

Économies d'agglomération et productivité des entreprises : estimation sur données individuelles françaises

Yoann Barbesol* et Anthony Briant**

La concentration spatiale persistante de certains secteurs d'activité s'explique difficilement sans admettre l'existence d'économies d'agglomération. L'estimation quantitative de l'ampleur de ces économies d'agglomération a récemment connu un regain d'intérêt du fait de la disponibilité croissante de données fines au niveau des entreprises. En utilisant trois fichiers administratifs concernant la comptabilité et l'emploi des entreprises françaises, nous mesurons dans cet article l'impact sur la productivité globale des facteurs de ces entreprises des *externalités d'urbanisation*, à savoir celles liées à la taille globale, à l'accessibilité et à la diversité industrielle de la zone d'implantation, et des *externalités de localisation*, qui se rapportent au niveau de spécialisation d'une zone d'activité.

Notre étude souligne tout d'abord la prégnance des premières : nous concluons à l'existence d'un impact positif et significatif de l'accessibilité d'une zone au reste du marché national ainsi que de la densité de son tissu économique sur la productivité moyenne des entreprises. À ces externalités d'urbanisation, s'ajoutent des effets locaux très significatifs de la spécialisation, suggérant l'existence d'externalités de localisation. Ainsi les entreprises sont, en moyenne, plus productives dans les zones où leur industrie est relativement plus concentrée.

En revanche, la diversité des activités économiques locales n'influerait que très peu sur la productivité. De même, le degré de concurrence locale n'aurait, en moyenne, pas d'effets sur la productivité des entreprises d'un *cluster*.

L'étude fine des voies de production, de diffusion et de captation de ces externalités constituerait certainement une extension utile à la nôtre ; nous suggérons un premier élément de réponse sans toutefois pouvoir conclure. Nous trouvons en effet que les agents qualifiés semblent jouer un rôle essentiel dans ces trois processus : une fois que l'on a tenu compte de la qualité de leur main-d'œuvre, les firmes les plus productives demeurent bien celles qui évoluent dans un environnement où la main-d'œuvre locale est la plus qualifiée, suggérant que les employés qualifiés sont les plus à même de capter et de diffuser des externalités de production. Toutefois la causalité de cette relation reste difficile à certifier.

* Insee, Dese/DEEE/MSE, 15, Boulevard Gabriel-Péri BP100 92244 Malakoff Cedex, yoann.barbesol@insee.fr

** École d'Économie de Paris / Paris-Jourdan Sciences Économiques (UMR 8545 CNRS-EHESS-ENPC-ENS), 48 Boulevard Jourdan, 75014 Paris, France, anthony.briant@m4x.org

Nous remercions Pierre-Philippe Combes, Laurent Gobillon, Simon Quantin et Sébastien Roux pour leurs conseils et commentaires, ainsi que Pauline Givord, Anne Epaulard et deux rapporteurs anonymes pour leur lecture attentive d'une première version de cet article. Anthony Briant remercie les membres du Département des Études Économiques d'Ensemble pour leur accueil durant la rédaction de cet article. Nous remercions également les participants au séminaire D3E du 10 décembre 2007 ainsi que ceux du séminaire Fourgeaud du 8 avril 2009 pour leurs remarques et commentaires, et plus particulièrement Miren Lafourcade pour avoir accepté de discuter une première version de cet article et Pierre Joly, rapporteur du séminaire Fourgeaud. Nous avons en outre bénéficié du travail de Sébastien Roux et de Simon Quantin pour la construction de la base de données appariée utilisée dans l'article : nous les remercions pour nous avoir aidés à comprendre comment les variables de cette base ont été construites.

La concentration spatiale des activités économiques, et plus particulièrement de certains secteurs d'activité, est un phénomène observé et étudié dans de nombreux pays (1). Si certaines activités sont contraintes dans leur lieu de production par des facteurs naturels (2), ces derniers ne sauraient expliquer l'ensemble des choix de localisation des entreprises. L'existence de cette agglomération spatiale peut paraître d'autant plus surprenante qu'elle génère des surcoûts pour l'entreprise en termes de loyers ou de salaires plus élevés. La permanence de zones, pour la plupart urbaines, où l'activité économique est fortement concentrée suppose donc l'existence de bénéfices à l'agglomération, d'économies d'agglomération. Comme l'a montré Starrett (1978), dans un espace homogène et avec des coûts à l'échange positifs, ces économies d'agglomération reposent sur l'existence d'une forme ou d'une autre d'indivisibilité, donc de rendements croissants. Elles se répartissent entre externalités pécuniaires, associées à la baisse des prix des produits et des facteurs de production, et externalités pures de production, qui renvoient aux gains de productivité.

Réduire les coûts de déplacement des biens, des hommes et des idées

Les fondements microéconomiques de l'agglomération des entreprises et des travailleurs sont multiples (3). Depuis les travaux d'Alfred Marshall (1890), il est coutume de penser que l'agglomération permet de réduire les coûts de déplacement des biens, des personnes ou des idées.

Dans un marché local plus large, il est possible pour une entreprise de se procurer une plus grande variété de consommations intermédiaires, générant des gains de productivité par une plus grande désintégration verticale et une plus grande spécialisation. La disponibilité d'une plus grande variété de biens de consommation attire également les consommateurs. Face à des rendements croissants, les entreprises profitent de la présence d'un nombre plus important de consommateurs facilement accessibles.

Dans un bassin d'emploi plus large, les travailleurs ont tendance à être plus spécialisés et donc plus productifs : face à un plus grand nombre de possibilités d'embauche, ils n'hésitent pas à se spécialiser. Une entreprise a donc plus

de facilité à trouver de bons candidats pour ces postes vacants : l'appariement entre les offres et demandes d'emploi est plus simple et a plus de chances d'être de meilleure qualité.

Enfin, tout échange d'information entre les entreprises, concernant les caractéristiques de la demande ou bien les technologies de production, est plus aisé lorsque la densité des acteurs économiques est plus élevée. Ainsi, la création, l'accumulation et la diffusion des connaissances se trouvent facilitées dans les zones les plus densément peuplées. Pour toutes ces raisons, les entreprises ont tendance à s'agglomérer.

À l'inverse, cette agglomération peut être la source de surcoûts. D'une part, le coût des facteurs de production, du travail, du capital ou bien du terrain, est plus élevé dans les zones où l'activité économique est concentrée. D'autre part, des externalités négatives de production, telles que la congestion des réseaux de transport ou toute forme de pollution, sont susceptibles de réduire la productivité des entreprises. Enfin, dans les zones de forte concentration d'un secteur, il est possible que la concurrence soit plus âpre.

Externalités d'urbanisation ou externalités de localisation ?

Si l'existence de bénéfices à l'agglomération des activités économiques est bien établie sur le plan théorique, la mesure empirique de leur ampleur est plus délicate. Un tel travail empirique se heurte d'abord à un problème d'identification. Il est effectivement difficile à partir d'une seule observable, la productivité des entreprises ou le salaire des travailleurs, de déterminer lequel des mécanismes évoqués plus haut est effectivement à l'œuvre (4).

Face à cette difficulté, l'ambition des travaux empiriques a donc été plus modeste. Les auteurs ont cherché à évaluer l'environnement écono-

1. Cf. Barlet, Briant et Crusson (2008) pour la France, Duranton et Overman (2005) pour le Royaume-Uni et Ellison et Glaeser (1997) pour les États-Unis par exemple.

2. Ellison et Glaeser (1999) estiment qu'au moins 20 % de la concentration géographique mesurée aux États-Unis peut s'expliquer par la présence d'avantages naturels, à entendre dans un sens large (ressources primaires, coût des facteurs, notamment travail, peu élevé, etc.).

3. Duranton et Puga (2004) offrent une revue détaillée de la littérature théorique sur ces fondements microéconomiques.

4. Rosenthal et Strange (2004) proposent une revue détaillée de la littérature sur l'identification et l'estimation des externalités d'agglomération.

mique le plus profitable à l'entreprise, en distinguant les externalités *d'urbanisation* des externalités *de localisation*, sans chercher à isoler un mécanisme en particulier.

La question est alors de savoir si la productivité d'une entreprise est plus grande lorsqu'elle se localise à proximité d'autres entreprises du même secteur d'activité (*externalités de localisation* ou de Marshall-Arrow-Romer) ou bien d'entreprises de secteurs d'activité différents et variés (*externalités d'urbanisation* ou de Jacobs). Dans le premier cas, les externalités ne sont supposées jouer qu'à l'intérieur d'un secteur, on parle donc d'*externalités intra-sectorielles*. Les entreprises profitent alors d'une plus grande spécialisation de leur zone d'implantation, conforme à leur activité. En revanche, concernant les externalités d'urbanisation, les gains de productivité proviennent des phénomènes de fertilisation croisée entre entreprises de secteurs différents (*externalités inter-sectorielles*). Ces dernières profitent alors d'un marché local plus large, présentant une plus grande diversité d'activités économiques. Soulignons que ces deux catégories ne sont pas exclusives l'une de l'autre, même si leurs effets sont étudiés séparément, et souvent mis en comparaison. Dans les deux cas, les effets mesurés sont des effets nets, où externalités positives et négatives se compensent partiellement. Ainsi, si la spécialisation sectorielle d'une zone peut permettre le partage des gains liés à une plus grande variété de consommations intermédiaires, elle est aussi susceptible de se traduire par une plus grande concurrence sur le marché du bien final. L'estimation ne mesure, en tout état de cause, que la résultante de ces deux effets contradictoires.

Mesurer l'ampleur des externalités d'urbanisation et de localisation sur la productivité des entreprises françaises

Notre article s'inscrit dans le prolongement de cette littérature empirique, encore peu fournie. En effet, si l'étude des externalités remonte à l'essor des sociétés industrielles et s'est enrichie depuis de nouveaux apports théoriques qui ont permis d'irriguer la réflexion sur l'organisation et la concentration des activités sur le territoire, la validation quantitative de ces idées a longtemps souffert de l'absence de données au niveau individuel. Faute de mieux, cette dernière a d'abord reposé sur l'analyse de données agrégées (5).

Cet obstacle à une analyse quantitative fine a été levé il y a peu. À partir de données américaines individuelles désormais accessibles, des économistes (Henderson, 2003) ont mené de nouvelles études quantitatives, jetant un éclairage neuf sur la thématique des externalités d'agglomération. Les données administratives françaises, pourtant plus riches par certains égards que les bases américaines, ont encore été peu exploitées dans ce cadre. Nous inspirant notamment de l'exemple de Combes, Duranton, Gobillon et Roux (2008), nous avons tenté de les mettre à profit (6) (cf. encadré) pour comparer *externalités d'urbanisation* et *externalités de localisation*. Autrement dit, nous cherchons à expliquer, pour un secteur donné, les disparités de productivité entre entreprises situées dans des zones d'emploi différentes par la taille et l'accessibilité de la zone d'une part et par l'organisation du secteur considéré dans la zone d'autre part.

Revue de la littérature

Comme le soulignent Rosenthal et Strange (2004), si des externalités de production existent, elles affectent la productivité globale des facteurs de l'entreprise, c'est-à-dire qu'elles permettent d'accroître la production, à ratio capital/travail donné.

Henderson (2003) est le premier à étudier l'impact des externalités d'agglomération sur la productivité globale des facteurs des entreprises. Il estime, sur données américaines, des fonctions de production au niveau de l'établissement en y introduisant des indicateurs d'agglomération (nombre de firmes du même secteur dans la ville, indice de diversité sectorielle dans la ville, etc.). Il montre notamment qu'un accroissement de 10 % du nombre d'établissements du même secteur dans la même ville accroît de 10 à 15 % la production d'un établissement préexistant dans ce secteur, à capital, travail et consommations intermédiaires inchangés. Par

5. L'estimation de l'ampleur des externalités d'agglomération reposait alors sur l'étude de la croissance de l'emploi local (cf. Glaeser, Kallal, Scheinkman et Schleifer (1992) et Henderson, Kuncoro et Turner (1995) pour des travaux fondateurs et Combes (2000a) puis Combes, Magnac et Robin (2004) pour des travaux sur données françaises).

6. Dans une perspective différente de la nôtre, Duranton, Martin, Mayer et Mayneris (2008) proposent quelques résultats sur l'impact des politiques de clusters (du type des pôles de compétitivité) sur la productivité des entreprises les constituant. Leurs données sont celles des Enquêtes Annuelles d'Entreprises pour les années 1996 à 2004. Martin, Mayer et Mayneris (2008) étendent ces résultats sur les mêmes données.

