éléments aléatoires sont alors considérés, un qui est spécifique à la région, l'autre qui est spécifique au pays, les deux étant supposés indépendants. Dans ce cadre, les termes aléatoires régionaux correspondant à un même pays peuvent être corrélés, et cette corrélation peut être estimée en même temps que les autres paramètres du modèle. Le fait de mettre plus de structure sur la façon dont se font les choix de localisation permet également d'adopter une structure plus sophistiquée de la composante aléatoire du modèle empirique. Les déterminants des choix de localisation à différentes échelles spatiales sont évalués séparément, une fois que la décomposition géographique a été choisie (le pays, puis la région ; l'Europe, puis le pays, puis la région ; ou encore la région, puis la zone d'emploi dans le cas de la France). Un test statistique permet d'évaluer lequel des deux modèles, imbriqué ou non imbriqué, est préférable. Finalement, la stratégie de modèle logit emboîté permet de contourner les problèmes de capacité de calcul, le temps nécessaire à l'estimation augmentant très fortement avec le nombre de choix possibles. La décomposition en deux étapes permet de réduire ce nombre pour chacune des étapes, et rend possible l'estimation qui ne pourrait se faire en une seule étape⁷.

Une autre alternative, adoptée dans certains travaux, consiste à prendre comme variable dépendante le nombre de choix de localisation réalisés dans une région, puis à appliquer des modèles de comptage tels que les modèles de Poisson et binomiaux négatifs, ou plus simplement les modèles Tobit. Cette stratégie permet de corriger le biais associé au fait que la variable dépendante est censurée à gauche, mais celle-ci est ensuite traitée comme une variable continue. Le principal avantage des modèles de comptage est qu'il n'existe plus aucune limite au nombre d'alternatives considérées. Cependant, ces modèles empiriques font des hypothèses de distribution de leurs résidus très spécifiques.

D'un point de vue plus économique, comme la plupart des travaux se concentrent sur les stratégies d'implantation des multinationales, et donc sur les Investissements Directs Étrangers (IDE), les variables d'agglomération classiques (densité, potentiel marchand, spécialisation, diversité, capital humain ou dotations factorielles) sont très souvent complétées par des variables de contexte, comme les aides régionales ou les taux d'imposition. Alors que la variable dépendante se réfère à un choix de localisation individuel, l'échelle à laquelle les effets d'agglomération sont capturés varie donc considérablement d'une étude à l'autre. Par exemple, Head *et al.* (1999) capturent les effets d'agglomération à l'échelle des 50 états américains, tandis que Guimaraes *et al.* (2000) travaillent à l'échelle des 308 "concelhos" portugais, qui sont beaucoup plus petits. En raison de la perspective régionale de ce rapport, nous ne présenterons pas dans la suite les études menées à l'échelle des pays⁸, mais leurs conclusions ne diffèrent pas significativement de celles obtenues à l'échelle régionale.

7. Le lecteur intéressé trouvera dans Train (2003) une présentation détaillée des techniques d'estimation associées à cette catégorie de modèles.

8. Deux illustrations représentatives des travaux menés à l'échelle des pays sont fournies par Disdier et Mayer (2004), qui comparent les économies d'agglomération dont bénéficient les multinationales françaises en Europe de l'Ouest et de l'Est, et Buch *et al.* (2006), qui comparent les choix de localisation des multinationales allemandes dans ces mêmes pays.

1.3 Biais d'endogénéité dans l'estimation des effets d'agglomération

S'il est important de clarifier le cadre théorique dans lequel s'inscrit l'estimation des effets d'agglomération, il est au moins aussi crucial d'établir la liste des biais d'endogénéité dont peuvent souffrir ces estimations. La stratégie d'estimation la plus utilisée, la méthode des Moindres Carrés Ordinaires (MCO) fait l'hypothèse que les variables explicatives ne sont pas corrélées avec le terme d'erreur aléatoire. Dans ce cadre, aucune variable omise susceptible d'influencer les performances économiques locales ne doit être corrélée avec les variables explicatives observées, sinon l'estimation des effets d'agglomération est biaisée. Or l'économie géographique montre que les variables capturant les effets d'agglomération dépendent assez directement des choix de localisation des entreprises et des travailleurs, et *vice versa*. Il y a donc peu de chances que cette hypothèse d'exogénéité soit vérifiée.

En économie géographique comme dans tout autre domaine, il existe deux sources potentielles de corrélation entre les variables explicatives et le terme résiduel et donc, deux sources d'endogénéité potentielles : les variables omises et la causalité inverse. Nous les illustrons dans le cas de l'estimation de l'effet de la densité sur la productivité, mais elles concernent presque toutes les autres variables véhiculant les effets d'agglomération.

1.3.1 Variables omises et causalité circulaire

Une variable omise est une variable qui a un impact sur le phénomène étudié, mais qui n'est pas intégrée à la spécification estimée. Les variables omises susceptibles de déterminer la productivité des entreprises sont très nombreuses. Par exemple, certains secteurs d'activité sont plus productifs que d'autres à l'échelle nationale, et sont en même temps surreprésentés dans les régions les plus denses. Il en va de même pour certaines dotations factorielles, qui ne peuvent pas être mesurées, mais qui augmentent (biens publics) ou diminuent (relief) la productivité. Si ces variables sont corrélées avec la densité, leur effet sur la productivité sera capté par le terme aléatoire, terme qui est alors corrélé avec la densité. Le coefficient estimé pour la densité est donc biaisé, puisque cette variable capte des effets qui ne reflètent pas directement les économies d'agglomération véhiculées par la concentration des activités économiques. Dans les exemples précédents, les zones denses sont plus productives non pas parce que la densité favorise les interactions locales qui rendent les travailleurs et les entreprises plus efficaces, mais parce que les zones denses bénéficient de meilleures infrastructures publiques, qui augmentent la productivité.

Notons qu'il est crucial pour la politique régionale d'estimer correctement l'impact de la densité. En effet, si ce sont les biens publics locaux, et non la densité elle-même, qui engendrent des gains de productivité, alors il vaut mieux investir dans les infrastructures locales que de chercher à densifier l'espace économique. Si l'inverse est vrai et que c'est la densité qui augmente la productivité, alors les investissements publics dans les régions périphériques resteront sans effet sur la productivité.

Les biais induits par les variables omises peuvent conduire les décideurs politiques à surestimer, autant qu'à sous estimer, les effets d'agglomération. L'effet positif de la densité est par exemple sous-estimé si l'on oublie d'inclure dans la spécification des variables négativement corrélées avec la densité, mais qui augmentent la productivité ou, ce qui est plus fréquent, des variables positivement corrélées avec la densité, mais qui réduisent la productivité. Pour ce dernier cas, Roback (1982) montre que seule une analyse minutieuse permet d'évaluer toutes les variables à inclure. Son modèle postule que certaines aménités de consommation, comme les biens culturels, les équipements de loisirs ou les restaurants, sont

à la fois surreprésentées dans les villes (donc corrélées positivement avec la densité) et valorisées par les salariés. Les ménages qui migrent vers les villes pour cette raison consomment également plus de logement, ce qui augmente les rentes foncières. Ces coûts incitent les entreprises à substituer au foncier de la main d'œuvre, ce qui réduit la productivité marginale du travail, et donc les salaires. Par conséquent, les aménités urbaines, bien qu'elles ne déterminent pas directement la productivité, doivent être incluses dans les équations de salaire ou de productivité, faute de quoi les effets d'agglomération induits par la densité, qui intègrent leur impact négatif indirect, sont sous-estimés.

Un autre biais induit par les variables omises peut être illustré par le fait que certaines catégories de travailleurs préfèrent des zones spécifiques, comme discuté dans Moretti (2004a) et étudié dans Combes *et al.* (2008a). Par exemple, les travailleurs qualifiés ont en général une plus forte préférence pour les aménités urbaines, et ils bénéficient souvent davantage des économies d'agglomération que les autres travailleurs. Dans ce cas, une partie de l'augmentation de la productivité observée dans les zones denses n'est due qu'à la surreprésentation des qualifiés dans ces zones. Si ce problème de sélection n'est pas neutralisé, par exemple à l'aide de variables de contrôle reflétant les compétences ou les aptitudes des salariés, la densité capture également cet effet de composition de la main d'œuvre, et les effets d'agglomération qu'elle véhicule sont surestimés.

La deuxième source d'endogénéité, la causalité inverse, se produit lorsque les travailleurs et/ou les entreprises choisissent de se localiser en fonction du salaire réel et/ou de la profitabilité attendue. Or si les chocs de productivité ne peuvent être appréhendés par l'économètre, parce qu'ils ne sont pas observables, ils sont en réalité pris en compte par les travailleurs et/ou les entreprises. Par exemple, si certaines autorités locales choisissent des politiques plus efficaces ou que certains élus prennent de meilleures décisions dans certaines régions que dans d'autres, ces informations ne sont pas enregistrées dans les bases de données officielles, mais elles sont transmises par le biais des médias ou de discussions informelles. Les agents utilisent cette information pour anticiper qu'ils retireront des gains plus élevés en se localisant dans ces régions, ce qui crée une corrélation positive entre la densité et le choc de productivité occasionné par l'efficacité de la gouvernance locale. Dans ce type de situation, ce n'est pas la densité qui augmente la productivité, mais la productivité qui augmente la densité, raison pour laquelle on parle de causalité inverse. Malheureusement, lorsqu'on estime par les MCO l'effet de la densité sur la productivité, celle-ci intègre déjà l'impact positif de la gouvernance locale. Dans le même esprit, certaines régions sont parfois touchées par des catastrophes naturelles, du type de l'ouragan Katrina à la Nouvelle Orléans. Ces chocs de productivité négatifs occasionnent le départ de nombreux agents économiques de la zone ou leurs choix futurs de ne pas s'y implanter, ce qui provoque une chute de la densité. Ici encore, l'estimation par les MCO de l'impact de la densité sur la productivité est biaisée à la hausse.

1.3.2 Techniques de panel et variables instrumentales

Heureusement, les deux sources de biais peuvent être neutralisées à l'aide de techniques économétriques appropriées. Une première stratégie, assez triviale, consiste à introduire toutes les variables de contrôle disponibles, pour peu qu'elles soient compatibles avec le cadre théorique considéré ou la simple intuition économique. Par exemple, l'Institut Géographique National répertorie tous les bâtiments administratifs et les caractéristiques topographiques du paysage français, parmi lesquels de nombreux éléments sont susceptibles d'affecter la productivité des entreprises (aéroports, gares, universités, hôpitaux, équipements

culturels, etc.). En ce qui concerne le biais associé au tri spatial des qualifications, on peut utiliser le niveau de formation des salariés ou leurs diplômes, qui sont assez souvent répertoriés dans les bases de données régionales.

Malheureusement, la liste des variables de contrôle est tellement longue qu'on en vient vite à douter de la possibilité de corriger ainsi le biais d'endogénéité. Lorsque des données de panel sont disponibles, il est donc préférable d'utiliser des effets fixes, c'est-à-dire des variables muettes qui prennent la valeur 1 pour toutes les observations correspondant à un groupe donné d'observations (secteur, zone, temps). S'il n'est plus nécessaire de recueillir des données permettant de multiplier les variables de contrôle, la contrepartie est qu'on ne peut pas non plus en identifier les effets, qui sont noyés dans le résidu. En d'autres termes, les indicatrices neutralisent un groupe d'effets, mais restent difficiles à interpréter en raison du grand nombre d'effets qu'elles recouvrent. Elles sont donc peu informatives pour les politiques économiques locales, mais elles permettent une estimation correcte de l'effet des autres variables incluses dans la spécification, à condition que celles-ci varient dans une autre dimension que celle neutralisée (temps ou espace par exemple, si on considère un effet fixe secteur). Il est important de noter, à la suite de Combes *et al.* (2011a), que l'utilisation d'effets fixes peut également changer l'interprétation des résultats. Nous reviendrons plus loin sur ce point.

Une des avancées les plus significatives rendues possibles par le développement des données individuelles de panel est la possibilité de prendre en compte des effets fixes individus. En effet, quand les bases de données permettent de suivre les mêmes travailleurs sur plusieurs années, il est possible d'identifier séparément l'impact des caractéristiques individuelles qui ne varient pas au cours du temps (sexe, origine sociale, nationalité, structure familiale, diplômes), et celui des variables d'agglomération qui présentent une certaine variabilité temporelle (âge, expérience). Les effets d'agglomération estimés sont alors nets des caractéristiques individuelles prises en considération. Cette stratégie d'identification, initialement proposée par Glaeser et Maré (2001), puis reprise par Moretti (2004b) et Combes *et al.* (2008a), permet de corriger le biais d'endogénéité associé au tri spatial des travailleurs.

La seconde stratégie permettant d'effectuer cette correction repose sur l'instrumentation. L'objectif est ici de trouver des variables, les "instruments", qui sont corrélées avec les variables explicatives suspectées d'endogénéité, mais pas avec les chocs affectant la variable dépendante, et dont les effets sont déjà intégrés dans le terme aléatoire. On remplace alors la variable explicative suspectée d'endogénéité par son prédicteur obtenu via sa régression par les MCO sur les instruments. Ce prédicteur n'est pas corrélé avec le terme aléatoire, car il est proportionnel aux instruments qui, par définition, ne lui sont pas corrélés. L'estimation par les MCO de l'impact du prédicteur sur la variable dépendante n'est donc pas biaisée, et correspond à l'impact réel de la variable explicative endogène sur la variable dépendante.

La difficulté associée à cette stratégie est bien sûr de trouver les bons instruments. Ces derniers doivent être corrélés avec la variable instrumentée, mais non-corrélés avec le terme aléatoire, alors que la variable instrumentée est elle-même corrélée avec le terme aléatoire, ce qui crée une tension entre les deux objectifs. Ciccone et Hall (1996) proposent d'utiliser des variables historiques pour instrumenter la densité, comme par exemple la population de la zone, retardée de quelques décennies. En raison de l'inertie associée aux processus d'urbanisation, les densités "retardées" et les densités actuelles sont en général assez bien corrélées. Toutefois, si cette corrélation est due à la géographie physique ou aux dotations factorielles intrinsèques des régions, les variables retardées sont de mauvais instruments, car ils sont corrélés avec les chocs de productivité actuels. Si au contraire ces chocs ne dépendent pas de ce type d'effets permanents,

et qu'ils sont par exemple induits par des biens publics locaux récents, ou des chocs technologiques ayant eu lieu après la date d'observation des variables retardées, ils satisfont les deux critères requis pour être de bons instruments a priori : ils sont bien corrélés avec la variable instrumentée, et ils ne sont pas corrélés avec les chocs de productivité actuels.

Il est également possible d'utiliser d'autres instruments que les variables retardées. Par exemple, Combes *et al.* (2010) proposent de compléter les densités "historiques" par des instruments capturant la nature géologique des sols. Ces variables ont pu conditionner l'installation des populations il y a très longtemps, ce qui fait qu'elles sont encore corrélées aux densités actuelles, mais elles ont moins de chances que les variables historiques d'être corrélées avec les chocs de productivité actuels.

Les études économétriques montrent cependant que l'instrumentation peut créer davantage de distorsions qu'elle n'en élimine, lorsque les instruments choisis sont mauvais. L'exogénéité des instruments peut heureusement être testée si l'on dispose de plus d'instruments que de variables à instrumenter, en utilisant des tests de suridentification. Ces tests présupposent que certains instruments sont exogènes et s'emploient à tester si, sous cette hypothèse, les autres instruments sont également exogènes. De toute évidence, quand tous les instruments appartiennent à la même famille et sont très corrélés entre eux, ces tests ne sont pas très convaincants. Imaginons, par exemple, que la densité actuelle soit instrumentée par les densités d'il y a 200 et 220 ans. La démonstration que la densité d'il y a 200 ans n'est pas corrélée avec les chocs actuels de productivité sous l'hypothèse que la densité d'il y a 220 ans n'est pas non plus corrélée avec ces chocs, ou *vice versa*, n'est pas très probante, car les densités d'il y a 200 et 220 ans sont vraisemblablement très corrélées entre elles.

Par conséquent, les études économétriques proposent de compléter les tests de suridentification par des tests d'instruments faibles. Ces tests permettent de savoir si l'instrument ajouté est suffisamment orthogonal aux autres instruments pour donner des informations supplémentaires permettant de prédire la variable instrumentée. S'il y a orthogonalité, supposer que l'instrument est exogène pour tester l'exogénéité des autres instruments est une réponse plus convaincante, que si le résultat du test est négatif.

Malheureusement, les études estimant les effets d'agglomération ne s'attaquent pas toutes sérieusement au problème de l'endogénéité. La plupart des études ne fournissent pas non plus de tests de suridentification, et rares sont celles qui proposent à la fois des tests de suridentification et d'instruments faibles.

Il existe cependant deux autres façons de procéder. La première consiste à utiliser des expériences naturelles qui affectent les variables explicatives de manière exogène, afin d'identifier leur effet causal sur la variable d'intérêt qui, elle, n'en dépend pas. Par exemple, Hanson (1996) ou Hanson (1997) étudient l'impact de l'ALENA sur les salaires mexicains, et Redding et Sturm (2008) celui de la division et de la réunification de l'Allemagne sur la croissance de la population urbaine. Le problème est que, par définition, les estimations obtenues correspondent aux effets d'agglomération occasionnés par l'expérience naturelle considérée, et il est difficile de les généraliser à d'autres situations, ou de savoir dans quelle mesure cette expérience est représentative de ce qui se produirait dans un autre contexte engendrant un changement identique de la variable explicative.

