

Transports Publics Urbains et Attractivité Internationale

Enjeux Théoriques et Evaluation Empirique de l'Impact du
Réseau Express Régional sur les Investissements Étrangers en
Île-de-France

Thierry Mayer¹
Corentin Trevien²

12 avril 2012

1. SciencesPo., thierry.mayer@sciences-po.fr

2. Insee-SciencesPo., corentin.trevien@insee.fr, ce rapport ne reflète pas la position de l'Insee et n'engage que ses auteurs. Il constitue la version provisoire d'un document de travail de l'Insee à paraître.

Le réseau de transport du Grand Paris ou le projet « Crossrail » à Londres illustrent l'importance croissante d'un système de transport efficace. Ils revêtent un rôle particulier dans la compétition entre grandes métropoles pour attirer les investisseurs internationaux. L'objectif de ce travail d'anticiper l'effet du Grand Paris Express sur la localisation des investissements directs à l'étranger (IDE) en Île-de-France. Nous présentons tout d'abord des éléments méthodologiques et théoriques sur les choix de localisation des entreprises et l'évaluation de l'impact économique des réseaux de transport. Pour éclairer les enjeux actuels, nous exploitons l'expérience du réseau express régional (RER) qui constitue le dernier investissement d'ampleur dans les transports publics ferrés de l'agglomération parisienne. Le RER a conduit, dans les années 1970, 1980 et, dans une moindre mesure, 1990, à une amélioration significative des temps de déplacement des franciliens pour les communes reliées. Cette expérience nous permet ainsi de déterminer l'impact de cette réduction des temps de parcours sur la localisation des investissements étrangers en France.

1 Éléments méthodologiques et théoriques

L'un des objectifs du projet de métro automatique du Grand Paris est d'augmenter l'attractivité du territoire francilien aux yeux des investisseurs étrangers. Ces agents économiques sont « sans attache » territoriale forte. Leurs choix de localisation prennent en compte de manière plus objective les caractéristiques d'un territoire. Ces entreprises offrent donc des situations plus « pures » pour identifier les déterminants de localisation. De plus, cette question se situant dans le cadre d'une réflexion sur une concurrence supposée entre grandes métropoles mondiales, il semble plus pertinent d'étudier les comportements d'entreprises multinationales, susceptibles de devoir choisir leur lieu d'implantation parmi ces grandes métropoles.

Plusieurs approches permettent de traiter cette question. L'objectif de cette revue de littérature est de proposer une grille de lecture de ces approches, de mettre en avant les problèmes méthodologiques soulevés et de présenter la manière dont les économistes ont tenté de résoudre ces problèmes. Pour simplifier, ces méthodologies peuvent être scindées en deux grands ensembles : les approches d'équilibre général et les approches d'équilibre partiel. La première approche est la plus ambitieuse et cherche à simuler l'impact de la construction de l'infrastructure de transport dans toute sa complexité. Les effets d'une telle construction sont en effet très nombreux et peuvent produire des effets antagonistes. Imaginons par exemple, la construction d'un nouvel équipement dans une région ou une métropole. L'amélioration des conditions de transport est susceptible d'augmenter la productivité des entreprises implantées autour des nouvelles gares. Cette amélioration doit donc permettre de développer l'attractivité de ces zones, et même éventuellement de la région dans son ensemble. Cependant, l'impact de cette infrastructure ne se résume pas à ce premier effet. Par exemple, il est très probable que la rente foncière croisse autour des nouvelles gares. Cette augmentation est à même d'entraîner un changement dans les incitations des entreprises étrangères et locales à se localiser dans ces territoires car elle entraîne un renchérissement des conditions de production. L'analyse en équilibre général cherche à prendre en compte l'ensemble de ces effets mais la complexité de cette approche oblige les chercheurs à imposer une structure forte et contraignante à leur modèle, tout en simplifiant de nombreux aspects du problème. Dans le cadre de la Nouvelle Économie Géographique, des modélisations permettant de déterminer l'impact des infrastructures de transport ont été proposées (voir Martin et Rogers, 1995 pour une des premières contributions). Autre limite, cette analyse doit souvent se restreindre à un nombre réduit de régions, le secteur des transports est alors simplement réduit à un paramètre de type iceberg³.

3. Les coûts « iceberg » permettent de modéliser simplement les coûts de transport en économie. À la manière des icebergs qui fondent durant leur voyage, lorsqu'une marchandise transite d'un pays A jusqu'à un pays B, seule

Dans ce cas, le recours à des techniques numériques est souvent indispensable, ce qui complique l'application empirique.

A l'inverse, l'approche d'équilibre partiel permet une plus grande flexibilité dans l'analyse économique des infrastructures de transport. Dans cette seconde démarche, on ne cherche plus à déterminer l'impact total ou « à l'équilibre » de cette construction, mais l'impact direct. Ceci revient à supposer implicitement que cette partie directe de l'impact est de premier ordre par rapport aux aspects indirects. L'approche est clairement moins ambitieuse mais permet de travailler plus finement sur les données et surtout de considérer des modèles moins contraints pour révéler l'effet causal de la politique.

La littérature académique sur la question se distingue en deux grands ensembles. D'une part, il existe un corpus important étudiant les décisions de localisation des investissements directs. Il existe d'autre part une littérature plus récente et moins fournie sur l'impact des réseaux de transport sur la croissance économique locale. Des progrès ont notamment permis d'améliorer l'identification du lien causal entre réseaux de transport et activité économique. Le projet de recherche se situe au point de rencontre de ces deux littératures et il convient dans un premier temps de bien identifier les enjeux méthodologiques et les développements qui ont été récemment accomplis dans ces deux champs.

En ce qui concerne les travaux sur la localisation des investissements directs, la plupart des avancées consistent à mieux comprendre le lien entre la théorie économique et l'estimation économétrique. Alors que les travaux des années 1980 et 1990 reposaient pour l'essentiel sur des modèles en forme réduite, la Nouvelle Économie Géographique a permis des avancées importantes (Head et Mayer, 2004). La vague récente de travaux en économie internationale, mettant en avant l'importance de l'hétérogénéité des firmes, a également porté sur l'analyse de la décision de localisation (Yeaple, 2009; Chen et Moore, 2010; Mayer et al., 2010).

Concernant l'analyse de l'influence locale des réseaux de transport sur la croissance démographique ou économique, les progrès ont été là encore importants. Le biais d'endogénéité dans une régression expliquant la croissance ou la productivité locale par l'existence ou la densité du réseau de transport est potentiellement très important. En effet, il est probable que le choix précis du tracé se fasse en fonction de la croissance passée ou, plus gênant, des perspectives de croissance des différents territoires. Dans ce cas, lorsque l'on compare sans précautions méthodologiques les territoires équipés d'infrastructures de transport aux autres territoires, les différences observées sont déterminées non seulement par l'effet propre de ces infrastructures mais aussi par les différences initiales entre ces territoires. Deux situations sont envisageables. Imaginons tout d'abord que les pouvoirs publics poursuivent des objectifs d'efficacité et choisissent de construire des infrastructures de transport dans les zones les plus prometteuses. Dans ce cas, si les infrastructures de transport évaluées n'avaient pas été construites, les zones reliées auraient connu une croissance économique et démographique plus forte que les zones qui n'ont pas été reliées. Une comparaison directe de ces deux catégories de zones conduirait ainsi à attribuer à tort à la seule infrastructure de transport l'impact conjoint des différences antérieures et l'infrastructure. Dans ce cas, les résultats obtenus conduiraient à surestimer l'effet de la politique étudiée. À l'inverse, si les pouvoirs publics tentent à travers la construction d'infrastructures de transport de compenser les difficultés écono-

une fraction de cette marchandise parvient dans le pays B. En pratique, les intermédiaires (douanes, transporteurs, etc.) qui permettent d'acheminer un bien d'un pays à l'autre se rémunèrent en prélevant directement une partie de ce bien. Plus la fraction du bien qui parvient dans le pays destinataire est importante, plus les coûts de transport sont faibles (ie. les intermédiaires ont prélevé une part plus faible du bien).

miques de territoires défavorisés, une comparaison directe pourrait conduire à sous-estimer l'impact économique d'un réseau de transport. Dans ces conditions, l'omission des handicaps initiaux des territoires reliés ne permet pas de déterminer l'effet propre de la politique publique étudiée. Il faut retenir de ces deux exemples que la comparaison directe des territoires équipés de l'infrastructure de transport évaluée et des autres territoires peut conduire à biaiser les résultats de l'évaluation. Il n'est pas possible de connaître a priori le sens de ce biais. Il convient donc d'adopter des méthodes permettant de se prémunir contre ce risque.

La littérature portant spécifiquement sur l'impact des infrastructures de transports de passagers sur les choix de localisation des entreprises est limitée. Il est cependant possible de construire une analyse structurelle en s'inspirant de champs proches mais qui n'évaluent pas l'impact d'une infrastructure du type du Grand Paris Express sur l'attractivité d'un territoire. En effet, les avancées récentes concernant l'évaluation économique des infrastructures de transport d'une part et, d'autre part, l'apport des modèles de localisation des entreprises sont riches d'enseignements. Nous nous proposons donc d'étudier successivement ces deux champs.

1.1 Les problèmes de causalité dans l'analyse de l'impact d'une infrastructure de transport

La question de l'évaluation économique des infrastructures de transport s'intègre dans le champ plus général de l'évaluation des politiques publiques. Nous allons donc exposer les conclusions importantes de cette littérature avant de nous pencher spécifiquement sur l'évaluation des infrastructures de transport. Il faut noter que nous nous intéresserons aux transports de marchandises et de passagers, pour lesquels les enjeux sont proches.

Le cadre général d'évaluation des politiques publiques

L'objectif de la première section de cette revue de littérature est de rappeler les éléments méthodologiques permettant de déterminer l'effet causal de la construction d'une infrastructure de transport sur l'activité économique. Cette question recoupe largement la littérature empirique d'évaluation des politiques publiques. Le cadre général de ces travaux a été défini par Rubin (1974) et des avancées importantes ont été proposées dans les dernières années pour répondre aux problèmes économétriques posés par l'évaluation des politiques publiques (voir Givord (2010) pour une présentation en français des avancées récentes dans ce domaine ou Imbens et Wooldridge (2008) pour une présentation plus formalisée). Le cadre standard pour ce type de question consiste à estimer l'effet causal d'un « traitement », ici, le fait d'être relié à une infrastructure de transport. Pour ce faire, on compare les « revenus » potentiels d'un territoire avec ou sans traitement. Plus concrètement, cela signifie dans notre cas que l'on souhaite comparer l'activité économique d'un territoire, selon qu'il est ou non équipé d'infrastructures de transport. L'effet causal du traitement correspond à la différence entre ces deux situations. Le problème fondamental de l'évaluation des politiques publiques réside dans le fait que l'on n'observe pas simultanément les revenus des individus avec et sans traitement. Dans notre cas, cela signifie que, pour une commune donnée, on ne connaît pas simultanément l'activité économique d'un même territoire avec ou sans infrastructure de transport.

Il faut absolument éviter une comparaison directe entre les zones équipées d'infrastructures de transport et les autres zones car elle peut conduire à des biais importants d'endogénéité. Comme nous l'avons expliqué précédemment, ce biais d'endogénéité provient du fait que l'écart des performances économiques entre les zones reliées aux infrastructures de transport et celles qui ne le sont

pas s'explique non seulement par la présence de ces infrastructures de transport mais aussi par d'autres phénomènes sous-jacents. Ces phénomènes peuvent correspondre aux performances économiques passées ou futures des zones, à l'effet d'autres politiques publiques, etc. Une évaluation naïve de l'impact des infrastructures de transport risque d'attribuer l'effet conjoint de l'infrastructure de transport et de ces phénomènes sous-jacents à la seule infrastructure. Ainsi, cette méthode naïve ne permet pas de dégager l'effet causal de l'infrastructure.

Une première solution consiste à intégrer des variables de contrôle. Dans notre cas, il peut s'agir de la population, du nombre d'emplois ou d'entreprises localisés sur un territoire. Cette première précaution permet de raisonner « toutes choses égales par ailleurs », c'est-à-dire d'estimer l'effet des infrastructures de transport sur l'activité économique, sans tenir compte de l'effet des autres variables de contrôle incluses dans le modèle. Cette première solution n'est cependant pas suffisante. L'observation d'un territoire est toujours imparfaite et incertaine. Dans notre cas précis, il est souvent malaisé de déterminer exactement les perspectives économiques d'un territoire, qui sont pourtant connues (ou supposées connues) des agents économiques et des acteurs politiques, lors de leurs décisions. Il est donc impossible d'introduire toutes les variables de contrôle qui expliquent le phénomène étudié et donc de résoudre complètement le problème d'endogénéité. Ces variables de contrôle ne sont pas suffisantes étant donné qu'il peut exister des phénomènes non observés ou observés avec erreur. Il est donc nécessaire d'utiliser des méthodes plus poussées.

Il existe de nombreuses méthodes économétriques permettant de répondre au problème d'endogénéité. Le choix d'une méthode particulière dépend souvent des particularités de la politique publique évaluée ou des données disponibles. Leur point commun est de chercher à déterminer la situation économique qui aurait prévalu si cette politique publique n'avait pas été mise en œuvre. Cette situation, appelée « contrefactuel », est comparée à l'activité économique observée, permettant ainsi d'établir l'effet causal de la politique publique étudiée. Ces méthodes s'appuient souvent sur des « expériences naturelles ». En économie, contrairement aux évaluations médicales, il n'existe pas, de manière générale, de dispositif expérimental permettant de disposer d'un groupe de contrôle et d'un groupe traité où les individus sont assignés de manière aléatoire. Une expérience naturelle a pour particularité de rendre plausible l'hypothèse selon laquelle une partie des territoires que l'on compare n'a pas été choisie selon des critères endogènes, c'est-à-dire liés au phénomène étudié, mais exogènes. Concrètement, il est possible de trouver un élément permettant d'expliquer le choix politique de construire une infrastructure de transport, mais qui n'est pas lié au phénomène étudié, c'est-à-dire l'activité économique des territoires.