DONNÉES

Dans cette étude, nous utilisons la même base de données que Combes, Duranton, Gobillon et Roux (2008), qui résulte, pour chaque année de 1994 à 2004, de la fusion de trois fichiers administratifs : la base de données *Sirene*, les fichiers de déclarations fiscales d'entreprises (RSI/BRN), les fichiers de déclaration de données sociales (DADS) au niveau établissement.

La base de données *Sirene* (*Système d'Identification du Répertoire des ENTreprises*) contiennent les identifiants des firmes et des établissements (respectivement les codes Siren et Nic), des informations sur leur classification économique (activité principale (APE) dans la catégorie NAF), ainsi que des informations sur la localisation géographique (code géographique). Le champ du répertoire utilisé pour cette étude est le champ ICS (Industrie, Commerce et Services). En sont exclues les activités agricoles, les activités financières et l'administration.

Les fichiers RSI (*Régime Simplifié d'Imposition*) et BRN (*Bénéfices Régime réel Normal*) contiennent les déclarations annuelles de comptes d'entreprise à l'administration fiscale. L'information concernant la production et le capital des entreprises est issue de ces fichiers fiscaux. Nous restreignons cette étude aux entreprises déclarant des impôts aux titres des bénéficiers industriels et commerciaux (BIC). Sont donc exclues les entreprises avec des revenus agricoles (RA) ou des bénéficiers non commerciaux (BNC). Seules celles remplissant les formulaires du Régime Normal (BRN) ou du Régime Simplifié (RSI) sont présentes. Les plus petites structures payant des impôts sous le régime allégé sont écartées. Cet ensemble représente près de 98 % du chiffre d'affaire des entreprises françaises et environ 65 % du nombre total des entreprises (données 2003). Les fichiers fiscaux nous permettent de reconstruire le chiffre d'affaire, la valeur ajoutée, le capital ainsi que l'investissement des entreprises. Rappelons encore que ces données ne sont disponibles qu'au niveau de l'entreprise et non pas au niveau de l'établissement.

Les fichiers DADS (*Déclarations Annuelles de Données Sociales*) permettent de reconstituer toutes les informations sur l'emploi de chaque établissement employant au moins un salarié. Les employeurs sont en effet tenus, annuellement et pour chaque salarié, de fournir un certain nombre d'informations : le nombre de jours travaillés, le nombre d'heures travaillées, sa rémunération et d'autres caractéristiques personnelles (sexe, âge, catégorie socioprofessionnelle). Ce fichier a été réorganisé afin de disposer d'informations sur l'emploi au niveau de l'établissement (nombre d'heures travaillées, masse salariale) par catégorie de qualification.

La base de données *Sirene* contient près de 3 millions d'établissements chaque année, tandis que le fichier DADS regroupe en moyenne 1,6 million d'établissements tous les ans. L'appariement des deux conduit à l'élimination des établissements sans salariés qui

sont en effet absents des fichiers DADS, ainsi que des établissements engagés dans les activités financières ou immobilières, absents eux de la base de données *Sirene*. Le panel qui en résulte est ensuite fusionné aux données fiscales RSI/BRN, qui contiennent eux entre 1,6 et 2 millions d'observations (firmes) par an. Dans la base de données issue de ce nouvel appariement ont disparu les firmes non soumises à l'impôt sur les sociétés (coopératives, associations, etc.) qui ne figurent pas dans les fichiers fiscaux.

Nous utilisons également les inventaires communaux pour l'année 1998 qui nous fournissent un certain nombre d'informations sur les caractéristiques géographiques, ainsi que sur les dotations, des communes françaises.

Un certain nombre d'opérations de sélection et de correction ont été menées afin d'extraire du fichier brut issu de l'appariement *Sirene*-BRN/RSI-DADS, une base de données propre à l'estimation. Plus précisément, nous construisons deux panels : un panel d'entreprises sur lequel sont menées les estimations ; un panel d'établissements servant à définir les différentes variables d'agglomération.

Traitements initiaux des données

Tout d'abord, les observations présentant des valeurs aberrantes pour une des variables clés (valeur ajoutée, capital, emploi) ont été supprimées. Nous supprimons également les quelques entreprises hors champs ICS (Industrie, Commerce, Services).

Ensuite, le suivi longitudinal des entreprises est particulièrement difficile. Il se peut que des firmes entrent et sortent du panel, sans que cela corresponde à une véritable création ou destruction d'entreprise (fusions et acquisitions, changement de localisation géographique, etc.). Lorsqu'une firme est observée de manière discontinue, nous conservons séparément l'ensemble des périodes continues d'observation. L'identifiant d'une entreprise devient donc son identifiant *Siren* couplé à l'année de début de la période d'observation (ci-après couple (*Siren*, début)). Nous disposons ainsi de 1 773 341 identifiants d'entreprise (*Siren*), correspondant à 2 198 714 périodes d'observations différentes, *i.e.* couples (*Siren*, début).

Il se peut en outre qu'une entreprise change de secteur d'activité principale au cours d'une période d'observation continue. Si, sur la période d'observation, une firme change plus de deux fois de NAF220, nous la supprimons. Ensuite, pour chacune de ces périodes, nous affectons donc chaque firme dans le secteur où elle est présente le plus longtemps. Il est difficile d'avoir des explications toujours très pertinentes sur ces changements de secteurs d'activité. Il est certain qu'une firme est rarement mono-active. Les variations dans l'activité déclarée ne sont certainement pas toujours dues à un changement drastique dans l'activité



ailleurs, il ne trouve pas d'effets significatifs des externalités d'urbanisation, autrement dit de la taille globale de la ville sur la productivité des établissements. D'un point de vue méthodologique, il contrôle l'hétérogénéité individuelle inobservée par l'introduction d'effets fixes individuels dans les fonctions de produc-

tion, contraignant ainsi son estimation à reposer sur les variations temporelles par période de 5 ans entre 1972 et 1992.

La disponibilité croissante de données individuelles d'entreprises a permis à l'étude d'Henrierson (2003) d'être reproduite sur d'autres

Encadré (suite)

des entreprises, mais plus à un rééquilibrage dans les différentes activités exercées. Nous atteignons là une des limites de la classification des entreprises par secteur d'activité, limite qu'il nous semble difficile de dépasser. Nous disposons alors d'un panel de 2 184 811 couples (*Siren*, début) (1 762 367 entreprises), correspondant à 9 186 699 observations *entreprise-année*.

Le choix du panel des établissements

Nous conservons l'ensemble des établissements correspondant à ces 1 762 367 entreprises. Cela représente 10 646 945 observations établissement-année.

Le choix du panel des entreprises

Comme nous ne nous intéressons pas aux entrées ou sorties de firmes d'un marché, nous ne gardons dans notre panel que les périodes d'observation supérieures ou égales à 2 ans consécutifs. De plus, les estimations de productivité à la Olley et Pakes (1996) et Levinsohn et Petrin (2003) nécessitent d'avoir des données retardées d'investissement ou de consommations intermédiaires non disponibles pour les entreprises présentes une année seulement.

Nous disposons de deux données d'emploi : l'une issue des BRN, l'effectif salarié directeur, l'autre des DADS, le nombre d'heures totales travaillées dans l'entreprise. La première mesure est peu satisfaisante, et parfois sans lien avec les données d'heures travaillées. Nous calculons donc un effectif salarié en équivalent temps plein en divisant le nombre d'heures travaillées par 1650. Cette valeur correspond à la moyenne de la durée annuelle de travail des salariés à temps complet (hors enseignants) calculée à partir des enquêtes *Emploi* 2003 et 2004 (Insee, 2006). Les entreprises employant strictement moins de cinq salariés ont été retirées du panel. Ce choix est important, puisqu'il conduit à la réduction de plus de moitié de la taille de notre échantillon. Néanmoins, ces entreprises représentent moins de 10 % de l'emploi total et de la valeur ajoutée chaque année. Ce processus de tri ne vise à éliminer que des données susceptibles de brouter nos estimations. Il est nécessaire de rappeler que nos indices d'agglomération ont été établis à partir du panel complet des établissements (entreprises de moins de 5 salariés comprises).

Nous ne conservons ensuite dans notre panel que les entreprises qui restent mono-établissement au cours

de la période d'observation. En terme économique, cette sélection est plus dommageable que la sélection précédente. Si les entreprises mono-établissement représentent environ 90 % du total des entreprises présentes dans l'échantillon, elles ne comptent que pour 50 à 60 % de l'emploi total et environ 50 % de la valeur ajoutée chaque année. Cette sélection est néanmoins justifiée pour les raisons suivantes. Les données issues des BRN-RSI ne permettent pas de calculer une productivité au niveau de l'établissement, dont seules les données d'emploi sont connues. Les données de production, de capital et de consommations intermédiaires ne sont disponibles qu'au niveau de l'entreprise. Or une même entreprise peut posséder plusieurs établissements, il est alors impossible de localiser de manière satisfaisante sa production et son capital. Quel est alors l'environnement industriel de la firme à considérer ? Celui de son siège, de son établissement le plus grand ? Ou bien une moyenne des environnements auxquels font face chacun de ces établissements ? Nous testerons la robustesse de nos résultats à cette sélection (cf. Tableau 8).

La sélection des entreprises mono-établissement permet de plus d'éviter un écueil dans les données : celui du dégroupage-regroupage. Pour des soucis de simplicité de déclaration, certaines entreprises pluri-établissements sont autorisées à ne déclarer l'emploi que dans un seul établissement. Dans ce cas, la localisation géographique de l'emploi des établissements est impossible (ce problème est également susceptible d'influencer le calcul de nos indicateurs d'agglomération. Tant que ce problème de dégroupage-regroupage est aléatoire dans l'espace, il ne biaise pas nos résultats). C'est donc pour cette raison que nous faisons le choix de restreindre notre échantillon aux entreprises mono-établissement uniquement. Plus précisément, nous nous restreignons aux entreprises qui restent mono-établissement tout au long de leur période d'observation. Désormais la localisation des facteurs et de la productivité ne pose plus de difficultés particulières.

Nous ne conservons enfin que 181 secteurs d'activités (en NAF220) qui correspondent, pour une année donnée, à l'activité principale d'au moins 100 firmes. Ces firmes ne sont présentes que dans un petit nombre de *clusters*. Au total, notre base de données contient 465 981 firmes, renvoyant à 3 242 626 observations *entreprise-année*. Ces firmes sont des firmes mono-établissement, présentes dans la base au moins deux ans consécutivement, ayant un effectif (en équivalent temps plein) supérieur à 5 salariés.

pays pour d'autres périodes. Ainsi Cingano et Schivardi (2004) proposent une étude à partir de données italiennes. Ils estiment d'abord une fonction de production par entreprise. Ils calculent ensuite un taux de croissance moyen sur la période 1986-1998 de la productivité moyenne dans un secteur et une zone donnés. Ils régressent en deuxième étape ces taux de croissance de la productivité sectorielle locale sur des variables d'agglomération pour l'année initiale. Ils montrent que la taille du marché local et la spécialisation de la zone sont les principaux déterminants de la croissance de la productivité sectorielle locale. Les autres déterminants envisageables, à savoir la diversité du tissu économique local ou le degré de concurrence sectorielle locale, n'ont pas d'impacts significatifs.

Une difficulté importante dans cette littérature empirique concerne les possibles biais d'endogénéité. Combes, Duranton, Gobillon et Roux (2008) s'interrogent sur le sens de la causalité entre productivité et densité des activités économiques. Les entreprises sont-elles effectivement plus productives dans les villes les plus denses ou bien les villes les plus productives deviennent-elles plus grandes ? Ces auteurs utilisent un ensemble de variables historiques et géologiques comme sources de variation dans les choix de localisation des entreprises, exogènes à la productivité contemporaine. Suivant la spécification et les instruments retenus, l'élasticité de la productivité à la densité varie entre 1,4 et 4,6 %.

Ces auteurs se concentrent sur les effets d'urbanisation et ne fournissent pas de résultats pour les externalités de localisation. Dans cette étude, nous étendons leurs résultats dans deux directions.

- Nous estimons les effets de localisation conjointement avec la mesure des effets d'urbanisation.

- Nous complétons leur étude par une analyse sectorielle. Il n'y a effectivement aucune raison de penser que les externalités d'urbanisation, comme les externalités de localisation, jouent avec la même force dans les différents secteurs.