L'utilisation des GMM est donc plus fréquente, dans le prolongement des travaux économétriques de Arellano et Bond (1991). Toutefois, elle requiert des séries temporelles et l'utilisation de la technique des variables instrumentales décrite plus haut, dans laquelle les instruments sont les niveaux et les différences retardés des variables instrumentées. Henderson (1997) et Henderson (2003) figurent parmi les premières études à avoir utilisé cette stratégie dans le but de quantifier les déterminants de l'emploi et de la pro-

ductivité aux États-Unis. Cette méthode soulève néanmoins deux questions. D’abord, il est impossible de donner une interprétation structurelle aux effets des instruments. Il est en général difficile d’expliquer la raison économique pour laquelle ils devraient être valides, et le sens dans lequel ils devraient modifier les estimations. Ensuite, les tests d’instruments faibles conçus pour les méthodes GMM n’existent pas encore. Or les tests de suridentification pour les GMM sont basés sur des instruments qui appartiennent tous à la même famille, et ce d’autant plus que la période sur laquelle portent les bases de données utilisées est courte, ce qui est souvent le cas pour l’étude des effets d’agglomération (souvent 15 ans au mieux). Par conséquent, ces tests fournissent peu d’information sur la validité réelle des instruments. À l’inverse, il est clair que les GMM facilitent la recherche d’instruments. En particulier, lorsque la variable dépendante et les variables explicatives sont spécifiques à un secteur d’activité, ni les variables historiques, ni la géologie ne constituent de bons instruments car ils sont de mauvais prédicteurs de la taille du marché locale. Les GMM constituent alors généralement la seule alternative possible. Un autre avantage des GMM est qu’ils peuvent être appliqués à des vecteurs de variables dépendantes, et plus seulement à une variable d’intérêt isolée. Par exemple, il est possible d’étudier simultanément les déterminants de différentes variables de performance locale, comme Combes *et al.* (2004) le font pour l’emploi moyen par entreprise et le nombre d’entreprises par région-secteur, dont on a vu qu’ils dépendent de façon endogène l’un de l’autre. De même, Graham *et al.* (2010) évaluent l’effet des externalités inter- et intra-sectorielles sur la productivité en estimant simultanément la relation inverse, où l’intensité des externalités est une fonction de la productivité.

Par souci d’économie, la discussion menée dans cette section a porté principalement sur le lien unissant la productivité et la densité des emplois. Il convient de garder à l’esprit que les biais d’endogénéité concernent potentiellement toutes les variables explicatives susceptibles de véhiculer les effets d’agglomération. Les potentiels marchands, la spécialisation des zones et la variété de compétences offertes par la main-d’œuvre peuvent souffrir d’un biais similaire, car elles résultent des choix de localisation des entreprises et des ménages, choix qui dépendent eux-mêmes de la productivité, de l’emploi et de l’innovation spécifique à la région. La densité est probablement la variable pour laquelle il est le plus facile de trouver des instruments, mais cette recherche peut s’avérer plus difficile pour les autres variables.

Les biais d’endogénéité sont également problématiques dans l’estimation des effets d’agglomération affectant la croissance locale, l’innovation ou les choix de localisations. Malheureusement, les études les négligent souvent pour ces variables d’intérêt. Au mieux, les auteurs choisissent comme instruments les variables explicatives retardées d’une période, ce qui n’est clairement pas suffisant. En ce qui concerne les variables potentiellement omises, certains auteurs incluent des indicatrices régionales, soit à une échelle géographique plus importante que celle envisagée pour les effets d’agglomération (8 indicatrices pour des grandes régions aux États-Unis dans Head *et al.* (1999) par exemple, deux indicatrices pour les districts de Lisbonne et Porto dans Guimaraes *et al.* (2000)), tandis que d’autres exploitent les séries chronologiques pour intégrer un effet fixe propre à chaque lieu (comme Hilber et Voicu (2010) le font pour la Roumanie, à l’échelle des régions NUTS2). Cela suppose cependant qu’il existe suffisamment de variation temporelle des effets d’agglomération pour pouvoir les identifier sur les seules évolutions temporelles des variables explicatives. Vu l’inertie des hiérarchies urbaines, l’absence de variabilité temporelle pourrait expliquer pourquoi de nombreuses caractéristiques régionales ne sont plus significatives lorsqu’on intègre des effets fixes dans la régression, comme dans Hilber et Voicu (2010). À l’inverse, ce résultat peut être aussi dû à l’omission de variables influençant les choix de localisation, comme les infrastructures de transport.

Chapitre 2

L'ampleur des effets d'agglomération

Les préoccupations méthodologiques détaillées dans la partie précédente peuvent paraître assez fastidieuses, de prime abord. Elles sont pourtant très importantes pour deux raisons. Tout d'abord, elles mettent en évidence le fait qu'aucune réponse théorique définitive ne peut être donnée concernant l'impact de la concentration des activités économiques sur le dynamisme économique local, que ce dernier soit mesuré par la productivité des entreprises ou les salaires, la croissance de l'emploi et la capacité à innover ou à attirer des nouvelles entreprises, *parce que la façon parfaite de l'évaluer n'a pas encore été trouvée*.

Ensuite, elles nous encouragent à rester prudent dans l'interprétation des résultats empiriques qui peuvent être obtenus sous des hypothèses économiques et économétriques très différentes. S'il existe de nombreuses quantifications effectives des effets d'agglomération, peu d'études choisissent la même façon de procéder dans des contextes différents, et ces études varient dans leur degré d'achèvement.

En dépit de ces réserves, un certain nombre de travaux présentent des estimations intéressantes et solidement étayées. Le Chapitre 2 de ce rapport a pour objectif de les examiner. Bien que la mesure des effets d'agglomération ait ciblé de nombreuses régions du monde, nous nous focalisons plus particulièrement sur l'Europe, puisque ce rapport a pour vocation d'éclairer la politique francilienne. Nous n'évoquons les autres travaux, notamment ceux portant sur les États-Unis, que lorsqu'ils présentent des avancées méthodologiques susceptibles d'être transposées ou répliquées dans le cadre européen.

Le Chapitre 2 s'organise de la façon suivante. Il fournit tout d'abord des fourchettes quantitatives permettant d'évaluer les gains de productivité et de salaire induits par la concentration géographique des activités (Section 2.1). Il se concentre ensuite sur les déterminants de la croissance locale, des choix de localisation des entreprises et de l'entrepreneuriat (Section 2.2). Il étudie enfin les spillovers technologiques et la manière dont ces derniers se propagent dans l'espace (Section 2.3).

2.1 Gains de productivité et de salaires induits par la taille des marchés

Il est maintenant bien établi que la densité locale des activités économiques augmente la productivité des entreprises et des travailleurs. Si l'on régresse les salaires ou la productivité totale des facteurs sur l'emploi local ou sur la densité d'emploi, les valeurs obtenues pour l'élasticité, que nous allons détailler ci-après, se situent dans une fourchette allant de 0,05 à 0,09, lorsqu'on utilise les MCO. Cela signifie que lorsque la densité double, la productivité des entreprises augmente de 3,5 à 6,5%. Or les écarts de densité au sein de l'espace européen, mesurés par le ratio du dernier décile (P90) et du premier décile (P10),

sont de l'ordre de 1 à 15 (c'est le cas pour les régions NUTS2). L'écart de productivité imputable aux différentiels spatiaux de densité peut donc atteindre $(15^{0,09} - 1) \times 100 \simeq 30\%$. Les écarts entre régions d'un même pays peuvent être tout aussi substantiels. En France continentale, à l'échelle des départements, le ratio inter-décile P90/P10 (respectivement inter-centile P99/P01) de la densité d'emplois s'élevait en 2004 à 12 (resp. 2650). Les économies d'agglomération à attendre d'une densification des activités sont donc supérieures de 25% (resp. 100%) pour le département des Bouches-du-Rhône (resp. de Paris), comparé au département du Lot (resp. de la Lozère). Ces chiffres ne reflètent cependant que la fourchette haute des estimations. Les effets d'agglomération estimés par des méthodes plus sophistiquées (GMM, techniques des variables instrumentales) sont beaucoup plus faibles, comme nous allons le voir.

2.1.1 Économies d'agglomération liées à la densité

À l'échelle de toute l'Europe, il n'existe que trois études ayant estimé l'élasticité de la productivité à la densité. Ciccone (2002) est la toute première étude à reproduire à l'échelle des régions NUTS3 l'analyse menée par Ciccone et Hall (1996) pour les comtés américains. Cependant, les bases de données européennes permettaient à l'époque de ne considérer que les régions françaises, allemandes, italiennes, espagnoles et britanniques. L'élasticité moyenne estimée pour ces régions est voisine de 0,05, mais il est intéressant de remarquer que les effets d'agglomération ne diffèrent pas beaucoup d'un pays à l'autre. Cette élasticité moyenne est donc un bon prédicteur des gains de productivité à attendre de la densification des emplois à l'intérieur des pays de la "vieille" Europe.

Deux études plus récentes ont élargi le spectre des pays considérés, mais au prix d'un niveau de détail géographique moindre. Brülhart et Mathys (2008) étudient 245 régions NUTS2 appartenant à 20 pays d'Europe de l'Est et de l'Ouest, sur la période 1980-2003 (1990-2003 pour les pays de l'Est), et huit grands secteurs d'activités relevant aussi bien de l'industrie que des services. Les auteurs utilisent la méthode des panels dynamiques évoquée au Chapitre 1 (GMM), pour neutraliser les problèmes d'endogénéité associés à l'estimation de l'effet de la densité. Malheureusement, le résultat des estimations varie considérablement avec la stratégie adoptée pour traiter cette endogénéité (méthode de moments en différences ou en systèmes). Les estimations confirment néanmoins la présence d'effets d'agglomération tout à fait significatifs, avec une estimation de l'élasticité à long terme d'environ 13%. Des régressions en coupe réalisées sur des sous-ensembles d'années indiquent par ailleurs que l'ampleur des effets d'agglomération semble avoir augmenté au cours du temps, ce qui est conforme à l'idée, avancée par les modèles d'économie géographique, que la réduction des barrières aux échanges constituerait le principal aiguillon de la polarisation.

Foster et Stehrer (2009) obtiennent des estimations plus proches de celles fournies par Ciccone (2002). Ils s'appuient sur un panel de 255 régions NUTS2 appartenant à 26 pays européens, et englobant six secteurs d'activités différents, dont l'agriculture, la sylviculture et la pêche, qu'ignoraient Brülhart et Mathys (2008). La stratégie d'instrumentation est également très différente, puisqu'elle repose sur l'instrumentation de la densité par la superficie et fait l'hypothèse que le niveau de compétence de la main d'œuvre régionale est exogène. Cette stratégie n'est donc pas très convaincante. En dépit de ces limites, l'étude produit un résultat intéressant pour la conception des politiques régionales européennes : les économies d'agglomération sont plus fortes dans les nouveaux États membres que dans les anciens, et cet écart va croissant.

Les travaux qui s'intéressent à un seul pays européen ont l'avantage d'utiliser des bases de données individuelles qui ne sont pas disponibles au niveau européen. Ces bases fournissent une information très précise sur la localisation des travailleurs et des entreprises, en même temps qu'elles permettent de purger l'estimation des effets d'agglomération de l'impact des caractéristiques individuelles (expérience, diplômes, âge, etc.). L'inconvénient est que les estimations obtenues ne peuvent être interprétées que pour le pays considéré et peuvent difficilement être extrapolées. Les décideurs politiques peuvent néanmoins faire l'hypothèse raisonnable que des ordres de grandeur similaires pourraient être observés dans les pays voisins, ou dans les pays dont l'économie se rapproche le plus de celle du pays d'étude, en particulier pour les variables conditionnant l'agglomération des activités (degré de mobilité de la main d'œuvre, coûts aux échanges, économies d'échelle, degré de différenciation des produits, etc.).

Les données individuelles de salaire fournies par les Déclarations Annuelles de Données Sociales françaises ont permis à Combes *et al.* (2008a) de mener l'analyse des effets d'agglomération la plus complète en matière de traitement de l'auto-sélection spatiale des travailleurs et de l'endogénéité. En ce qui concerne l'Italie, l'étude de Cingano et Schivardi (2004) est une des premières à quantifier les effets d'agglomération sur la productivité totale des facteurs en mesurant cette dernière à l'échelle individuelle des établissements. Mais elle ne tient pas compte de l'endogénéité potentielle des variables explicatives locales. Le travail de Mion et Naticchioni (2009), qui répliquent Combes *et al.* (2008a) pour l'Italie, pallie en partie cette limite en corrigeant, au niveau individuel, le biais de sélection lié à la répartition non aléatoire des travailleurs entre les provinces italiennes. Pour le Royaume-Uni, Graham (2007) et Graham *et al.* (2010) s'intéressent également à la productivité totale des facteurs des établissements, mais seule la seconde étude adopte une stratégie d'instrumentation fondée sur les GMM. Des études plus anciennes réalisées par Fingleton (2003) et Rice *et al.* (2006) estiment les économies d'agglomération au Royaume-Uni, à l'aide de données de salaires et de revenus agrégés par régions (districts et régions NUTS3 respectivement). L'utilisation de données agrégées par région semble être la seule stratégie actuellement disponible pour l'Espagne. Viladecans-Marsal (2004) s'appuie néanmoins sur une nomenclature spatiale des régions très détaillée, comprenant 331 unités, pour y quantifier les économies d'agglomération, en utilisant elle-aussi des techniques de GMM. À notre connaissance, un grand nombre de pays européens, dont l'Allemagne, n'ont pas encore fait l'objet d'estimations précises de l'impact de la densité des activités sur la productivité ou les salaires.

Une méta-analyse de l'ensemble de ces travaux, fournie par Melo *et al.* (2009), permet de souligner la sensibilité des résultats au pays ou aux secteurs étudiés, à la façon dont les économies d'agglomération sont spécifiées, ainsi qu'à la présence d'effets fixes régionaux ou de variables de contrôle permettant d'appréhender les qualifications de la main d'œuvre, ce qui invite à la prudence. Les auteurs constatent néanmoins que l'impact de la taille des régions sur leur productivité est significativement positif dans la plupart des cas, et proche de 0,05 en moyenne, en l'absence d'instrumentation. Le biais d'endogénéité imputable aux variables omises ou à la causalité inverse varie peu selon les études, et il avoisine 20%. L'instrumentation fait ainsi chuter l'élasticité de la productivité des entreprises à la densité de 0,05 à 0,04. Cet ordre de grandeur est valide aussi bien à l'intérieur des pays (et ce, même lorsque des instruments très différents sont utilisés, comme dans Combes *et al.* (2010)), que pour des ensemble de pays (les régions européennes par exemple dans Ciccone (2002)). Notons que Ciccone et Hall (1996) ne trouvent presque aucune différence entre les effets d'agglomération estimés par les MCO et les estimations instrumentées pour les États-Unis, ce qui peut modifier les implications de politique économique.

Congestion et estimations spécifiques à un secteur d'activité

Les modèles d'économie géographique prédisent qu'une forte concentration géographique des activités conduit, au delà d'un certain point, à des phénomènes de congestion qui peuvent finir par réduire la productivité du travail. L'effet de la densité sur la productivité des entreprises devrait donc être concave. Il est surprenant de constater que la plupart des études ne fournissent aucune estimation de ce degré de concavité, puisqu'elles s'appuient sur des spécifications log-linéaires du même type que l'équation (1.3). On peut imaginer que l'introduction d'un terme quadratique a sans doute été tentée, mais qu'elle n'a pas conduit à des estimations significativement négatives de la variable de densité au carré. Même si cette non significativité peut tout simplement provenir d'une insuffisante variabilité temporelle des données, qui permet difficilement l'identification des termes quadratiques, les implications de politique économique établies en l'absence d'effets de seuil démontrés ne peuvent être que tranchées : densifier les espaces a nécessairement un impact positif sur la productivité.

Quelques études mettent tout de même en évidence un impact concave de la taille, mais elles le font à l'échelle des villes ou des microrégions. Par exemple, Au et Henderson (2006) mettent en évidence une courbe en cloche reliant la productivité et la population des villes. Ils en concluent que la plupart des villes chinoises se trouvent actuellement sur la partie gauche de la courbe, et qu'elles sont donc trop petites au regard des gains de productivité maximum qui pourraient être obtenus en accroissant la taille des villes.

À l'échelle des régions européennes, peu d'effets de seuil ont été signalés. Pour le Royaume-Uni, Graham (2007) s'appuie sur une stratégie originale pour estimer les rendements décroissants de l'agglomération, en utilisant la congestion routière. Cinq des neuf secteurs d'activité considérés dans l'étude sont caractérisés par des effets de densité concaves significatifs. Par ailleurs, l'étude démontre que, lorsque la congestion du réseau routier est prise en compte, l'élasticité de densité augmente dans sept des neuf secteurs considérés. Ceci confirme l'intuition, développée dans le Chapitre 1 du rapport, qu'en l'absence de variables de contrôle, la densité capture l'impact *net* des forces d'agglomération et de dispersion. Quand ces dernières sont en partie contrôlées, l'impact positif de la densité doit nécessairement augmenter. Graham (2007) estime que la congestion du réseau routier, à elle-seule, peut représenter jusqu'à 30% des effets d'agglomération. Enfin pour l'Espagne, Martinez-Galarraga *et al.* (2008) estiment que l'élasticité de la productivité du travail à la densité d'emploi, proche de 0,08 sur de la période 1860-1900, est tombée à 0,07 sur la période 1914-1930, à 0,04 sur la période 1965-1979 et est devenue non significative sur la période 1985-1999 : l'avantage productif dont bénéficiaient les régions NUTS3 espagnoles les plus denses a progressivement disparu.

Il faut ici souligner que, malheureusement, les études empiriques n'estiment que très rarement les effets d'agglomération par secteur d'activité, alors que les informations sectorielles s'avèreraient très utiles pour les décideurs publics. Une première explication possible réside dans la difficulté à trouver des instruments propres aux secteurs étudiés. Une autre explication, plus directe, est tout simplement l'absence d'information sectorielle disponible pour les variables d'intérêt considérées, à savoir la productivité ou les salaires. Brülhart et Mathys (2008) et Foster et Stehrer (2009) constituent des exceptions notables de ce point de vue, à l'échelle des régions européennes. Ces deux études mettent en évidence des effets d'agglomération significatifs dans tous les secteurs considérés, à l'exception de l'agriculture, pour laquelle la densité régionale a un impact négatif sur la productivité, ce qui est tout à fait conforme à l'intuition.

Compte tenu de la part des terres dans la production agricole, et de l'impact positif de la densité sur le prix des terrains, les zones les moins denses constituent nécessairement les meilleures localisations pour ce secteur. Graham (2007) trouve des effets d'agglomération plus importants pour les activités de services que pour les secteurs industriels au Royaume-Uni.