Dans la suite de ce document, nous présentons plusieurs méthodes particulières permettant l'évaluation causale des infrastructures de transport. Ces méthodes s'attachent à reconstruire un groupe de contrôle et un groupe de traitement crédibles (voir l'évaluation empirique du RER dans la suite du document) ou à démontrer l'existence de variables instrumentales, afin de répondre concrètement au problème d'endogénéité.

Les réponses apportées à l'impact des transports sur l'activité économique

Pour résoudre le problème d'endogénéité, les économistes ont récemment proposé des méthodes astucieuses et pertinentes pour les évaluations des infrastructures de transport. En particulier, les travaux de Michaels (2008); Duranton et Turner (2012); Donaldson (2010), utilisent les plans initiaux de réseaux autoroutiers ou ferroviaires comme variable instrumentale. La méthode des variables instrumentales permet de corriger le biais inhérent à l'évaluation des infrastructures de transport. Ce biais, appelé biais d'endogénéité dans la littérature économique, provient du fait

que la variabilité des performances économiques entre les zones reliées aux infrastructures de transport et celles qui ne le sont pas s'explique non seulement par la présence de ces infrastructures de transport mais aussi par d'autres phénomènes sous-jacents et des différences antérieures. La méthode des variables instrumentales s'appuie sur l'existence de variables, dites instrumentales, qui expliquent le phénomène étudié, ici le rôle économique des transports, sans être liées directement aux phénomènes économiques sous-jacents ou aux différences initiales dont on veut exclure l'effet dans notre analyse.⁴ Sans entrer dans les détails techniques de cette méthode, il faut retenir que, correctement appliquée, elle permet d'isoler l'impact causal du phénomène étudié en excluant l'effet des phénomènes sous-jacents inobservés qui brouillent l'analyse⁵.

Dans les travaux de Michaels (2008); Duranton et Turner (2012); Donaldson (2010), précédemment mentionnés, les plans initiaux de réseaux autoroutiers ou ferroviaires constituent de bonnes variables instrumentales pour deux raisons. D'une part, ces schémas directeurs expliquent bien le réseau actuel. D'autre part, les auteurs démontrent qu'ils sont si anciens que les chances de causalité inverse sont minimales. En d'autres termes, ils ont été conçus à une époque suffisamment éloignée pour que leur lien avec les perspectives économiques ou démographiques de territoires étudiés soit négligeable. Plus concrètement, ces plans ont été établis pour des motifs qui ne sont pas liés à la dynamique démographique et économique actuelle des territoires desservis. Ces articles utilisent également les segments du réseau qui n'ont finalement pas été construits pour des raisons « accidentelles » (par exemple des difficultés budgétaires conjoncturelles) pour mener un test de placebo. Ainsi, les auteurs démontrent que les régions situées à proximité d'une infrastructure figurant sur le schéma directeur mais finalement non construites ne connaissent pas une croissance supérieure aux autres régions non reliées. En d'autres termes, il n'existe pas de lien entre le phénomène étudié, ici la croissance d'une région, et la variable instrumentale choisie. L'hypothèse d'exogénéité du plan initial est donc validée. Il est à noter que les résultats issus de ce type de travaux trouvent un impact du réseau de transport très différent des évaluations naïves. Il apparaît donc fondamental d'utiliser ce type de techniques pour une évaluation valide. La suite de cette section expose en détail trois articles s'attachant à déterminer l'effet causal des infrastructures de transport.

Duranton et Turner (2012) évaluent l'impact du réseau routier sur l'évolution locale de l'emploi aux États-Unis. Ils utilisent trois instruments pour expliquer la structure du réseau autoroutier de 1983 et évaluer son impact sur la croissance de l'emploi dans les différentes aires urbaines pendant les vingt années suivantes. Le premier instrument est le plan autoroutier de 1947, qui prévoyait un schéma d'ensemble du réseau visant à connecter les grands centres de population de l'époque et à assurer un objectif de défense nationale des années d'après-guerre, ainsi qu'à connecter les points de passage vers le Canada et le Mexique. Dès lors, ce plan initial a un bon pouvoir explicatif du réseau plusieurs décennies plus tard, mais les auteurs démontrent qu'il a peu de chances d'être directement lié à la croissance urbaine de 1983 à 2003. Le deuxième instrument correspond au nombre de kilomètres de voies ferrées en 1898 dans les différentes aires urbaines. L'idée de cet instrument est qu'il était moins coûteux de construire une autoroute sur l'emprise d'une voie ferrée abandonnée. L'existence d'une voie ferrée en 1898 favorise donc la construction d'une autoroute mais n'est pas liée à la croissance de l'emploi. Le troisième instrument est constitué des grandes routes d'exploration des États-Unis du 16^e au 19^e siècle. Ces trois instruments expliquent

4. Deux conditions sont nécessaires pour s'assurer de la validité d'une variable instrumentale. La condition de rang requiert que cette variable explique la variable d'intérêt, ici la construction d'une infrastructure de transport. La condition d'exclusion nécessite que la variable instrumentale soit indépendante des phénomènes sous-jacents susceptibles de biaiser les estimations.

5. Pour une explication détaillée de la méthode des variables instrumentales, on pourra consulter Givord (2010).

correctement la structure du réseau autoroutier en 1983, tout en satisfaisant raisonnablement l'hypothèse d'exogénéité. Les résultats montrent que l'élasticité de la croissance locale par rapport à la taille du réseau autoroutier est de l'ordre de 15 %. Cela signifie qu'un doublement du nombre de kilomètres d'autoroutes conduit à une augmentation de 15 % de l'emploi d'une ville. Les estimations instrumentales sont plus élevées que les estimations naïves dans ce cas. Cela tendrait ainsi à montrer que les routes ont été construites plutôt dans des zones à faible croissance. Les auteurs mesurent également l'impact du réseau ferroviaire de 1898 sur la croissance urbaine sur 1920-2000. Ils obtiennent une élasticité particulièrement importante de 30 %, soit du même ordre de grandeur.

Donaldson (2010) s'intéresse à l'impact de la construction d'une infrastructure de transport sur la croissance du revenu, en étudiant le cas du réseau ferroviaire indien entre 1870 et 1930, où la quasi totalité du réseau fut construite. L'auteur explique que la première cause de construction du réseau était militaire, car il permettait de détecter et de mater les potentiels soulèvements de population à travers le pays. Dans une équation naïve, Donaldson (2010) estime que l'arrivée du réseau ferré dans un district augmente le revenu par tête de 16 %. Le modèle utilisé par l'auteur permet d'éliminer les différences entre territoires qui ne varient pas dans le temps, c'est-à-dire toutes les raisons permanentes qui expliqueraient qu'un district soit à la fois plus productif que les autres et plus susceptible d'être traversé par le réseau ferré. La méthode sélectionnée permet également de s'assurer que les différences variables entre districts ne biaisent pas les estimations. En effet, les décideurs politiques ont pu décider de construire des voies de chemin de fer dans les endroits promis à une forte croissance. Pour traiter ce problème, l'auteur procède à des tests de falsification. Il s'appuie pour cela sur 40 000 km de voies ferrées dont la construction était projetée mais n'a jamais été entreprise. Cet abandon provient d'éléments exogènes. Par exemple, le chef du gouvernement indien décida en 1868 d'un plan de construction sur trente ans, entamé tout juste avant son retrait en 1869. Son successeur, choisissant une approche économique plus prudente, arrêta les travaux. Le plan de 1868 ne fut donc pas respecté. Il est ainsi possible déterminer si les districts traversés par ces lignes projetées mais jamais construites présentent une croissance des revenus plus importante que les autres districts. Cette croissance est nulle, suggérant ainsi que le choix des tracés du plan de 1868 ne répondait pas à une logique économique. L'instrument est donc valide. D'autres tests de falsification viennent confirmer l'hypothèse d'exogénéité du projet de réseau ferré, c'est-à-dire son indépendance vis-à-vis des perspectives de croissance du revenu par tête.

Enfin, la confrontation avec un modèle théorique structurel permet de dépasser le simple stade du constat et d'expliquer les phénomènes dégagés. Ce modèle prévoit que la construction du réseau ferroviaire modifie le revenu par tête selon un canal précis. Cette construction permet une baisse des coûts de transport, qui accroît le commerce à l'intérieur du pays. Le gain lié à l'accroissement des échanges est synthétisé par une variable très simple, qui correspond à l'opposé du taux d'ouverture de la région⁶. L'introduction de cette variable dans la régression annule tout à fait l'impact de la variable de réseau ferré sur le revenu. Cela signifie que cet impact s'explique uniquement par les mécanismes du modèle et provient donc entièrement de la réduction des coûts de transport de marchandise, et de la baisse de la part de la production locale dans la consommation qui en résulte.

Henderson et al. (2011) utilisent également l'infrastructure de transport historique comme instrument de l'infrastructure actuelle pour le cas de la décentralisation de l'activité industrielle parmi les villes chinoises. Ils montrent que la présence et l'étendu du réseau ferré en 1990, prédit par son niveau de 1962, a joué un rôle important dans la baisse de la concentration spatiale de l'activité

6. Le taux d'ouverture d'une région correspond à la moyenne des importations et des exportations ramenée à la production totale.

industrielle dans les centres-villes chinois. Ainsi, les autoroutes radiales comme circulaires contribuent à l'étalement de la population alors que transports en commun, comme le bus, conduisent à une relative densification. Les réseaux ferrés, quant à eux, contribuent à une décentralisation massive des activités industrielles.

1.2 Comment estimer les facteurs de décision d'un choix de localisation ?

Après avoir étudié les conditions d'une évaluation valide des infrastructures de transport, nous abordons maintenant les moyens pour intégrer cette question au champ des choix de localisation.

L'importance donnée à la concurrence entre les métropoles dans la localisation des entreprises conduit à s'intéresser tout particulièrement aux entreprises de services lorsqu'il est question d'attractivité. Ce constat pourrait conduire à favoriser l'étude de cette catégorie d'entreprises dans l'évaluation des projets semblables au Grand Paris Express. Pourtant, il serait regrettable d'oublier l'industrie dans cette analyse, l'Île-de-France restant la première région industrielle du pays⁷. D'autant plus que ces deux types d'entreprises doivent être distingués dans l'évaluation de l'impact des transports de passagers sur les choix de localisation des entreprises. Les transports de voyageurs jouent sur les coûts de toutes les entreprises en modifiant le prix du foncier ou en améliorant les déplacements de la main-d'œuvre. Par contre, seules les entreprises de service (services aux entreprises ou sièges sociaux) peuvent bénéficier des transports de passagers pour l'accès à leurs clients. L'industrie est beaucoup plus dépendante des infrastructures de transport de marchandises, dont il sera également question, pour vendre sa production. Il est important de noter cette distinction pour l'interprétation de nos résultats.

Première étape : établir un modèle théorique des déterminants des choix de localisation

Les éléments qui entrent en compte dans les choix de localisation des entreprises concernent à la fois le processus de production et la capacité à écouler les produits. Head et Mayer (2004) exposent ces différents éléments. Les entreprises calculent leur profit espéré dans chaque zone où elle peuvent s'implanter. Elles choisissent ensuite la zone qui maximise leur profit. Ainsi, pour un territoire donné, ce choix dépend des coûts auxquels les entreprises font face ainsi que du potentiel marchand. L'équation de profit π_r d'une entreprise qui s'implanterait dans un territoire r s'écrit de la façon suivante :

$$\pi_r = \frac{c_r^{1-\sigma}}{\sigma} M_r - F_r,$$

Dans ces équations, c_r correspond au coût marginal et F_r au coût fixe associés au territoire r . M_r est le potentiel marchand à la Krugman, il est liée aux ventes que l'entreprise peut espérer réaliser si elle s'implante dans le territoire r . Dans le cadre du même modèle, le potentiel marchand peut être décomposé de la manière suivante :

$$M_r = \sum_j \frac{\phi_{rj} E_j}{G_j}.$$

Dans cette seconde équation, le potentiel marchand est défini comme la somme des marchés potentiels de l'entreprise sur chaque territoire j . Il est fonction de trois éléments : ϕ_{rj} représente

7. Voir le site <http://industrie-iledefrance.org/24-panorama-de-l-industrie-regionale.html>

l'accessibilité de la région r à la région j , E_j les dépenses des consommateurs de la région j et G_j la concurrence à laquelle la firme doit faire face dans le territoire j . Cette concurrence est le fait de ses rivales qui peuvent être localisées dans d'autres territoires mais qui vendent dans le territoire j . Ainsi, pour maximiser le potentiel marchand lors du choix de localisation, l'entreprise devra arbitrer entre concurrence et importance de la demande. L'effet de l'accessibilité est ambigu, d'une part, il facilite l'accès de l'entreprise aux autres territoires mais, inévitablement, il facilite également l'accès des autres entreprises et intensifie la concurrence. Il faut noter que l'amélioration des transports collectifs de voyageurs modifie le potentiel marchand pour certaines entreprises, susceptibles d'utiliser ces moyens de transport pour accéder à leurs clients. Il s'agit par exemple des entreprises de services. Dans le cas de l'industrie, ce sont les transports de marchandises qui déterminent l'accès aux clients. Un projet comme le Grand Paris Express ne devrait donc pas avoir de conséquence pour le potentiel marchand des entreprises industrielles. Pour les entreprises de service, l'effet est ambigu, d'une part, des transports de voyageurs plus efficaces améliorent l'accès aux autres marchés mais ils entraînent également une intensification de la concurrence.

Les coûts fixes et variables sont déterminés par le coût du travail et le coût des autres facteurs de production (consommations intermédiaire, foncier, etc.). Les coûts sont également définis par la productivité de ces facteurs. Pour toutes les entreprises, l'amélioration des transports de voyageurs peut modifier les coûts de nombreuses manières : meilleur accès aux consommations intermédiaires, à une main-d'œuvre moins cher ou plus qualifiée, installation dans un territoire où le foncier est moins cher, augmentation de la productivité. Par exemple, un projet comme le grand Paris Express est susceptible de diminuer les coûts des entreprises par l'amélioration de l'accès aux fournisseurs de services, des étapes de la production utilisant le transport de voyageurs ou du fonctionnement du marché du travail. Il peut également les augmenter, par une renchérissement du foncier à proximité des nouvelles gares par exemple. Pour les coûts autant que pour le potentiel marchand, l'effet d'une telle infrastructure est incertain. L'analyse empirique permettra de déterminer où et pour qui l'amélioration des transports de voyageurs favorise ou décourage l'implantation des entreprises.