Ces extensions ont cependant un coût : celui du manque d'instruments. En effet, nous ne disposons pas des instruments que ces auteurs proposent. De plus, si ceux-ci permettent d'in-

strumenter les variables d'urbanisation, les instruments pour les variables de localisation font encore défaut.

Stratégie d'estimation et biais potentiels

La littérature propose deux stratégies d'estimation des externalités d'agglomération à partir de données individuelles d'entreprises.

La première stratégie, en une étape, consiste à introduire directement les indicateurs d'agglomération dans la fonction de production. Henderson (2003) met en œuvre cette stratégie (sur 3 000 entreprises environ) en introduisant en plus un effet fixe individuel. Ce dernier permet de tenir compte de l'hétérogénéité individuelle inobservable, et donc d'éliminer le biais qui résulterait de l'omission de certaines variables individuelles. Il estime des coefficients technologiques propres à chaque secteur d'activité. Cette stratégie pose plusieurs difficultés. Avec un nombre important de données, l'estimation conjointe de coefficients technologiques au niveau sectoriel et d'effets fixes individuels est difficile à mettre en œuvre. L'identification ensuite repose sur les variations inter-temporelles qui s'avèrent très faibles pour les variables d'urbanisation. Enfin, si les externalités d'agglomération contribuent en partie à l'effet fixe individuel, les estimations peuvent être biaisées à la baisse.

Nous privilégions donc une stratégie alternative qui vise, dans une première étape, à construire un indicateur de la productivité sectorielle locale à partir de l'estimation de la productivité individuelle de chaque firme ; puis, dans une seconde étape, à régresser cette productivité sectorielle locale sur des variables d'agglomération. Cette stratégie a l'avantage d'une plus grande souplesse dans la première étape d'estimation de la fonction de production. Cependant, si une corrélation existe entre les variables d'agglomération et les variables individuelles (observables ou inobservables), la seconde étape peut-être biaisée (biais de causalité inverse ou biais de sélection spatiale). C'est la raison pour laquelle Combes, Duranton, Gobillon et Roux (2008) instrumentent cette deuxième étape.

La productivité sectorielle locale est calculée au niveau des *clusters* ; ils correspondent à un secteur en NAF220 (*s*), dans une zone d'emploi *z*, une année *t* donnée. Un *cluster* est donc défini et indicé par un triplet *zst*.

Première étape : estimer une fonction de production au niveau de l'entreprise

Nous estimons donc dans une première étape une fonction de production au niveau de l'entreprise i , de secteur d'activité s (7), implantée dans la zone d'emploi z (8) :

$$va_{it} = \theta_{st} + \alpha_s l_{it} + \sum_{q=2,3} \gamma_s^q sh_{iq,t} + \beta_s k_{it} + u_{it} \quad (1)$$

Les variables observables individuelles considérées dans la fonction de production sont les suivantes :

- le logarithme de la valeur ajoutée va_{it} ;
- l'emploi l_{it} mesuré par le logarithme du nombre d'heures travaillées ;
- la part $sh_{iq,t}$ des heures travaillées par les employés de type q (9) ;
- le logarithme du stock de capital k_{it} (10).

Les coefficients technologiques, *i.e.* les élasticités de la valeur ajoutée au travail et au capital, sont supposés être spécifiques au secteur d'activité de la firme en NAF220.

L'effet fixe secteur-temps θ_{st} nous permet de tenir compte de tous les chocs macroéconomiques sectoriels et non locaux. Il élimine donc de notre mesure de productivité tous les déterminants spécifiques au secteur d'activité, et notamment les effets de croissance technologique différenciés entre secteurs.

La productivité de l'entreprise est alors mesurée par le résidu estimé u_{it} de l'équation précédente.

Corriger du potentiel biais de simultanéité

Nous estimons dans un premier temps cette fonction de production par Moindres Carrés Ordinaires (MCO). Cependant, l'estimation d'une telle fonction de production est potentiellement soumise à plusieurs sources d'erreurs comme le soulignent Griliches et Mairesse (1995).

La simultanéité entre la productivité et le choix des inputs est très certainement la principale source de biais dans cette première étape. Ce biais est présent s'il existe des déterminants

inobservables de la productivité de l'entreprise qui influencent le choix des facteurs de production (par exemple l'âge de la firme, sa structure financière, son appartenance à un groupe, les qualités d'encadrement de sa direction, etc.) ou bien si la firme subit un choc de productivité, partiellement anticipé par l'entrepreneur qui influence le choix du niveau d'emploi et de capital. Ce choc n'étant pas observable pour l'économètre, l'erreur est alors mécaniquement corrélée aux variables explicatives individuelles.

La littérature a proposé différentes solutions pour corriger ce biais. Nous en retenons deux dans ce qui suit : la méthode d'Olley et Pakes (1996) (OP) et celle de Levinsohn et Petrin (2003) (LP) (pour une revue de littérature sur les fonctions de production et les grandes familles d'estimateurs les concernant, cf. Akerberg, Benkard, Berry et Pakes, 2007).

Ces méthodes sont dites *méthodes à indicateurs*. Olley et Pakes (1996) et Levinsohn et Petrin (2003) proposent de trouver des indicateurs (*proxies*) au choc anticipé de productivité. Olley et Pakes (1996) supposent que ce choc anticipé de productivité suit un processus de Markov et, qu'à capital donné, il détermine de façon unique le niveau des investissements. Levinsohn et Petrin (2003) font l'hypothèse qu'à capital donné, ce choc détermine uniquement le niveau des consommations intermédiaires. Dans les deux cas, les auteurs utilisent ces relations entre productivité et investissement (ou

7. Les secteurs considérés dans cette étude sont ceux de la Nomenclature d'Activité Française à 3 chiffres (224 groupes pour la NAF 2003), dite NAF220. Le système statistique public dispose de deux nomenclatures d'activité différentes : la Nomenclature d'Activité Française (NAF) et la Nomenclature d'Économique de Synthèse (NES). Cette dernière vise à refléter, autant que possible, le comportement d'agents confrontés à leur marché, alors que la NAF fait intervenir d'autres critères tels que les spécificités techniques de production ou l'organisation en filière de production. Il nous semble que la NAF est donc mieux adaptée à l'estimation d'une fonction de production.

8. Les zones d'emploi ont été définies par l'Insee en 1983 (révisées en 1994 et 1999). Il s'agit de zones géographiques à l'intérieur desquelles la plupart des actifs travaillent et résident. Leur contour a donc été déterminé de manière à réduire les migrations domicile-travail trans-frontalières. La taille moyenne d'une zone d'emploi est de 1 570 km². La plus petite zone d'emploi (Vitry-sur-Seine, 45 km²) est environ 140 fois plus petite que la plus grande zone d'emploi (Toulouse, 6 208 km²).

9. Ces catégories, introduites afin de contrôler de la qualité de la main-d'œuvre employée par la firme, sont définies au moyen de la Nomenclature des Professions et Catégories Socioprofessionnelles (CSP). Suivant Burnod et Chenu (2001), les CSP permettent de regrouper (partiellement) les travailleurs en fonction de leur niveau de qualification. On définit ainsi trois catégories de travailleurs : ($q = 3$) renvoient aux travailleurs très qualifiés, ($q = 2$) aux travailleurs qualifiés, enfin ($q = 1$) rassemble les travailleurs non-qualifiés, catégorie prise comme référence dans notre estimation.

10. Dans notre base de données, cette mesure est la somme des actifs tangibles et intangibles mesurés au coût historique, actifs financiers exclus.

consommations intermédiaires) pour approximer le choc anticipé.

Notons que les *méthodes à effets fixes* sont également utilisées pour corriger du biais de simultanéité. L'introduction d'un effet fixe individuel dans la fonction de production est censée capter l'ensemble des inobservables corrélées aux choix des facteurs. Cela revient également à supposer que le choc anticipé de productivité est fixe dans le temps. L'estimation d'un modèle à effet fixe repose alors sur l'estimateur intra-, l'estimateur en première différence ou en différence longue suivant les hypothèses faites sur le choc idiosyncratique. Cependant, du fait d'un ajustement lent du stock de capital, le coefficient relatif à cette variable est généralement très faible. Henderson (2003) qui utilise cette méthode trouve un coefficient compris entre 0,02 et 0,1 (11). Nous n'avons pas retenu cette approche par la suite.

Mesure de productivité sectorielle locale

Nous disposons donc de trois mesures individuelles de productivité, chacune renvoyant à une méthode d'estimation particulière (MCO, OP, LP). Nous calculons ensuite, à partir de chacune de ces trois mesures individuelles, une productivité moyenne par *cluster* « zone-secteur-temps » (TFP_{zst}). Celle-ci est égale à la moyenne (12) des productivités individuelles :

$$TFP_{zst} = \frac{1}{N_{zst}} * \sum_{i \in (z,s,t)} \hat{u}_{it} \quad (2)$$

où N_{zst} est le nombre d'établissements dans le *cluster* zst .

Les trois méthodes précédentes ne permettent pas forcément de tenir compte de l'hétérogénéité

sectorielle locale inobservable. Pour cette raison, nous développons une quatrième méthode d'estimation en introduisant un effet fixe zone-secteur-temps (F_{zst}) dès la première étape (équation 3). Dans ce cas, les méthodes d'estimation proposées par Olley et Pakes (1996) et Levinsohn et Petrin (2003) étant difficiles à manipuler avec autant d'effets fixes, nous estimons la fonction de production par MCO. Cet effet fixe F_{zst} constitue un quatrième indicateur de productivité sectorielle locale.

$$v_{ait} = \theta_{st} + \alpha_s l_{it} + \sum_{q \in 2,3} \gamma_s^q . sh_{iqt} + \beta_s k_{it} + F_{zst} + u_{it} \quad (3)$$

Nous disposons à ce stade de quatre mesures de productivité sectorielle locale (TFP_{zst}), les trois premières obtenues à partir de mesures de productivités individuelles, la quatrième estimée directement au niveau du *cluster*. Sachant qu'aucune des méthodes économétriques utilisées dans l'estimation de la fonction de production n'est complètement satisfaisante, elles nous permettent surtout de tester la sensibilité de nos résultats aux différents biais. Ces quatre mesures apparaissent fortement corrélées, bien qu'elles ne corrigent pas toutes les mêmes biais (cf. tableau 1).

La distribution des productivités est très comparable d'une méthode économétrique à une autre (cf. tableau 2). L'introduction dans la première étape des effets fixes secteur-année implique

11. Pour relâcher l'hypothèse de fixité, Blundell et Bond (2000) proposent d'ajouter un processus AR(1). Ils estiment alors un modèle dynamique de panel. Les variables retardées de capital et de travail sont utilisées comme instruments, ce qui est valide sous certaines hypothèses sur la corrélation temporelle des termes idiosyncratiques.

12. Nous avons également testé la robustesse de nos résultats à l'utilisation d'une moyenne pondérée par l'emploi.

Tableau 1
Corrélation entre mesures de productivité sectorielle locale

	La productivité TFP_{zst} est évaluée par...			
	... MCO	... la méthode d'Olley et Pakes	... la méthode de Levinsohn et Petrin	... la méthode à effet fixe
MCO	1			
Olley et Pakes	0,83	1		
Levinsohn et Petrin	0,92	0,80	1	
Effet fixe	0,96	0,79	0,88	1

Lecture : toutes les corrélations sont significatives au seuil de 1 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Le nombre d'observations est de 242 178.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

que ces mesures de la productivité soient centrées sur zéro. Rappelons donc que l'identification repose dans cette étude sur la variabilité entre zones d'emploi pour un secteur et une année donnés.

Notons également que le nombre de *clusters* est plus petit pour les méthodes d'Olley et Pakes (1996) et Levinsohn et Petrin (2003). Pour ces deux méthodes, l'échantillon des entreprises est en effet plus restreint. La méthode d'Olley et Pakes, par exemple, nécessite d'utiliser les données d'investissement. Nous ne pouvons donc calculer une productivité que pour les années à partir de 1995, réduisant d'autant le nombre de *clusters* étudiés.