L'étude de Combes *et al.* (2011b) est la seule à notre connaissance à évaluer simultanément l'existence d'effets de seuil de la densité sur la productivité, et à comparer la force de ces effets entre secteurs. Les auteurs y effectuent une analyse cliométrique des effets d'agglomération sur la productivité moyenne du travail des départements français entre 1860 et 2000. Le fait de disposer d'une période aussi longue permet un traitement minutieux de l'endogénéité, puisqu'il est possible d'utiliser comme instruments les variables explicatives retardées de presque 200 ans (le premier recensement remonte en France à 1801). Les auteurs estiment que l'élasticité de la productivité à la densité est de 0,09 en moyenne sur la période 1860-2000. Cette moyenne cache cependant d'importantes disparités entre les secteurs. L'élasticité de la productivité à la densité est négative pour le secteur agricole (-0,11) : les déséconomies d'agglomération sont donc confirmées pour ce secteur. En revanche, la densité a un impact significativement positif sur la productivité des autres secteurs, avec une élasticité plus élevée pour l'industrie (+0,13) que pour les services (+0,07), ce qui contredit le résultat obtenu par Graham (2007) pour le Royaume-Uni. Combes *et al.* (2011b) montrent de plus qu'il existe des effets de seuil liés à la taille des marchés, mais qu'ils transitent davantage par les potentiels marchands, dont l'impact positif sur la productivité tend à diminuer drastiquement sur la période, et pas par la densité, dont le coefficient est au contraire très stable au cours du temps.

2.1.2 Diffusion spatiale des économies d'agglomération liées à la densité

La déclin rapide dans l'espace des économies d'agglomération véhiculées par la taille des marchés est une autre conclusion solidement étayée par la littérature académique : les économies d'agglomération ne débordent pas beaucoup des frontières régionales. Une illustration de ce résultat est fournie par Arzaghi et Henderson (2008) pour les agences de publicité de Manhattan. Ces dernières connaissent une décroissance spatiale extrêmement rapide de leurs économies d'agglomération, qui sont observées principalement dans un rayon de 500 mètres. Même si cet exemple est certainement trop extrême pour être représentatif de secteurs d'activités plus traditionnels comme l'industrie manufacturière, il n'en reste pas moins que les effets d'agglomération sont rarement significatifs au-delà de 100 kilomètres à l'échelle régionale.

La première façon d'apprécier le rôle de la diffusion spatiale des effets d'agglomération consiste à comparer l'élasticité de la densité locale et celle du potentiel marchand. En effet, dans la définition du potentiel marchand, l'activité des zones est escomptée en fonction de leur proximité relative de la zone d'étude. En général, cet escompte spatial se fait en divisant l'activité de la zone par la distance à la zone observée, ce qui revient à présupposer une décroissance spatiale assez forte des effets d'agglomération. L'utilisation de la fonction inverse suppose en effet que l'impact d'une activité située à 20 kilomètres est quatre fois plus faible que celui d'une activité située à 5 kilomètres ; il est 10 fois plus faible pour une activité située à 100 kilomètres que pour une activité située à 10 kilomètres, et ainsi de suite. Cette forme fonctionnelle est rarement rejetée du point de vue empirique, qu'elle soit estimée à l'échelle de toute l'Europe, comme dans Head et Mayer (2006), ou à l'intérieur des pays, comme dans Combes *et al.* (2008a), Combes *et al.* (2010) ou Combes *et al.* (2011b).

Head et Mayer (2006) estiment l'effet du potentiel marchand sur les salaires de 58 régions NUTS1.

Ils estiment l'effet de deux types de potentiel marchands. Le premier est une forme réduite à la Harris (1954), calculé comme la somme des PIB régionaux (y compris celui de la région d'observation) divisés par la distance entre régions. Le second est un potentiel marchand "réel", plus structurel i.e. corrigé des différentiels de prix dont le Chapitre 1 a souligné l'importance. Les auteurs montrent que les deux types de potentiels marchands affectent positivement les salaires régionaux, et qu'ils ont des pouvoirs explicatifs similaires. Combes *et al.* (2008a) et Combes *et al.* (2010) distinguent les effets d'agglomération véhiculés par la densité locale et le potentiel marchand externe aux zones d'emploi françaises : les deux variables ont un impact positif significatif, même lorsqu'elles sont instrumentées et que les spécifications intègrent des variables de contrôle des compétences de la main d'oeuvre locale. D'une manière générale, l'intégration des potentiels marchands dans les spécifications estimées affecte peu l'impact de la densité locale, qui reste stable et très significatif. Selon Combes *et al.* (2011b) la productivité semble être plus sensible aux différentiels de potentiels marchands (élasticité de 0,16) que de densité (élasticité de 0,09). Cette moyenne cache une forte hétérogénéité entre les secteurs. L'élasticité de la productivité au potentiel marchand est très élevée dans l'agriculture (+0,28), élevée dans l'industrie (+0,13) et non significative dans les services. Dans l'agriculture, la proximité des marchés conserve donc un rôle important, en raison du caractère périssable de nombreux produits. Elle est moins déterminante dans l'industrie, en raison de la baisse importante des coûts de transport des biens manufacturés qui s'est opérée en France sur la période 1860-2000. Elle est inopérante dans les services aux personnes, pour lesquels la proximité des consommateurs reste indispensable.

Pour les régions européennes de niveau NUTS2, Foster et Stehrer (2009) intègrent, à côté de la densité, un potentiel marchand externe calculé avec la fonction exponentielle, comme celui supposé par Hanson (2005) dans l'équation (1.5). Cela revient à supposer une diffusion encore plus contrainte des effets d'agglomération dans l'espace qu'avec la fonction inverse. Non seulement cette forme fonctionnelle n'est pas rejetée, mais les tests de diverses fonctions exponentielles montrent que seuls ceux ayant la plus forte décroissance spatiale présentent des effets significatifs.

La seconde stratégie permettant de tester la décroissance spatiale des économies d'agglomération consiste à scinder la variable de potentiel marchand en plusieurs composantes (locale, régionale, nationale, internationale, par exemple), au lieu de considérer une seule variable agrégée. Par exemple, pour les régions NUTS3, Ciccone (2002) constate que la production des régions contigües a une incidence positive sur la productivité de la région étudiée. Mais il ne donne pas l'estimation correspondante, et ne teste pas l'impact des régions situées au-delà du périmètre de contigüité.

Les deux premières études ayant fourni une quantification réelle basée sur cette seconde stratégie sont donc celles de Rosenthal et Strange (2003) et de Desmet et Fafchamps (2005). Mais elles ne s'intéressent malheureusement pas à la productivité des entreprises. Nous en détaillerons donc les résultats plus loin. Il est néanmoins utile à ce stade de décrire la stratégie qu'ils ont proposée pour estimer l'intensité avec laquelle les effets d'agglomération décroissent dans l'espace. En effet, leur méthodologie a ensuite été adaptée par d'autres auteurs ayant analysé la productivité ou les salaires. L'idée est simple : elle consiste à considérer des zones concentriques autour de la région étudiée, et à intégrer leur taille (en termes d'emplois, par exemple) comme variable explicative. Dans Rosenthal et Strange (2003), le périmètre central est situé à l'intérieur d'un cercle de 1,5 kilomètre de rayon à partir du centre de la zone, qui correspond à son code postal. Trois zones sont ensuite étudiées. La première contient les activités situées dans un rayon de 1,5 à 8 kilomètres, la seconde dans un rayon de 8 à 15 kilomètres et la troisième dans

un rayon de 10 à 24 kilomètres. Dans Desmet et Fafchamps (2005), la première zone contient les activités situées dans un rayon de 5 kilomètres de la commune d'observation, la seconde celles situées dans un rayon de 5 à 10 kilomètres, la troisième celles situées de 10 à 20 kilomètres, et ainsi de suite, tous les 10 kilomètres jusqu'à 100 kilomètres.

Suivant une stratégie similaire, Rice *et al.* (2006) observent qu'au Royaume-Uni, les externalités d'agglomération se réduisent sensiblement avec la distance. Les marchés lointains ont bien une incidence sur la productivité et les salaires locaux, mais l'impact de ceux situés dans une courbe isochrone de 40 à 80 minutes de temps de déplacement est quatre fois inférieur à celui des marchés situés à moins de 40 minutes. Les zones situées dans un rayon de 80 à 120 minutes ont un impact deux fois plus faible que celles situées dans un rayon de 40 à 80 minutes. Pour l'essentiel, on n'observe plus d'impact des marchés au-delà de 80 minutes. Rosenthal et Strange (2008) obtiennent des gradients spatiaux encore plus importants pour les salaires des villes américaines. Les marchés situés dans un rayon de 0 et 8 kilomètres ont un impact quatre à cinq fois supérieur à celui des marchés situés dans un rayon de 8 à 40 kilomètres. En ce qui concerne les zones périphériques (de 40 à 80 kilomètres et de 80 à 150 kilomètres), les effets sont encore plus faibles, voire très souvent négligeables.

Quelques très rares travaux ont adopté une approche plus structurelle pour mesurer l'effet des potentiels marchands sur les salaires, et étudié la propagation des externalités d'agglomération qu'ils engendrent. L'étude de Hanson (2005) est pionnière à cet égard. L'auteur estime l'équation de salaire nominal (1.5) évoquée au Chapitre 1, pour 3075 comtés américains sur la période 1970-1980. Il montre qu'une hausse de 10% du potentiel marchand réel accroît les salaires nominaux de 2,6% en moyenne. Cet effet positif tend même à se renforcer au cours du temps, l'élasticité passant à 3,7% sur la période 1980-1990. Les économies d'agglomération véhiculées par les potentiels marchands réels sont donc importantes, mais restent cependant très localisées. Dans une version préliminaire de son travail, l'auteur simule l'effet d'une baisse de revenu de 10% affectant les comtés de l'État de l'Illinois, et montre que l'onde de choc négative véhiculée par le potentiel marchand réel s'éteint assez rapidement au-delà d'un périmètre de 75 kilomètres. Toutefois, ce faible gradient spatial est en partie conditionnel à l'espace considéré. En Italie, dans le cadre du même modèle structurel, l'onde de choc engendrée par une baisse de revenu de 10% en provenance de la province du Latium se fait encore sentir au-delà de 200 kilomètres (Mion (2004)).

2.1.3 Ampleur des externalités de localisation et d'urbanisation

La question de l'intensité relative des externalités de localisation et d'urbanisation est ancienne, mais elle a été réactualisée par Glaeser *et al.* (1992). À l'origine, de nombreuses études considéraient l'activité totale de la zone et l'activité de la zone dans le secteur étudié comme deux déterminants simultanés de la productivité locale. Comme le met en évidence Combes (2000), cette stratégie soulève un sérieux problème d'identification, puisque les deux variables sont fortement corrélées. Lorsque les marchés locaux sont assez grands, les secteurs ont plus de chance d'y être représentés (par rapport à leur taille dans les autres régions). Or ce qu'il est important d'appréhender pour mesurer les effets d'agglomération véhiculés par un secteur est plus un effet de composition, à taille globale de l'économie locale donnée, qu'un effet de taille du secteur proprement dit. Aux côtés de la densité globale d'activités dans la zone, il est donc préférable d'inclure dans la spécification la part du secteur d'activité présent dans la zone, plutôt que sa taille absolue.

À titre d'exemple, Brülhart et Mathys (2008) intègrent comme déterminant de la productivité sectorielle des régions européennes la densité du secteur d'activité, mais il est difficile d'évaluer si cette variable mesure un effet de taille du secteur ou un effet de la densité totale d'emplois (lorsqu'elle est intégrée simultanément, la densité relative aux autres secteurs n'est pas significative, alors qu'elle le serait vraisemblablement si elle était intégrée seule).

En général, lorsque la densité totale et la spécialisation sont considérées simultanément, l'une et l'autre affectent de manière significative la productivité. Cingano et Schivardi (2004) montrent par exemple que c'est le cas pour l'Italie. Ils constatent également que la décroissance des effets dans l'espace est très forte, la spécialisation dans les régions voisines n'ayant aucun impact sur la productivité locale.

En France, Combes *et al.* (2008a) estiment séparément l'impact de la spécialisation pour 99 secteurs d'activité différents : il est significatif dans 94 d'entre-eux, et il est plus élevé dans les services aux entreprises et dans certains secteurs comme les instruments médicaux et fibres artificielles. Ce résultat conforte les conclusions de Henderson (2003) pour les États-Unis : les externalités d'agglomération sont plus marquées dans les secteurs de haute technologie. Plus intéressant, Combes *et al.* (2008a) montrent que si la densité d'emplois explique une grande partie des disparités spatiales de productivité, ce n'est pas le cas de la spécialisation. Les deux variables ont donc des implications très différentes en matière de développement local : modifier la densité totale affecte la productivité *et* l'ampleur des inégalités régionales, mais c'est nettement moins le cas pour la spécialisation.

Dans le prolongement de l'intuition développée par Jacobs (1969), il est intéressant d'étudier le rôle particulier de la diversité sectorielle. Aucun consensus n'émerge cependant pour cette variable, dont l'effet n'est pas très stable, ni robuste aux changements de spécifications. La diversité a donc un impact sur la productivité tantôt significativement positif, tantôt significativement négatif, ou pas significatif du tout, comme dans Combes *et al.* (2008a) (pour la France), Henderson (2003) (pour les États-Unis) et Cingano et Schivardi (2004) (pour l'Italie). Malgré les intuitions très intéressantes qui sous-tendent cette variable, l'effet empirique ne semble donc pas vraiment avéré. Cela tient peut être à la façon dont on évalue la diversité, souvent à l'aide d'indices d'Herfindahl calculés sur la base de la part de chaque secteur d'activité dans la zone, et à l'aide de données très agrégées. Or certaines branches d'activité peuvent profiter de la présence d'un ensemble de secteurs très différents, sans pour autant bénéficier d'externalités provenant de tous les secteurs présents dans l'économie. C'est malheureusement l'hypothèse associée à l'indice d'Herfindahl. Pour contrer ce problème, Moretti (2004b) utilise une mesure de la "proximité" des secteurs d'activité, et observe qu'aux États-Unis, les économies d'agglomération sont plus importantes entre les secteurs économiquement proches qu'entre les secteurs éloignés. La bonne façon de tester le rôle de la diversité n'a cependant pas encore véritablement été trouvée.

2.1.4 Économies d'agglomération liées au tri spatial des travailleurs

Si l'on reconnaît maintenant la place importante qui revient au capital humain dans la croissance (Lucas (1988)), l'économie géographique met en lumière son rôle dans la structuration des espaces économiques. La littérature récente est extrêmement florissante sur la question de l'influence des travailleurs qualifiés, de leur distribution dans l'espace et des externalités positives qu'ils sont susceptibles d'exercer sur les autres travailleurs. Comme détaillé au Chapitre 1, cette question soulève cependant un épineux problème d'identification. Lorsqu'on régresse la productivité moyenne locale sur la part des

qualifiés employés dans la région, deux effets sont appréhendés simultanément : l'externalité potentielle engendrée par les qualifiés sur les autres travailleurs de la région mais aussi, plus directement, leur productivité intrinsèque, qui est plus élevée, indépendamment de la région observée. En utilisant des données agrégées, il n'est pas possible de distinguer ces deux effets.

Si on utilise en revanche des données individuelles, il est possible de connaître les aptitudes du travailleur, ainsi que le niveau de compétences moyen des autres travailleurs localisés dans la même région. Grâce à la variabilité individuelle des données, on peut ainsi distinguer l'effet des compétences propres, de l'externalité de capital humain. Pourtant, selon Ciccone et Peri (2006), il reste un problème d'interprétation qui découle du fait que l'équation de salaire classique, comme celle proposée pour les États-Unis par Moretti (2004b), confond l'externalité positive et l'effet de productivité marginale décroissante. Ciccone et Peri (2006) s'efforcent d'identifier séparément les deux effets, et obtiennent des résultats contrastés. Alors que Moretti (2004b) conclut qu'il existe bien un effet externe positif des diplômés de l'enseignement supérieur sur la productivité, Ciccone et Peri (2006) montrent que les externalités liées à la formation universitaire sont en moyenne très faibles. Rosenthal et Strange (2008), qui utilisent l'approche classique, mettent aussi en évidence un rôle positif joué par les diplômés de l'enseignement supérieur collaborant avec les entreprises aux États-Unis. Ils considèrent la variable à différentes distances de chaque site de travail, comme ils le font pour la densité. Ils démontrent que les externalités de capital humain déclinent sensiblement avec la distance : elles sont 3,5 fois plus importantes pour les collaborateurs localisés dans un rayon de 0 à 8 kilomètres que pour ceux localisés dans un rayon de 8 à 40 kilomètres. En Europe, Rice *et al.* (2006) mesurent l'impact sur les salaires et la productivité locale de la part des salariés diplômés de l'enseignement supérieur et concluent qu'il est significativement positif. Toutefois, comme la spécification n'est pas estimée au niveau individuel, il est impossible de dissocier les rôles joués par la sélection des travailleurs et par les externalités de capital humain. En France, Combes *et al.* (2011b) montrent que le niveau d'éducation a aussi un impact positif sur la productivité, en particulier l'éducation supérieure. Le fait de considérer le capital humain rend non significatif l'effet du potentiel marchand sur la période récente.

Lorsqu'on dispose simultanément de données sur la productivité et les salaires, on peut évaluer la part des gains de productivité liés à l'agglomération qui se répercute sur les salaires. Ce type d'identification n'est à ce jour malheureusement disponible que pour les États-Unis et le Royaume-Uni. Moretti (2004b) constate ainsi que les écarts de productivité estimés entre villes américaines riches et pauvres en capital humain sont similaires aux différences observées pour les salaires, ce qui plaide pour une compensation quasi totale des gains de productivité. Rice *et al.* (2006) décomposent les revenus moyens des régions britanniques en un indice de productivité et un indice d'occupation professionnelle. Ils constatent que les deux tiers environ de la variance spatiale des rémunérations peuvent être attribués aux variations de la productivité.

En France et dans quelques autres pays européens, les données individuelles permettent d'identifier séparément les effets d'agglomération véhiculés par la taille des marchés et ceux induits par les compétences individuelles. Combes *et al.* (2008a) considèrent des effets fixes individuels qui englobent un grand nombre de caractéristiques individuelles, et ils les complètent par des variables comme l'âge et l'âge au carré (qui mesurent l'expérience du travailleur), qui varient au cours du temps, et des effets fixes associés aux localités

(zones d'emploi)¹. Ils parviennent tout d'abord à identifier une externalité positive émanant des cadres supérieurs. Ils montrent ensuite que les caractéristiques individuelles expliquent plus de la moitié des disparités spatiales de salaires. L'élasticité de la productivité à la densité est très affectée par la prise en considération des effets fixes individuels : elle passe de 0,04 (une fois purgée du biais de causalité inverse), à seulement 0,02 après la prise en compte des effets de composition locale de la main d'œuvre.