Deuxième étape : choisir la méthode économétrique pour estimer les déterminants du choix de localisation

La modélisation économétrique des choix de localisation a connu récemment des avancées importantes. Un débat récent a porté sur l'intérêt du recours aux données individuelles. Parmi les deux grandes familles de modèles économétriques disponibles, les modèles Logit conditionnel et les modèles de Poisson, seule la première catégorie permet d'intégrer les caractéristiques des entreprises. Ces modèles consistent à expliquer le choix d'une firme qui se localise dans une commune donnée. Pour expliquer ce choix, il est possible d'inclure des informations sur les entreprises comme sur les territoires⁸. Cependant, ces modèles réclament des données détaillées et posent des problèmes d'estimations lorsque les territoires que les entreprises peuvent choisir sont trop nombreux. La deuxième catégorie de modèles, les modèles de Poisson, correspond à des modèles de comptage. Il s'agit ici de compter le nombre d'entreprises qui s'installent dans un territoire et d'expliquer cet

8. Ces modèles reposent sur des hypothèses fortes. Tout d'abord, ils supposent « l'indépendance des alternatives non pertinentes ». Cette hypothèse signifie que lorsqu'une entreprise détermine sa préférence entre un territoire A et un territoire B, les caractéristiques de tout autre territoire C ne doivent pas entrer en considération. En outre, les données concernant une même commune sont répétées pour toutes les entreprises. Cette agrégation des variables explicatives peut conduire à une sous-estimation des écarts-types dans l'estimation. Cet aspect a longtemps été éludé dans cette littérature.

effectif par les caractéristiques de ce territoire. Les hypothèses sont plus faibles mais la prise en compte de l'hétérogénéité des firmes, d'une importance croissante dans la littérature théorique, est plus difficile.

Les différences entre ces deux types de modèles ne concernent pas seulement le type de données utilisées ; elles reposent également sur la modélisation théorique des choix de localisation des entreprises (Schmidheiny et Brülhart, 2011). Les modèles Logit conditionnel reposent sur l'hypothèse qu'il existe un nombre fixé d'entreprises qui se répartissent ensuite entre les différentes localisations possibles. Ainsi, si un territoire améliore ses caractéristiques, attirant d'avantage d'entreprises, cela entraînera nécessairement une diminution de même ampleur du nombre d'entreprises localisées dans d'autres territoires. Dans ce premier cas, une modification des caractéristiques d'un territoire conduit uniquement à une réallocation spatiales des entreprises au sein de la région étudiée. Au contraire, dans les modèles de Poisson, l'amélioration des caractéristiques du territoire permet d'augmenter le nombre de firmes qui choisissent de s'y établir sans affecter aucunement les autres territoires. Il n'y a alors plus de concurrence entre les territoires au sein de la région étudiée et l'effet observé est uniquement dû à une augmentation de l'attractivité extra-régionale du territoire dont les caractéristiques s'améliorent.

Schmidheiny et Brülhart (2011) insistent sur le fait que la réalité économique est située entre ces deux cas extrêmes. Les modèles de Logit imbriqué permettent de combiner les caractéristiques des deux premiers modèles. Ils reposent sur l'arbre de choix des entreprises et supposent ainsi qu'une firme détermine sa localisation selon un processus en deux étapes. Tout d'abord, elle effectue un choix entre plusieurs régions, par exemple, entre Greater London et l'Île-de-France. Ensuite, elle définit sa localisation plus précise au sein de la région choisie, par exemple, une commune en Île-de-France. L'avantage de ce modèle est d'offrir une situation intermédiaire entre le modèle de Logit conditionnel, qui intègre uniquement la concurrence entre territoires, et le modèle de Poisson, qui ne prend en compte que les effets internes d'une amélioration des conditions offertes aux entreprises. Il permet de déterminer le poids relatif de chacun de ces deux facteurs. L'autre intérêt est de pouvoir tester plusieurs arbres de choix possibles pour la localisation des entreprises. Ainsi, il est possible de déterminer si une entreprise choisit d'abord un pays ou un type de territoire, comme une région-capitale, avant de choisir sa localisation précise. Pour cela on détermine le modèle qui correspond le mieux aux données disponibles.

Il serait passionnant de pouvoir mener une étude modélisant la concurrence entre les métropoles européennes de l'Union Européenne dans d'attraction des entreprises. Nous pourrions alors répondre entièrement à la question de l'attractivité induite par la modification du réseau de transport, en intégrant la dimension internationale de la question. Une telle étude permettrait également de déterminer le nombre total d'entreprises attirées en Île-de-France par l'amélioration des transports et pas uniquement d'étudier leur répartition dans les communes de région parisienne. Pour cela, il faudrait disposer de données sur l'implantation des investisseurs étrangers dans au moins deux grandes régions-capitales de l'Union Européenne, par exemple l'Île-de-France et Greater London, sur longue période. De telles données permettraient d'évaluer l'impact des extensions ou des créations d'infrastructures de transport sur le volume total des investissements dirigés vers chaque métropole. Nous ne disposons pas de telles données à l'heure actuelle, même si une collaboration avec le centre d'économie spatiale de la LSE semble possible dans l'avenir pour mener ce type de projet.

Concrètement, les estimations économétriques de ces modèles s'appuient sur l'équation de profit présentée dans la partie précédente. Pour une entreprise i s'implantant dans un territoire

r , le profit s'écrit $\pi_{ri} = x'_r\beta + u_{ri}$. Le vecteur x_r contient toutes les caractéristiques observables de la commune déterminant la localisation de l'entreprise. Ces variables ont été détaillées dans la section précédente (le salaire, la productivité de la main-d'œuvre, le prix du foncier, le prix des consommations intermédiaires, la demande locale, le niveau de concurrence, l'accessibilité aux autres territoires, etc.). u_{ri} est un terme d'erreur propre à l'entreprise i et au territoire r , supposé indépendant des variables explicatives du modèle. L'entreprise choisit le territoire qui lui procure le profit maximal. Les modèles Logit et Poisson peuvent être présentés à partir de cette équation de profit. Dans le modèle Logit conditionnel, la probabilité qu'une entreprise choisisse le territoire r s'écrit de la façon suivante :

$$P_r = \frac{\exp(x'_r\beta)}{\sum_i \exp(x'_i\beta)}$$

Le vecteur des paramètres β est estimé par maximum de vraisemblance. On peut ainsi déduire l'espérance du nombre de firmes qui choisissent de s'implanter dans la région r comme le produit de la probabilité qu'une entreprise s'implante dans la région P_r et le nombre total d'entreprises n . Ainsi, $E(n_r) = nP_r$. Pour le modèle de Poisson, l'hypothèse principale est :

$$E(n_r) = \exp(\alpha + x'_r\beta)$$

L'espérance du nombre de firmes qui choisissent de s'implanter dans la territoire r étant ainsi définie, on peut estimer le modèle par la méthode du maximum de vraisemblance. Il faut alors supposer que la variable correspondant au nombre d'entreprises dans le territoire r suit une loi de Poisson. Schmidheiny et Brülhart (2011) montrent que les modèles de comptage de type Poisson, les modèles multinomiaux de type Logit conditionnel ou imbriqué donnent des estimateurs identiques quand on n'inclut que des variables propres aux territoires. Par contre, c'est l'interprétation et la prévision qui diffèrent, en lien avec les considérations théoriques présentées dans les paragraphes précédents.

Enfin des progrès récents ont porté sur la résolution du biais d'endogénéité lié à l'estimation de variables déterminées localement comme les salaires, la demande locale ou encore les infrastructures de transport. Liu et al. (2010) ont montré qu'une méthode de type « control function » permettait une analyse causale plus raffinée que le reste de la littérature. Il s'agit d'une façon alternative d'implémenter la méthode des variables instrumentales. Les auteurs montrent que l'utilisation d'une variable instrumentale, qui n'est pas déterminée localement, permet d'estimer l'effet causal du salaire sur les choix de localisation des firmes. Ils montrent également que l'utilisation d'une fonction de contrôle conduit à des résultats sensiblement différents de ceux obtenus par des méthodes plus basiques. Cela montre donc que ces problèmes d'endogénéité ne concernent pas seulement l'évaluation des infrastructures de transport mais également les autres variables qui entrent dans les choix de localisation des entreprises.

Troisième étape : intégrer l'impact des infrastructures de transport dans le choix de localisation

Il existe un certain nombre d'études qui incluent des variables d'infrastructure de transport dans un modèle économétrique estimant les déterminants du choix de localisation. Nous synthétisons ci-dessous les résultats de ces études, dont il faut noter qu'elles ne sont pas très nombreuses et, surtout, qu'elles ne s'intéressent que marginalement à l'impact des réseaux de transport collectif. La plupart des études citées examinent si la présence d'un aéroport ou d'une autoroute importante

affecte la probabilité de localisation des entreprises dans la région. Les infrastructures de type métro sont quasiment absentes de l'analyse. Il faut également noter qu'aucun article, ou presque, ne cherche à traiter les problèmes d'endogénéité précédemment soulevés.

Coughlin et Segev (2000) étudient l'impact de l'existence d'une autoroute majeure sur la localisation des filiales étrangères dans les différents comtés américains entre 1989 et 1994. L'impact est positif et très significatif.

Cheng et Kwan (2000) sont parmi les seuls à inclure une variable mesurant la densité du réseau ferré dans la localisation des investissements étrangers entre les différentes provinces chinoises, sur la période 1985-1995. Ils utilisent également une variable de densité du réseau routier, dont l'effet estimé est significatif, à la différence du réseau ferré qui semble n'avoir aucun impact. Toutefois, ces résultats sont à prendre précaution car d'autres auteurs trouvent le résultat inverse pour le même pays. En effet, Kang et Lee (2007) s'intéressent à la localisation des investissements coréens en Chine. Leur base de données s'étend de 1988 à 2002 pour l'ensemble des provinces chinoises. Les auteurs incluent la densité ferroviaire locale dans leur modèle et, à la différence de la densité du réseau autoroutier, ils obtiennent un impact significatif et important pour cette variable.

Strauss-Kahn et Vives (2009) proposent une étude qui est probablement l'une des plus pertinentes pour notre question. Ils étudient les déterminants de la relocalisation des sièges de firmes américaines entre 1996 et 2001, entre différentes aires urbaines. La base de données contient près de 30 000 sièges sociaux, dont une part non négligeable change de localisation (36 sur les 500 plus grands par exemple). Les auteurs soulignent en particulier une tendance des entreprises à migrer du nord vers le sud du pays. L'analyse repose sur un modèle de Logit emboîté. Les entreprises choisissent d'abord si elles souhaitent changer de localisation, puis, en cas de changement, elles déterminent la nouvelle aire urbaine d'implantation. Dans leur modèle économétrique, ils introduisent des variables indicatrices portant sur l'existence d'une plateforme aéroportuaire de petite ou de grande taille, qui constitue leur seul variable d'infrastructure de transport. L'impact est important, un petit aéroport augmente l'attractivité de l'aire urbaine de 40%, un grand de 90%. De plus, la présence d'un grand aéroport réduit la probabilité de départ de l'aire urbaine de 40%. Toutefois, l'éventuel biais d'endogénéité n'est pas pris en compte. De plus, des variables importantes, comme le lien entre la spécialisation de l'aire urbaine et le domaine d'activité du siège social, l'importance des secteurs des services aux entreprises n'ont pas été intégrées à l'analyse. Davis et Henderson (2008) utilisent une instrumentation fondée sur la méthode des moments généralisés pour expliquer la localisation de nouveaux sièges sociaux à l'intérieur des Etats-Unis. Ils montrent que ce type d'établissement présente une disposition plus forte que le reste de l'économie à l'agglomération spatiale et soulignent l'importance prépondérante de la présence de services aux entreprises et de services financiers à proximité.

Py et Hatem (2009) proposent une analyse des choix de localisation des entreprises multinationales sur la période 2002-2006. Cet article compare les choix de localisation selon le secteur d'activité ou la fonction de l'établissement considéré. Aucune variable concernant les infrastructures de transport n'est intégrée. Il est néanmoins intéressant de noter l'impact variable du potentiel marchand, particulièrement dans l'optique d'une évaluation d'une infrastructure du type du Grand Paris Express. En effet, les filiales des services aux entreprises et des logiciels et services informatiques sont les plus sensibles au potentiel marchand. Concernant les fonctions au sein des entreprises, les sièges, les établissements de prestation de services et de distribution sont les plus réactifs aux changements de la taille du marché. Defever (2006) avait mené une analyse similaire sur 11 000 investissements dans 23 pays d'Europe sur la période 1997-2002 et trouvait des ré-

sultats très similaires. Les activités de siège social, de distribution et de marketing sont les plus sensibles à la taille de la demande locale, et les moins sensibles aux coûts du travail.

Holl (2004a,b) fournit une étude très détaillée sur la période récente portant sur l'impact de l'accessibilité à une autoroute en Espagne (1980-1994) et au Portugal (1986-1996). Elle estime un modèle de choix de localisation des entreprises du secteur manufacturier en s'appuyant sur des données géolocalisées des créations d'entreprises au niveau de la commune. Le modèle économétrique estimé contient des variables indicatrices correspondant à la distance à l'autoroute (moins de 10 km, 10-20 km, 20-30 km, etc.). Les résultats révèlent un impact important de l'accès direct à une infrastructure routière. Pour le cas espagnol, les créations d'entreprises industrielles augmentent de 14 % dans les communes situées à moins de 10 km d'une autoroute.