La deuxième étape : expliquer les différences de productivité moyenne entre *clusters*

Soient donc :

- θ_{st} est un jeu d'indicatrices secteur-année (13) ;
- URB_{zt} et LOC_{zst} respectivement des indicateurs d'externalités d'urbanisation et de localisation ;
- X_{zt} un ensemble de caractéristiques de la zone d'emploi susceptibles d'accroître la productivité des entreprises qui y sont localisées, mais non liées à une quelconque externalité. Nous estimons donc :

$$TFP_{zst} = \theta_{st} + URB_{zt} \cdot \beta + LOC_{zst} \cdot \gamma + X_{zt} \cdot \rho + v_{zst} \quad (4)$$

Les variables d'urbanisation, présentées plus en détails ci-dessous sont : la densité en emploi de la zone ; sa superficie ; son potentiel marchand ; la diversité de ses activités. Concernant les variables de localisation, il s'agit de : la spécialisation ; la dotation locale en qualifications ; le nombre d'établissements du *cluster*.

Les élasticités β et γ sont globales. Selon cette hypothèse, les externalités d'agglomération (d'urbanisation et de localisation) jouent, dans un premier temps, avec la même force dans tous les secteurs. Cette hypothèse sera levée par la suite, dans le cadre d'une analyse sectorielle.

Nous estimons cette deuxième étape par moindres carrés pondérés, les poids étant le nombre d'entreprises par *cluster* dans l'échantillon. Ceci permet de donner plus de poids aux *clusters* où la productivité sectorielle locale est la mieux estimée (14).

Cette seconde étape souffre également d'un certain nombre de biais potentiels.

Biais de variables omises : on peut imaginer qu'un certain nombre de dotations locales (ressources naturelles, infrastructures publiques...), spécifiques à un secteur ou non, accroissent la productivité des entreprises de la zone, sans pour autant renvoyer à une quelconque forme d'externalités.

13. Ces indicatrices permettent de soustraire aux variables explicatives leurs moyennes par secteur et année.

14. Nous avons essayé cette régression sans poids, ou avec des poids différents (par exemple, la part de l'emploi des entreprises de l'échantillon dans l'emploi total de la zone d'emploi). Cela ne change pas les résultats.

Tableau 2
Statistiques descriptives portant sur les logarithmes des productivités sectorielles locales TFP_{zst}

	La productivité TFP_{zst} est évaluée par...			
	... MCO	... la méthode d'Olley et Pakes	... la méthode de Levinsohn et Petrin	... la méthode à effet fixe
Nombre	311 698	242 209	311 643	311 698
Moyenne	- 0,04	- 0,04	- 0,05	0
Écart-type	0,35	0,34	0,38	0,36
Minimum	- 6,90	- 7,78	- 7,28	- 7,35
25 ^e percentile	- 0,17	- 0,18	- 0,20	- 0,13
Médiane	- 0,03	- 0,03	- 0,04	0,01
75 ^e percentile	0,11	0,11	0,12	0,15
Maximum	5,32	5,13	4,78	14,25

Lecture : Nombre : nombre de clusters « secteur – zone d'emploi – date » (*zst*) présents dans l'échantillon ; le nombre théorique possible est $181 \cdot 341 \cdot 11 = 678\ 931$.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

La théorie traditionnelle du commerce international (Heckscher-Ohlin) repose sur l'existence de telles différences de dotations, et conclut à la spécialisation des pays ou des régions. Des considérations institutionnelles, notamment fiscales, peuvent également influencer les choix de localisation des entreprises (Rathelot et Sillard (2008) pour la France et Duranton, Gobillon et Overman (2007) pour le Royaume-Uni). Néanmoins, les différentiels de fiscalité locale n'influencent pas directement les différentiels de productivité.

Néanmoins, pour tenir compte de ces différences de dotations, nous introduisons dans nos régressions un ensemble de variables propres à chaque zone d'emploi, susceptibles d'influencer les performances des entreprises. Trouver de telles variables est une question délicate. Dans notre étude, nous utilisons des caractéristiques communales, fixes dans le temps (15). Il s'agit des caractéristiques suivantes : être au bord de la mer, être dans un massif montagneux, posséder ou non des monuments historiques ou culturels, et posséder un point d'eau (plan d'eau, étang, etc.). Enfin, nous introduisons une mesure de la distance moyenne de la zone d'emploi à la bretelle d'autoroute la plus proche. Cette dernière, renvoyant à l'existence d'infrastructures publiques favorisant la productivité des entreprises, peut s'avérer endogène. Dans les tableaux de résultats, l'introduction de ces variables apparaît sous la mention « Caractéristiques des zones d'emploi ».

Choix endogène de localisation : il peut exister des variables inobservables individuelles, corrélées à la productivité, qui influencent le choix de localisation des entreprises. Il s'agit alors d'une forme de (auto-)sélection des firmes dans l'espace susceptible de biaiser les résultats de la deuxième étape. À titre d'exemple, il est possible que les entrepreneurs les plus talentueux se regroupent spatialement pour des raisons sans rapport avec les externalités d'agglomération (qualité de vie, offre culturelle locale, etc.). Dans ce cas, les entreprises les plus productives seront agglomérées sans que cela ne soit la conséquence d'externalités positives d'agglomération (cf. Nocke (2006) pour un modèle théorique). Nous ne traitons pas ce problème de (auto-)sélection spatiale dans cette étude. Combes, Duranton et Gobillon (2008) proposent une analyse détaillée de ces phénomènes d'auto-sélection dans l'espace des travailleurs en fonction de leurs qualifications, dans le cadre d'une analyse des disparités spatiales de salaires en France.

Sélection dans le marché : Melitz et Ottaviano (2008) proposent un modèle dans lequel, en

présence d'un coût fixe d'entrée sur le marché et de concurrence entre des entrants hétérogènes (en termes de productivité), seules les firmes les plus productives survivent. Les auteurs montrent que ce mécanisme d'éviction des entreprises les moins productives est d'autant plus fort que le marché est de grande taille. Il introduit in fine une corrélation positive entre taille du marché et productivité, susceptible de biaiser à la hausse l'effet des externalités d'urbanisation.

Combes, Duranton, Gobillon et Roux (2008) utilisent les variables retardées (de 150 ans) de densité en emploi et de potentiel marchand, ainsi qu'un ensemble de caractéristiques géologiques des sous-sols pour corriger des biais de variables omises dans une estimation instrumentée, par double moindres carrés. La logique est simple : ces variables sont censées déterminer les choix de localisation actuelle des entreprises sans pour autant affecter leur productivité. Notons que le biais de sélection spatiale est plus difficile à corriger. Dans un article récent, Combes, Duranton, Gobillon, Puga et Roux (2009) développent une méthodologie originale pour distinguer les phénomènes d'agglomération des phénomènes de sélection dans le marché.

Dans cette étude, nous ne disposons pas des instruments historiques et géologiques utilisés par ces auteurs. Néanmoins, une solution simple pour corriger partiellement cette hétérogénéité spatiale inobservable est d'introduire des indicatrices régionales, au-delà des caractéristiques propres à chaque zone d'emploi et fixes dans le temps présentées plus haut. Notre intuition est qu'une partie des phénomènes d'auto-sélection spatiale ou de sélection dans le marché se jouent à un niveau spatial supérieur à celui de la zone d'emploi, par exemple la région. De plus, les différentes zones d'emploi d'une même région française sont susceptibles d'avoir des caractéristiques, en termes d'avantages naturels, assez proches. Cette solution ne règle pas complètement tous les problèmes d'endogénéité évoqués plus haut, mais nous semble néanmoins être un bon compromis.

Sources et ampleur des externalités d'urbanisation

La première question est donc de savoir si la productivité des entreprises est plus forte

15. Issues des inventaires communaux, pour l'année 1998. Nous calculons la moyenne de ces variables au niveau de la zone d'emploi, pondérée par la taille respective des communes.

là où les activités économiques sont davantage agglomérées. Pour ce faire, nous régressons la productivité sectorielle locale sur la densité en emploi dans la zone d'emploi, définie pour une zone d'emploi donnée z à la date t par :

$$\text{Densité en emploi}_{zt} = \frac{\text{Emploi}_{zt}}{\text{Superficie}_z} \quad (5)$$

où :

- Emploi_{zt} est le nombre de salariés mesuré en équivalent temps plein (ETP) (16) dans la zone z à la date t ;

- Superficie_z est la surface de cette zone en km^2 .

En 2004, la moyenne de la densité en emploi s'établit à 60 salariés en équivalent temps plein (ETP) par km^2 , mesure cohérente avec la densité de population en France, environ 112 habitants par km^2 (cf. tableau 3, ligne *Densité en emploi*). De manière non surprenante, l'emploi apparaît

très inégalement réparti sur le territoire. Pour preuve, la moitié des zones d'emploi françaises en 2004 ont une densité inférieure à 12,1 salariés ETP par km^2 .

Les zones d'emploi ont avant tout été définies par l'Insee sur un critère économique, à savoir un espace géographique à l'intérieur duquel la plupart des actifs résident et travaillent (17). Les zones d'emploi sont par conséquent de taille assez variable (cf. tableau 3, ligne *Superficie*). D'ailleurs, les zones d'emploi les plus denses sont systématiquement les plus petites. La corrélation entre le logarithme de ces deux variables s'établit à - 0,62 dans nos données. Ainsi à densité égale, les différences dans l'étendue des marchés locaux peuvent être importantes, nous jugeons donc nécessaire, en complément de la variable de densité, d'introduire la superficie de

16. Plus précisément, nous disposons du nombre d'heures totales travaillées dans la zone d'emploi (cf. encadré) que nous divisons par 1 650 (Insee, 2006).

17. Voir pour plus de détails la rubrique Définitions et Méthodes du site internet de l'Insee.

Tableau 3
Statistiques descriptives sur les indicateurs d'agglomération
(externalités d'urbanisation et de localisation)

	Moyenne	Écart-type	Minimum	25 ^e percentile	Médiane	75 ^e percentile	Maximum
Externalités d'urbanisation							
Emploi	3 1675,4	5 7021,8	1 891,6	9 047,2	16 895,1	33 173,6	693 283,8
Nombre d'établissements	3 069,6	5 243,5	234	1 041	1 721	3 081	72 706
<i>Densité en emploi</i>	59,4	382,8	0,94	6,9	12,1	24,4	6 577,6
<i>Superficie</i>	1 569,7	986,7	44,9	837,5	1 420,9	2 066,5	6 207,7
<i>Potentiel marchand</i>	119,8	192,8	34,1	53,7	67,8	103,8	1 948,1
<i>Diversité de la zone d'emploi</i> (évaluée au niveau NAF220)	29,1	8,6	5,7	24,0	30,0	35,3	51,5
<i>Diversité de la zone d'emploi</i> (évaluée au niveau NAF60)	12,4	2,6	3,9	10,9	12,5	14,2	18,7
Externalités de localisation (évaluées au niveau NAF220)							
Emploi	363,3	1 210,9	0,09	30,4	96,9	303,7	69 256,5
Nombre d'établissements	36,3	131,7	1	3	9	28	8 340
<i>Spécialisation</i>	1,79	6,27	0,0	0,44	0,87	1,51	574,52
<i>Dotation locale en qualifications</i>	0,91	0,47	0,00	0,64	0,87	1,10	11,81
Externalités de localisation (évaluées au niveau NAF60)							
Emploi	985,7	3 433,6	0,1	75,7	264,9	844,2	183 433,8
Nombre d'établissements	96,2	339,8	1	6	18	71	14 184
<i>Spécialisation</i>	1,25	2,25	0	0,42	0,82	1,33	101,37
<i>Dotation locale en qualifications</i>	0,87	0,38	0	0,65	0,85	1,04	5,55

Lecture : les variables sont présentées dans le corps du texte. Les statistiques descriptives sont données pour l'année 2004. Les variables pour les externalités d'urbanisation sont calculées sur les 341 zones d'emploi métropolitaines françaises. Les variables pour les externalités de localisation (au niveau NAF220) sont calculées sur 27 943 clusters pour l'année 2004. Les variables pour les externalités de localisation (au niveau NAF60) sont calculées sur 10 809 clusters pour l'année 2004.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

la zone d'emploi comme variable supplémentaire. Cela a l'avantage de faciliter l'interprétation : nous pouvons ainsi étudier l'accroissement de la densité, à superficie donnée. D'un point de vue économétrique, cette spécification est équivalente à l'introduction simultanée du nombre de salariés dans la zone ($Emploi_{zt}$) et de la superficie.

Nous régressons donc le logarithme de la productivité sectorielle locale (dans ses 4 évaluations) sur les logarithmes de la densité en emploi et de la superficie de la zone d'emploi (cf. tableau 4). Dans la partie du haut, les régressions incorporent des indicatrices « secteur-année » et les « contrôles zone » qui informent donc des dotations factorielles (avantages naturels) d'une zone d'emploi. L'identification repose alors sur les différences (en coupe) entre zones d'emploi pour un secteur et une année donnés. Dans la partie du bas, nous ajoutons des indicatrices régionales : l'identification repose alors sur les comparaisons, pour un secteur et

une année donnés, entre zones d'emploi d'une même région.