Combes *et al.* (2008a) soulignent également que les aptitudes individuelles ne sont pas réparties de façon aléatoire sur le territoire français. Les salariés les plus qualifiés, même en termes de caractéristiques non observables, ont tendance à se concentrer dans les zones les plus denses, et la corrélation entre les effets fixes individuels et les effets fixes zone s'élève à 0,29. La corrélation entre les effets fixes individuels et la densité est quant à elle de 0,44. C'est la raison fondamentale pour laquelle le contrôle des caractéristiques individuelles a tant d'influence sur l'estimation de l'élasticité de la productivité à la densité. Il existe un important biais de variable omise lorsque ces caractéristiques sont ignorées. Comme elles sont corrélées positivement à la densité, les économies d'agglomération sont alors surestimées de moitié.

Mion et Naticchioni (2009), qui reproduisent l'étude de Combes *et al.* (2008a) sur données italiennes, parviennent à des conclusions similaires, mais plus mesurées. La corrélation entre les effets fixes individuels et la densité est encore significativement positive, mais de 0,21 seulement. La prise en compte des caractéristiques individuelles atténue l'élasticité des salaires à la densité, dont l'effet net des externalités de capital humain est plus faible qu'en France : en Italie l'élasticité diminue de 0,022 à 0,007.

De toute évidence, l'interaction entre les caractéristiques individuelles et la localisation a de nombreuses implications de politique économique. Par exemple, elle modifie l'interprétation de l'évolution temporelle des disparités de rémunération entre régions. Duranton et Monastiriotis (2002) montrent que si les rendements de l'éducation et la distribution des salariés étaient restés stables au Royaume-Uni sur la période 1982-1997, l'écart de revenus entre le Nord et le Sud aurait pu être comblé, alors qu'il a en réalité très fortement augmenté.

Enfin, quelques études ont examiné si les économies d'agglomération étaient plus importantes pour certains types de salariés ou d'entreprises que pour d'autres. Par exemple, Bacolod *et al.* (2009b) confirment l'idée intuitive que les rendements de l'éducation supérieure sont plus élevés en milieu urbain. Dans le même esprit, Combes *et al.* (2012) établissent que les entreprises les plus efficaces bénéficient d'économies d'agglomération supérieures aux autres. Par exemple, les entreprises situées dans le premier quartile de la distribution des productivités profitent trois fois moins de la densité que celles situées dans le dernier quartile. Les auteurs constatent également que les grandes entreprises tirent un plus grand avantage de la densité (+50% pour les entreprises de plus de 100 salariés par rapport aux PME de 6 à 10 salariés).

D'autres études ont cherché à mieux identifier les externalités d'agglomération véhiculées par le capital humain. Par exemple, Rosenthal et Strange (2003) constatent que le nombre d'heures travaillées aux États-Unis est une fonction décroissante de la densité pour les salariés les moins qualifiés, et croissante de la densité pour les très qualifiés. De plus, l'effet est plus prononcé pour les jeunes salariés. Par ailleurs, le nombre d'heures travaillées par les jeunes plus qualifiés est sensible à la proximité de leurs semblables. Ces schémas sont compatibles avec la présence de personnes consacrant plus de temps et d'efforts au travail dans les villes, et avec la présence d'importantes externalités urbaines de capital humain.

Bacolod *et al.* (2009a) identifient les types de compétences individuelles ayant les plus forts rendements,

1. L'effet de la localisation est identifié séparément de celui des effets individuels grâce à la présence de salariés mobiles entre les zones.

en fonction de la taille des villes aux États-Unis. Ils concluent que seules les compétences cognitives et la capacité à évoluer en société sont mieux rétribuées dans les grandes villes. La motricité et la force physique y sont en revanche moins bien rétribuées. À l'aide d'une base de données françaises originale décrivant les pratiques de communication en milieu professionnel, Charlot et Duranton (2004) établissent que dans les grandes villes riches en diplômés de l'enseignement supérieur, les collaborateurs communiquent davantage, ce qui a un effet positif sur leur salaire. Ce facteur explique entre 13 et 22% de la "prime" salariale associée aux grandes villes, qui disposent d'un plus grand nombre de diplômés de l'enseignement supérieur.

Malheureusement, les études consacrées aux canaux micro-économiques par lesquels transitent les externalités de capital humain restent rares, et ne concernent que très peu les régions européennes. Ces informations seraient pourtant riches d'enseignements dans la perspective du Grand Paris. La région francilienne est en effet celle qui attire le plus grand nombre de cadres, et d'après le recensement de la population de 2006, la part des cadres y a augmenté en moyenne de 2,3 points entre 2001 et 2006 (3.8 points pour le seul département de Paris).

2.2 Dynamiques de l'emploi et choix de localisation des entreprises

L'étude des déterminants de la croissance locale a connu un regain d'intérêt suite aux travaux menés par Glaeser *et al.* (1992) et Henderson *et al.* (1995). Ces deux études font l'hypothèse implicite que les économies d'agglomération dont bénéficie la productivité se répercutent sur l'emploi, dont elles favoriseraient donc aussi la croissance. La raison qui pourrait expliquer pourquoi ces auteurs, et bien d'autres, ont préféré travailler sur l'emploi plutôt que sur la productivité est sans doute liée au fait que cette dernière, tout comme les salaires, n'a été mesurée à des niveaux géographiques fins (villes, bassins d'emploi) et pour différents secteurs, que très récemment.

Comme nous l'avons déjà souligné, il faut toutefois bien garder à l'esprit que la hausse de la productivité n'occasionne une croissance de l'emploi local que sous certaines conditions régissant les élasticités de l'offre et la demande de main-d'œuvre. Cingano et Schivardi (2004) montrent par exemple qu'une hausse de la productivité locale a pénalisé l'emploi en Italie. Ils observent des signes contraires pour certains déterminants communs de la productivité et de la croissance de l'emploi, alors que ces deux variables sont construites à partir de la même base de données. De même, Mameli *et al.* (2008) illustrent le fait que des signes contradictoires peuvent apparaître selon la classification industrielle adoptée. Ils sont par exemple positifs pour la classification à trois chiffres, et négatifs pour celle à deux chiffres. De tels changements peuvent avoir une interprétation économique (les économies d'agglomération peuvent n'opérer qu'au niveau sectoriel le plus fin), mais ils peuvent aussi relever du pur artefact statistique.

Il est donc compliqué pour la recherche de se concentrer sur une seule de ces deux variables, et d'extrapoler les gains d'agglomération dont elle peut bénéficier à l'autre variable. Bien sûr, il est utile pour les décideurs politiques d'étudier les effets d'agglomération propres à l'une et à l'autre de ces variables, mais chaque étude prise séparément ne permet pas de déduire grand-chose sur les évolutions à attendre de l'autre variable, ou sur les canaux sous-jacents par lesquels transitent les externalités d'agglomération. Il faut donc être conscient qu'on n'améliore pas forcément la productivité de la même manière que l'emploi, et que l'amélioration de l'une des deux variables peut être préjudiciable à l'autre. Une autre difficulté associée à l'étude des dynamiques locales de croissance est que les salaires locaux et la production locale en constituent des déterminants importants, mais que ces variables sont très endogènes et doivent donc

être soigneusement instrumentées. Sinon, l'interprétation des effets n'a plus de sens. Malheureusement, certains travaux sont peu circonspects à cet égard.

2.2.1 Emploi total, spécialisation et diversité

Nous avons vu au Chapitre 1 que les déterminants de la croissance de l'emploi sont très semblables à ceux considérés pour la productivité ou les salaires. L'impact de l'emploi total sur la croissance de l'emploi industriel est un premier exemple de la variété des résultats qui peuvent être obtenus dans la littérature. Par exemple, Combes (2000) observe qu'en France, la taille de l'économie locale renforce la croissance de l'emploi du secteur manufacturier, mais nuit à la croissance des activités de services. En Espagne, selon Viladecans-Marsal (2004), la taille de l'économie locale n'a pas d'impact significatif dans trois des six secteurs considérés dans l'étude, tandis qu'une courbe en cloche est visible dans les trois autres secteurs. Blien *et al.* (2006), qui étendent les résultats obtenus par Blien et Suedekum (2005), dégagent pour l'Allemagne des conclusions un peu différentes de celles obtenues pour la France : la taille des marchés a un impact positif aussi bien sur la croissance de l'emploi industriel que sur celle des activités de services. Nous ne disposons que de deux études récentes pour l'Italie : la première considère l'industrie et les services (Mameli *et al.* (2008)), la seconde se limite aux services aux entreprises (Micucci et Giacinto (2009)). Les deux concluent que l'emploi total stimule la croissance de l'emploi industriel.

Comme nous l'avons mentionné plus haut, la question de la décroissance spatiale des effets d'agglomération est essentielle pour la détermination d'une politique régionale. Pour les États-Unis, Desmet et Fafchamps (2005) montrent que les secteurs de l'industrie et de la construction, contrairement aux services, ont vu leur emploi s'éloigner progressivement des centres vers la proche périphérie. Ils régressent la croissance de l'emploi local dans un secteur donné, sur l'emploi localisé à diverses distances : les coefficients sont négatifs pour les distances inférieures à 20 kilomètres et sont ensuite légèrement positifs pour des distances situées de 20 à 70 kilomètres. Les services présentent un tout autre visage : l'emploi y a progressé plus rapidement dans les centres d'affaires spécialisés, et plus lentement dans les zones voisines. Les coefficients sont positifs à des distances inférieures à 5 kilomètres, et légèrement négatifs à des distances comprises entre 5 et 20 kilomètres.

Le gradient spatial associé aux économies d'agglomération dont peut bénéficier l'emploi a été rarement étudié pour les économies européennes. Viladecans-Marsal (2004) est une exception notable pour l'Espagne : dans trois des six secteurs considérés dans son étude, elle met en évidence un effet significativement positif de l'emploi des zones voisines sur la croissance locale. Dans le même esprit, Solé-Ollé et Viladecans-Marsal (2004) montrent que la croissance des centres-villes des grandes zones urbaines, qu'elle soit mesurée en termes de population ou en termes d'activité économique, a un effet positif sur la croissance des banlieues. L'analyse menée par Micucci et Giacinto (2009) pour l'Italie fait aussi apparaître un impact significatif de la taille des localisations distantes.

Nous avons souligné dans la section précédente que l'impact de la diversité sur la productivité n'était pas robuste. Son impact sur la croissance de l'emploi industriel est également très variable d'une étude à l'autre. Selon Combes (2000), en France, la diversité stimule la croissance de l'emploi dans le secteurs des services, mais la pénalise dans la plupart des secteurs industriels, même s'il existe quelques exceptions. En Espagne, Viladecans-Marsal (2004) estime un impact significativement positif de la diversité dans trois secteurs, mais significativement négatif dans un autre et non significatif dans les deux derniers. Pour

l'Allemagne, Blien *et al.* (2006) estiment un effet positif de la diversité sur la croissance de l'emploi dans l'industrie et les services, mais l'effet est plus fort dans le premier secteur. Enfin, la diversité semblerait avoir un effet positif sur la dynamique des emplois en Italie, selon Mameli *et al.* (2008).

Le rôle de la spécialisation est difficile à évaluer, car il ne peut pas être dissocié du processus de retour vers la moyenne : la variable permettant de mesurer le degré de spécialisation, généralement la part de l'emploi industriel dans l'emploi total, est corrélée à l'emploi industriel initial, qui capte la convergence vers la moyenne, et à l'emploi total, qui capte le rôle de la taille de l'économie locale. Il a été au départ suggéré que l'effet de la spécialisation locale pouvait être identifié en utilisant un effet non linéaire de l'emploi sectoriel, qui entre dans la spécification à la fois en logarithmes et en niveaux. Toutefois, cela rend l'interprétation difficile lorsque les deux effets agissent en sens contraire, comme par exemple dans Henderson *et al.* (1995), et il est préférable d'éviter cette stratégie. Par conséquent, l'impact de la spécialisation, dont Combes (2000) et Mameli *et al.* (2008) démontrent qu'il est négatif dans l'industrie et les services, respectivement en France et en Italie, peut s'expliquer par des effets de retour à la moyenne faisant plus que compenser de possibles effets positifs de la spécialisation. Cela pourrait également être le cas dans l'étude de Blien *et al.* (2006), qui estiment un modèle dynamique complet de l'emploi industriel. Les auteurs observent une légère tendance négative de la spécialisation sur l'emploi, cohérent avec la co-existence d'effets de retour vers la moyenne et d'agglomération intra-sectoriels. Mais ces effets d'agglomération ne seraient pas assez forts pour renforcer la croissance de manière pérenne, et conduire à une dynamique suffisamment puissante pour engendrer la concentration totale de l'emploi industriel dans une seule zone. Viladecans-Marsal (2004) est une des rares études qui estime, pour l'Espagne, un impact significativement positif de la spécialisation, mais dans deux secteurs d'activité seulement.

Pour mesurer l'effet de la spécialisation, Glaeser *et al.* (1992) ont popularisé l'utilisation de la taille moyenne des entreprises du secteur d'activité dans la zone. L'interprétation qu'ils ont donnée à cette variable, en termes de concurrence, est particulièrement trompeuse. En effet, les entreprises implantées dans une zone ne subissent pas la seule concurrence des autres entreprises situées dans la zone, mais aussi celle émanant des entreprises localisées sur les marchés externes (tant pour les inputs que pour les débouchés). Cela rend la notion de concurrence difficile à cerner, puisque nous ne savons pas sur quels marchés elle s'applique, soit en termes de lieux, soit en termes de biens. Pour autant, l'effet de la taille moyenne des entreprises est intéressant à étudier en soi, car cette variable reflète l'intensité des économies d'échelle internes aux entreprises, dont les modèles d'économie géographique soulignent l'importance. Les deux études citées plus haut, Combes (2000) pour la France et Blien *et al.* (2006) pour l'Allemagne, montrent toutes deux que la présence de grandes entreprises amoindrit la croissance de l'emploi, aussi bien dans l'industrie que dans les services. Pour appréhender différemment le rôle joué par la taille des entreprises locales, Combes (2000) introduit un indice d'Herfindahl local permettant de capturer l'hétérogénéité de la distribution des établissements au sein de la zone. Il constate que cette variable est également préjudiciable à la croissance locale. Il en ressort que la structure qui favorise le plus la croissance de l'emploi en France est composée de petites entreprises de tailles relativement similaires.

2.2.2 Modèles dynamiques et durée des chocs

Une question cruciale pour la conception des politiques régionales est celle de la durée des chocs positifs engendrés par les effets d'agglomération : affectent-ils les performances économiques locales de

façon pérenne? Plus la vitesse d'ajustement est élevée, plus l'impact du choc se fait sentir rapidement, mais moins ses effets sont durables. La disponibilité de séries chronologiques dans les bases de données régionales a favorisé l'analyse de ces effets dynamiques. Les études s'appuient sur l'estimation de modèles auto-régressifs complets, comme celui proposé par Henderson (1997), l'une des toutes premières études dynamiques effectuées pour les villes américaines. L'analyse des dynamiques locales est très intéressante, car elle permet également de traiter les problèmes d'endogénéité en utilisant les techniques d'estimation GMM. Par ailleurs, les effets à court terme des variables d'agglomération peuvent être distingués de leurs effets à long terme.

Par exemple, Blien *et al.* (2006) montrent, pour l'Allemagne, que l'impact positif observé à court-terme pour la diversité se dissipe rapidement au fil du temps, dans l'industrie comme dans les services. Aucun effet à long terme n'est observé. De même, la taille de l'entreprise a un effet positif significatif à court terme, mais pas à long terme, quel que soit le secteur considéré.

Combes *et al.* (2004) élargissent le cadre d'analyse utilisé par Henderson (1997) et ventilent l'emploi local par secteur d'activité en deux composantes : l'emploi moyen par entreprise et le nombre d'entreprises dans le secteur. Ils estiment un modèle Vecteur Auto-Régressif instrumenté par des méthodes de type GMM. Du point de vue des implications de politique économique, cela permet de déterminer si les caractéristiques locales influencent les deux composantes de la croissance locale, la marge intensive et extensive des firmes, de la même manière et sur la même durée. La réponse est bien évidemment non, ce qui met en lumière l'arbitrage auquel doivent faire face les élus pour stimuler l'emploi local : les politiques favorisant la croissance des entreprises existantes ne sont pas nécessairement les mêmes que celles qui encouragent l'implantation ou la création de nouvelles entreprises. Pour être plus précis, un grand nombre d'établissements de tailles différentes influence positivement la croissance des établissements existants, alors que la création de nouvelles entreprises est favorisée dans les lieux où il n'y a qu'un petit nombre d'établissements de tailles similaires. Un grand bassin d'emploi régional avec un petit nombre de secteurs d'activité de tailles voisines favorise la création d'entreprises *et* la croissance des établissements existants. Une autre conclusion de cette étude est que le processus d'ajustement démontre une plus grande inertie aux États-Unis qu'en France, où les valeurs retardées cessent d'être significatives après un an, alors que des effets retardés de six ou sept sont mis en évidence par Henderson (1997). Les estimations en panel dynamique montrent donc que les externalités statiques dominent les effets de long terme en France, contrairement aux États-Unis. Par ailleurs, alors que les effets fixes zone-secteur expliquent l'essentiel de la variation spatiale de la taille des établissements, le nombre d'établissements est davantage déterminé par la structure économique locale. Les politiques économiques favorisant la création de nouvelles entreprises semblent donc plus efficaces.

Fuchs (2011) reproduit sur données allemandes l'étude de Combes *et al.* (2004). Il souligne l'influence positive de la diversité sur la marge intensive et extensive des entreprises, alors que les résultats relatifs à la spécialisation ne sont pas tranchés. Des établissements de tailles similaires favorisent la croissance des flux d'implantation d'entreprises, mais pénalisent la croissance des entreprises existantes, comme en France. Finalement, Fuchs (2011) montre également que l'impact de la structure industrielle locale sur la dynamique de l'emploi ne diffère pas pour les petites et les grandes entreprises.