Il faut retenir de cette revue de la littérature que la localisation des sièges sociaux est très sensible à la présence locale de firmes de services aux entreprises et de services financiers. La construction d'une infrastructure comme le Grand Paris Express devrait justement conduire à une augmentation de la productivité et de la présence des fournisseurs de service. En effet, l'amélioration des conditions de transport améliore l'accès aux clients et les incite à une localisation proche des gares, plus adaptée ou intéressante économiquement. Dès lors, l'attractivité pour les sièges sociaux internationaux serait renforcée à travers l'amélioration de l'approvisionnement en biens et services intermédiaires.

2 Méthode d'évaluation de l'impact du réseau express régional

La première partie a permis de poser les jalons théoriques du lien entre transports urbains et localisation d'entreprises. Dans la suite de ce rapport, nous proposons une évaluation empirique de l'impact du RER sur la localisation des investissements étrangers en Île-de-France. Cette évaluation permet d'illustrer et de confronter aux données une partie des arguments avancés dans la partie précédente.

2.1 Rapide histoire du Réseau Express Régional

Pour évaluer de façon précise l'impact du Réseau express régional, il est nécessaire d'étudier le processus de décision qui a conduit à la mise en place de ce réseau. En effet, la méthode permettant de résoudre le biais d'endogénéité, dont l'origine a été détaillée dans la première partie, repose sur l'histoire de la mise en place du RER et plus particulièrement sur le schéma directeur de 1965 (voir Collardey (1999) pour plus de détails sur l'histoire du RER).

Un réseau ferré d'Île-de-France construit en étoile

Construit principalement durant le XIXe siècle, le réseau ferré d'Île-de-France est le reflet du réseau ferré national. Des origines jusqu'à la constitution de la SNCF en 1937, il est construit et géré par plusieurs compagnies privées indépendantes, selon un système de concessions. Chacune d'elle se voit attribuer une zone spécifique où elle constitue un réseau relativement indépendant des autres réseaux, avec une gare parisienne propre en tête de ligne. Par exemple, il s'agit de la gare du Nord pour le Compagnie des chemins de fer du nord ou la gare d'Austerlitz pour la Compagnie Paris-Orléans. Ce contexte historique explique pourquoi chaque gare parisienne est un cul de sac. Cette situation est particulièrement pénalisante pour le trafic de banlieue. En effet, les usagers

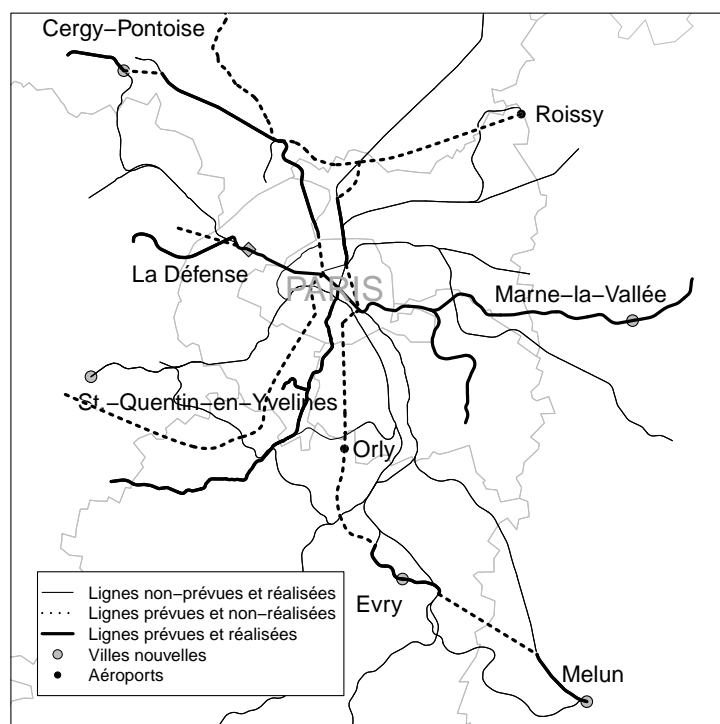


Figure 1 – Comparaison du SDAURP de 1965 et du réseau ferré d'Île-de-France en 2005

sont contraints à de longs changements pour rejoindre le réseau du métropolitain et les autres gares.

Le schéma directeur de 1965 : un projet très ambitieux

Cette structure du réseau explique pourquoi l'idée d'une amélioration du réseau ferré d'Île-de-France n'est pas nouvelle. Déjà, dans les années 1930, le projet Langevin proposait de relier les lignes de l'agglomération parisienne pour éviter les ruptures de charge entre les gares parisiennes. Ce projet n'avait pas abouti, si ce n'est à l'électrification de la ligne de Sceaux, qui constitue aujourd'hui la branche sud de la ligne B du RER. C'est véritablement le SDAURP (Schéma directeur d'aménagement et d'urbanisme de la région de Paris)⁹ de 1965 qui va lancer l'idée d'une vaste modernisation du système ferroviaire de la région capitale.

Pour contrôler l'expansion de Paris, les pouvoirs publics décident de construire des villes nouvelles à 30 kilomètres de Paris. Ces nouveaux centres urbains doivent permettre d'éviter l'asphyxie de la partie centrale de l'agglomération en drainant l'augmentation de la population. Un nouveau système ferroviaire est imaginé pour relier ces villes nouvelles à Paris. Très ambitieux, il envisage d'ajouter des centaines de kilomètres de lignes nouvelles au réseau ferré hérité du XIXe siècle, encore parfois desservi par des trains à vapeur. Ce projet permet ainsi de relier efficacement les neuf villes nouvelles, prévues par le SDAURP, au centre de l'agglomération. À la différence des lignes existantes, le nouveau réseau traverse Paris et relie les gares entre elles.

9. Le SDAURP est disponible à l'adresse suivante : <http://www.driea.ile-de-france.developpement-durable.gouv.fr/presentation-du-sdaurp-de-1965-a970.html> et le projet du RER est exposé pages 184 à 190

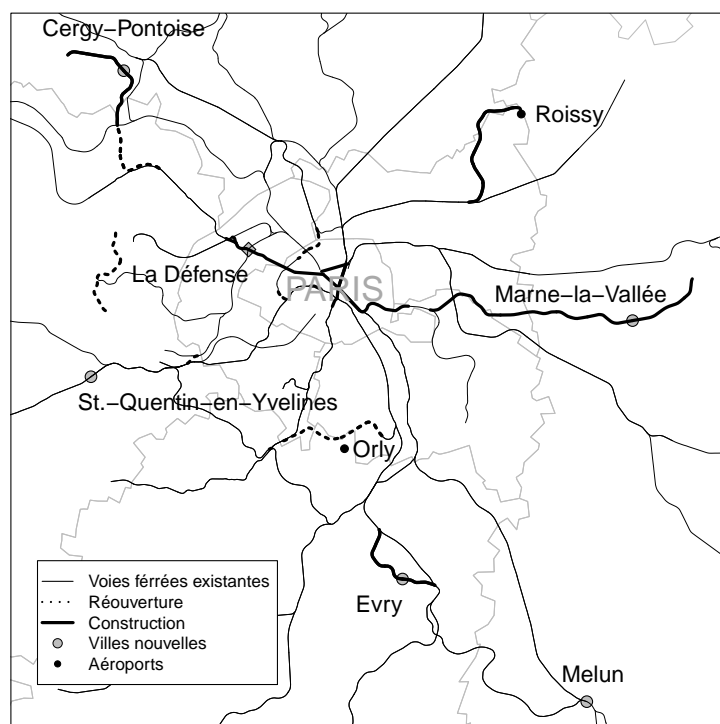


Figure 2 – Construction et réouvertures de lignes en Île-de-France entre 1975 et 2005

Le futur réseau express régional est décrit de la manière suivante dans le document du schéma directeur de 1965 : « *Moyen de transport à haute capacité, le réseau express régional doit nécessairement drainer les principaux centres d'emploi. Les lignes proposées convergent vers Paris et le traversent. Si ces lignes sont dessinées d'une extrémité à l'autre de la future agglomération, c'est moins pour permettre d'aussi longs déplacements, qui resteront peu nombreux, que pour supprimer les ruptures de charge. [. . .] A l'extérieur de Paris, les lignes suivent les axes d'urbanisation prévus par le schéma directeur, traversant les nouveaux centres urbains ou y aboutissant.* »

La construction d'un réseau express régional plus modeste

La crise économique des années 1970 a conduit à une mise en œuvre plus modeste de ce plan. Seulement six villes nouvelles ont été fondées : Cergy-Pontoise, Marne-la-Vallée, Melun-Sénart, Evry et Saint-Quentin-en-Yvelines. Les grandes transversales nord-sud initialement prévues n'ont pas été mises en place (voir figure 1) et le RER a été développé en s'appuyant largement sur le réseau existant. Peu de lignes nouvelles ont été réalisées. Ces constructions concernent le centre de l'agglomération et le voisinage des villes nouvelles et des aéroports (voir figure 2). Les réalisations plus importantes ont lieu dans les années 1970 et 1980. A l'exception de l'ouverture du RER E, les années 1990 correspondent principalement à une extension du réseau sur les lignes existantes en grande banlieue (voir figure 3).

Quatre tunnels ont été creusés dans Paris : la Défense – Nation pour le RER A, Luxembourg – Gare du Nord pour RER B, Gare d'Orsay – Invalides pour le RER C et Haussmann-Saint-Lazare – Gare de l'Est pour le RER E. Le RER D utilise quant à lui les infrastructures des lignes A et B. A l'extérieur de Paris, quatre branches nouvelles ont été construites pour desservir les villes nouvelles de Marne-La-Vallée, Evry et Cergy-Pontoise ainsi que l'aéroport Paris-Charles-de-Gaulle.

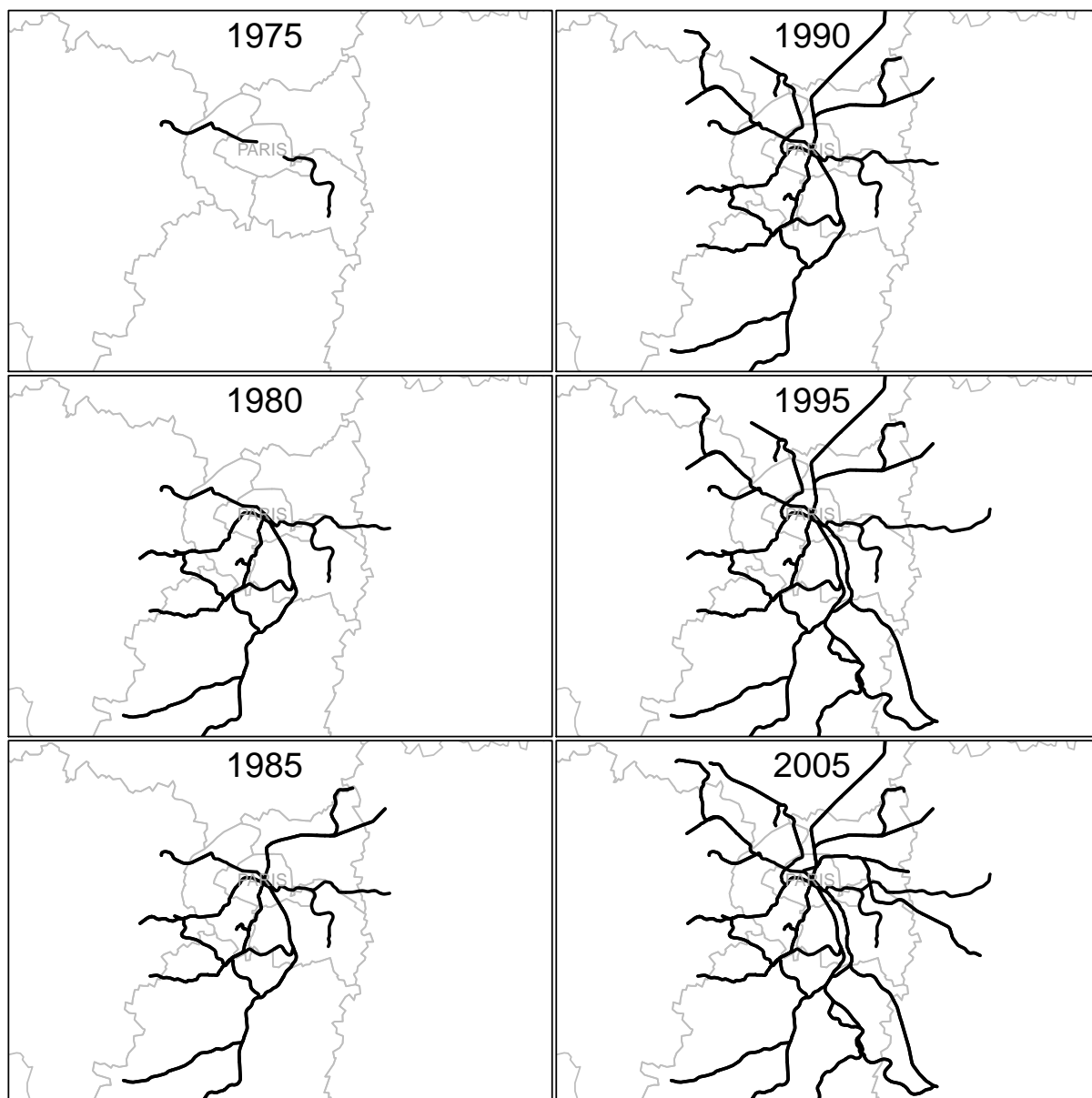


Figure 3 – Extension du réseau RER entre 1975 et 2005

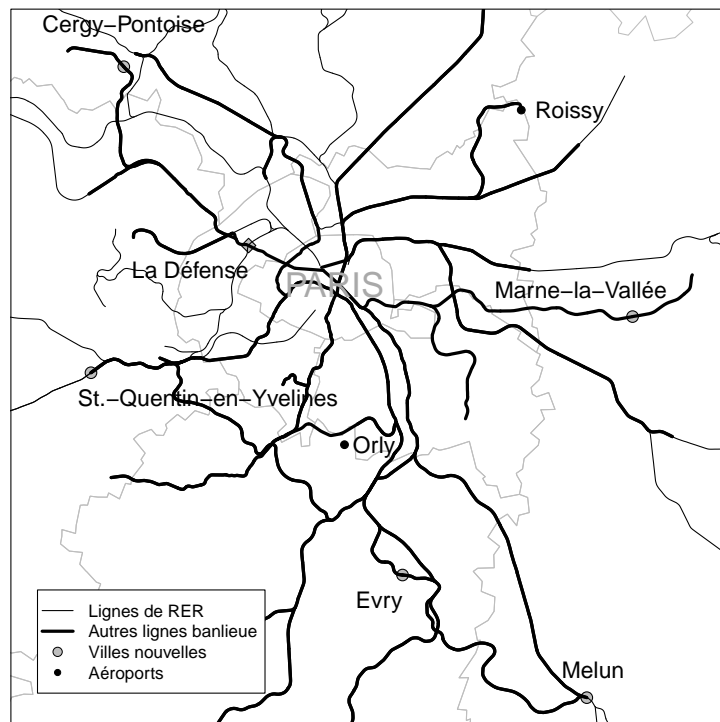


Figure 4 – Réseau ferré d’Île-de-France en 2005

Une portion de la ligne Grande Ceinture a été rouverte pour desservir l’aéroport d’Orly. À l’exception de ces lignes nouvelles, les 587 km du RER sont principalement composées d’infrastructures préexistantes au projet de réseau express régional. Ainsi, les voies situées entre les pôles d’intérêt majeur, comme les villes nouvelles, les aéroports et la Défense, ont été converties en ligne du RER. À l’inverse, les autres lignes n’ont pas connu une modernisation de même ampleur et restent aujourd’hui exploitées par des trains de banlieue (voir figure 4).