Les externalités d'urbanisation dépendent en premier lieu de la taille et de la densité de l'économie locale

Quelle que soit la mesure de la productivité retenue, les résultats sont convergents et de l'ordre de 3 à 4 % sans les indicatrices régionales et de 2 à 3 % avec les indicatrices régionales. Ces résultats sont en premier lieu conformes à ceux de Combes, Duranton, Gobillon et Roux (2008) qui régressent une productivité locale, non spécifique à un secteur, sur la densité en emploi dans la zone d'emploi. Ces auteurs trouvent une élasticité variant entre 1,4 % et 4,6 % suivant les spécifications et instruments utilisés. Réintroduire l'hétérogénéité sectorielle ne change donc pas le résultat moyen. De plus, bien que nous n'instrumentions pas cette deuxième étape, le biais, s'il existe, reste faible. Remarquons que Melo,

Tableau 4

Les entreprises sont plus productives dans les zones d'activité économique les plus denses

	La productivité TFP_{zst} est évaluée par...			
	... MCO	... la méthode d'Olley et Pakes	... la méthode de Levinsohn et Petrin	... la méthode à effet fixe
Estimation sans indicatrices régionales				
Ln Densité en emploi _{zt}	0,033*** (0,004)	0,035*** (0,004)	0,041*** (0,005)	0,036*** (0,004)
Ln Superficie _z	- 0,002 (0,004)	0,0004 (0,004)	0,001 (0,005)	- 0,002 (0,004)
Nombre d'observations	311 698	242 209	311 643	311 698
R ² ajusté	0,129	0,127	0,16	0,205
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui
Estimation avec indicatrices régionales				
Ln Densité en emploi _{zt}	0,022*** (0,003)	0,026*** (0,003)	0,03*** (0,004)	0,025*** (0,003)
Ln Superficie _z	0,011*** (0,003)	0,014*** (0,003)	0,018*** (0,004)	0,013*** (0,003)
Nombre d'observations	311 698	242 209	311 643	311 698
R ² ajusté	0,151	0,145	0,182	0,225
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par 4 méthodes) sur les logarithmes de la densité en emploi et de la superficie de la zone d'emploi, avec introduction d'indicatrices propres à chaque couple secteur-année, ainsi que des caractéristiques géographiques de zones (présentées dans le corps du texte). Les deux régressions diffèrent par l'introduction ou non d'indicatrices propres à chaque région française. Les coefficients estimés se lisent comme des élasticités.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi.

***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

Graham et Noland (2009) concluent également dans leur méta-analyse que les biais d'endogénéité sont généralement de faibles ampleurs dans les études internationales sur le sujet. Notons enfin que ces effets sont robustes à l'exclusion de l'échantillon de la région Île-de-France.

Cette élasticité de 3 % signifie qu'un doublement de la densité en emploi (qui correspond environ au déplacement d'un écart-type dans la distribution du logarithme de cette variable) correspond à un accroissement de 3 % de la productivité des entreprises, à secteur donné. Cet effet est donc loin d'être négligeable, lorsqu'on le compare à la progression annuelle moyenne de la productivité des entreprises françaises : 2,2 % par an en moyenne sur la période 1993-1999 (Crépon et Duhautois, 2003).

Ces résultats sont en accord avec la littérature sur le sujet. Ainsi Rosenthal et Strange (2004) concluent que, quel que soit le pays ou la période considérée, l'ampleur des externalités d'urbanisation s'établit dans un segment 3 % - 11 %. Il n'est pas surprenant que dans un pays développé comme la France, la valeur trouvée se trouve au bas de l'intervalle. L'urbanisation française est un processus ayant atteint un certain équilibre, les firmes ont donc partiellement internalisées les bénéfices qu'elles pouvaient tirer de cette agglomération.

Nous pensons que la comparaison des spécifications avec et sans indicatrices régionales fournit un bon intervalle, de 2 % à 4 %, pour l'ampleur des externalités d'urbanisation. En effet, si la spécification sans indicatrices régionales est susceptible d'être biaisée vers le haut (pour les raisons évoquées précédemment), la spécification avec indicatrices régionales tend certainement à sous-estimer l'ampleur des externalités d'urbanisation. En effet, l'identification y repose uniquement sur les différences entre zone d'emploi d'une même région. Cependant, les différences de densité moyenne entre régions expliquent en partie les différences de productivité moyenne entre régions (*i.e.* les indicatrices régionales). Pour preuve, une régression de la densité en emploi sur le jeu d'indicatrices régionales donne un R^2 de 48 %.

L'effet ici mis en évidence est linéaire. Si tel est réellement le cas, et si aucun frein à la migration des entreprises n'existait, on devrait observer une complète agglomération de l'activité économique en France. Nous avons tenté d'introduire des non-linéarités dans cette régression, par exemple le carré du logarithme

de la densité. Ces variables n'apparaissent pas significatives. Ceci suggère que même les zones d'emploi les plus denses en France, en région parisienne, n'ont pas atteint la masse critique à partir de laquelle les effets négatifs de l'agglomération l'emportent sur les effets positifs. Ce résultat est néanmoins à prendre avec précaution.

L'accessibilité au reste du marché national est également un déterminant important de la productivité des entreprises

Les zones d'emploi ne sont pas des territoires isolés ; elles sont au contraire fortement liées les unes aux autres par des flux migratoires et/ou des flux de commerce. Ces interactions entre zones sont au cœur des modèles de la Nouvelle Économie Géographique. Leur estimation empirique a montré que ces interactions pouvaient largement influencer le salaire des travailleurs ou la profitabilité des entreprises (Head et Mayer, 2004 ; Head et Mayer, 2006). En effet, les firmes bénéficient certes d'externalités d'agglomération au sein de leur zone d'implantation, mais peuvent également tirer parti des marchés installés dans des zones voisines. Un indicateur classique de ces externalités entre zones est le potentiel marchand (ou potentiel de marché), traditionnellement représenté comme la somme pondérée de la densité en emploi des zones voisines, les poids étant les inverses des distances géographiques entre barycentres des zones d'emploi.

$$\text{Potentiel marchand}_{zt} = \sum_{z' \neq z} \frac{\text{Densité en emploi}_{z't}}{\text{Distance}_{zz'}} \quad (6)$$

où $\text{Distance}_{zz'}$ représente la distance à vol d'oiseau entre les barycentres des zones d'emploi z et z' .

La moyenne du potentiel de marché s'établit à près de 120 pour l'année 2004 (cf. tableau 3, ligne *Potentiel marchand*), mais cette variable est, comme la densité, très inégalement répartie sur le territoire. La moitié des zones d'emploi françaises ont en effet un potentiel de marché inférieur à 67,8, soit près de la moitié de la moyenne (18).

18. Notre mesure du potentiel marchand considère la France comme un territoire isolé du reste du monde. Il se peut que certaines zones d'emploi des régions frontalières françaises jouissent d'une meilleure accessibilité au reste de l'Europe, ce qui n'est pas pris en compte par cette variable.

Il est d'autant plus important de tenir compte de cette variable que les zones denses sont, en France, souvent proches les unes des autres. Pour preuve la corrélation entre la variable de densité et de potentiel de marché (prises en logarithme) s'élève à 0,62 dans notre échantillon. Cette auto-corrélation spatiale de la variable de densité est donc susceptible de biaiser à la hausse l'effet sur la productivité de la seule densité, tel qu'il est mesuré dans le tableau 3. Cette variable de potentiel marchand est une manière simple de neutraliser l'effet de cette auto-corrélation spatiale. On s'attend donc à ce que son

introduction réduise l'impact de la densité, tel qu'il a été mesuré précédemment.

L'effet de l'introduction de cette variable est bien celui attendu, puisque l'élasticité à la densité en emploi a été réduite de 1,5 à 2 fois, s'établissant entre 2 et 3 % (cf. tableau 5). L'élasticité de la productivité au potentiel marchand est significativement positive, comprise entre 3,5 et 4 %. Non seulement la densité du marché locale est-elle importante, mais une bonne accessibilité au reste du marché national est également un déterminant majeur de la productivité des entrepri-

Tableau 5
L'accessibilité au reste du marché national est également un déterminant important de la productivité des entreprises

	La productivité TFP_{zst} est évaluée par...			
	... MCO	... la méthode d'Olley et Pakes	... la méthode de Levinsohn et Petrin	... la méthode à effet fixe
Estimation sans indicatrices régionales				
Ln Densité en emploi $_{zt}$	0,022*** (0,005)	0,025*** (0,004)	0,029*** (0,006)	0,025*** (0,005)
Ln Superficie $_z$	0,016*** (0,004)	0,018*** (0,004)	0,019*** (0,005)	0,018*** (0,005)
Ln Potentiel marchand $_{zt}$	0,037*** (0,005)	0,035*** (0,004)	0,038*** (0,005)	0,038*** (0,005)
Ln Diversité de la zone d'emploi $_{zt}$	-0,01 (0,006)	-0,008 (0,006)	-0,005 (0,008)	-0,009 (0,007)
Nombre d'observations	311 698	242 209	311 643	311 698
R ² ajusté	0,143	0,138	0,172	0,217
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui
Estimation avec indicatrices régionales				
Ln Densité en emploi $_{zt}$	0,02*** (0,003)	0,024*** (0,003)	0,03*** (0,004)	0,023*** (0,003)
Ln Superficie $_z$	0,018*** (0,003)	0,019*** (0,003)	0,021*** (0,004)	0,019*** (0,003)
Ln Potentiel marchand $_{zt}$	0,019** (0,008)	0,016** (0,008)	0,006 (0,011)	0,018** (0,008)
Ln Diversité de la zone d'emploi $_{zt}$	-0,013*** (0,005)	-0,011** (0,005)	-0,009 (0,006)	-0,012** (0,005)
Nombre d'observations	311 698	242 209	311 643	311 698
R ² ajusté	0,152	0,145	0,182	0,226
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par 4 méthodes) sur les logarithmes de la densité en emploi et de la superficie de la zone d'emploi, du potentiel marchand, de l'indice de diversité de la zone d'emploi, avec introduction d'indicatrices propres à chaque couple secteur-année, ainsi que des caractéristiques géographiques de zones (présentées dans le corps du texte). Les deux régressions diffèrent par l'introduction ou non d'indicatrices propres à chaque région française. Les coefficients estimés se lisent comme des élasticités.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi.

***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

ses. La réduction de l'effet de la densité consécutive à l'introduction du potentiel marchand est du même ordre de grandeur que celle liée à l'introduction des indicatrices régionales au tableau 4. Ceci s'explique simplement par le fait que les différences de potentiel marchand sont avant tout des différences régionales. La régression de cette variable sur le jeu d'indicatrices régionales donne un R^2 de 89 %. Ceci explique également que, lorsque le potentiel marchand est introduit au côté des indicatrices régionales (cf. tableau 5, partie basse), son effet soit réduit de moitié pour s'installer à environ 2 %.

Il est intéressant de noter que le pouvoir explicatif de notre modèle a augmenté dans chacune des régressions (mais évidemment plus dans la spécification sans indicatrices régionales), suggérant que la variable de potentiel marchand a bien un effet au-delà de sa simple corrélation avec la variable de densité. Comme nous l'évoquons plus haut, les zones d'emploi les plus denses en France sont également celles ayant une meilleure accessibilité au reste du marché national, ces deux effets se conjuguent donc pour accroître la productivité des entreprises. Cet effet persiste, mais avec une moindre ampleur, une fois que l'on a introduit des indicatrices régionales.

La diversité du tissu industriel local joue peu, et plutôt négativement, sur la productivité des entreprises

Si la densité des activités économiques prend déjà partiellement en compte la présence d'externalités d'urbanisation, la répartition de l'emploi au sein des différents secteurs présents dans la zone peut être fort différente, à densité de l'emploi et accessibilité données. Il est donc usuel d'introduire un indice de diversité du tissu industriel. La corrélation de la variable de diversité avec les deux précédentes n'est que de 0,18 pour la densité et 0,05 pour le potentiel de marché, suggérant bien que deux zones de même densité et de même accessibilité peuvent présenter des organisations sectorielles différentes.