Dans les années 1990, certaines études se sont penchées sur les déterminants du chômage à l'échelle des régions, un enjeu important pour l'Union européenne compte tenu des fortes disparités observées entre pays. Il est surprenant de constater qu'il n'existe aucun lien véritable entre ces travaux précurseurs

et les développements récents de l'économie géographique empirique. Par exemple, Jurajda et Terrell (2009) étudient l'impact du capital humain sur le taux de chômage local dans certains pays d'Europe de l'Est (République tchèque, Hongrie, Ukraine), mais ne confèrent aucun rôle à la densité des activités économiques, au potentiel marchand, à la spécialisation ou à la diversité. Les problèmes d'endogénéité liés aux choix de localisation endogènes des agents et aux variables potentiellement omises ne sont pas non plus considérés.

2.2.3 Choix de localisation, densité et accès aux marchés

Le premier facteur dont il est possible de démontrer le rôle positif et quasi-systématique sur le choix de localisation des IDE est, comme prévu par la théorie, la taille du marché local. Head *et al.* (1999) capturent cette taille par la valeur totale de la production locale et Guimaraes *et al.* (2000) par celle de la production industrielle et l'emploi dans les services. Ces variables ont toutes un effet positif sur la probabilité de localisation des IDE. Devereux *et al.* (2007) ne considèrent l'impact de la taille qu'en association avec l'effet d'autres variables, comme les aides régionales susceptibles de bénéficier aux entreprises souhaitant s'implanter au Royaume-Uni. Ces aides, en moyenne, n'ont presque aucun effet sur les choix de localisation des entreprises : les auteurs montrent qu'une subvention de près de 100000£ n'occasionne qu'une augmentation de la probabilité pour une entreprise de choisir la région aidée de 1%. Cet effet très modeste cache cependant de fortes disparités liées à la taille des régions : une subvention de 100000£, assortie de l'implantation de 10 nouvelles firmes accroît la probabilité pour une entreprise de choisir la région aidée de 6.7%. Les aides régionales sont donc proportionnellement plus efficaces dans les régions qui sont déjà les plus attractives pour les firmes. Ce résultat est très intéressant pour les politiques d'aide au développement régional : elles ne rendent une région plus attractive aux yeux des entreprises que si ces dernières peuvent simultanément y bénéficier d'économies d'agglomération.

Guimaraes *et al.* (2000) considèrent également le rôle joué par la distance aux principales villes du Portugal dans les choix de localisation, variable qui préfigure les potentiels marchands utilisés de manière plus systématique depuis les années 2000. Par exemple, Head et Mayer (2004) comparent l'influence des potentiels marchands "nominaux" (i.e. en forme réduite) et "réels" (i.e. structurels) sur les décisions d'implantation des filiales japonaises au sein des régions NUTS2. Ils constatent que ces deux variables ont un impact significativement positif sur la localisation des IDE, même lorsqu'on considère un grand nombre de variables de contrôle additionnelles. Les études de Basile *et al.* (2008) et Basile *et al.* (2009) étendent l'analyse des choix de localisation des multinationales à différentes nationalités et à 50 régions (47 pour les seconds) de huit pays (cinq pour les seconds) de l'Union Européenne : France, Allemagne, Italie, Espagne, Irlande, Royaume-Uni, Portugal et Suède (les trois derniers ne sont pas étudiés dans Basile *et al.* (2009)). Le potentiel marchand nominal est significatif, même lorsque la taille locale est considérée. Cependant, il n'est important que pour les multinationales européennes, l'effet étant non significatif pour les autres entreprises.

Comme dans les travaux sur les déterminants de la productivité, la fonction d'escompte utilisée pour calculer le potentiel marchand (inverse à la distance dans la plupart des cas) implique une décroissance spatiale rapide des effets d'agglomération. Rosenthal et Strange (2003) montrent que le gradient des créations d'entreprises (domestiques ou étrangères) aux États-Unis décroît effectivement rapidement avec la distance : l'effet de l'emploi propre au secteur d'activité dans un rayon de 1,5 kilomètre est de 10 à

1 000 fois supérieur (selon la spécification) à l'effet de l'emploi situé dans un rayon de 3 à 7 kilomètres. Au-delà de 7 kilomètres, l'atténuation des effets d'agglomération liés à la taille des marchés est moins prononcée.

À une échelle géographique plus fine, les villes espagnoles, Arauzo-Carod et Viladecans-Marsal (2009) montrent qu'en Espagne, la décroissance spatiale des externalités d'agglomération dont bénéficient les IDE est également très forte, et croît avec le niveau technologique du secteur d'activité. En Italie, Basile (2004) décèle même des déséconomies d'agglomération émanant des provinces voisines, alors que la taille de la province d'accueil a bien un effet positif. L'auteur distingue par ailleurs les acquisitions réalisées par des entreprises étrangères et créations ex-nihilo. Ces dernières sont plus utiles pour évaluer les effets d'agglomération, car les entreprises de ce type ont plus de latitude dans leurs choix de localisation. L'étude montre que l'effet du nombre d'établissements de la province d'accueil n'est positif que pour les premières. Toutefois, la demande locale, mesurée par la consommation d'électricité, exerce une influence positive sur les deux types d'IDE.

L'impact du potentiel marchand sur les choix de localisation est donc très consensuel. Il est par ailleurs confirmé lorsque les données sont ventilées selon différentes dimensions. Par exemple, Crozet *et al.* (2004) montrent que l'accès aux marchés a un impact très positif sur les IDE en France, quel que soit le pays d'origine de l'entreprise². Ce résultat est conforté pour l'Allemagne par Spies (2010), qui montre que le potentiel marchand a un impact positif sur les IDE, indépendamment du secteur d'appartenance des multinationales. Enfin, une étude récente menée par Pusterla et Resmini (2007) se concentre sur les choix de localisation des IDE au sein des régions NUTS2 de quatre pays d'Europe de l'Est (Bulgarie, Hongrie, Pologne et Roumanie). L'emploi industriel local et les potentiels marchands contribuent tous deux à renforcer l'attractivité des régions, même si l'impact est plus particulièrement fort pour IDE des secteurs de faible technologie, comparé aux secteurs de haute-technologie.

2.2.4 Choix de localisation et composition industrielle locale

Les travaux qui s'intéressent aux choix de localisation des IDE s'interrogent presque tous sur l'impact d'une variable rarement considérée pour la productivité ou la croissance locales : la présence au sein de la région d'accueil d'autres entreprises étrangères. Cette présence a un effet positif sur les IDE car elle peut refléter certaines caractéristiques non observables de la région d'accueil qui la rendent aussi plus attractive aux yeux des nouveaux IDE, ou encore l'efficacité des entreprises étrangères à tisser des réseaux d'affaires qui seront ensuite exploités par les nouveaux entrants. Mais une seconde interprétation est tout à fait possible : la présence d'entreprises étrangères peut être préjudiciable à l'entrée de nouveaux IDE, qui craignent la concurrence des entreprises en place, aussi bien sur le marché des inputs locaux que sur le marché des produits. La littérature tend à montrer que l'effet négatif de la concurrence semble être assez systématiquement dominé.

Par exemple, Head et Mayer (2004) mettent en évidence un effet positif des IDE passés sur les choix de localisation des filiales japonaises dans les régions européennes. En Italie, l'étude de Basile (2004) conclut, de manière similaire, que les investissements passés favorisent les IDE entrants, que ces derniers prennent la forme d'acquisitions ou de créations ex-nihilo. La présence de multinationales dans les provinces voisines

2. Dans l'optique d'une comparaison avec Devereux *et al.* (2007), il est intéressant de remarquer que ces auteurs montrent qu'un doublement de la Prime d'Aménagement du Territoire, qui constitue la principale aide à la localisation d'entreprises en France, n'accroît la probabilité pour une entreprise de se localiser dans la région aidée que de 4% seulement.

de la province d'accueil influencent également positivement les acquisitions. Spies (2010), Pusterla et Resmini (2007) et Barrios *et al.* (2006) mettent en évidence des effets similaires pour l'Allemagne (à l'échelle sectorielle), les pays d'Europe de l'Est (pour les secteurs de haute-technologie ou de basse haute-technologie) et l'Irlande, respectivement. L'étude de Basile *et al.* (2009) permet de nuancer quelque peu cette conclusion : l'effet positif de la présence étrangère est visible pour les IDE européens, mais pas pour les autres (ce qui contredit le résultat de Basile *et al.* (2008), qui montrent que l'effet est positif pour les uns comme pour les autres). Crozet *et al.* (2004) confortent l'idée que la présence étrangère n'affecte pas tous les IDE de la même façon : les effets les plus significatifs pour la France concernent les IDE en provenance du Japon, du Royaume-Uni, de la Belgique et des États-Unis. Enfin, Devereux *et al.* (2007) obtiennent un effet significativement positif des filiales étrangères sur les IDE, mais aussi sur les investissements des entreprises locales, avec un impact plus marqué pour les premiers que pour les seconds.

L'effet potentiel des réseaux sociaux ou d'affaires a également été étudié, mais de manière indirecte, au travers de la distance séparant la localisation des filiales et le pays d'origine où réside le siège social de l'entreprise. Les résultats sont conformes à l'intuition : la distance affecte négativement les IDE à destination de la France (Crozet *et al.* (2004), Lafourcade et Paluzie (2009)) ou les IDE européens à destination de l'Europe (Basile *et al.* (2008)). La proximité culturelle joue également un rôle important : d'une manière générale, la probabilité pour une filiale de se localiser dans un pays dont les habitants parlent la même langue que le pays d'accueil est également plus élevée.

Dans le même esprit que pour l'étude des déterminants de la productivité, les travaux examinent également les effets de la composition industrielle locale, avec le même type de variables que celles utilisées pour la productivité. La spécialisation, mesurée par exemple par le nombre de secteurs d'activité présents dans la zone, a une influence positive sur le choix de localisation des IDE, à l'échelle de toute l'Europe (Head et Mayer (2004)), ou de certains pays (par exemple le Portugal dans Guimaraes *et al.* (2000), ou la France dans Crozet *et al.* (2004)). Les créations d'entreprises bénéficient également d'externalités de spécialisation, par exemple dans les villes (Arauzo-Carod et Viladecans-Marsal (2009)) ou les régions (Costa-Campi *et al.* (2004)) espagnoles, et au Royaume-Uni (Devereux *et al.* (2007)), où l'effet positif augmente avec le degré de concentration spatiale du secteur d'activité.

Devereux *et al.* (2007) mettent également en évidence un impact positif de la diversité industrielle sur la localisation des IDE au Royaume-Uni, tandis qu'en Espagne, cet effet dépend du secteur étudié (Arauzo-Carod et Viladecans-Marsal (2009)). En ce qui concerne l'Irlande, Barrios *et al.* (2006) constatent que la diversité a un impact significativement positif sur les IDE depuis les années 1980, mais pas avant, et seulement pour les entreprises de haute technologie, pour laquelle la spécialisation n'a en revanche pas d'impact. À l'inverse, la diversité ne joue aucun rôle pour les entreprises de basse technologie, alors que la spécialisation encourage ce type d'IDE. L'étude récente de Hilber et Voicu (2010) montre que la spécialisation a une incidence positive sur les IDE à destination de la Roumanie, tant pour les entreprises domestiques que pour les entreprises étrangères, mais que seul l'effet positif sur les premières résiste à l'introduction d'effets fixes régionaux. Il en va de même pour la spécialisation des régions voisines, qui bénéficie aux entreprises nationales. La diversité n'a un impact positif significatif que lorsque les effets fixes régionaux ne sont pas présents dans la spécification.

Nous avons déjà noté que Guimaraes *et al.* (2000) distinguaient le rôle spécifique joué par la concentration de l'industrie et des services, l'effet des services étant plus prononcé que celui de l'industrie. L'importance des services est confirmée pour les choix de localisation des IDE à destination des régions

d'Europe de l'Est notamment. Selon Cieřlik (2005), la concentration des services encourage l'implantation des IDE dans les r'egions NUTS3 polonaises (49 r'egions sont 'etudi'ees), et la m'eme conclusion pr'evaut pour les r'egions NUTS3 roumaines (21 r'egions sont 'etudi'ees), m'eme lorsque des effets fixes r'egion sont int'egr'es aux sp'ecifications : une augmentation de 10% de la densit'e des emplois tertiaires au sein d'une r'egion roumaine fait augmenter de 11,9% les chances d'y attirer un investisseur 'etranger (Hilber et Voicu (2010)).

2.2.5 Choix de localisation, co'uts du travail et entrepreneariat

Une autre variable d'eterminant le choix de localisation des IDE est bien 'evidemment le co'ut de la main d'oeuvre locale. De prime abord, l'intuition 'economique sugg'ere que des salaires 'elev'es devraient d'ecourager l'implantation des IDE. Il est malheureusement difficile de confirmer cette intuition par l'etude empirique, car cette assertion ne peut 'etre v'erifi'ee qu'aux qualifications ou aptitudes des travailleurs donne'es. Or les bases de donne'es existantes ne permettent g'eneralement pas de mesurer ces aptitudes, et donc le rendement r'eel du travail, comme le n'ecessiterait une interpretation correcte des r'esultats. C'est pourtant bien le co'ut du travail rapport'e au nombre d'unit'es de main d'oeuvre efficace que la th'eorie pr'ekoniserait d'utiliser comme d'eterminant des choix de localisation (cf. Chapitre 1). Lorsque la productivit'e du travail n'est pas prise en compte, un impact positif des salaires sur les choix de localisation peut tout simplement traduire l'attrait que r'epresente pour les multinationales la pr'esence de personnel qualifi'e.

Dans le m'eme esprit, certains articles consid'ereent que les taux de ch'omage r'egionaux conditionnent la localisation des IDE, mais la encore la corr'elation empirique est difficile 'a interpreter. Un taux de ch'omage 'elev'e peut refl'eter un exc'es d'offre sur le march'e du travail local, qui pousse les salaires 'a la baisse, ou au contraire, des niveaux de salaires trop 'elev'es, si c'est ce qui provoque le ch'omage. Lorsque les salaires sont int'egr'es dans la sp'ecification, le taux de ch'omage peut refl'eter la caract'ere d'efavoris'e d'une r'egion et une demande de main d'oeuvre faible, si des variables de contr'ole ne purgent pas l'impact de ces 'elements dans la sp'ecification.

'A l'echelle des r'egions europ'eenes, Head et Mayer (2004) et Basile *et al.* (2009) ne d'ec'elent aucun effet significatif des salaires et du ch'omage sur les IDE³. Ces r'esultats doivent s'interpreter 'a la lumiere des mises en garde que nous venons d'evoker, car la productivit'e du travail n'est pas consid'eree dans ces travaux. Le taux de ch'omage n'a toutefois toujours aucune incidence sur la localisation des IDE dans les travaux men'es par Basile *et al.* (2008) et Basile *et al.* (2009), qui prennent en compte le niveau d'education. Mais ce dernier n'a pas non plus d'impact significatif sur les choix de localisation, ce qui sugg'ere que l'education constitue une mesure tr'es imparfaite des aptitudes r'elles de la main d'oeuvre 'a l'echelle des r'egions. Crozet *et al.* (2004) n'obtiennent pas non plus d'effets d'agglom'eration li'es au ch'omage en France pour l'ensemble des IDE, mais cette moyenne est trompeuse : des taux de ch'omage 'elev'es d'ecouragent les IDE en provenance de la Belgique, des Pays-Bas ou de l'Italie, mais encouragent les IDE en provenance des 'Etats-Unis. Pour les IDE des autres pays, le ch'omage est non significatif.

L'impact du ch'omage et du niveau d'etude secondaire ne sont pas robustes 'a l'introduction d'effets fixes r'egionaux en Roumanie, comme le montrent Hilber et Voicu (2010), alors que le nombre de conflits sociaux, suppos'es refl'eter les dysfonctionnements des march'es locaux du travail, ne sont jamais significatifs. En

3. Lorsqu'il est introduit seul dans la sp'ecification, le taux de ch'omage a n'eanmoins un impact n'egatif significatif sur les IDE en provenance de l'UE, selon Basile *et al.* (2008).

Pologne, Cieřlik (2005) ne trouve aucun effet significatif des salaires et du niveau d'éducation sur les IDE, et un effet négatif du chômage. En Italie, Basile (2004) constate que les coûts unitaires du travail (une mesure indirecte de la productivité) et le chômage découragent significativement les IDE, qu'ils prennent la forme d'acquisitions ou de création ex-nihilo. En Allemagne, selon Spies (2010), ces variables ont un impact négatif significatif dans un seul secteur - dénommé "Autres services" - sur les quatre étudiés. Dans l'étude de Pusterla et Resmini (2007), les salaires n'ont d'impact négatif que sur les secteurs de faible technologie lorsque la spécification inclut des variables de contrôle pour l'éducation. Enfin, Devereux *et al.* (2007) décèlent un effet négatif des salaires des peu qualifiés, un effet positif des salaires des plus qualifiés, et un effet négatif du chômage, mais à peine significatif, sur les choix de localisation des IDE à destination du Royaume-Uni.

Enfin, au-delà de l'implantation des IDE, les déterminants des choix de localisation des créateurs d'entreprise sont primordiaux pour l'analyse des effets d'agglomération. Malheureusement, comme l'observent Glaeser *et al.* (2010), les études consacrées à cette analyse restent encore très rares, surtout pour l'Europe. Figueiredo *et al.* (2002) constitue une exception notable. Les auteurs y analysent les choix de localisation des créateurs d'entreprises au Portugal. Ils constatent que, comme pour les IDE, l'activité économique passée est le critère le plus important. Par ailleurs, les auteurs soulignent un second résultat très intéressant : les externalités d'agglomération ne bénéficient qu'aux créateurs d'entreprises originaires de la zone, et pas aux autres. La connaissance qu'ils peuvent avoir du marché, ainsi que les effets de réseau qui découlent de cet avantage informationnel, constituent sans nul doute les vecteurs indispensables des externalités d'agglomération.