2.2 Quel groupe de traitement, quel groupe de contrôle pour évaluer le RER ?

Dans la première partie, nous avons exposé en quoi les questions d’endogénéité peuvent biaiser les estimations de l’effet causal d’une infrastructure de transport. Nous allons maintenant présenter la méthode que nous proposons pour éviter ce biais. La démarche de notre méthode consiste à comparer des communes qui présentent les mêmes perspectives économiques à l’exception de leur intégration au réseau express régional. Pour cette raison, les villes nouvelles, la Défense et les aéroports de Paris ont été retirées de notre étude. D’une part, ces territoires présentaient des perspectives économiques particulièrement favorables, notamment pour l’implantation d’entreprises étrangères. Ce sont donc des cas particuliers en Île-de-France. D’autre part, il n’existe pas d’exemple de ville nouvelle ou d’aéroport exclu du réseau express régional et pouvant servir de groupe de contrôle. Il n’est donc pas possible de trouver des communes de contrôle pour ces territoires. Ils ne peuvent donc pas être inclus dans notre étude. Malgré cela, notre méthode permettant de déterminer l’impact causal de la construction d’une gare RER sur les implantations d’entreprises, les résultats obtenus pour les communes de peuvent être généralisé, avec précautions, à l’ensemble de l’Île-de-France.

Pour des raisons similaires, il n’est pas possible d’introduire Paris dans notre modèle. Tout

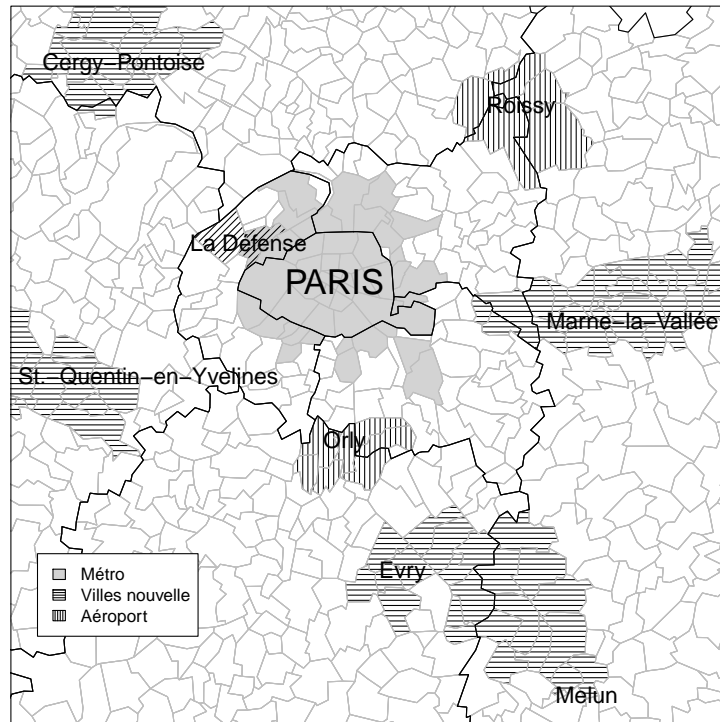


Figure 5 – Communes exclues du groupe de contrôle

d'abord, la commune de Paris est équipée depuis la première moitié du XXe siècle d'un métro efficace. La construction de réseau express régional n'a pas apporté une amélioration des conditions de transport aussi importante qu'en banlieue. L'effet éventuel de cette nouvelle infrastructure serait donc plus difficile à dégager. Ensuite, Paris occupe en Île-de-France une position unique. Il n'est pas possible de trouver une commune comparable, même en utilisant des variables de contrôle.

Enfin, plusieurs communes de proche banlieue ont été reliées au réseau de métro de Paris dans les années 1970, 1980 et 1990. Ces nouvelles infrastructures n'ont probablement pas été sans effet sur les choix de localisation des entreprises. Afin de bien circonscrire notre article à l'étude de l'impact du RER et parce que nos hypothèses identifiantes seraient moins convaincantes pour ces communes, elles ont également été exclues de nos estimations (les communes exclues de notre analyse sont représentées dans la carte 5).

Une fois ces restrictions prises en compte, quatre groupes de contrôle ont été construits à partir des communes restantes, qui sont situées entre Paris et les pôles majeurs de la région parisienne ou en lointaine banlieue. Ils reposent tous sur l'expérience naturelle présentée dans le paragraphe précédent. Il faut noter que le plan initial de 1965 et le réseau actuel présentent un point commun intéressants : les villes nouvelles et les aéroports sont, comme prévu, reliés au réseau. Par contre, on remarque des différences importantes concernant l'itinéraire utilisé pour relier les grands pôles (Voir figure 1). En effet, le SDAURP prévoyait la construction ex nihilo de lignes de RER. Des communes sans gares auraient dû être reliées au RER dans les zones situées entre Paris et les pôles économiques majeurs de la région. A l'inverse, le réseau effectivement construit utilise des lignes préexistantes. Ainsi, des communes de banlieue, qui devaient être initialement exclues du réseau y ont été finalement relié. Elles ont ainsi bénéficié d'une amélioration de leur déserte ferroviaire grâce à la transformation d'une ligne de banlieue en ligne RER. Ces différences entre

le schéma initial et le réseau actuel indiquent que le choix des communes desservies par le RER entre Paris et les pôles économiques majeurs n'est pas dicté par des perspectives de croissance démographique ou économiques mais par des contraintes techniques. Cela nous conduit à considérer que le choix des gares intermédiaires est indépendant du phénomène économique étudié et que les communes conservées dans notre étude ne posent pas de problème de biais d'endogénéité. Nous proposons quatre groupes de contrôle permettant d'exploiter ces éléments, ils sont présentés dans le tableau 1. Les communes retenues dans chacun des groupes sont présentées dans la carte 6.

2.3 Le modèle économétrique

La méthode économétrique doit prendre en compte des particularités importantes de la politique étudiée. Ainsi, le traitement s'étale dans le temps : le RER se déploie progressivement en région parisienne, les communes sont traitées les unes après les autres. En outre, dans la première partie, on a indiqué que la maximisation du profit de l'entreprise guide son choix de localisation. Le profit de l'entreprise est déterminé par de nombreux éléments : le salaire, la productivité de la main-d'œuvre, le prix du foncier, le prix des consommations intermédiaires, la demande locale, le niveau de concurrence, l'accessibilité aux autres territoires, etc. Ces paramètres varient selon les communes, il est donc nécessaire de contrôler ces différences dans nos estimations pour ne pas biaiser nos résultats. Malheureusement, nous ne disposons pas directement de ces variables mais d'autres variables qui en sont des approximations : le nombre d'entreprises pour la concurrence, la population active au lieu de travail pour la disponibilité de la main d'œuvre, la densité pour la disponibilité foncière, le taux de chômage et la population pour l'importance de la demande locale.

Compte tenu de ces spécificités, la méthode des différence-de-différences nous semble la plus adaptée. La première différence correspond à l'évolution communale des implantations d'entreprises avant et après la mise en place du RER. La seconde différence consiste à comparer cette évolution avec celle des communes où le RER n'a pas été mis en place. L'hypothèse centrale de ce modèle, appelée hypothèse de « tendance commune », est que l'évolution des implantations d'entreprises dans les communes reliées au réseau express régional aurait été similaire en l'absence de la construction d'une gare, à celle qu'on observe dans les communes n'ayant pas été reliées. Cette hypothèse ne peut pas être formellement vérifiée mais on peut s'assurer qu'elle est valable dans les années qui précèdent la mise en place du RER. Ce test est présenté dans la dernière partie. Le modèle s'écrit donc de la manière suivante :

$$\log(Y_{i,t} + 1) = \delta RER_{i,t} + \beta x_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t}$$

La variation du nombre d'entreprises résultant de la mise en place du RER est susceptible de dépendre la taille de la commune. Par exemple, si l'ouverture d'une gare a bien un effet sur la démographie d'entreprise, il est probable que l'augmentation du nombre d'entreprises sera plus importante dans une commune qui compte déjà 10 000 entreprises que dans une commune qui n'en compte que 100. Pour cette raison, la variable dépendante $Y_{i,t}$ a subi une transformation en logarithme. On supprime ce problème d'échelle en s'intéressant ainsi à l'augmentation du nombre d'entreprises en pourcentage et non pas en valeur absolue.

Des effets fixes communaux α_i et temporels α_t ont été ajoutés à notre modèle. Ils permettent de neutraliser les différences communales constantes dans le temps et les variations conjoncturelles communes à toute l'Île-de-France. Cette spécification permet de limiter l'hétérogénéité inobservée aux caractéristiques des communes qui évoluent dans le temps. Pour cette raison, les variables

Table 1 – Description des quatre groupes de communes étudiées

Communes exclues de tous les groupes	
Paris, communes équipées d'une station de métro, villes nouvelles, communes aéroportuaires, La Défense	
Groupe 1 : communes d'Île-de-France	
Le groupe de contrôle est constitué de toutes les communes qui ne sont pas situées à proximité d'une gare RER.	
<i>Avantages</i>	<i>Inconvénients</i>
Ce groupe contient un nombre très important de communes et sa constitution est particulièrement simple.	Le groupe de contrôle comprend des communes potentiellement très différentes des communes reliées au RER.
Groupe 2 : communes disposant d'une gare en 2005	
Le groupe de contrôle est composé des communes disposant d'une gare du réseau transilien qui n'est pas desservie par le RER.	
<i>Avantages</i>	<i>Inconvénients</i>
Toutes les communes comparées sont reliées au réseau ferré, elles ont donc des caractéristiques proches. De plus, les communes traitées ont vu leur gare desservie par le RER parce qu'elles étaient situées entre Paris et un pôle économique d'importance. Les autres n'ont pas eu cette chance. Le traitement est donc indépendant du phénomène étudié.	Ce groupe correspond moins précisément à l'expérience naturelle dégagée et peut contenir des communes n'ayant qu'une chance faible de traitement.
Groupe 3 : communes situées à moins de 10 km d'une station RER	
<i>Avantages</i>	<i>Inconvénients</i>
Les communes voisines des communes équipées d'une gare sont susceptibles d'avoir les caractéristiques les plus proches.	L'installation d'une gare RER peut avoir un effet négatif sur la communes située dans le voisinage des gares sans être à proximité immédiate, par effet d'aspiration. Cet effet pourrait conduire à surestimer l'effet du RER pour ce groupe.
Groupe 4 : communes reliées au RER et qui n'auraient pas dû l'être ou communes exclues du RER alors que le 1965 prévoyait de les relier	
<i>Avantages</i>	<i>Inconvénients</i>
Ce groupe correspond le plus fidèlement à l'expérience économique présentée dans le paragraphe précédent.	Peu de communes sont dans cette situation. L'effet du RER pourrait être difficilement détectable.