Selon Jacobs (1969), la diversité sectorielle d'une zone peut être la source d'externalités si les idées et les technologies circulent d'un secteur à un autre. Ces industries s'enrichissent alors de cette interpénétration, en étant ainsi au contact d'un plus grand nombre d'innovations technologiques et/ou de processus de production différents. Duranton et Puga (2001) proposent une formalisation théorique à l'existence

de gains liés à une plus grande diversité. Notre indicateur de diversité est défini comme l'inverse d'un indicateur d'Herfindhal :

$$\text{Diversité de la zone d'emploi}_{zt} = \frac{1}{\sum_{s \in S_{zt}} \left(\frac{\text{Emploi}_{zst}}{\text{Emploi}_{zt}} \right)^2} \quad (7)$$

où S_{zt} représente l'ensemble des secteurs économiques (en NAF220) présents dans la zone d'emploi z à la date t . L'indicateur oscille entre 1 si l'emploi local est concentré dans un seul secteur et le nombre total de secteurs, si l'emploi est équi-réparti entre chaque (19).

Cette variable n'est pas significative dans la spécification sans indicatrices régionales (cf. tableau 5, ligne *Diversité de la zone d'emploi*_{zt}). Elle est significativement négative, mais avec une valeur faible des élasticités (de l'ordre de 1 %) dans la spécification avec indicatrices régionales. Ceci suggère que la diversité du tissu industriel local n'a que peu d'impact, et plutôt négatif contrairement aux théories jacobiniennes, sur la productivité des entreprises.

Sources et ampleur des externalités de localisation

Nous introduisons maintenant les variables liées aux externalités de localisation, autrement dit liées à l'organisation d'un secteur particulier (en NAF220) dans une zone d'emploi donnée. Nous ne conservons dans cette partie qu'une seule mesure de la productivité sectorielle locale, celle donnée par la méthode à effets fixes (20). Nous ne conservons également que la spécification avec indicatrices régionales.

Les entreprises sont, en moyenne, plus productives dans les zones où leur industrie est relativement plus concentrée

En accord avec l'hypothèse de Marshall-Arrow-Romer, la concentration d'entreprises d'un

19. Mori, Nishikimi et Smith (2008) préfèrent à cet indice d'Herfindhal le nombre de secteurs dans la zone ou l'indicateur de diversité relative introduit par Duranton et Puga (2000). Les résultats sont robustes à l'utilisation de l'une ou l'autre de ces mesures alternatives.

20. Nous avons vérifié que l'une ou l'autre de nos mesures de productivité donnait des résultats similaires.

même secteur d'activité dans une zone géographique est susceptible de générer, à leur profit, des effets externes positifs. Ces effets transitent soit par l'échange d'idées et de technologies, soit par le partage des gains liés à un marché plus large pour les facteurs de production, notamment le facteur travail, ou pour les consommations intermédiaires. On mesure généralement la spécialisation relative d'une zone z dans l'activité s (en NAF220) par la part du secteur s dans l'emploi local, rapportée à la part de l'emploi national dans ce secteur.

$$Spécialisation_{zst} = \frac{Emploi_{zst} / Emploi_{zt}}{Emploi_{st} / Emploi_t} \quad (8)$$

Cet indicateur renseigne sur la relative spécialisation de la zone z dans l'industrie s par rapport à la spécialisation de l'économie nationale dans ce secteur. Ainsi, si pour une zone et une indus-

trie données, cet indicateur prend une valeur supérieure à 1, cela signifie que l'emploi de l'industrie s est sur-représenté dans la zone z , par rapport à son poids dans l'économie nationale. Il est tout d'abord intéressant de noter que les zones les plus denses n'ont pas de schéma de spécialisation particulier. En effet, la corrélation de la variable de densité et de spécialisation s'établit à - 0,13. Au contraire, les zones d'emploi spécialisées se trouvent plus souvent en zones rurales.

La moyenne de cette variable s'établit en 2004 à 1,79 (cf. tableau 3). Cette statistique simple rend compte de la spécialisation locale en France. Néanmoins cette moyenne est essentiellement tirée par quelques zones d'emploi fortement spécialisées. En effet, les trois quarts des zones d'emploi françaises ont des indices de spécialisation inférieurs à 1,51.

Pour expliquer la productivité, nous introduisons, au côté des variables d'urbanisation, notre

Tableau 6
L'influence des externalités de localisation

Estimation	(A)	(B)	(C)	(D)
Externalités d'urbanisation				
Ln Densité en emploi _{zt}	0,023*** (0,003)	0,024*** (0,003)	0,023*** (0,003)	0,022*** (0,004)
Ln Superficie _z	0,019*** (0,003)	0,021*** (0,003)	0,02*** (0,003)	0,018*** (0,004)
Ln Potentiel marchand _{zt}	0,018** (0,008)	0,012 (0,009)	0,013 (0,009)	0,012 (0,009)
Ln Diversité de la zone d'emploi _{zt}	- 0,012** (0,005)	- 0,012** (0,005)	- 0,012** (0,005)	- 0,012** (0,005)
Externalités de localisation				
Ln Spécialisation _{zst}		0,021*** (0,002)	0,021*** (0,002)	0,02*** (0,003)
Ln Dotation locale en qualifications _{zst}			0,007*** (0,002)	0,007*** (0,002)
Ln Nombre d'établissements _{zst}				0,002 (0,003)
Nombre d'observations	311 698	311 698	311 698	311 698
R ² ajusté	0,226	0,233	0,233	0,233
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) sur les logarithmes de la densité en emploi, de la superficie de la zone d'emploi, du potentiel marchand et de l'indice de diversité (colonne A) progressivement complétée par l'introduction du logarithme de la spécialisation (colonne B), le logarithme de la dotation locale en qualifications (colonne C), le logarithme du nombre d'établissements (colonne D). Nous avons vérifié que l'une ou l'autre de nos mesures de productivité donnent des résultats similaires. Les coefficients estimés se lisent comme des élasticités.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi.

***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

indicateur de spécialisation de la zone d'emploi (cf. tableau 6, colonne B). Cette variable a un effet significativement positif sur notre mesure de la productivité, avec une élasticité de 0,02 (21). À secteur donné, les entreprises localisées à proximité d'entreprises du même secteur enregistrent une productivité moyenne supérieure. L'introduction de cette variable n'a pas d'effet marqué sur les coefficients des autres variables de la régression et accroît légèrement le pouvoir explicatif du modèle. Ceci suggère que cet effet de spécialisation joue au-delà des effets d'urbanisation mesurés précédemment. L'écart-type du logarithme de cette variable vaut 1,15. Ainsi un doublement de la spécialisation locale, *i.e.* un doublement de l'emploi sectoriel local à emploi total de la zone donné (22), induit un accroissement de 2 % en moyenne de la productivité des entreprises du secteur considéré.

Si les clusters dotés d'une main-d'œuvre relativement plus qualifiée sont ceux où les entreprises sont les plus productives, la causalité reste difficile à évaluer

Même si nous tenons bien compte de la qualité de la main-d'œuvre employée par les entreprises dans nos régressions de première étape, il se peut que la productivité estimée pour un *cluster* soit positivement corrélée avec la proportion locale de travailleurs qualifiés. En effet, il est envisageable que les travailleurs les plus éduqués soient également ceux qui tirent le plus parti d'échanges informels, autrement dit qu'ils soient les agents par lesquels se diffusent le plus efficacement ces externalités d'agglomération dans l'économie locale. Combes et Duranton (2006) développent ainsi un modèle où la rotation de l'emploi qualifié d'une entreprise à une autre est le support des transferts d'innovation. Il est donc nécessaire d'introduire, dans nos régressions de deuxième étape, la part relative des employés qualifiés dans l'emploi sectoriel local. Nous construisons ici un indicateur rendant compte d'une dotation locale en main-d'œuvre qualifiée supérieure à la qualification moyenne dans le secteur au niveau national.

$$Dotation\ locale\ en\ qualifications_{zst} = \frac{Emp_{zst}^3 / Emp_{zst}}{Emp_{st}^3 / Emp_{st}} \quad (9)$$

où Emp_{zst}^3 est l'emploi qualifié (groupe $q = 3$, cf. *supra* note 9.) dans le *cluster* zone-secteur-année

et Emp_{zst} l'emploi total dans le *cluster*. Pour un secteur donné, une valeur de cet indicateur supérieure à 1 indique que les établissements de la zone emploient une main-d'œuvre relativement plus qualifiée qu'un établissement moyen du même secteur en France. Notons que les qualifications sont inégalement réparties sur le territoire (cf. tableau 3), comme le soulignent Combes, Duranton et Gobillon (2008). La moyenne de cette variable s'établit à 0,91 et sa médiane à 0,87.

Nous introduisons dans la spécification précédente cette mesure de la dotation locale en qualifications (cf. tableau 6, colonne C). À secteur donné, les entreprises situées dans les *clusters* où la part du travail qualifié est la plus élevée sont également celles qui enregistrent la productivité moyenne la plus forte. Néanmoins son introduction n'apporte pas de pouvoir explicatif à notre modèle. Au contraire, elle réduit très légèrement l'impact de la variable de densité, avec laquelle elle est positivement corrélée (0,22). Ainsi, il est difficile de mettre en évidence un effet causal d'une plus grande présence de travailleurs qualifiés dans un *cluster* sur la productivité de ce *cluster* au-delà de l'effet de la densité.

Ceci montre notamment qu'une prise en compte adéquate de la qualité de la main-d'œuvre dans la fonction de production de la première étape est particulièrement importante. En effet, si l'on ne corrige pas de la qualité de la main-d'œuvre en première étape, il est possible en seconde étape de trouver un lien positif entre productivité et qualité de la main-d'œuvre, mais ce pourrait n'être qu'un pur effet de composition, non lié à une quelconque externalité. Ce résultat fait écho au résultat de Combes, Duranton et Gobillon (2008) qui montrent qu'une part importante des disparités spatiales de salaires en France s'explique par une distribution non aléatoire dans l'espace des travailleurs qualifiés. Ces derniers se trouvent majoritairement dans les zones les plus denses.

La concurrence n'a, en moyenne, pas d'effets sur la productivité des entreprises d'un cluster

La naissance, l'accumulation et la transmission d'informations sont susceptibles de fluctuer

21. Il est intéressant de remarquer que cet effet est robuste à l'introduction ou non du jeu d'indicatrices régionales.

22. Combes (2000b) montre clairement que l'interprétation des externalités marshalliennes doit être celle-ci : un accroissement de la spécialisation locale à emploi total de la zone donné, contrairement aux interprétations fournies par Glaeser, Kallal, Scheinkman et Schleifer (1992) et Henderson, Kuncoro et Turner (1995).

aussi bien selon le nombre d'établissements présents dans la zone, que suivant leurs tailles respectives. De plus, si le bien est vendu localement, la distribution de taille des firmes peut constituer une mesure du degré de concurrence locale. Celle-ci peut très largement influencer le prix des produits des firmes. Nous appréhendons la concurrence dans une zone par le nombre d'établissements dans le *cluster*, une fois que l'on a tenu compte de la spécialisation locale (23). Nous ne trouvons pas d'effet systématique de la concurrence, telle que nous l'appréhendons, sur la productivité des entreprises (cf. tableau 6, colonne D).

Existe-t-il des différences sectorielles dans la force des externalités d'urbanisation et de localisation ?

Nous cherchons ici à comprendre si les externalités d'agglomération mises en évidence dans les parties précédentes jouent avec la même force dans différentes industries. Nous estimons donc la spécification (D) du tableau 5 séparément pour les 27 industries manufacturières (NAF60 15 à 36) et de services (NAF60 64 à 74) en NAF60. Nous avons retenu ce niveau de classification sectorielle plus agrégé pour des raisons de précision des estimations. En effet, nous aurions pu estimer cette spécification pour chaque secteur en NAF220. Cependant, le nombre de *clusters zst* sur lequel repose alors l'estimation est particulièrement faible pour certains secteurs.