2.3 Technopôles et diffusion spatiale des externalités technologiques

Les externalités d'agglomération constituent donc un déterminant majeur du choix de localisation des IDE. De plus, les régions sont d'autant plus attractives qu'elles accueillent déjà un nombre élevé d'entreprises, notamment du même secteur d'activité. Cet effet positif est le signe que les bénéfices engendrés par les groupements d'entreprises ont un effet qui dépasse en intensité la pression de la concurrence. Face à ce constat, les interventions publiques hésitent entre deux optiques : lutter contre cette tendance au nom de l'équité spatiale, ou au contraire, encourager la formation de pôles industriels spécialisés afin de maximiser les externalités. Cette seconde stratégie s'inspire des expériences particulièrement réussies des districts marshalliens, comme celui de la Silicon Valley, qui a largement contribué au succès de l'industrie de haute technologie américaine. De tels pôles technologiques peuvent émerger de façon spontanée. Mais les décideurs publics peuvent aussi encourager leur apparition ou leur développement à l'aide de dépenses d'infrastructures, de politiques d'éducation et de recherche, ou de subventions aux entreprises. C'est sur ce mode que la politique industrielle française a misé en 2005 sur les "pôles de compétitivité". Cette politique ambitieuse, dotée de 1,5 milliard d'euros pour la période 2009-2011, cible les aides publiques sur des groupements localisés d'entreprises dynamiques, notamment en matière de R&D. Les pôles sont définis explicitement comme "une association d'entreprises, de centres de recherche et d'organismes de formation engagés dans une démarche partenariale, destinée à dégager des synergies autour de projets innovants, conduits en commun".

Sur le papier, la politique des pôles de compétitivité est assurément une bonne idée, mais il convient d'en réexaminer l'opportunité à la lumière des travaux ayant analysé les externalités technologiques en-

gendrées par la proximité spatiale et la diffusion imparfaite dans l'espace des connaissances et des innovations.

2.3.1 Production des connaissances et spillovers informationnels aux États-Unis

Jaffe (1989) est l'une des premières études à avoir mis en évidence une véritable relation causale de la proximité géographique sur l'innovation aux États-Unis. L'auteur étudie la manière dont les brevets déposés par des entreprises privées sont liés à la R&D réalisée dans les universités voisines, en distinguant cet effet de celui de la R&D menée au sein des entreprises, et en s'attaquant au problème de la causalité inverse. Il montre que la recherche effectuée dans les universités situées dans les mêmes états que les entreprises privées accroît de manière significative leurs dépôts de brevets. Le fait que les entreprises et les universités soient localisées au même endroit au sein de l'État, et donc "co-localisées", n'a cependant pas d'impact significatif sur le dépôt de brevets.

À la suite de Jaffe (1989), la littérature a remarqué à juste titre que le dépôt d'un brevet n'était pas nécessairement synonyme d'innovation. L'étude de Acs *et al.* (1992) tente de pallier cette limite en utilisant des variables d'enquêtes sur l'innovation, et confirme le rôle décisif de la proximité sur l'amélioration des connaissances. Les conclusions de Jaffe (1989) ne sont pas seulement confirmées, mais aussi renforcées : la recherche universitaire encourage plus l'innovation que le dépôt de brevets et la co-localisation des universités et des entreprises privées a bien un effet significativement positif sur les innovations.

Comme pour les déterminants de la productivité ou de la croissance, l'étendue spatiale des retombées technologiques de l'agglomération est une question cruciale. Adams et Jaffe (1996) et Adams (2002) capturent ces effets à l'aide de variables de R&D calculées à différentes distances du périmètre étudié, et ils comparent la force respective de leurs effets. La première étude estime que l'impact de la recherche localisée à plus de 150 kilomètres représente à peine 20% de celui de la recherche effectuée dans un rayon inférieur à 150 kilomètres. La seconde met en évidence des effets significatifs dans un rayon de 300 kilomètres, mais pas au-delà. De plus, les spillovers universitaires semblent plus localisés que les spillovers industriels. Les externalités technologiques sont également plus faibles, et encore plus localisées, pour les nouveaux produits que pour les brevets. Des conclusions similaires sont obtenues par Anselin *et al.* (1997), qui travaillent au niveau des aires métropolitaines américaines (les MSA).

La littérature poursuit l'analyse des externalités technologiques en étudiant l'interaction entre la taille des entreprises et leur capacité à bénéficier de spillovers technologiques. Selon Acs *et al.* (1994), même si toutes les entreprises bénéficient en moyenne de spillovers, les grandes firmes profiteraient davantage de la R&D privée, tandis que les PME exploiteraient principalement les résultats de la R&D universitaire ou publique.

La composition sectorielle locale joue un rôle ambigu sur la production des externalités technologiques : aux États-Unis, Feldman et Audretsch (1999) mettent en évidence un effet très positif de la diversité sur l'innovation. La spécialisation ne semble avoir aucun impact sur le taux d'innovation, ce qui peut paraître surprenant au regard du succès emblématique rencontré par certains pôles technologiques, comme celui de la Silicon Valley.

Enfin, toujours pour les États-Unis, le dépôt de brevets est plus important dans les zones où les emplois sont plus denses : doubler la densité d'une ville accroît son taux de brevets par habitant de 20% selon Carlino *et al.* (2007), qui montrent de surcroît que cet effet est concave.

2.3.2 Spillovers informationnels et technologiques en Europe

Certains auteurs ont tenté de reproduire ces études pour l'Europe, mais malheureusement, sans jamais complètement s'attaquer au problème des variables manquantes et de causalité inverse. Bottazzi et Peri (2003) examinent par exemple le rôle joué par la R&D sur les spillovers technologiques. Ils étudient le nombre de brevets déposés dans 86 régions européennes pour la période 1977-1995. Le nombre de brevets y est analysé comme une fonction de l'intensité de la R&D menée dans les régions situées dans un rayon de 0 à 300 kilomètres, de 300 à 600 kilomètres, de 600 à 900 kilomètres, de 900 à 1300 kilomètres ou de 1300 à 2000 kilomètres. La R&D est instrumentée par la densité régionale en 1930, ce qui permet de neutraliser le biais d'endogénéité associé à cette variable. Les retombées de la R&D sont ressenties dans un rayon de 300 kilomètres de la région source, mais les spillovers technologiques s'épuisent rapidement au-delà. Ainsi, en deçà de 300 kilomètres, les auteurs estiment qu'un doublement des ressources consacrées à la R&D dans une région accroît ses dépôts de brevets de 80% à 90%. L'effet n'est plus que de 2-3% pour les régions situées à 300 kilomètres et les spillovers technologiques disparaissent complètement au-delà de ce rayon.

Des conclusions similaires sont obtenues par Greunz (2003), qui examine encore plus minutieusement l'effet de la distance à l'aide de données couvrant 153 régions européennes couvrant la période 1989-1996. L'auteur fait l'hypothèse que les retombées interrégionales des spillovers de connaissance diminuent proportionnellement à l'inverse du carré de la distance, et il dissocie les retombées émanant des voisins directs, des voisins du deuxième ordre (les voisins des voisins), et ainsi de suite. L'activité locale des régions, de même que celles des régions appartenant aux trois premiers ordres de voisinage, affecte de manière significative le nombre de brevets déposés, avec une intensité décroissante selon les ordres. Les retombées émanent principalement du secteur privé, mais les universités de la région et des régions voisines du premier ordre contribuent aussi à encourager le dépôt des brevets. Plus important, la proximité technologique des régions accroît l'intensité des spillovers technologiques.

Parent et Riou (2005) confirment ces conclusions sur un échantillon de 335 régions réparties dans neuf pays européens pour la période 1989-1999, mais ils substituent à la distance le temps de transport entre les régions. Bode (2004) construit un potentiel marchand externe de R&D et montre qu'il a un impact significatif, mais faible, en Allemagne. Il distingue par ailleurs les régions à forte et basse intensité de R&D et constate que seules les dernières bénéficient de spillovers technologiques. Enfin, Maggioni *et al.* (2007) étudient un sous-ensemble de 109 régions européennes, et mettent en évidence que, outre le fait qu'elle engendre des spillovers technologiques, une spécialisation dans les secteurs de haute technologie produit un nombre plus élevé de brevets par habitant. Ils soulignent également combien il est pénalisant d'être situé à la périphérie de l'Europe (mesurée par la distance par rapport à Bruxelles) : ces régions bénéficient de retombées technologiques bien plus faibles.

Quant aux rares études menées en France, elles portent essentiellement sur la politique des pôles de compétitivité. Duranton *et al.* (2010) estiment qu'il faudrait doubler la taille des clusters (plus précisément, doubler le nombre de travailleurs du secteur dans la zone) pour augmenter la productivité de leurs entreprises de 5%⁴. Or moins de 2% des entreprises françaises connaissent une telle variation annuelle et, en moyenne, il faut plus de 10 ans pour qu'elles fassent l'expérience d'un tel changement. Cette moyenne masque cependant un effet concave de la taille des clusters. Ces conclusions sont intéressantes à plus d'un

4. Briant et Barbesol (2010), à l'aide d'une méthodologie différente, obtiennent une valeur de 2%, à emploi total donné.

titre pour les décideurs publics. Il semblerait que l'État ait du mal à identifier les secteurs, les régions ou les projets susceptibles de tirer la croissance future, et qu'il soit difficile d'engendrer des externalités technologiques à coups de subventions. Le marché pousse naturellement les entreprises à former des pôles, et les aides publiques ne jouent qu'un rôle secondaire dans les choix de localisation. Il semble donc assez vain de constituer une infrastructure administrative, forcément contraignante, en vue d'attirer les entreprises dans un cluster qu'elles n'auraient pas déjà choisi d'elles-mêmes, et qui ne résulterait pas de leurs choix de localisation optimaux (Martin *et al.* (2011b)). C'est aussi ce qui ressort de l'évaluation des Systèmes Productifs Locaux (SPL) menée par Martin *et al.* (2011a). La politique des SPL mise en œuvre en France en 1999 et 2000 était certes moins ambitieuse que celle des pôles de compétitivité, mais elle reposait sur des principes similaires. Or l'appartenance à un SPL n'a eu presque aucun effet visible sur la productivité des entreprises qui les ont rejoints. Les auteurs montrent même que les effets d'aubaines ont été forts : les SPL ont attiré des entreprises en déclin, à la recherche d'un second souffle, et ont souvent bénéficié aux groupements d'entreprises déjà constitués.

2.3.3 Proximité géographique, brevets et citations

Jaffe *et al.* (1993) et Jaffe et Trajtenberg (2002) procèdent différemment et ne cherchent pas à estimer une fonction de production d'innovation, en y incluant des retombées spatiales. Leur démarche consiste à comparer le lieu où les brevets sont déposés et le lieu où ces brevets sont exploités, en utilisant les citations de brevets. Le degré de concentration spatiale des citations est comparé à un taux de citations de référence reflétant la distribution spatiale des secteurs susceptibles d'utiliser les brevets, c'est à dire le taux de citations que l'on aurait observé en l'absence d'externalités technologiques. Les données américaines font apparaître que les citations de brevets sont significativement plus nombreuses que celles qui prévaudraient en l'absence de spillovers. Ce résultat est valide à l'échelle du pays, des états et des aires métropolitaines (les MSA). Les spillovers technologiques décroissent avec la durée de vie du brevet, mais le gradient est assez faible. À l'échelle des aires métropolitaines, le taux de citations des brevets de moins de 10 ans déposés par les grandes entreprises est de 21,9 (8,8 si on exclut les auto-citations), contre 3,6 pour la distribution de référence. En ce qui concerne les brevets de plus de 10 ans, le taux de citation est presque deux fois plus faible : il tombe à 13,3 (8,7 si on exclut les auto-citations), contre 1,3 pour la distribution de référence.

Ces résultats trouvent un écho favorable dans de très nombreuses études effectuées sur données américaines, comme par exemple Thompson (2006). En ce qui concerne l'Europe, l'étude de Maurseth et Verspagen (2002) est originale, car c'est l'une des premières à s'appuyer sur un modèle gravitaire pour étudier les citations de brevets réalisées par 112 régions européennes, sur la période 1979-1996. Deux conclusions intéressantes émergent. La distance réduit considérablement les flux des citations entre régions. De plus, les citations entre régions d'un même pays ou parlant la même langue, sont également plus élevées, à distance donnée. Pour ce qui a trait à la structure industrielle, les citations se produisent aussi plus souvent entre les régions spécialisées dans certains secteurs industriels, et notamment celles qui entretiennent des liens technologiques (évalués à l'aide d'une mesure de "distance technologique"). Des travaux plus spécifiques, consacrés par exemple au secteur des biotechnologies confirment le rôle important joué par la distance technologique, soit via les liens input-output des entreprises (Plunket (2009), pour la France) comme vecteur des externalités technologiques engendrées par la proximité spatiale, soit

via l'appartenance à un même domaine technologique (Plunket (2009), pour la région francilienne).

Fischer *et al.* (2006), Fischer *et al.* (2009a) et LeSage *et al.* (2007) confirment l'importance de cette proximité technologique en utilisant des modèles de Poisson et des techniques d'économétrie spatiale. Verspagen et Schoenmakers (2004) constatent également que les citations au sein des entreprises multinationales sont plus fréquentes lorsque les établissements sont plus proches dans l'espace, et cette tendance se retrouve également pour les multinationales.

Maggioni *et al.* (2007) étudient également le rôle joué par la distance physique *et* technologique dans la diffusion des spillovers, dans le cadre d'un modèle gravitaire appliqué à un ensemble de 109 régions européennes. Les deux types de distance affectent de manière significative le co-dépôt de brevets entre les régions : doubler la distance diminue de moitié le nombre de brevets co-déposés. Les auteurs trouvent également un effet négatif supplémentaire pour les régions qui ne sont pas contigües. Le degré de périphéralité géographique (mesuré par la distance à Bruxelles) n'a pas d'influence négative sur les brevets co-déposés. La co-appartenance à un réseau de recherche européen l'augmente en revanche de 0,3%.

Enfin, Fischer *et al.* (2009b) proposent un exercice comparable à celui mené par Jaffe *et al.* (1993) pour l'Europe. Ils illustrent tout d'abord le fait qu'il existe une tendance claire à la concentration spatiale des dépôts de brevets aux niveaux régional et national. Puis, ils comparant les lieux de dépôt et de citation des brevets, et montrent que les citations ont environ sept fois plus de chances de provenir de la même région, que les citations qui reflèteraient la distribution des secteurs utilisateurs (2,6 fois plus de chances si on exclut les auto-citations). Les citations ont également 2,7 fois plus de chances de provenir du même pays que les brevets d'origine (1,7 fois plus de chances si on exclut les auto-citations).

Conclusion : Quelles leçons pour le réseau du Grand Paris ?

De nombreuses études empiriques ont ainsi évalué l'ampleur des différentes sources d'économies d'agglomération. Les ordres de grandeur décrits dans ce rapport convergent pour les effets d'agglomération véhiculés par la densité : un doublement de cette variable permet d'accroître la productivité des entreprises et/ou les salaires de l'ordre de 1,4 à 2,5%. Les effets de composition de la main d'œuvre locale expliquent près de 50% des effets d'agglomération, et les biais d'endogénéité les atténuent d'environ 20%. Lorsque ni l'un ni l'autre ne sont pris en compte, la fourchette haute des estimations se situent donc de 3,5 à 6,5%. Les dynamiques locales (emploi, innovation) et les choix de localisation sont également très sensibles à la proximité spatiale, mais les ordres de grandeur diffèrent assez considérablement d'une étude à l'autre.

Ces éléments de quantification, qu'il est impossible d'extrapoler sans beaucoup de prudence, permettent néanmoins d'éclairer la question de l'impact d'une infrastructure comme celle du Grand Paris de plusieurs façons.

2.3.4 Les infrastructures de transport : créations ou déplacements d'activités ?

Tout d'abord, le fait qu'il existe une relation positive entre le dynamisme économique local et la densité des activités économiques (au sens large, qu'elle émane de la taille du marché local ou des marchés avoisinants), peut conduire à une sous-estimation de l'effet potentiellement bénéfique des infrastructures de transport (Venables (2007)). En effet, le modèle canonique de l'économie urbaine, tel que décrit par exemple par Fujita (1989), prédit que la baisse des coûts de transport est un des deux déterminants majeurs de l'augmentation de la taille des villes, avec la hausse des revenus⁵. Depuis le début des années 1990, de très nombreuses études empiriques ont tenté d'évaluer l'impact des infrastructures de transport sur la croissance des villes, sans toutefois parvenir à produire des résultats consensuels (Lafourcade et Mayer (2009)). L'étude de Duranton et Turner (2012), la plus aboutie dans ce domaine, estime qu'une hausse de 10% du stock d'autoroutes interfédérales des aires urbaines américaines (les MSA) aurait accru leur emploi de 1,5% sur la période 1983-2003. Ce gain, s'il paraît faible de prime abord, doit être réévalué à la lumière des économies d'agglomération additionnelles procurées par la hausse du volume d'emplois, tant en termes de densité, que de croissance locale.

Toutefois, la baisse des coûts de transports engendrée par la construction d'un projet d'investissement comme celui du Grand Paris est une arme à double tranchant. Les infrastructures de transport peuvent en effet tout aussi bien attirer de nouvelles activités économiques que les encourager à fuir vers les autres régions qu'elles contribuent à mieux connecter, comme l'a montré par exemple le modèle de Martin et Rogers (1995). Elles peuvent notamment contribuer au phénomène d'étalement urbain, souvent associé à une baisse de la densité de population au centre des villes et à une extension spatiale de la périphérie, caractérisée par le développement d'un habitat individuel plus fragmenté, et donc moins dense. Si la densité des activités est à l'origine d'importants effets d'agglomération, l'impact potentiellement bénéfique du Grand Paris est à nuancer. De récents travaux montrent effectivement que le développement des réseaux

5. La condition dite d'Alonso-Muth prédit en effet que la taille de la ville, mesurée par la distance séparant la frontière de la ville au centre (Central Business District ou CBD), où sont supposés concentrés tous les emplois, croît avec le revenu des habitants et décroît avec le coût de transport entre leur lieu de résidence et le CBD. Ce résultat découle de la condition d'équilibre des marchés fonciers à la frontière de la ville.

de transport est une des principales sources de l'étalement urbain. Les estimations de Baum-Snow (2007) et Baum-Snow (2010) montrent par exemple qu'aux États-Unis, chaque autoroute traversant le centre-ville aurait un impact marginal sur le taux de croissance de sa population de -18%. Si le système d'autoroutes interfédérales n'avait pas été construit, la population des centres-villes américains serait ainsi en moyenne supérieure de 8% à son chiffre actuel.