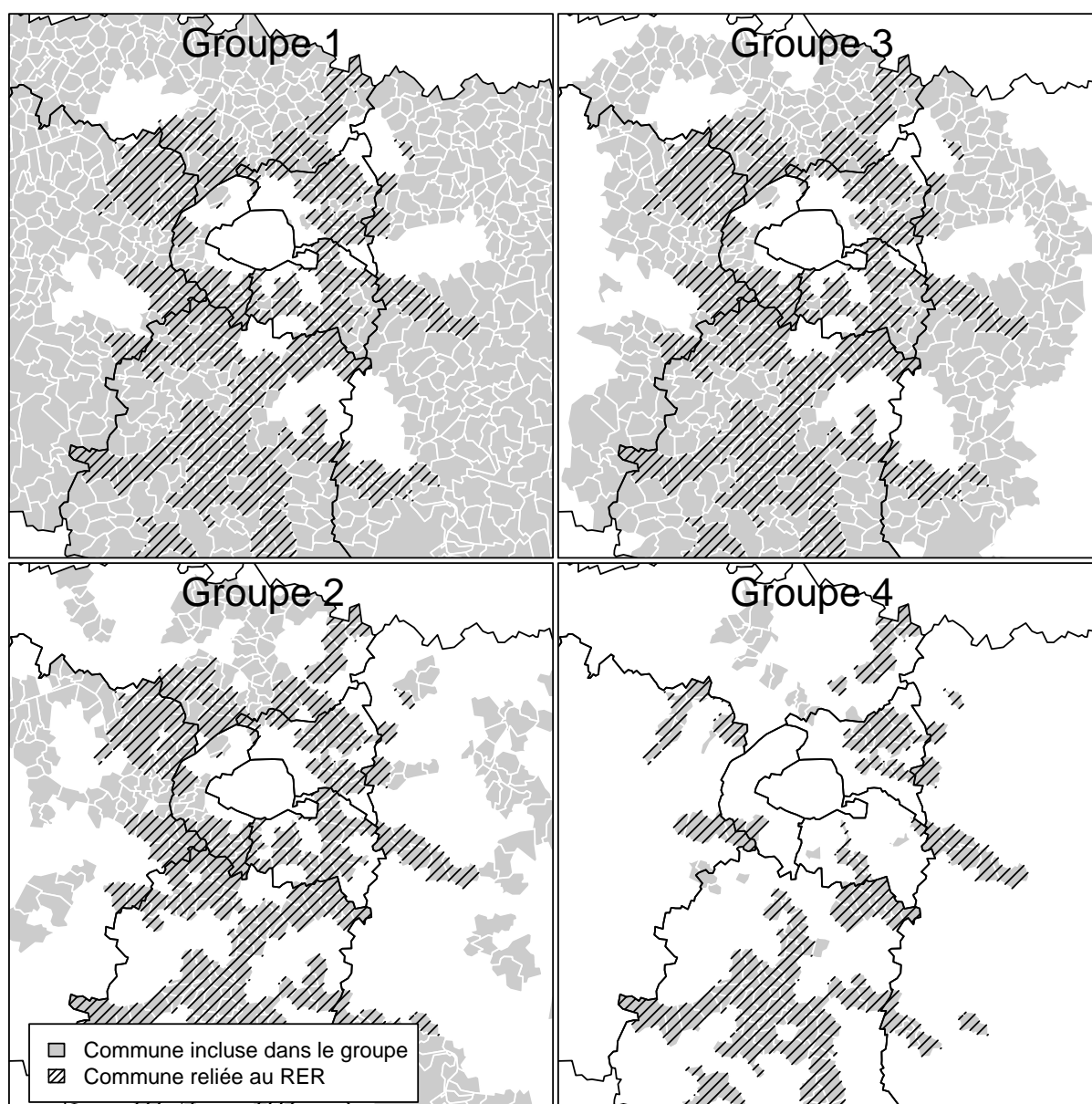


Figure 6 – Communes retenues pour chacun des quatre groupes

de contrôle $x_{i,t}$ sont propres à chaque commune et varient temporellement. Plusieurs variables de traitement $RER_{i,t}$ ont été testées. Les différentes possibilités sont présentées dans la partie suivante.

Enfin, les tests de robustesse proposés par Bertrand et al. (2004), permettant de vérifier la validité des estimateurs fournis par la méthode des différence-de-différences, seront présentés dans une version ultérieure de ce travail.

3 Les données

Afin d'évaluer le modèle présenté dans la section précédente, nous avons exploité des données portant sur les entreprises, les communes et les infrastructures de transports en Île-de-France, des années 1970 aux années 2000. Nous détaillons dans cette partie les données utilisées.

3.1 Comment mesurer l'évolution des investissements étrangers en Île-de-France ?

Jusqu'au milieu des années 1990, une part importante des investissements étrangers en France étaient soumise à une autorisation administrative préalable. Les investisseurs devaient faire enregistrer leurs prises de participation dans des entreprises françaises auprès de la direction du Trésor du Ministère de l'Économie et des Finances. Les données administratives résultant de cette obligation ont été recueillies par l'ancien service statistique du Ministère de l'Industrie, le Sessi. Nous disposons de ces données entre 1974 et 1996. Après cette date, l'assouplissement des contraintes administratives pesant sur les investisseurs étrangers a conduit à un recueil moins systématique. Les données issues de cette source administratives sont donc moins fiables. Ces données contiennent des informations sur les entreprises (secteurs d'activité, département, adresse, date de l'investissement, identifiant entreprise Siren) ainsi que sur l'investisseur (pays d'origine, montant investi).

Elles ont été complétées par des fichiers issus du répertoire Sirene de l'Insee, qui contient la liste de l'ensemble des entreprises et des établissements actifs au 1er janvier entre 1974 et 2004. Ces fichiers permettent d'améliorer la fiabilité des données sur les entreprises (secteur d'activité, commune d'activité, date de création) et de déterminer l'ensemble des établissements d'une entreprise. Nous procédons comme Crozet et al. (2004) pour déterminer la localisation de capitaux étrangers. Selon cette méthode, toute implantation d'établissement par une entreprise détenue par au moins 10 % de capitaux étrangers dans une période de 10 ans qui suit la prise de participation correspond à l'implantation d'un établissement à capitaux étrangers.

Nous disposons également de données d'emploi à partir de 1977 pour les entreprises affiliées à l'Unedic. Ces données pourront éventuellement être utilisées dans des développements ultérieurs visant à lier investissements étrangers et emploi.

Les figures 7 et 8 présentent des statistiques descriptives issues des données d'entreprise utilisées dans cette étude. Les ruptures de série de 1986-87 et 1999-2000, dues à des changements de champ ou des modification du répertoire Sirene, ressortent facilement. Ces figures mettent en avant une croissance constante des implantations et du stock d'établissements en Île-de-France. Concernant les investissements étrangers, on remarque une très nette diminution des implantations enregistrées à partir du milieu des années 1990, due à l'extinction progressive de l'obligation de déclaration des prises de participation étrangère. Les données présentées après 1995, qui ne sont pas utilisées dans les estimations économétriques, correspondent aux implantations des entreprises

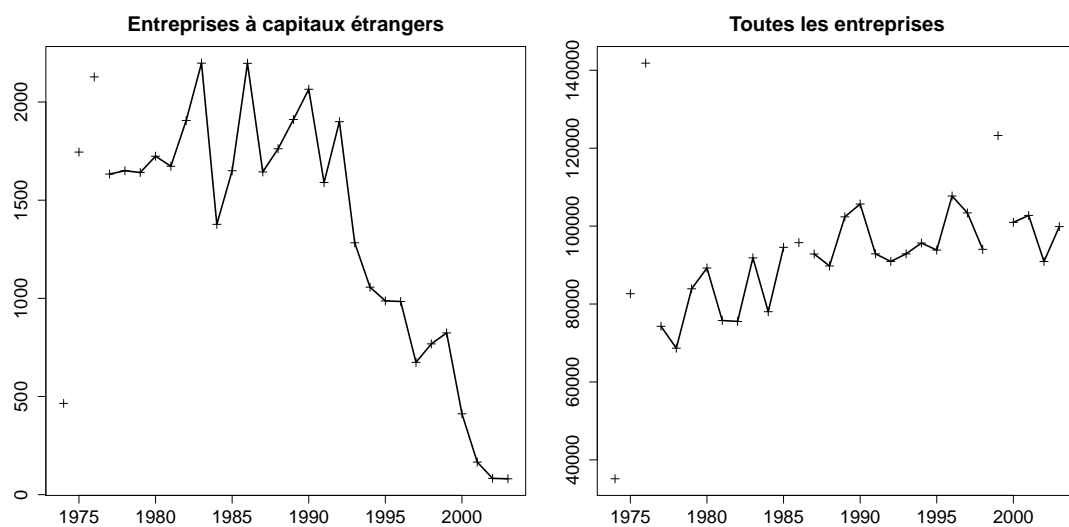


Figure 7 – Implantation d'entreprises et d'entreprises à capitaux étrangers en Île de France entre 1974 et 2003

Note de lecture : les points ne sont pas reliés pour les première années, où la constitution de la base Sirene entraîne des variations importante dans le suivi de la démographie d'entreprise, ainsi que pour les années où les changements de champ empêchent d'étudier les variations annuelles.

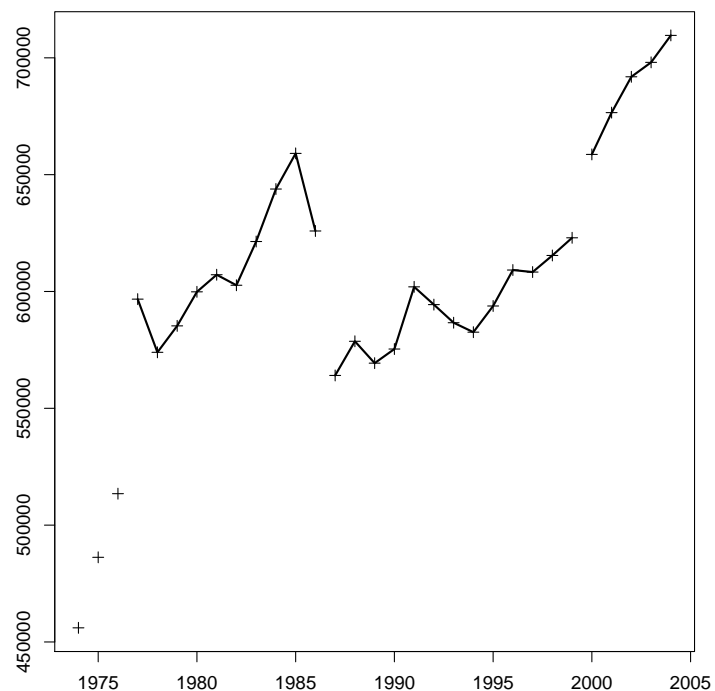


Figure 8 – Stock d'entreprises en Île de France entre 1974 et 2004

Table 2 – Statistiques descriptives sur l'emploi et la population des communes des groupes d'étude

		Effectif	Pop. 1968	Pop. 1999	Emp. 1968	Emp. 1999
Groupe 1	RER	150	18067	21994	5184	6994
	Hors RER	1011	1899	2840	620	875
Groupe 2	RER	150	18067	21994	5184	6994
	Hors RER	295	4807	6725	1578	2216
Groupe 3	RER	150	18067	21994	5184	6994
	Hors RER	449	3192	4634	986	1447
Groupe 4	RER	87	14189	17923	4041	5206
	Hors RER	86	7054	9589	2072	3350

Note de lecture : le groupe 4 contient moins de communes traitées parce que les communes pour lesquelles le schéma directeur de 1965 prévoit une gare RER et qui ont été effectivement équipées sont exclues.

pour lesquelles la prise de participation est antérieure à 1995 mais qui créent des établissements après cette date.

3.2 Les variables de localisation

Nous avons choisi de retenir la commune comme unité d'étude car la plupart des données dont nous disposons ne sont pas disponibles à une échelle plus précise. Pour chaque commune, nous disposons des données issues des recensements de 1968, 1975, 1982, 1990, 1999 et 2006. Elle apportent des informations sur la population résidente (composition sociale, âge), sur les logements (résidences principales et secondaires, logements vacants) et de la population active (compté sur le lieu de travail).

Le tableau 2 présente différentes statistiques descriptives portant sur les communes des groupes d'étude présentés dans la partie précédente. Tant pour l'emploi que pour la population, ces statistiques soulignent des écarts importants entre les communes reliées au RER et les autres. Ces différences sont particulièrement importantes au sein des groupes 1 et 3. Elles plaident pour la présence d'effets fixes communaux et de variables de contrôle dans le modèle. On notera également que le groupe de contrôle est toujours plus important que le groupe traité.

3.3 Quelle variable de traitement ?

Une part importante de notre travail a porté sur la prise en compte de l'amélioration des conditions de transport due au RER. La constitution de ces données pose deux difficultés importantes. Ainsi, il a fallu déterminer l'ensemble des communes reliées au RER, à chaque date, ainsi que l'importance de l'amélioration des transports liée au RER.

Quelles sont les communes reliées au réseau RER ?

Plusieurs méthodes peuvent être imaginées pour déterminer les communes reliées au RER. La plus simple consiste à prendre en compte uniquement les communes qui abritent une station RER en leur sein. Cette solution n'est pas satisfaisante car une commune peut être située à proximité immédiate d'une station de RER sans pour autant que la station soit située sur son territoire. Nous avons donc choisi d'affiner cette approche naïve en ne considérant pas uniquement la station

RER mais son voisinage. Concrètement, le voisinage d'une station RER est défini comme tout lieu situé à moins d'une certaine distance d'une station RER, par exemple 1 000 mètres. On s'intéresse ensuite à la surface de chaque commune située dans le voisinage d'une gare RER, c'est-à-dire à moins de 1 000 mètres d'une station.

Cette méthode est à l'origine de deux nouvelles difficultés. D'une part, il faut définir la distance qui est pertinente pour déterminer l'accessibilité à une station RER. Deux variantes, à 300 et 1 000 mètres, ont été testées. D'autre part, il faut déterminer si c'est la surface totale ou la part de la commune située à proximité d'une station qui est pertinente pour notre analyse. Ces deux versions ont également été estimées. Les résultats avec les différentes variables proposées sont présentés en annexe.

Comment déterminer l'amélioration de l'accessibilité ?

La présence d'une station RER ou d'une gare ne permet pas de résumer toute l'information concernant l'accessibilité d'une commune en transport ferré. Nous avons donc entamé la construction d'une variable d'accessibilité plus précise qui sera intégrée aux versions ultérieures de l'article. L'idée générale consiste à calculer le temps de parcours d'une commune vers l'ensemble des autres communes reliée au réseau ferré d'Île-de-France et ceci pour chaque année en 1969 et 2009. En considérant le réseau ferroviaire comme un graphe, il est possible de calculer le temps de parcours minimum entre chacun des points de ce graphe. On agrège ensuite ces données dans un indice d'accessibilité dont la forme exacte reste à définir. Plusieurs spécifications ont été testées pour l'instant sans que nous ayons pu déterminer la plus appropriée.

Le réseau de train de banlieue peut être représenté comme un graphe. En effet, chaque station sur chaque ligne de chemin de fer peut être vue comme un « sommet » de ce graphe et ces sommets sont reliés entre eux par des « arrêtes ». Il existe deux types d'arêtes. Le premier correspond à une voie de chemin de fer entre deux stations voisines d'une même ligne. La « valeur » d'une telle arrête est le temps de trajet en train entre ces deux stations. Le deuxième type d'arête correspond aux correspondances, plus précisément, au trajet à pied entre les quais de deux lignes différentes d'une même station.