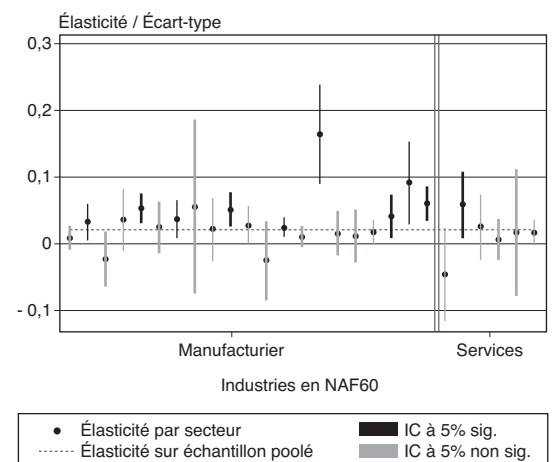
Les effets mesurés jusqu'ici sont des effets moyens, tous secteurs confondus. Cependant, il n'y a, *a priori*, aucune raison de penser que les externalités d'urbanisation ou de localisation jouent avec la même force dans chaque secteur. À ce sujet, Henderson (2003) montre qu'aux États-Unis les externalités de localisation jouent avec plus de force dans les secteurs high-tech que dans les secteurs plus traditionnels. Nous nous concentrons dans cette partie sur les élasticités correspondantes aux variables de densité et de spécialisation locale, qui ont le plus grand pouvoir explicatif.

Pour éviter de présenter un ensemble fastidieux de tableaux, pour chacune de ces deux variables, nous reportons chaque fois sur un graphique la collection des 27 valeurs d'élasticité, chacune correspondante à une industrie en NAF60 (cf. graphiques I et II). Les industries sont classées suivant leur ordre dans la classification en NAF60 tel qu'elle est définie par l'Insee. Les

industries manufacturières apparaissent donc à droite des industries de service. Nous reportons également la valeur de l'élasticité moyenne sur l'échantillon *poolé* trouvée dans le tableau 6, ainsi que les intervalles de confiance à 5 % pour chacune des élasticités. Ces intervalles de confiance sont gris clair lorsque l'élasticité n'est pas significative au seuil de 5 %, noirs sinon.

Trois remarques nous semblent intéressantes. Tout d'abord, les externalités d'urbanisation et de localisation ne semblent concerner qu'un petit nombre d'industries manufacturières et de service. Pour un nombre assez important d'entre elles, l'élasticité mesurée n'est pas significative. Ensuite lorsque l'effet est significatif, il joue

Graphique I
Hétérogénéité de l'élasticité de la productivité à la densité en emploi (variable d'urbanisation *Densité en emploi_{zt}*)



Lecture : le graphique reporte les 27 valeurs de l'élasticité de la productivité à la densité en emploi, chacune correspondante à une industrie en NAF60. La valeur 0,011 de l'élasticité moyenne sur l'échantillon *poolé* trouvée dans le tableau 7, ainsi que les intervalles de confiance à 5 % pour chacune des élasticités. Ces intervalles de confiance sont gris clair lorsque l'élasticité n'est pas significative au seuil de 5 %, noirs sinon.

La régression utilisée est celle du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) sur les logarithmes de la densité en emploi, de la superficie de la zone d'emploi, du potentiel marchand et de l'indice de diversité progressivement complétée par l'introduction du logarithme de la spécialisation, le logarithme de la dotation locale en qualifications, le logarithme du nombre d'établissements.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

23. Nous avons également essayé d'introduire d'autres variables de concurrence, telle que la taille moyenne des établissements dans la zone ou l'inverse d'un indice de concentration de l'emploi de type Herfindhal

$$C_{zst} = \frac{1}{\sum_{j \in E_{zst}} \left(\frac{Emp_{jt}}{Emp_{zst}} \right)^2} . \text{ Ceci ne change}$$

pas le résultat.

bien positivement, que ce soit les effets d'urbanisation (densité) que de localisation (spécialisation locale) (24).

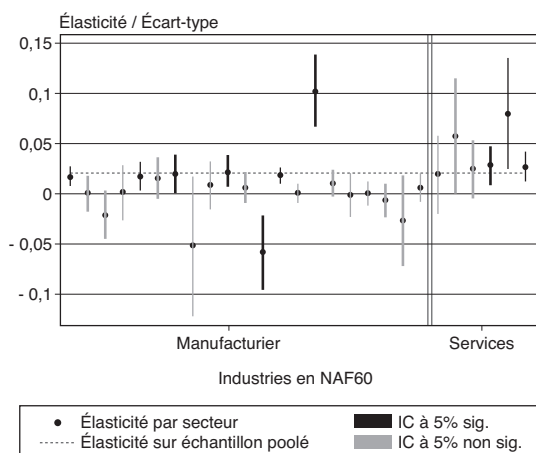
Contrairement à ce que l'on pourrait penser, les effets d'urbanisation ne semblent pas avoir d'impact plus systématique sur les industries de services que les industries manufacturières. Seule une industrie de service (Activités immobilières) enregistre une élasticité à la densité significativement positive. Au contraire, les effets de localisation semblent être importants pour la moitié des industries de service. Enfin, il est intéressant de remarquer que l'industrie qui enregistre à la fois la plus grande élasticité à la densité et à la spécialisation locale est l'industrie de la fabrication des machines de bureau et de matériel informatique. Cet exemple est intéressant car cette industrie nécessite à la fois un grand nombre de consommations intermédiaires et un accès facile au consommateur (ce qui peut justifier sa localisation dans les zones les plus denses) ainsi qu'une main-d'œuvre assez spé-

cialisée (ce qui peut justifier sa concentration spatiale au sein des zones les plus denses). Pour ce secteur, externalités d'urbanisation et externalités de localisation jouent simultanément.

Les résultats restent sensiblement les mêmes lorsque les estimations sont menées à un niveau sectoriel plus agrégé

Nous avons jusqu'ici supposé que les externalités de localisation se produisent au sein d'un secteur en NAF220. Cependant, il est possible que des externalités existent entre firmes qui n'appartiennent pas uniquement au même secteur en NAF220, mais également à un niveau plus agrégé (cf. tableau 7, l'analogue du tableau 6 quand les productivités sectorielles locales et les variables d'agglomération ne sont plus calculées au niveau NAF220, mais au niveau NAF60). Les résultats ne changent pas vraiment. Une fois les indicatrices régionales prises en compte, les variables de densité et de spécialisation restent celles qui ont le pouvoir explicatif le plus fort. L'élasticité de la densité s'établit à 1,8 et 2,6 %, alors que l'élasticité de la variable de spécialisation locale est autour de 2,5-3 %.

Graphique II
Hétérogénéité de l'élasticité de la productivité à la spécialisation (variable de localisation Spécialisation_{zst})



Lecture : le graphique reporte les 27 valeurs de l'élasticité de la productivité à la spécialisation, chacune correspondante à une industrie en NAF60. La valeur 0,025 de l'élasticité moyenne sur l'échantillon poolé trouvée dans le tableau 7, ainsi que les intervalles de confiance à 5 % pour chacune des élasticités. Ces intervalles de confiance sont gris clair lorsque l'élasticité n'est pas significative au seuil de 5 %, noirs sinon.

La régression utilisée est celle du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) sur les logarithmes de la densité en emploi, de la superficie de la zone d'emploi, du potentiel marchand et de l'indice de diversité progressivement complétée par l'introduction du logarithme de la spécialisation, le logarithme de la dotation locale en qualifications, le logarithme du nombre d'établissements.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

Nos résultats sont robustes à l'introduction des établissements d'entreprises pluri-établissements

Nous n'avons jusqu'ici gardé que les entreprises mono-établissement de plus de cinq salariés ETP. L'exclusion des entreprises pluri-établissements trouve sa justification dans le fait qu'en France, les données de production, valeur ajoutée et capital ne sont connues qu'au niveau de l'entreprise et non pas de l'établissement. Il ne nous est donc pas possible d'estimer une productivité pour chaque établissement d'une entreprise pluri-établissement. Inversement, s'il nous est possible d'avoir une productivité pour l'entreprise pluri-établissement, comment connaître celle de chacun de ses établissements ? Nous étudions deux cas polaires (cf. tableau 8) : le premier dans lequel tous les établissements d'une firme pluri-établissement se voient affecter la productivité de l'entreprise dans son ensemble, le second dans lequel la productivité de l'entreprise est réallouée à chacun de ces établissements au prorata de l'emploi. Nos résultats sont robustes dans ces deux spé-

24. Seule l'industrie de la métallurgie présente une élasticité à la spécialisation locale significativement négative.

cifications. La densité et la spécialisation locale restent les deux déterminants majeurs des différences spatiales de productivité. L'élasticité de la densité est légèrement réduite à 1,5 %.

Entre consommations intermédiaires et facteurs de production

Les données fiscales que nous utilisons dans cette étude restent imparfaites pour aborder le processus de production de l'entreprise, et donc sa productivité. L'une des limites de ces données tient à la mesure des facteurs de production, travail et capital (25).

Dans la liasse fiscale, la ligne « Charges externes et autres charges » regroupe l'ensemble des prestations facturées à l'entreprise par l'extérieur et qui ne sont pas toujours, *stricto sensu*, des consommations intermédiaires. Il peut

s'agir de la facturation du travail intérimaire, ou de la location de capital (comme par exemple le terrain ou les bâtiments), etc. Ce montant a jusqu'ici été retiré de la valeur de la production pour calculer la valeur ajoutée. Une mesure adéquate de la valeur ajoutée ne devrait pas tenir compte de ces dépenses, qui ne sont pas de consommations intermédiaires, mais bien un coût du capital ou du travail.

Néanmoins, mener une estimation correcte de la fonction de production nécessiterait de pouvoir reconstruire les quantités de travail ou de capital correspondantes à ces coûts. Les données à disposition ne le permettent pas. Néanmoins,

25. Une autre limite importante concerne les effets de prix des produits et des facteurs sur la mesure de productivité des entreprises. Foster, Haltiwanger et Syverson (2008) fournissent une avancée récente sur la correction des effets de prix dans la mesure de productivité. Néanmoins, ce point est au-delà des ambitions de cet article et ne peut être traité avec les données à notre disposition.

Tableau 7
Évaluation « au niveau NAF60 » de l'influence sur la productivité des externalités d'agglomération (urbanisation et localisation)

Estimation	(A)	(B)	(C)	(D)
Externalités d'urbanisation				
Ln Densité en emploi _{zt}	0,023*** (0,003)	0,023*** (0,003)	0,018*** (0,003)	0,011 (0,007)
Ln Superficie _z	0,018*** (0,003)	0,018*** (0,003)	0,014*** (0,003)	0,006 (0,006)
Ln Potentiel marchand _{zt}	0,02*** (0,007)	0,013 (0,008)	0,018** (0,007)	0,015* (0,009)
Ln Diversité de la zone d'emploi _{zt}	- 0,025*** (0,008)	- 0,022*** (0,008)	- 0,026*** (0,007)	- 0,025*** (0,008)
Externalités de localisation				
Ln Spécialisation _{zst}		0,03*** (0,004)	0,029*** (0,004)	0,025*** (0,003)
Ln Dotation locale en qualifications _{zst}			0,041*** (0,005)	0,042*** (0,005)
Ln Nombre d'établissements _{zst}				0,008 (0,006)
Nombre d'observations	119 936	119 936	119 936	119 936
R ² ajusté	0,387	0,403	0,408	0,408
Indicatrice par secteur-année	Oui	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) sur les logarithmes de la densité en emploi, de la superficie de la zone d'emploi, du potentiel marchand et de la diversité (colonne A) progressivement complétée par l'introduction du logarithme de la spécialisation (colonne B), le logarithme de la dotation locale en travail qualifié (colonne C), le logarithme du nombre d'établissements (colonne D). Nous avons vérifié que l'une ou l'autre de nos mesures de productivité donnent des résultats similaires.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi. ***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

Tableau 8

Une évaluation de la sensibilité des résultats à l'exclusion des entreprises pluri-établissements

	Les établissements d'une firme pluri-établissement se voient affecter la productivité de l'entreprise dans son ensemble			La productivité de la firme est réallouée à chacun de ses établissements au prorata de l'emploi		
Estimation	(A)	(B)	(C)	(D)	(E)	(F)
Externalités d'urbanisation						
Ln Densité en emploi _{zt}	0,017*** (0,003)	0,018*** (0,003)	0,028*** (0,003)	0,016*** (0,002)	0,015*** (0,002)	0,016*** (0,003)
Ln Superficie _z	0,015*** (0,003)	0,017*** (0,003)	0,029*** (0,004)	0,01*** (0,002)	0,014*** (0,002)	0,015*** (0,004)
Ln Potentiel marchand _{zt}		0,003 (0,009)	0,003 (0,009)		0,013** (0,006)	0,01* (0,006)
Ln Diversité de la zone d'emploi _{zt}		- 0,007 (0,004)	- 0,008* (0,004)		- 0,009** (0,004)	- 0,009*** (0,004)
Externalités de localisation						
Ln Spécialisation _{zst}			0,022*** (0,003)			0,015*** (0,003)
Ln Dotation locale en qualifications _{zst}			- 0,00007 (0,003)			0,002 (0,002)
Ln Nombre d'établissements _{zst}			- 0,012*** (0,003)			- 0,002 (0,006)
Nombre d'observations	374 956	374 956	374 956	374 956	374 956	374 956
R ² ajusté	0,302	0,302	0,307	0,147	0,147	0,154
Indicatrice par secteur-année	Oui	Oui	Oui	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) d'abord sur les seuls logarithmes de la densité en emploi et de la superficie de la zone d'emploi, (colonne A et D), d'abord complétée par l'introduction des logarithmes du potentiel marchand et de l'indice de diversité (colonnes B et E), puis par les logarithmes de la spécialisation, de la dotation locale en qualifications et du nombre d'établissements (colonnes C et F). Les estimations incorporent des effets fixes temporels et 5 caractéristiques des zones d'emploi fixes dans le temps. Les coefficients estimés se lisent comme des élasticités.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi.