L'effet estimé des infrastructures ferroviaires est cependant différent de celui des autoroutes : selon Baum-Snow *et al.* (2011), si le réseau ferroviaire intra-urbain n'a aucun impact sur la population des villes chinoises, il contribue en revanche à en décentraliser les activités productives, en particulier la production manufacturière⁶. Plus intéressant dans la perspective du Grand Paris, la forme des réseaux de transport revêt une importance fondamentale. Les infrastructures en anneau autour des villes favoriseraient encore plus la décentralisation de leur production, en particulier de leur production manufacturière⁷.

2.3.5 L'effet du Grand Paris : un scénario... parmi d'autres possibles

Il est donc clair qu'une infrastructure comme le Grand Paris va entraîner des relocalisations d'entreprises et/ou de salariés qui vont directement affecter la densité des activités et donc, du fait des économies d'agglomération liées à la densité, la productivité de la région francilienne. Il est cependant possible de réaliser des simulations de l'effet que pourrait avoir le Grand Paris. Il est important de préciser que les chiffres que nous allons fournir ici n'ont *aucun souci de réalisme*, puisqu'ils sont fondés sur des hypothèses fictives concernant l'ampleur des relocalisations attendues. Le scénario que nous allons étudier permet cependant d'illustrer le type de conclusions qui pourraient être tirées d'une étude économétrique sérieuse de la question.

L'Île de France est subdivisée en 20 zones d'emploi qui ont été redéfinies en 2010 par l'INSEE, et qui sont illustrées sur la Figure 2.1.

Le premier constat que l'on peut faire est, qu'à l'heure actuelle, les disparités de densité entre les zones d'emploi franciliennes sont très marquées. Dans la zone d'emploi de Paris, la densité est supérieure à 5000 salariés par km². Elle est déjà environ dix fois plus faible dans les grandes zones d'emploi qui lui sont contiguës, à l'exception d'Orly où elle ne l'est que de trois fois. La densité peut être encore dix fois plus faible dans les zones d'emploi les plus périphériques par rapport à cette première ceinture.

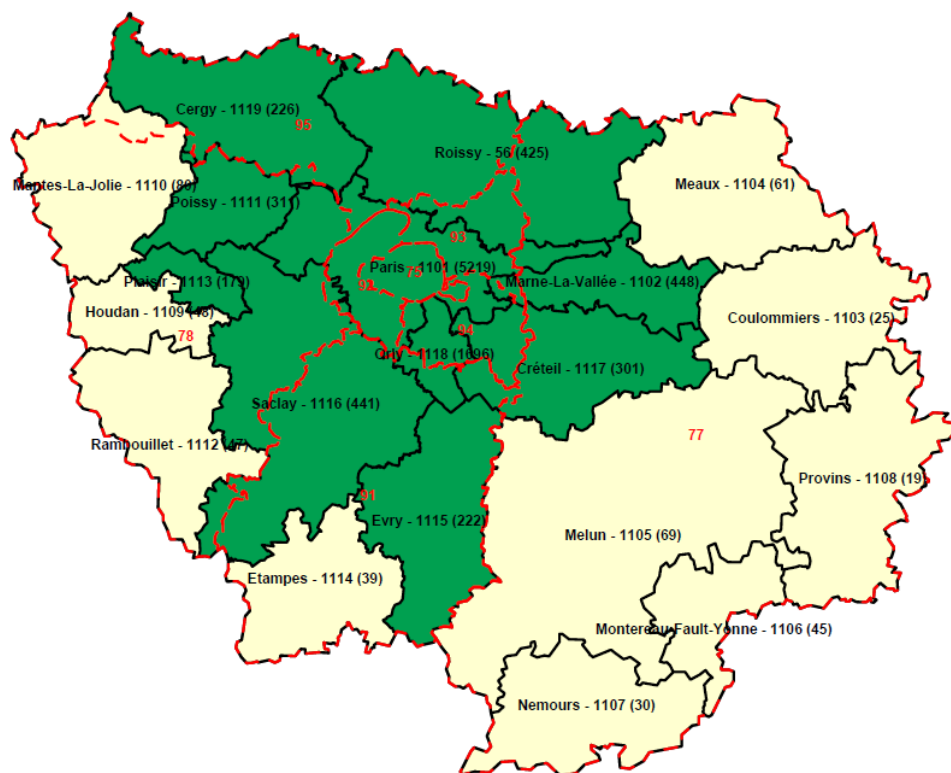
Ces différences de densité impliquent, toutes choses égales par ailleurs, de grandes différences de productivité pour les salariés, même pour ceux travaillant dans des zones d'emploi très proches. Si on prend comme référence une élasticité de la productivité à la densité égale à 0,02, qui correspond à l'estimation de Combes *et al.* (2008a) pour la France, un salarié qui se localiserait à Paris aurait une productivité supérieure de quasiment 6% à un salarié strictement identique, mais qui se localisait à Créteil, simplement parce que la densité y est 17 fois plus faible. S'il se localisait à Créteil, il aurait une productivité supérieure de quasiment 5% à celle d'un salarié identique se localisant à Nemours, qui est presque dix fois moins dense.

Si l'on compare les dix zones d'emploi franciliennes les plus denses (en vert sur la carte) aux dix zones

6. Chaque ligne ferroviaire additionnelle transitant par un centre-ville chinois réduirait le taux de croissance de son PIB total de 13%, et celui de sa production manufacturière de 20%.

7. D'après Baum-Snow *et al.* (2011) la construction de toute nouvelle rocade autour d'une ville chinoise aurait un impact de -24% sur le taux de croissance de la population du centre, de -0.57 log point de % sur le taux de croissance de son PIB et de -1 log point de % sur le taux de croissance de sa production industrielle.

FIGURE 2.1 – Les zones d’emploi franciliennes



Notes : (i) en trait noir épais, les frontières des zones d’emploi : le nom de chaque zone d’emploi figure avec son code et la densité d’emplois (en nombre d’actifs occupés par km² en 2007) est entre parenthèses ; (ii) en rouge, la frontière des départements, avec le numéro correspondant ; (iii) en vert, les 10 zones d’emploi les plus denses.

les moins denses, l'écart de densité est d'environ 1 à 20. Un même salarié qui se relocaliserait d'une zone d'emploi francilienne périphérique à une zone d'emploi centrale augmenterait donc sa productivité d'un peu plus de 6%. Sachant que les dix zones d'emploi centrales regroupent environ 5,205 millions de salariés, alors que les dix périphériques n'en regroupent que 325000, cet écart de productivité implique que le PIB de l'Île de France, soit 542 milliards d'euros en 2007, se ventile en 512 milliards d'euros pour les zones centrales et 30 milliards d'euros pour les zones périphériques.

Examinons le scénario suivant qui est, rappelons-le, totalement fictif : quel serait l'impact en termes de productivité et de PIB de l'arrivée de 500000 nouveaux salariés dans les 10 zones d'emploi centrales ?⁸ La densité d'emploi y progresserait donc de 9,6%. Ainsi, chaque salarié verrait sa productivité augmenter de 0,18%. Cela peut paraître faible. Néanmoins, appliqué à un PIB de 512 milliards d'euros, cela représente un gain de près d'1 milliard d'euros (940 millions plus précisément). De plus, cette augmentation de productivité, du fait de l'endogénéité des choix de localisation, va attirer de nouveaux salariés. Selon Combes *et al.* (2008a), cet afflux entraîne une augmentation de l'élasticité de la productivité à la densité de 20% : elle passe donc de 0,020 à 0,024. La hausse du PIB liée à l'augmentation de la densité due à l'arrivée de 500000 nouveaux salariés dans les zones centrales est donc de 1,128 milliards d'euro, effets direct et indirect compris. Et ce gain ne concerne que les salariés qui étaient auparavant dans les zones centrales.

Deux questions cruciales se posent ensuite. La première concerne l'origine des 500000 salariés, la seconde leur niveau de qualification. Pour ce qui est de la première question, vu la très faible densité d'activités dans les zones non centrales, ces 500000 salariés ne proviendraient vraisemblablement pas de l'Île-de-France. La densité moyenne d'emploi française hors Île-de-France est d'environ 40 actifs occupés par km², soit un écart de densité d'environ 24 avec les zones centrales franciliennes. Si les 500000 salariés proviennent d'une zone ayant la densité et la productivité moyenne française hors IDF (soit environ 63000 euros par salarié), leur productivité va augmenter de 6,6%. Sachant que le PIB hors IDF était de 1344,8 milliards en 2007, cela représente un gain de PIB considérable, légèrement supérieur à deux milliards d'euros. Ainsi, si l'accroissement de la densité est déjà profitable pour les salariés en place, il l'est encore bien plus pour les salariés provenant des autres régions françaises et ce second effet est relativement fort (et il s'agit bien de la simple variation de productivité des salariés en question, pas d'un déplacement de PIB d'une région française à une autre).

Une estimation de l'élasticité de la productivité à la densité permet donc de dégager des ordres de grandeur des gains à attendre des relocalisations d'activités induites par le Grand Paris. Cependant, les chiffres donnés plus haut ne peuvent être considérés sérieusement : une analyse plus sophistiquée doit avant être effectuée. Tout d'abord, en ce qui concerne l'origine des salariés, supposer qu'ils viennent de zones où la densité et la productivité reflètent la moyenne nationale est une hypothèse plutôt favorable. Il faudrait en amont prendre en compte les différences de mobilité des salariés des zones urbaines et rurales, afin de déterminer les gains des salariés ayant la plus forte probabilité de se relocaliser en Île-de-France. Typiquement, si les salariés proviennent de villes où la densité, et donc la productivité, sont bien plus élevées que la moyenne, les gains en pourcentage de productivité pour les arrivants seront plus faibles, car le différentiel de densité sera aussi plus faible. Mais, par ailleurs, comme ces salariés sont plus productifs, le gain de PIB apporté par chacun sera aussi plus élevé. Les gains des salariés en place resteront, en

8. Ce chiffre implique que le Grand Paris occasionnerait un afflux de travailleurs égal à environ 10% de la population active occupée de l'Île-de-France, ce qui correspond déjà à un choc massif.

revanche, les mêmes.

Dans le même ordre d'idées, le Grand Paris n'a pas vocation a priori à attirer les salariés dans Paris intra-muros. Ainsi, il ne faudrait pas considérer l'impact des nouveaux salariés en direction de toutes les zones d'emploi franciliennes les plus denses, et il faudrait exclure de ce groupe Paris intra-muros. Cela abaisserait le différentiel de densité entre les zones d'origine et de destination, et donc les gains de productivité, à la fois pour les salariés déjà présents et pour les nouveaux arrivants. De plus, on ne peut exclure le fait qu'une partie des 500000 nouveaux salariés dans cette zone ne provienne pas de Paris intra-muros, c'est à dire de zones plus denses que les zones d'attraction du Grand Paris. Pour ceux-là, cette relocalisation vers l'extérieur de Paris, moins dense, se traduirait par une baisse de la productivité, et donc des pertes de PIB qu'il faudrait déduire des calculs précédents. Finalement, et à l'inverse, les zones d'attraction du Grand Paris sont a priori plus restreintes que les 10 zones d'emploi que nous avons considérées, et les communes les plus périphériques pourraient en être exclues. Cela jouerait alors en sens inverse, puisqu'on exclut des pôles d'attraction du Grand Paris des zones peu productives.

Un deuxième aspect sur lequel insiste la littérature, et que nous avons largement discuté dans ce rapport, provient du fait que presque la moitié du différentiel de productivité entre les zones d'emploi denses et moins denses provient du fait que les salariés n'y ont pas les mêmes qualifications. Clairement, l'Île-de-France bénéficie d'une population extrêmement qualifiée. Or les calculs effectués plus haut ont supposé que les nouveaux arrivants gardaient leur qualification d'origine. Le problème est que la littérature n'est pas encore parvenue à déterminer si les zones les plus denses attireraient simplement des travailleurs plus qualifiés, auquel cas il n'y a pas de gain supplémentaire à attendre, ou si elles rendent les salariés plus qualifiés, par des effets d'apprentissage. Dans ce cas, les gains à attendre des relocalisations vers une zone plus dense pourraient être jusqu'à deux fois supérieurs.

Si l'on va dans cette direction, il faudrait néanmoins aussi commencer par affiner les mesures des élasticités de la productivité à la densité, en les évaluant pour chaque catégorie socioprofessionnelle prise séparément. Par exemple, Combes *et al.* (2012) montrent que les entreprises les plus efficaces ont des gains de productivité dus à la densité plus importants que celles qui sont moins efficaces. Les salariés qui sont les plus qualifiés devraient aussi avoir des gains plus élevés lorsqu'ils se relocalisent d'une zone à l'autre. Ainsi, les gains d'un projet comme le Grand Paris pourraient être évalués non seulement en fonction du volume d'emplois susceptibles de se relocaliser, mais aussi de sa composition en termes de catégories socioprofessionnelles.

Finalement, nous n'avons évoqué pour ces calculs que le rôle de la densité. Mais notre revue de littérature souligne aussi que le potentiel marchand, par exemple, a un impact positif sur la productivité, en sus de la densité. Le potentiel marchand étant plus élevé en Île-de-France qu'ailleurs, un accroissement du nombre de salariés représenterait de nouveaux gains, aussi bien pour les salariés en place que pour les nouveaux arrivants. De même, notre rapport a également montré que les salariés qualifiés exerçaient des externalités positives sur les autres salariés. Ainsi, les salariés qui viendraient se localiser dans le Grand Paris, zone nettement plus qualifiée en moyenne que le reste de la France, bénéficieraient de ces externalités, ce qui représenterait des additionnels.

En conclusion, il est encore utile d'insister sur le fait que *les calculs effectués dans cette conclusion ne sont pas des prévisions des effets d'agglomération à attendre du Grand Paris*. Ils illustrent seulement le fait que les estimations fournies par la littérature académique pourraient conduire,

après une réflexion sur la mise au point de scénarii de relocalisations raisonnables d'activités, à des fourchettes basses et hautes des gains de productivité engendrés par ces relocalisations. De plus, si la littérature mesurant de tels bénéfices a énormément progressé au cours de la dernière décennie, *les travaux empiriques s'intéressant aux coûts de l'urbanisation restent encore très embryonnaires*. De même, *les déterminants des choix de localisation des ménages et des entreprises sont encore peu évalués*, alors qu'ils sont au cœur de l'analyse de l'impact d'un projet de transport de grande envergure, comme celui du Grand Paris.

Bibliographie

- Zoltan J. ACS, David B. AUDRETSCH et Maryann P. FELDMAN : Real effects of academic research : Comment. *American Economic Review*, 82(1):363–367, 1992.
- Zoltan J. ACS, David B. AUDRETSCH et Maryann P. FELDMAN : R&D spillovers and recipient firm size. *Review of Economics and Statistics*, 100(1):15–29, 1994.
- James D. ADAMS : Comparative localization of academic and industrial spillovers. *Journal of Economic Geography*, 2(3):253–278, 2002.
- James D. ADAMS et Adam B. JAFFE : Bounding the effects of R&D : An investigation using matched establishment-firm data. *Rand Journal of Economics*, 27(4):700–721, 1996.
- Luc ANSELIN, Attila VARGA et Zoltan ACS : Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics*, 42(3):422 – 448, 1997.
- Josep-Maria ARAUZO-CAROD et Elisabet VILADECANS-MARSAL : Industrial location at the intra-metropolitan level : The role of agglomeration economies. *Regional Studies*, 43(4):545–558, 2009.
- Manuel ARELLANO et Steve BOND : Some tests of specification for panel data : Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2):277–297, 1991.
- Mohammad ARZAGHI et J. Vernon HENDERSON : Networking off Madison Avenue. *Review of Economic Studies*, 75(4):1011–1038, 2008.
- Chun-Chung AU et Vernon HENDERSON : Are Chinese cities too small? *Review of Economic Studies*, 73(3):549–576, 2006.
- David B. AUDRETSCH et Maryann P. FELDMAN : R&D spillovers and the geography of innovation and production. *American Economic Review*, 86(3):630–640, 1996.
- Marigee BACOLOD, Bernardo S. BLUM et William C. STRANGE : Skills in the city. *Journal of Urban Economics*, 65(2):136 – 153, 2009a.
- Marigee BACOLOD, Bernardo S. BLUM et William C. STRANGE : Urban interactions : soft skills versus specialization. *Journal of Economic Geography*, 9(2):227–262, 2009b.
- Salvador BARRIOS, Holger GÖRG et Eric STROBL : Multinationals’ location choice, agglomeration economies, and public incentives. *International Regional Science Review*, 29(1):81–107, 2006.
- Roberto BASILE : Acquisition versus greenfield investment : The location of foreign manufacturers in Italy. *Regional Science and Urban Economics*, 34(1):3 – 25, 2004.
- Roberto BASILE, Davide CASTELLANI et Antonello ZANFEI : Location choices of multinational firms in Europe : The role of EU cohesion policy. *Journal of International Economics*, 74(2):328 – 340, 2008.
- Roberto BASILE, Davide CASTELLANI et Antonello ZANFEI : National boundaries and the location of multinational firms in Europe. *Papers in Regional Science*, 88(4):733–748, 2009.
- Nathaniel BAUM-SNOW : Suburbanization and transportation in the monocentric model. *Journal of Urban Economics*, 62(3):405–423, 2007.
- Nathaniel BAUM-SNOW : Changes in transportation infrastructure and commuting patterns in u.s. metropolitan areas, 1960-2000. *American Economic Review Papers & Proceedings*, 100(2):378–382, 2010.