Dans une même station, il existe autant de sommet que de lignes. Si toutes les lignes étaient représentées par un même sommet, cela signifierait que le temps de changement entre deux lignes est nul, ce qui n'est pas réaliste. La valeur de ce second type d'arête correspond donc au temps de marche d'une ligne à l'autre. Pour plus de précision, cette valeur varie en fonction du type de station et du type de lignes reliées. Premièrement, plus il y a de lignes connectées dans la station, plus le temps de connexion est élevé. Deuxièmement, le temps de connexion est plus important lors d'une connexion avec un train de banlieue ou entre un métro et un RER.

Nos données contiennent le Réseau express régional, les trains de banlieue, le tramway 2, le métro et Orlyval. Les services d'autobus n'ont pas été inclus, notre étude se concentrant en priorité sur le transport ferroviaire. Nous ajoutons aussi quelques subtilités afin de prendre en compte les trains sans arrêt, ainsi que les correspondances obligatoires pour relier certaines stations d'une même ligne quand la ligne contient plusieurs branches.

Une fois notre graphe défini, le calcul de la durée du trajet entre deux stations peut être considéré comme un « problème du plus court chemin ». L'algorithme de Dijkstra propose une solution à ce problème et fournit le chemin le plus bas coût entre un sommet et tous les autres

sommets. L'algorithme utilise la matrice d'adjacence, qui contient l'ensemble des arrêtes du graphe et ainsi que la valeur de chacune des arrêtes. Il permet de calculer la matrice de distance qui correspond au temps minimum entre tout couple de stations du réseau.

4 Les résultats

Le modèle exposé ci-dessus a été estimé grâce aux données présentées dans la partie précédente. Plusieurs variantes sont proposées, en fonction de l'éloignement à Paris, du secteur d'activité ou de la variable RER choisie. Les estimations ont été réalisées pour les entreprises étrangères ainsi que pour l'ensemble des entreprises. Les résultats présentés dans cette section correspondent uniquement au coefficient associé à la variable RER.

Concernant la variable de traitement, permettant d'estimer l'impact du RER, plusieurs spécifications ont été testées. Elles sont présentées en annexe. Nous avons retenu la variable qui explique le mieux le comportement des entreprises, tout en étant simple à interpréter. Il s'agit de la surface de la commune située à moins de 1 km d'une station RER. Cette surface est normalisée pour être égale à 1 si l'ensemble du pourtour de la gare est situé sur la commune¹⁰. Si une station est à cheval entre deux communes, la valeur de la variable *RER* pour chacune des deux communes correspondra à la part du voisinage de la station située sur chaque commune. Les deux communes se « partagent » la station RER. Les coefficients obtenus s'interprètent de la façon suivante : un coefficient de valeur x signifie que les implantations d'entreprise augmentent de x pourcents avec l'implantation d'une station RER dont le voisinage à 1 km est situé en totalité sur la commune. Les résultats sont présentés pour les quatre groupes d'études exposés précédemment.

En outre, il n'est pas exclu que les communes qui ne sont pas reliées au réseau soient négativement affectées par l'implantation du RER en Île-de-France. Par effet d'aspiration, elles peuvent voir les implantations d'entreprises diminuer sur leur sol. Étant donné que ces communes sont utilisées comme contrefactuel, le coefficient estimé correspond alors à la somme de l'effet pur du réseau express dans les communes reliées et de l'effet dépressif dans les communes non reliées. Pour déterminer si l'ouverture du RER entraîne plutôt une création nette d'entreprises ou un transfert d'entreprises déjà existantes, il faudrait pouvoir disposer de données d'IDE à l'extérieur à l'Île-de-France pour mettre en place le protocole exposé dans la première partie de ce rapport. En d'autres termes, il faudrait pouvoir déterminer le nombre total d'implantations d'entreprises étrangères en Île-de-France en l'absence de RER pour savoir lequel des deux effets domine. Nos données ne permettent clairement pas de répondre à cette question.

Enfin, il faut noter que la politique publique étudiée correspond principalement à l'amélioration d'une infrastructure déjà existante. Les lignes nouvelles sont presque toutes situées dans les communes exclues de notre analyse. Ainsi, l'impact estimé est nécessairement plus limité que celui qu'on aurait obtenu en évaluant une infrastructure nouvelle, comme le Grand Paris Express.

4.1 Quel impact du RER sur les implantations d'entreprises ?

Les résultats de l'évaluation empirique de l'effet du RER sont présentés pour les implantations des entreprises à capitaux étrangers ainsi que pour les implantations de l'ensemble des entreprises. Ces résultats ont été calculés pour toute l'Île-de-France puis détaillés pour la proche et la grande

10. En pratique, on divise la surface communale à moins de 1 km d'une station par la surface d'un disque de 1 km de rayon. Cette manipulation permet une interprétation plus aisée de la variable d'intérêt.

Table 3 – Impact du RER sur les implantations d'entreprises

	Groupe 1	Groupe 2	Groupe 3	Groupe 4
<i>Ensemble des entreprises</i>				
Ensemble Île-de-France	0,093*** (0,016)	0,043*** (0,013)	0,061*** (0,014)	0,025 (0,018)
Proche banlieue	0,011 (0,013)	0,001 (0,012)	0,011 (0,013)	0,011 (0,019)
Grande banlieue	0,11*** (0,019)	0,066*** (0,016)	0,082*** (0,018)	0,039* (0,023)
<i>Entreprises à capitaux étrangers</i>				
Ensemble Île-de-France	0,053*** (0,013)	0,041** (0,019)	0,050*** (0,016)	0,085*** (0,030)
Proche banlieue	0,068* (0,042)	0,049 (0,045)	0,068* (0,042)	0,112 (0,080)
Grande banlieue	0,042*** (0,013)	0,025 (0,021)	0,035** (0,018)	0,021 (0,032)

*Note : les valeurs des coefficients estimés sont présentées selon le groupe de commune, la zone géographique considérée et la catégorie d'entreprise. L'écart-type est indiqué entre parenthèses. Les étoiles correspondent à la significativité des estimations : *** à 1 %, ** à 5 % et * à 10 %. L'absence d'étoile signifie que le coefficient n'est pas significatif.*

Aide à la lecture : un coefficient de valeur \times signifie que l'installation d'une gare RER entraîne une augmentation de \times points de pourcentage des implantations d'entreprises.

banlieue (voir tableau 3). Par proche banlieue, on entend ici les trois départements de la petite couronne (Hauts-de-Seine, Seine-Saint-Denis et Val-de-Marne) et par grande banlieue, les départements de la grande couronne (Seine-et-Marne, Yvelines, Essonne et Val-d'Oise).

Concernant l'ensemble des entreprises, l'estimation de l'impact du RER varie de façon importante selon le groupe de contrôle. Pour le groupe 4, on trouve même un impact non significatif. Pour les autres groupes, la construction d'une gare RER entraîne une augmentation comprise entre 4 et 9 % des implantations d'établissements. Cet effet apparaît négligeable en proche banlieue et plus important en grande banlieue où il est compris entre 4 et 11 %. L'impact estimé est similaire pour les entreprises étrangères. L'effet pour l'ensemble de l'Île-de-France est également compris entre 4 et 8 %, selon les groupes. Par contre, le modèle ne permet pas de tirer de conclusions concernant la répartition géographique de ces nouvelles entreprises étrangères.

4.2 L'évolution temporelle de l'effet du RER

Afin d'étudier l'évolution temporelle de l'effet du RER, nous avons détaillé les résultats présentés dans la section précédente par périodes de trois ans. Les résultats présentés dans le tableau 4 nous permettent ainsi de déterminer si l'impact positif du RER sur l'implantation des établissements dure dans le temps mais également si la condition de tendance commune, cruciale pour la validité de notre modèle, est vérifiée ex ante.

L'impact du RER est tardif et stable dans le long terme. En effet, on ne détecte pas d'effet significatif avant huit ans pour l'ensemble des entreprises. Concernant les entreprises à capitaux étrangers, les effets sont significatifs moins tardivement, c'est-à-dire à partir de 5 ans et se stabi-

Table 4 – Évolution dans le temps de l'impact du RER sur les implantations d'entreprises

	Groupe 1	Groupe 2	Groupe 3	Groupe 4
<i>Ensemble des entreprises</i>				
Plus de 10 ans avant RER	−0,081*** (0,026)	−0,007 (0,02)	−0,024 (0,023)	0,021 (0,025)
8 à 10 ans avant RER	−0,001 (0,033)	−0,006 (0,025)	0,001 (0,029)	−0,02 (0,029)
5 à 7 ans avant RER	−0,085*** (0,029)	−0,022 (0,023)	−0,047* (0,026)	−0,011 (0,025)
2 à 4 ans avant RER	<i>réf.</i>	<i>réf.</i>	<i>réf.</i>	<i>réf.</i>
1 an avant à 1 ans après RER	0,029 (0,026)	0,006 (0,02)	0,014 (0,023)	−0,03 (0,023)
2 à 4 ans après RER	0,01 (0,026)	0,021 (0,02)	0,012 (0,023)	−0,004 (0,024)
5 à 7 ans après RER	0,031 (0,026)	0,007 (0,02)	0,014 (0,023)	0,001 (0,024)
8 à 10 ans après RER	0,121*** (0,027)	0,056*** (0,021)	0,078*** (0,024)	0,006 (0,027)
Plus de 10 ans après RER	0,143*** (0,023)	0,072*** (0,019)	0,093*** (0,021)	0,012 (0,028)
<i>Entreprises à capitaux étrangers</i>				
Plus de 10 ans avant RER	−0,041** (0,018)	−0,014 (0,027)	−0,042* (0,023)	−0,017 (0,041)
8 à 10 ans avant RER	−0,003 (0,021)	0,012 (0,031)	−0,004 (0,027)	−0,006 (0,041)
5 à 7 ans avant RER	0,034* (0,019)	0,058** (0,027)	0,038 (0,024)	0,053 (0,036)
2 à 4 ans avant RER	<i>réf.</i>	<i>réf.</i>	<i>réf.</i>	<i>réf.</i>
1 an avant à 1 ans après RER	−0,001 (0,017)	0,014 (0,025)	0,004 (0,022)	0,022 (0,034)
2 à 4 ans après RER	0,046*** (0,018)	0,024 (0,026)	0,036 (0,023)	0,051 (0,038)
5 à 7 ans après RER	0,082*** (0,018)	0,081*** (0,026)	0,088*** (0,023)	0,175*** (0,039)
8 à 10 ans après RER	0,069*** (0,02)	0,065** (0,03)	0,072*** (0,026)	0,137*** (0,047)
Plus de 10 ans après RER	0,081*** (0,018)	0,074*** (0,028)	0,09*** (0,024)	0,091* (0,048)

*Note : les valeurs des coefficients estimés sont présentées selon la catégorie d'entreprise. L'écart-type est indiqué entre parenthèses. Les étoiles correspondent à la significativité des estimations : *** à 1 %, ** à 5 % et * à 10 %. L'absence d'étoile signifie que le coefficient n'est pas significatif.*

Aide à la lecture : un coefficient de valeur x signifie que l'installation d'une gare RER entraîne une augmentation de x points de pourcentage des implantations d'entreprises.

Table 5 – Impact du RER sur les implantations d'entreprises selon le secteur d'activité

	Groupe 1	Groupe 2	Groupe 3	Groupe 4
Industrie	0,008 (0,015)	0,016 (0,017)	0,018 (0,016)	0,044* (0,025)
Construction	0,091*** (0,016)	0,067*** (0,017)	0,079*** (0,017)	0,007 (0,024)
Commerce	0,034** (0,015)	0,023 (0,016)	0,024 (0,016)	0,03 (0,023)
Transports	0,143*** (0,013)	0,12*** (0,017)	0,126*** (0,016)	0,109*** (0,027)
Activités financières	-0,005 (0,011)	-0,013 (0,015)	-0,006 (0,014)	-0,019 (0,023)
Activités immobilières	-0,084*** (0,014)	-0,047*** (0,018)	-0,059*** (0,016)	0,02 (0,027)
Services aux entreprises	0,299*** (0,016)	0,137*** (0,017)	0,197*** (0,017)	0,057** (0,025)
Services aux particuliers	0,134*** (0,015)	0,07*** (0,016)	0,098*** (0,016)	0,058** (0,025)
Éducation, santé, action sociale	0,008 (0,015)	0,016 (0,017)	0,018 (0,016)	0,044* (0,025)

*Note : les valeurs des coefficients estimés sont présentées selon le secteur d'activité de l'entreprise. L'écart-type est indiqué entre parenthèses. Les étoiles correspondent à la significativité des estimations : *** à 1 %, ** à 5 % et * à 10 %. L'absence d'étoile signifie que le coefficient n'est pas significatif.*

Aide à la lecture : un coefficient de valeur \times signifie que l'installation d'une gare RER entraîne une augmentation de \times points de pourcentage des implantations d'entreprises.

lisent ensuite. Concernant la validité de l'hypothèse de tendance commune, le groupe 3 et surtout le groupe 1 posent problème. Pour que le modèle en différence-de-différences soit valide et ne souffre pas d'un biais d'endogénéité, il est nécessaire que l'évolution des implantations d'entreprises dans le groupe de contrôle soit identique à celle du groupe traité en l'absence de traitement. Si cette hypothèse n'est pas vérifiable après la mise en place du traitement, elle l'est avant. Ainsi, pour les groupes 1 et 3, on remarque que les communes qui n'ont pas été reliées au RER attiraient moins les entreprises avant même la construction des infrastructures. Ils faut donc considérer les résultats estimés sur ces groupes avec prudence. A l'inverse, la validité des groupes 2 et 4 est confirmée.

4.3 L'hétérogénéité des effets du RER sur les secteurs d'activité

L'effet de l'implantation d'une station RER dans une commune n'est pas le même selon le secteur considéré (voir le tableau 5). Les implantations d'établissements à capitaux étrangers étant des événements assez rares à l'échelle d'une commune, les résultats détaillés par secteur ont pu être calculés uniquement pour les implantations de l'ensemble des entreprises.