***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : entreprises pluri-établissements incluses.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

nous présentons les résultats équivalents à ceux du tableau 6 sans la soustraction dans le calcul de la valeur ajoutée de la ligne « Charges externes et autres charges » (cf. tableau 9). N'étant pas en mesure d'introduire comme facteurs de production le capital ou le travail (intérimaire) correspondants, la productivité n'est mesurée qu'approximativement. Notre *a priori* est que la location de capital ou de travail est plus répandue dans les zones d'emploi denses que dans les zones peu denses. Aussi les effets mesurés jusqu'ici, notamment concernant les externalités d'urbanisation fournissent une valeur basse.

À l'inverse, ce test de robustesse en fournit une valeur haute.

Comme attendu, dans ce test de robustesse l'élasticité de la densité augmente légèrement passant de 2 % à 3,2 % (cf. tableau 9). De même, les élasticités de la spécialisation augmentent légèrement. Néanmoins, les changements restent assez faibles. Aussi, si un biais existe du fait de différences spatiales dans le recours à la location de capital (notamment de terrain) ou de travail, son effet sur l'estimation précédente reste très limité. □

Tableau 9

Une évaluation de la sensibilité des résultats au mode de calcul de la valeur ajoutée

Estimation	(A)	(B)	(C)
Externalités d'urbanisation			
Ln Densité en emploi _{zt}	0,032*** (0,004)	0,032*** (0,004)	0,032*** (0,007)
Ln Superficie _z	0,023*** (0,003)	0,027*** (0,004)	0,029*** (0,005)
Ln Potentiel marchand _{zt}		0,008 (0,009)	0,00006 (0,011)
Ln Diversité de la zone d'emploi _{zt}		-0,011** (0,006)	-0,011** (0,006)
Externalités de localisation			
Ln Spécialisation _{zst}			0,029*** (0,004)
Ln Dotation locale en qualifications _{zst}			0,001 (0,002)
Ln Nombre d'établissements _{zst}			0,0003 (0,005)
Nombre d'observations	311 698	311 698	311 698
R ² ajusté	0,22	0,22	0,233
Indicatrice par secteur-année	Oui	Oui	Oui
Indicatrice par région	Oui	Oui	Oui
Caractéristiques des zones d'emploi	Oui	Oui	Oui

Lecture : régressions du logarithme de la productivité sectorielle locale (calculée par la seule méthode à effets fixes) sur les logarithmes de la densité en emploi, de la superficie de la zone d'emploi, d'abord complétée par l'introduction du logarithme du potentiel marchand et de l'indice de diversité (colonne B) puis par l'introduction des logarithmes de la spécialisation, de la dotation locale en qualifications et celui du nombre d'établissements (colonne C). Les contrôles ajoutés sont des effets fixes temporels et 5 caractéristiques des zones d'emploi temps fixes dans le temps. Les coefficients estimés se lisent comme des élasticités.

La matrice de variance-covariance est supposée bloc-diagonale, chaque bloc correspondant à une zone d'emploi.

***, **, * indiquent respectivement une significativité au niveau de 1 %, 5 % et 10 %.

Champ : les seules entreprises mono-établissement de plus de 5 salariés en équivalent temps plein.

Source : calculs des auteurs à partir de la base de données résultant, pour chaque année de 1994 à 2004, de la fusion de la base de données Sirene, des fichiers de déclarations fiscales d'entreprises (RSI/BRN), des fichiers de déclaration de données sociales (DADS) au niveau établissement (cf. encadré).

BIBLIOGRAPHIE

Ackerberg D., Benkard C.L., Berry S. et Pakes A. (2007), « Econometric Tools for Analyzing Market Outcomes », in *Handbook of Econometrics*, J.J. Heckman et E. Leamer (éds.), vol. 6A, chap. 63, pp. 4171-4276, Elsevier Science Publisher.

Barlet M., Briant A. et Crusson L. (2008), « Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu », *document de travail*, n° G2008-09, D3E, Insee.

Blundell R. et Bond S. (2000), « GMM Estimation with Persistent Panel Data: an Application to Production Functions », *Econometric Reviews*, vol. 19, n° 3, pp. 321-340.

Burnod G. et Chenu A. (2001), « Employés qualifiés et non-qualifiés : une proposition d'aména-

gement de nomenclature des catégories socio-professionnelles », *Travail et Emploi*, pp. 86-87.

Cingano F. et Schivardi F. (2004), « Identifying the Sources of Local Productivity Growth », *Journal of European Economic Association*, vol. 2, n° 4, pp. 720-742.

Combes P.P. (2000a), « Economic Structure and Local Growth: France, 1984-1993 », *Journal of Urban Economics*, vol. 47, n° 3, pp. 329-355.

Combes P.-P. (2000b), « Marshall-Arrow-Romer Externalities and City Growth », *CERAS working paper*, n° 99-06.

Combes P.-P. et Duranton G. (2006), « Labour Pooling, Labour Poaching, and Spatial Clustering », *Regional Science and Urban Economics*, vol. 36, n° 1, pp. 1-28.

- Combes P.-P., Duranton G. et Gobillon L. (2008)**, « Spatial Wage Disparities: Sorting Matters ! », *Journal of Urban Economics*, vol. 63, n° 2, pp. 723-742.
- Combes P.-P., Duranton G., Gobillon L. et Roux S. (2008)**, « Estimating Agglomeration Economies with History, Geography, and Worker Effects », à paraître dans *The Economics of Agglomeration*, Edward Glaeser (éd.), University of Chicago Press.
- Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, et S. Roux (2009)**, « The Productivity Advantages of Large Markets: Distinguishing Agglomeration from Firm Selection », mimeo
- Combes P.-P., Magnac T. et Robin J.-M. (2004)**, « The Dynamics of Local Employment in France », *Journal of Urban Economics*, vol. 56, n° 2, pp. 217- 243.
- Crépon B. et Duhautois R. (2003)**, « Ralentissement de la productivité et réallocations d'emplois : deux régimes de croissance », *Economie et Statistique*, n° 367, p. 69-82.
- Duranton G., Gobillon L. et Overman H. (2007)**, « Assessing the Effects of Local Taxation using Microgeographic Data », *CEP Discussion Paper*, n° 748, Centre for Economic Performance, London School of Economics and Political Science, Londres.
- Duranton G., Martin P., Mayer T. et Mayneris F. (2008)**, « Les pôles de compétitivité. Que peut-on en attendre ? », *collection de Cepremap*, Centre pour la Recherche Économique et ses Applications, Éditions *Ens* Rue d'Ulm
- Duranton G. et Overman H.G. (2005)**, « Testing for Localisation Using Micro-Geographic Data », *Review of Economic Studies*, vol. 72, n° 4, pp. 1077-1106.
- Duranton G. et Puga D. (2000)**, « Diversity and Specialisation in Cities: Why, Where and When Does it Matter ? », *Urban Studies*, vol. 37, n° 3, pp. 533-555.
- Duranton G. et Puga D. (2001)**, « Nursery Cities: Urban Diversity, Process Innovation and the Lifecycle of Product », *American Economic Review*, vol. 91, n° 5, pp. 1454-1477.
- Duranton G. et Puga D. (2004)**, « Micro-foundations of Urban Agglomeration Economies », in *Handbook of Urban and Regional Economics*, J.V. Henderson et J.-F. Thisse (éds.), vol. 4, *Cities and Geography*, chap. 48, pp. 2063-2117, North-Holland, Amsterdam.
- Ellison G. et Glaeser E.L. (1997)**, « Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach », *Journal of Political Economy*, vol. 105, n° 5, pp. 889-927.
- Ellison G. et Glaeser E.L. (1999)**, « The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration ? », *American Economic Review*, vol. 89, n° 2, pp. 311-316.
- Foster L., Haltiwanger J. et Syverson C. (2008)**, « Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability ? », *American Economic Review*, vol. 98, n° 1, pp. 394-425.
- Glaeser E., Kallal H., Scheinkman J. et Schleifer A. (1992)**, « Growth in Cities », *The Journal of Political Economy*, vol. 100, n° 6, pp. 1126-1152.
- Griliches Z. et Mairesse J. (1995)**, « Production Functions: the Search for Identification », *NBER Working Paper*, n° 5067.
- Head K. et Mayer T. (2004)**, « The Empirics of Agglomeration and Trade », in *Handbook of Urban and Regional Economics*, J.V. Henderson et J.-F. Thisse (éds.), vol. 4, *Cities and Geography*, chap. 59, pp. 2609-2669, North-Holland, Amsterdam.
- Head K. et Mayer T. (2006)**, « Regional Wage and Employment Responses to Market Potential in the EU », *Regional Science and Urban Economics*, Special Issue: *New Economic Geography : Closing the Gap Between Theory and Empirics*, vol. 36, n° 5, pp. 573-594.
- Henderson J.V. (2003)**, « Marshall's Scale Economies », *Journal of Urban Economics*, vol. 53, n° 1, pp. 1-28.
- Henderson V., Kuncoro A. et Turner M. (1995)**, « Industrial Development in Cities », *The Journal of Political Economy*, vol. 103, n° 5, pp. 1067-1090.
- Insee (2006)**, « Les salariés à temps complet travaillent 1650 heures dans l'année », *Insee Première*, n° 1066.
- Jacobs J. (1969)**, *The Economy of Cities*, Random House, New York.

- Levinsohn J. et Petrin A. (2003)**, « Estimating Production Functions Using Inputs to Control for Unobservables », *Review of Economic Studies*, vol. 70, n° 2, pp. 317-342.
- Marshall A. (1890)**, *Principles of Economics*, 8^e edition, MacMillan, Londres.
- Martin P., Mayer T. et Mayneris F. (2008)**, « Spatial Concentration and Firm-Level Productivity in France », *Discussion Paper Series*, n° 6858, CEPR.
- Melitz M. J. et Ottaviano G. I. P. (2008)**, « Market Size, Trade, and Productivity », *Review of Economic Studies*, vol. 75, n° 1, pp. 295-316.
- Melo, P. C., D. J. Graham, et R. B. Noland (2009)**, « A meta-analysis of estimates of urban agglomeration economies », *Regional Science and Urban Economics*, vol. 39, n° 3, pp 332–342.
- Mori T., Nishikimi K. et Smith T.E. (2008)**, « The Number-Average Size Rule: a New Empirical Relationship between Industrial Location and City Size », *Journal of Regional Science*, vol. 48, n° 1, pp. 165-211.
- Nocke, V. (2006)**, « A Gap for Me : Entrepreneurs and Entry », *Journal of the European Economic Association*, 4(5), 929–956.
- Olley S. et Pakes A. (1996)**, « The Dynamics of Productivity in the Telecommunications Equipment Industry », *Econometrica*, vol. 64, n° 6, pp. 1263-1297.
- Rathelot R. et Sillard P. (2008)**, « The Importance of Local Corporate Taxes in Business Location Decisions: Evidence From French Micro Data », *Economic Journal*, vol. 118, n° 527, pp. 499-514.
- Rosenthal S.S. et Strange W.C. (2004)**, « Evidence on the Nature and the Sources of Agglomeration Economies », in *Handbook of Urban and Regional Economics*, J.V. Henderson et J.-F. Thisse (éds.), vol. 4, *Cities and Geography*, chap. 49, pp. 2119-2171, North-Holland, Amsterdam.
- Starrett D. (1978)**, « Market Allocations of Location Choice in a Model with Free Mobility », *Journal of Economic Theory*, vol. 17, n° 1, pp. 21–37.
-