- Nathaniel BAUM-SNOW, Loren BRANDT, J. Vernon HENDERSON et Matthew A. TURNER : Roads, railways and decentralization of Chinese cities. *Processed, Brown University*, 2011.
- Uwe BLIEN et Jens SUEDEKUM : Local economic structure and industry development in germany, 1993-2001. *Economics Bulletin*, 17:1–8, 2005.
- Uwe BLIEN, Jens SUEDEKUM et Katja WOLF : Productivity and the density of economic activity. *Labour Economics*, 13(4):445–458, 2006.
- Eckhardt BODE : The spatial pattern of localized R&D spillovers : an empirical investigation for Germany. *Journal of Economic Geography*, 4(1):43–64, 2004.
- Laura BOTTAZZI et Giovanni PERI : Innovation and spillovers : Evidence from European regions. *European Economic Review*, 47:687–710, 2003.
- Anthony BRIANT et Yoann BARBESOL : Économies d’agglomération et productivité des entreprises : estimation sur données individuelles françaises. *Economie et Statistique*, (3):419–420, 2010.
- Marius BRÜLHART et Nicole A. MATHYS : Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics*, 38(4):348 – 362, 2008.
- C. M. BUCH, J. KLEINER, A. LIPPONER et Farid TOUBAL : Determinants and effects of foreign direct investment : evidence from german firm-level data. *Economic Policy*, 20:52 – 110, 2006.
- Gerald A. CARLINO, Satyajit CHATTERJEE et Robert M. HUNT : Urban density and the rate of invention. *Journal of Urban Economics*, 61(3):389 – 419, 2007.
- Dennis CARLTON : The location and employment choices of new firms : An econometric model with discrete and continuous endogenous variables. *Review of Economics and Statistics*, 65:440–449, 1983.
- Sylvie CHARLOT et Gilles DURANTON : Communication externalities in cities. *Journal of Urban Economics*, 56(3):581–613, 2004.
- Antonio CICCONE : Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227, 2002.
- Antonio CICCONE et Robert E. HALL : Productivity and the density of economic activity. *American Economic Review*, 86(1):54–70, 1996.
- Antonio CICCONE et Giovanni PERI : Identifying human capital externalities : Theory with an application to US cities. *Review of Economic Studies*, 73:381–412, 2006.
- Andrzej CIEŚLIK : Regional characteristics and the location of foreign firms within poland. *Applied Economics*, 37(8):863–874, 2005.
- Federico CINGANO et Fabiano SCHIVARDI : Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2(4):720–742, 2004.
- Pierre-Philippe COMBES : Economic structure and local growth : France, 1984–1993. *Journal of Urban Economics*, 47(3):329–355, 2000.
- Pierre-Philippe COMBES, Gilles DURANTON et Laurent GOBILLON : Spatial wage disparities : Sorting matters! *Journal of Urban Economics*, 63:723–742, 2008a.
- Pierre-Philippe COMBES, Gilles DURANTON et Laurent GOBILLON : The identification of agglomeration economies. *Journal of Economic Geography*, 11:253–266, 2011a.
- Pierre-Philippe COMBES, Gilles DURANTON, Laurent GOBILLON, Diego PUGA et Sébastien ROUX : The productivity advantages of large markets : Distinguishing agglomeration from firm selection. *Econometrica*, à paraître, 2012.
- Pierre-Philippe COMBES, Gilles DURANTON, Laurent GOBILLON et Sébastien ROUX : Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. GLAESER, éditeur : *Agglomeration Economics*, pages 15–65. Chicago University Press, Chicago, IL, 2010.
- Pierre-Philippe COMBES et Miren LAFOURCADE : Transport costs : Measures, determinants, and regional policy implications for France. *Journal of Economic Geography*, 5(3):319–349, 2005.

- Pierre-Philippe COMBES, Miren LAFOURCADE, Jacques-François THISSE et Jean-Claude TOUTAIN : The rise and fall of spatial inequalities in France : A long-run perspective. *Explorations in Economic History*, 48:243–271, 2011b.
- Pierre-Philippe COMBES, Thierry MAGNAC et Jean-Marc ROBIN : The dynamics of local employment in France. *Journal of Urban Economics*, 56(2):217–243, 2004.
- Pierre-Philippe COMBES, Thierry MAYER et Jacques-François THISSE : *Economic Geography : The integration of Regions and Nations*. Princeton University Press, New Jersey, 2008b.
- Maria Teresa COSTA-CAMPI, Agustí SEGARRA-BLASCO et Elisabet VILADECANS-MARSAL : The location of new firms and the life cycle of industries. *Small Business Economics*, 22(3-4):265–281, 2004.
- Matthieu CROZET et Miren LAFOURCADE : *La nouvelle économie géographique*. La Découverte, Paris, 2009.
- Matthieu CROZET, Thierry MAYER et Jean-Louis MUCCHIELLI : How do firms agglomerate ? A study of FDI in France. *Regional Science and Urban Economics*, 34(1):27 – 54, 2004.
- Klaus DESMET et Marcel FAFCHAMPS : Changes in the spatial concentration of employment across US counties : a sectoral analysis 1972-2000. *Journal of Economic Geography*, 5(3):261–284, 2005.
- Michael P. DEVEREUX, Rachel GRIFFITH et Helen SIMPSON : Firm location decisions, regional grants and agglomeration externalities. *Journal of Public Economics*, 91(3-4):413 – 435, 2007.
- Anne-Celia DISDIER et Thierry MAYER : How different is Eastern Europe ? structure and determinants of location choices by French firms in Eastern and Western Europe. *Journal of Comparative Economics*, 32(1):280 – 296, 2004.
- Avinash K. DIXIT et Joseph E. STIGLITZ : Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308, 1977.
- Gilles DURANTON, Philippe MARTIN, Thierry MAYER et Florian MAYNERIS : *The Economics of Clusters : Lessons from the French Experience*. Oxford University Press, New Jersey, 2010.
- Gilles DURANTON et Vassilis MONASTIRIOTIS : Mind the gaps : The evolution of regional earnings inequalities in the UK 1982–1997. *Journal of Regional Science*, 42(2):219–256, 2002.
- Gilles DURANTON et Matthew A. TURNER : Urban growth and transportation. 2012.
- Maryann FELDMAN : *The Geography of Innovation*. Kluwer Academic Publishers, Boston, 1994.
- Maryann P. FELDMAN et David B. AUDRETSCH : Innovation in cities : Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2):409–429, 1999.
- Octávio FIGUEIREDO, Paulo Guimar AES et Douglas WOODWARD : Home-field advantage : location decisions of portuguese entrepreneurs. *Journal of Urban Economics*, 52(2):341 – 361, 2002.
- Bernard FINGLETON : Increasing returns : evidence from local wage rates in great britain. *Oxford Economic Papers*, 55(4):716–739, 2003.
- Manfred FISCHER, Thomas SCHERNGELL et Eva JANSENBERGER : The geography of knowledge spillovers between high technology firms in Europe : Evidence from a spatial interaction modelling perspective. *Geographical Analysis*, 38:288–309, 2006.
- Manfred FISCHER, Thomas SCHERNGELL et Eva JANSENBERGER : Geographic localisation of knowledge spillovers : evidence from high-tech patent citations in Europe. *The Annals of Regional Science*, 43:839–858, 2009a.
- M.M. FISCHER, T. SCHERNGELL et E. JANSENBERGER : Patents, patent citations and the geography of knowledge spillovers in Europe. In C. KARLSSON, A.E. ANDERSSON, P.C. CHESHIRE et R.R. STOUGH, éditeurs : *New Directions in Regional Economic Development*, pages 331–345. Springer Verlag, Berlin, 2009b.
- Neil FOSTER et Robert STEHRER : Sectoral productivity, density and agglomeration in the Wider Europe. *Spatial Economic Analysis*, 4(4):427–446, 2009.

- Michaela FUCHS : The determinants of local employment dynamics in Western Germany. *Empirical Economics*, 40(1):177–203, 2011.
- Masahisa FUJITA : *Urban Economic Theory : Land Use and City Size*. Cambridge University Press, Cambridge, 1989.
- Edward L. GLAESER, Heidi KALLAL, José A. SCHEINKMAN et Andrei SCHLEIFER : Growth in cities. *Journal of Political Economy*, 100(6):1126–1152, 1992.
- Edward L. GLAESER et David C. MARÉ : Cities and skills. *Journal of Labor Economics*, 19(2):316–342, 2001.
- Edward L. GLAESER, Stuart S. ROSENTHAL et William C. STRANGE : Urban economics and entrepreneurship. *Journal of Urban Economics*, 67(1):1 – 14, 2010.
- Daniel J. GRAHAM : Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, 62(1):103–120, 2007.
- Daniel J. GRAHAM, Patricia S. MELO, Piyapong JIWATTANAKULPAISARN et Robert B. NOLAND : Testing for causality between productivity and agglomeration economies. *Journal of Regional Science*, 50(5):935–951, 2010.
- Lydia GREUNZ : Geographically and technologically mediated knowledge spillovers between European regions. *The Annals of Regional Science*, 37:657–680, 2003.
- Paulo GUIMARAES, Octávio FIGUEIREDO et Douglas WOODWARD : Agglomeration and the location of foreign direct investment in portugal. *Journal of Urban Economics*, 47(1):115 – 135, 2000.
- Gordon H. HANSON : Economic integration, intraindustry trade, and frontier regions. *European Economic Review*, 40(3–5):941–949, 1996.
- Gordon H. HANSON : Increasing returns, trade, and the regional structure of wages. *Economic Journal*, 107(440):113–133, 1997.
- Gordon H. HANSON : Market potential, increasing returns, and geographic concentration. *Journal of International Economics*, 67(1):1–24, 2005.
- Chauncy HARRIS : 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, 1954.
- C.Keith HEAD, John C. RIES et Deborah L. SWENSON : Attracting foreign manufacturing : Investment promotion and agglomeration. *Regional Science and Urban Economics*, 29(2):197 – 218, 1999.
- Keith HEAD et Thierry MAYER : Market potential and the location of japanese investment in the European union. *Review of Economics and Statistics*, 86(4):959–972, 2004.
- Keith HEAD et Thierry MAYER : Regional wage and employment responses to market potential in the eu. *Regional Science and Urban Economics*, 36(5):573–595, 2006.
- Judith K. HELLERSTEIN, David NEUMARK et Kenneth R. TROSKE : Wages, productivity, and worker characteristics : Evidence from plant-level production functions and wage equations. *Journal of Labour Economics*, 17(3):409–446, 1999.
- Elhanan HELPMAN : The size of regions. In David PINES, Efraim SADKA et Itzhak ZILCHA, éditeurs : *Topics in Public Economics. Theoretical and Applied Analysis*, pages 33–54. Cambridge University Press, New York, NY, 1998.
- J. Vernon HENDERSON : Externalities and industrial development. *Journal of Urban Economics*, 42(3):449–470, 1997.
- J. Vernon HENDERSON : Marshall’s economies. *Journal of Urban Economics*, 53(1):1–28, 2003.
- J. Vernon HENDERSON, Ari KUNCORO et Matt TURNER : Industrial development in cities. *Journal of Political Economy*, 103(5):1067–1090, 1995.

- Christian A. L. HILBER et Ioan VOICU : Agglomeration economies and the location of foreign direct investment : Empirical evidence from Romania. *Regional Studies*, 44(3):355–371, 2010.
- Jane JACOBS : *The Economy of Cities*. Random House, New York, 1969.
- Adam B. JAFFE : Real effects of academic research. *American Economic Review*, 79(5):957–970, 1989.
- Adam B. JAFFE et Manuel TRAJTENBERG : *Patents, Citations and Innovations : A Window on the Knowledge Economy*. MIT Press, Cambridge, Mass, 2002.
- Adam B. JAFFE, Manuel TRAJTENBERG et Rebecca HENDERSON : Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577–598, 1993.
- Stephan JURAJDA et Katherine TERRELL : Growth in cities. *Regional unemployment and human capital in transition economies*, 17(2):241–274, 2009.
- Paul R. KRUGMAN : Increasing returns and economic geography. *Journal of Political Economy*, 99(3):484–499, 1991.
- Paul R. KRUGMAN et Anthony J. VENABLES : Globalization and the inequality of nations. *Quarterly Journal of Economics*, 110(4):857–880, 1995.
- Miren LAFOURCADE et Thierry MAYER : Des routes pavées de bonnes intentions. *CEDD Références économiques*, 9:1–4, 2009.
- Miren LAFOURCADE et Elisenda PALUZIE : European integration, FDI and the geography of French trade. *Regional Studies*, 45(4):419–439, 2009.
- James P. LESAGE, Manfred M. FISCHER et Thomas SCHERNGELL : Knowledge spillovers across Europe : Evidence from a poisson spatial interaction model with spatial effects. *Papers in Regional Science*, 86(3):393–421, 2007.
- Robert E. Jr LUCAS : On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42, 1988.
- Mario A. MAGGIONI, Mario NOSVELLI et Teodora Erika UBERTI : Space versus networks in the geography of innovation : A European analysis. *Papers in Regional Science*, 86(3):471–493, 2007.
- Francesca MAMELI, Alessandra FAGGIAN et Philip MCCANN : Employment growth in italian local labour systems : Issues of model specification and sectoral aggregation. *Spatial Economic Analysis*, 3(3):343–360, 2008.
- Alfred MARSHALL : *Principles of Economics*. Macmillan, London, 1890.
- Philippe MARTIN, Thierry MAYER et Florian MAYNERIS : Public support to clusters : A firm level study of French local productive systems. *Regional Science and Urban Economics*, 41(2):108 – 123, 2011a.
- Philippe MARTIN, Thierry MAYER et Florian MAYNERIS : Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, 69(2):182–195, 2011b.
- Philippe MARTIN et Carol Ann ROGERS : Industrial location and public infrastructure. *Journal of International Economics*, 39(3–4):335–351, 1995.
- Julio MARTINEZ-GALARRAGA, Elisenda PALUZIE, JORDI et Daniel A. TIRADO : Agglomeration and labour productivity in spain over the long term. *Econometrica*, 2(3):195–212, 2008.
- Per Botolf MAURSETH et Bart VERSPAGEN : Knowledge spillovers in Europe : A patent citations analysis. *Scandinavian Journal of Economics*, 104(4):531–545, 2002.
- Daniel MCFADDEN : Econometric models for probabilistic choice among products. *Journal of Business*, 53(3):S13–S29, 1980.
- Dennis MCFADDEN : Conditional logit analysis of qualitative choice behavior. In Paul ZAREMBKA, éditeur : *Frontier in Econometrics*, pages 105–142. Academic Press, New York, 1974.

- Patricia C. MELO, Daniel J. GRAHAM et Robert B. NOLAND : A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342, 2009.
- Giacinto MICUCCI et Valter di GIACINTO : The producer service sector in Italy : Long-term growth and its local determinants. *Spatial Economic Analysis*, 4(4):391–425, 2009.
- Giordano MION : Spatial externalities and empirical analysis : The case of Italy. *Journal of Urban Economics*, 56(1):97–118, 2004.
- Giordano MION et Paolo NATICCHIONI : The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, 42:28–55, 2009.
- Enrico MORETTI : Human capital externalities in cities. In Vernon HENDERSON et Jacques-François THISSE, éditeurs : *Handbook of Regional and Urban Economics*, volume 4, pages 2243–2291. North-Holland, Amsterdam, 2004a.
- Enrico MORETTI : Workers' education, spillovers, and productivity : Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690, 2004b.
- NATIONS UNIES : *World Urbanization Prospects : The 2007 Revision*. Nations Unies, New York, 2008.
- Ariel PAKES et Zvi GRILICHES : Patents and r&d at the firm level : A first look. In Zvi GRILICHES, éditeur : *R&D, Patents, and Productivity*. University of Chicago Press, Chicago, 1984.
- Olivier PARENT et Stéphane RIOU : Bayesian analysis of knowledge spillovers in European regions. *Journal of Regional Science*, 45(4):747–775, 2005.
- Anne PLUNKET : Firms' inventiveness and localized vertical r&d spillovers. *Journal of Innovation Economics*, 23(4):147–170, 2009.
- Michael E. PORTER : *The Competitive Advantage of Nations*. Free Press, New York, 1990.
- Fazia PUSTERLA et Laura RESMINI : Where do foreign firms locate in transition countries ? An empirical investigation. *The Annals of Regional Science*, 41:835–856, 2007.
- Stephen REDDING et Daniel STURM : The costs of remoteness : Evidence from German division and reunification. *American Economic Review*, 98(5):1766–1797, 2008.
- Stephen REDDING et Anthony J. VENABLES : Economic geography and international inequality. *Journal of International Economics*, 62(1):63–82, 2004.
- Patricia RICE, Anthony J. VENABLES et Eleonora PATACCHINI : Spatial determinants of productivity : Analysis for the regions of Great Britain. *Regional Science and Urban Economics*, 36(6):727–752, 2006.
- Jennifer ROBACK : Wages, rents and the quality of life. *Journal of Political Economy*, 90(6):1257–1278, 1982.
- Stuart S. ROSENTHAL et William C. STRANGE : Geography, industrial agglomeration, and agglomeration. *Review of Economics and Statistics*, 85(2):377–393, 2003.
- Stuart S. ROSENTHAL et William C. STRANGE : Evidence on the nature and sources of agglomeration economies. In Vernon HENDERSON et Jacques-François THISSE, éditeurs : *Handbook of Regional and Urban Economics*, volume 4, pages 2119–2171. North-Holland, Amsterdam, 2004.
- Stuart S. ROSENTHAL et William C. STRANGE : The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389, 2008.
- Paul SAMUELSON : The transfer problem and transport costs : Analysis of effects of trade impediments. *Economic Journal*, 64:264–289, 1954.
- AnnaLee SAXENIAN : *Regional Advantage : Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, MA, 1994.
- Albert SOLÉ-OLLÉ et Elisabet VILADECANS-MARSAL : Central cities as engines of metropolitan area growth. *Journal of Regional Science*, 44(2):321–350, 2004.

- Julia SPIES : Network and border effects : Where do foreign multinationals locate in Germany? *Regional Science and Urban Economics*, 40(1):20 – 32, 2010.
- Daniel F. SPULBER : *Global Competitive Strategy*. Cambridge University Press, Cambridge, 2007.
- Peter THOMPSON : Patent citations and the geography of knowledge spillovers : Evidence from inventor- and examiner-added citations. *Review of Economics and Statistics*, 88(2):383–389, 2006.
- Kenneth TRAIN : *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, MA, 2003.
- Anthony J. VENABLES : Evaluating urban transport improvements : cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of transport economics and policy*, 41(2):173–188, 2007.
- Bart VERSPAGEN et Wilfred SCHOENMAKERS : The spatial dimension of patenting by multinational firms in Europe. *Journal of Economic Geography*, 4(1):23–42, 2004.
- Elisabet VILADECANS-MARSAL : Agglomeration economies and industrial location : City-level evidence. *Journal of Economic Geography*, 5(4):565–582, 2004.
- J VÉRON : *L'Urbanisation du monde*. La Découverte, Collection Repères, Paris, 2006.