Les résultats confirment les prévisions théoriques de la première partie puisque les secteurs des services aux entreprises et aux particuliers répondent le plus à la mise en place du RER. L'augmentation de la productivité des entreprises par l'amélioration des conditions de transport est particulièrement sensible dans ces secteurs. En effet, à la différence du secteur industriel, les transports de passagers peuvent jouer un rôle important dans le processus de production de ces entreprises en facilitant l'accès aux clients. Le secteur de la construction est également encouragé

par l'évolution des conditions de transport liée au RER. Ces résultats pourraient suggérer que la mise en place d'une gare RER augmente la construction de nouveaux bâtiments. Cependant, il n'est pas certain que les entreprises du bâtiment implantées à proximité des gares RER exercent leur activité au même endroit. Ces résultats ne permettent donc pas de conclure directement que la construction d'une gare RER stimule la construction. Il faudrait disposer de données complémentaires portant sur les mises en chantier pour être en mesure de tirer de telles conclusions. Les implantations d'entreprises de transports augmentent également significativement. A l'inverse, le secteur immobilier est significativement défavorisé par la mise en place du RER. Nous nous assurerons dans une version ultérieure de l'article que ces résultats peuvent bien être attribués à l'effet du RER et pas aux changements dans l'enregistrement de ce secteur dans le répertoire Sirene au cours de la période étudiée. Aucun autre secteur ne présente des coefficients significatifs dans un nombre suffisant de groupes pour que l'on puisse conclure à un effet significatif du RER.

5 Conclusion

L'objectif de ce rapport est de fournir des arguments solides pour chiffrer l'impact espéré de la construction du réseau de transport du Grand Paris en termes d'attractivité. La principale question porte sans doute sur la question de la compétition entre grandes métropoles européennes, pour attirer les entreprises multinationales, en particulier leurs sièges sociaux, ou leurs centres de recherche / design / marketing... Dans cette perspective, il faudrait pouvoir estimer si la construction d'une infrastructure de transport collectif "de type métro" améliore la probabilité qu'à l'Ile-de-France d'attirer les investissements par rapport à la région du Grand Londres, de Berlin ou de Barcelone par exemple. Mener cette analyse directement est malheureusement impossible dans l'état actuel des données disponibles. En effet, cela impliquerait de disposer des choix de localisation d'un grand nombre d'entreprises sur tout ou partie du territoire européen, et sur une période de temps assez longue pour mesurer les effets d'investissements dans les infrastructures de transport qui ont eu lieu à différentes périodes dans différents pays. Il n'existe à notre connaissance aucune base de données disposant de l'ensemble de ces caractéristiques.¹¹

Dès lors, nous avons choisi une stratégie différente, reposant sur l'impact en termes d'attractivité du dernier investissement important en infrastructure de transport comparable, à savoir l'expérience du RER au cours des années 1970, 1980 et 1990. Il existe deux sources de problèmes potentiels au fait de prendre cette expérience du RER comme illustrative de ce qui devrait se passer avec la construction du Grand Paris Express. La première est liée à la *validité externe* de notre approche. En effet, l'échelle géographique est différente puisque l'on s'intéresse à un effet intra-francilien dans le cas du RER. Il nous semble logique de penser que la mobilité géographique des firmes est plus faible entre grandes métropoles européennes qu'entre différentes communes à l'intérieur de l'Ile-de-France. Dès lors, les impacts que nous obtenons sont très certainement une fourchette haute de l'impact que l'on peut attendre en termes de surcroît d'attractivité internationale. De plus la période étudiée est différente puisqu'elle porte sur des décisions effectuées principalement dans les années 1980 et 1990. Le biais est ici certainement inverse car il est fort probable que les entreprises soient plus mobiles aujourd'hui et dans les années à venir qu'il y a une trentaine d'année. Au total, il est difficile de régler difficilement ce problème de validité externe. Une chose est certaine, si le RER n'avait eu aucun impact sur la localisation des investissements étrangers en Ile-de-France, il aurait été très improbable que la construction du Grand Paris Express

11. Il semble possible de construire une telle base à partir de données nationales, mais cela suppose de coordonner le travail de différentes équipes de recherche ayant accès à chacune des sources nationales généralement confidentielles, ce qui dépasse largement le cadre de ce rapport.

réussisse à faire préférer Paris à Londres ou Berlin à ces mêmes investissements. Même si nos estimations doivent être extrapolées avec beaucoup de précisions, il semble que c'est aujourd'hui ce qui peut se faire de plus rigoureux pour répondre à la question posée.

Le deuxième problème envisageable porte sur la *validité interne* de notre approche. Il s'agit là de s'assurer que l'impact du RER que nous mesurons soit bien un impact causal, et non pas une simple corrélation statistique entre la construction d'infrastructures de transport et l'attractivité d'un territoire. Sur ce point, il est possible d'être assez confiant. Nous avons identifié plusieurs groupes de contrôle constitués de communes non reliées au réseau du RER, comparées à celles qui ont vu le RER y arriver, avant et après sa construction. Cette méthode cherche à mener une comparaison entre communes qui auraient selon toute vraisemblance connu la même situation si le RER n'avait pas été construit. Pour cela nous utilisons principalement le fait que le réseau finalement construit n'a pas exactement suivi le plan initial de 1965. En raison de la crise des années 1970, de nombreux tracés qui devaient être construits ex-nihilo ont été abandonnés, et à l'inverse le tracé effectif a utilisé des parties de voies existantes qui n'auraient pas dû bénéficier de cette amélioration de leur desserte. La comparaison entre les communes "traitées" et celles qui auraient pas dû l'être mais ne l'ont pas été ressemble donc bien à une expérience naturelle permettant d'identifier un effet causal du RER. Trois autres groupes de contrôles sont aussi testés pour établir une certaine robustesse des résultats.

Nos résultats montrent un impact non-négligeable du RER sur l'implantation des entreprises. **Pour l'ensemble des entreprises, la construction d'une gare RER entraîne un surcroît d'attractivité de 4 à 9%, ces chiffres sont de 5 à 10% pour les entreprises à capitaux étrangers.** L'impact est plus fort en proche qu'en grande banlieue pour ces dernières. On peut également noter que l'effet n'est pas immédiat. En effet, il apparaît comme étant statistiquement significatif environ 5 ans après la construction. Il est d'ailleurs important de noter qu'il n'y a pas de différence significative *avant* la construction de la gare entre les communes traitées et notre groupe de contrôle préféré. Enfin, il existe des différences importantes entre secteurs d'activité. Conformément aux attentes, ce sont les entreprises de services (aux entreprises et aux particuliers) qui montrent les impacts les plus importants. En effet, c'est pour ce groupe d'entreprises que l'accès aux clients et donc la productivité des employés est sans doute la plus fortement augmentée par l'arrivée du RER. Pour revenir à l'applicabilité au Grand Paris de nos résultats, ce résultat est certainement assez encourageant. En effet, notre revue de littérature a montré que la localisation des sièges sociaux était très sensible à la présence locale de firmes de services aux entreprises. Ces dernières étant les plus sensibles au réseau de transport, il peut donc s'enclencher un effet d'agglomération cumulatif entre ces deux groupes d'entreprises, à la faveur de la création du réseau.

Par ailleurs il peut exister un certain nombre de raisons qui peuvent expliquer que l'effet que nous mesurons à partir de la construction du RER est plus faible que ce qui sera observé avec le réseau du Grand Paris. Par exemple, nos estimations portent sur la construction du RER, qui a constitué essentiellement en une amélioration du réseau de voies ferrées existantes, ainsi que différentes interconnexions de parties du réseau au sein de la ville de Paris. Il est très probable que la construction ex-nihilo d'un réseau nouveau ait un impact plus fort que l'amélioration d'une infrastructure dans une large mesure déjà existante. Il est très difficile de chiffrer la différence d'effet entre la construction et amélioration, car aucune étude à notre connaissance ne distingue entre les deux cas.

Références

- Marianne Bertrand, Esther Duflo et Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.
- Maggie Xiaoyang Chen et Michael O. Moore. Location decision of heterogeneous multinational firms. *Journal of International Economics*, 80(2):188–199, 2010.
- Leonard K. Cheng et Yum K. Kwan. What are the determinants of the location of foreign direct investment? the chinese experience. *Journal of International Economics*, 51(2):379–400, 2000.
- Bernard Collardey. *Les trains de banlieue, tome II de 1938 à 1999*. Editions La Vie du Rail, 1999.
- Cletus C. Coughlin et Eran Segev. Location determinants of new foreign-owned manufacturing plants. *Journal of Regional Science*, 40(2):323–351, 2000.
- Matthieu Crozet, Thierry Mayer et Jean-Louis Mucchielli. How do firms agglomerate? a study of fdi in france. *Regional Science and Urban Economics*, 34(1):27–54, 2004.
- James C. Davis et J. Vernon Henderson. The agglomeration of headquarters. *Regional Science and Urban Economics*, 38(5):445–460, 2008.
- Fabrice Defever. Functional fragmentation and the location of multinational firms in the enlarged europe. *Regional Science and Urban Economics*, 36(5):658–677, 2006.
- Dave Donaldson. Railroads of the raj : Estimating the impact of transportation infrastructure. Nber working papers, National Bureau of Economic Research, Inc, 2010.
- Gilles Duranton et Matthew A. Turner. Urban growth and transportation. *Review of Economic Studies*, à paraître, 2012.
- Pauline Givord. Méthodes économétriques pour l'évaluation des politiques publiques. Documents de travail de la dese, Insee, 2010.
- Keith Head et Thierry Mayer. Market potential and the location of japanese investment in the european union. *The Review of Economics and Statistics*, 86(4):959–972, 2004.
- Adelheid Holl. Manufacturing location and impacts of road transport infrastructure : empirical evidence from spain. *Regional Science and Urban Economics*, 34(3):341–363, 2004a.
- Adelheid Holl. Transport infrastructure, agglomeration economies, and firm birth : Empirical evidence from portugal. *Journal of Regional Science*, 44(4):693–712, 2004b.
- Guido M. Imbens et Jeffrey M. Wooldridge. Recent developments in the econometrics of program evaluation. Nber working papers, National Bureau of Economic Research, Inc, 2008.
- Sung Jin Kang et Hong Shik Lee. The determinants of location choice of south korean fdi in china. *Japan and the World Economy*, 19(4):441–460, 2007.
- Xuepeng Liu, Mary E Lovely et Jan Ondrich. The location decisions of foreign investors in china : Untangling the effect of wages using a control function approach. *The Review of Economics and Statistics*, 92(1):160–166, 2010.
- Philippe Martin et Carol Ann Rogers. Industrial location and public infrastructure. *Journal of International Economics*, 39(3-4):335–351, 1995.

- Thierry Mayer, Isabelle Mejean et Benjamin Nefussi. The location of domestic and foreign production affiliates by french multinational firms. *Journal of Urban Economics*, 68(2):115–128, 2010.
- Guy Michaels. The effect of trade on the demand for skill : Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701, 2008.
- Loriane Py et Fabrice Hatem. Internationalisation et localisation des services : une analyse sectorielle et fonctionnelle appliquée aux firmes multinationales en europe. *Économie et Statistique*, 426(1):67–95, 2009.
- Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- Kurt Schmidheiny et Marius Brülhart. On the equivalence of location choice models : Conditional logit, nested logit and poisson. *Journal of Urban Economics*, 69(2):214–222, 2011.
- Vanessa Strauss-Kahn et Xavier Vives. Why and where do headquarters move ? *Regional Science and Urban Economics*, 39(2):168–186, 2009.
- Stephen Ross Yeaple. Firm heterogeneity and the structure of u.s. multinational activity. *Journal of International Economics*, 78(2):206–215, 2009.

A Annexes : quelle variable de traitement choisir pour déterminer l'effet du RER ?

Table 6 – Impact du RER sur les implantations d'entreprise selon la variable d'intérêt utilisée

	Groupe 1	Groupe 2	Groupe 3	Groupe 4
<i>Ensemble des entreprises</i>				
Indicatrice RER à 1 km	0,091*** (0,015)	0,027** (0,012)	0,045*** (0,014)	-0,003 (0,015)
Indicatrice RER à 300 m	0,088*** (0,017)	0,03** (0,013)	0,049*** (0,015)	0,01 (0,017)
Surface de la com. à moins 1 km d'une RER	0,093*** (0,016)	0,043*** (0,013)	0,061*** (0,014)	0,025 (0,018)
Surface de la com. à moins 300 m d'une RER	0,076*** (0,014)	0,036*** (0,011)	0,051*** (0,013)	0,023 (0,015)
Part de la com. à moins de 1 km d'une RER	0,159*** (0,032)	0,058** (0,025)	0,097*** (0,029)	0,067* (0,04)
Part de la com. à moins de 300 m d'une RER	1,199*** (0,294)	0,4* (0,231)	0,715*** (0,266)	0,328 (0,329)
<i>Entreprises à capitaux étrangers</i>				
Indicatrice RER à 1 km	0,048*** (0,011)	0,029 (0,018)	0,046*** (0,015)	0,065** (0,026)
Indicatrice RER à 300 m	0,023* (0,014)	0 (0,021)	0,019 (0,018)	0,027 (0,031)
Surface de la com. à moins 1 km d'une RER	0,07*** (0,012)	0,054*** (0,018)	0,068*** (0,016)	0,101*** (0,029)
Surface de la com. à moins 300 m d'une RER	0,067*** (0,011)	0,054*** (0,016)	0,066*** (0,014)	0,092*** (0,026)
Part de la com. à moins de 1 km d'une RER	0,065*** (0,023)	0,03 (0,035)	0,057* (0,03)	0,199*** (0,065)
Part de la com. à moins de 300 m d'une RER	0,654*** (0,208)	0,389 (0,309)	0,614** (0,27)	1,179** (0,513)

*Note : les valeur des coefficients estimés sont présentées selon le groupe de commune et la variable de traitement. L'écart-type est indiqué entre parenthèses. Les étoiles correspondent à la significativité des estimations : *** à 1 %, ** à 5 % et * à 10 %. L'absence d'étoile signifie que le coefficient n'est pas significatif.*

Aide à la lecture : un coefficient de valeur x signifie que l'installation d'une gare RER entraîne une augmentation de x points de pourcentage des implantations d'entreprises.