HAL
open science

Impact du Grand Paris Express sur la localisation des investissements directs étrangers

Pierre-Henri Bono, Quentin David, Rodolphe Desbordes, Loriane Py

To cite this version:

Pierre-Henri Bono, Quentin David, Rodolphe Desbordes, Loriane Py. Impact du Grand Paris Express sur la localisation des investissements directs étrangers. [Rapport de recherche] Laboratory for Interdisciplinary Evaluation of Public Policies. 2017, pp.92. hal-03455330

HAL Id: hal-03455330

https://sciencespo.hal.science/hal-03455330

Submitted on 29 Nov 2021

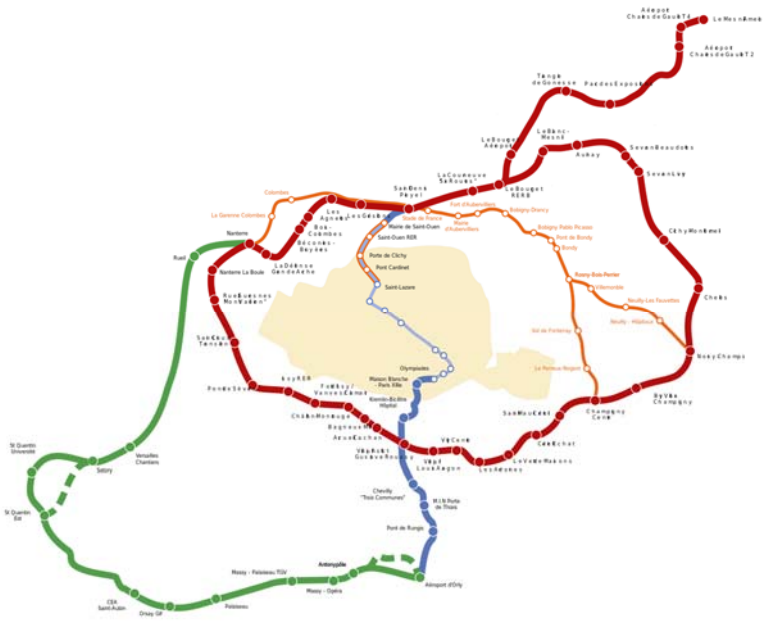
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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

SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

2017

Impact du Grand Paris Express sur la localisation des investissements directs étrangers



Ce rapport est une version d'étape d'un travail de recherche toujours en cours. Les conclusions générales du rapport ne sont pas remises en cause. Cependant, de nouvelles méthodes ont été mises en œuvre permettant une plus grande robustesse des conclusions obtenues ici.

Les opinions émises dans ce rapport sont propres aux auteurs et n'engagent pas nécessairement la position du LIEPP, de Sciences Po, de SKEMA Business School, de l'université de Lille 1 ou de la Banque de France.

Le laboratoire Interdisciplinaire d'Évaluation des Politiques Publiques (LIEPP) est un centre d'excellence soutenu par l'Agence Nationale de la Recherche dans le cadre des « Investissement d'Avenir » (ANR-11-LABX-0091, ANR-11-IDEX-0005-02).

Pierre-Henri Bono
Quentin David
Rodolphe Desbordes
Loriane Py

SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

Impact du Grand Paris Express sur la localisation des investissements directs étrangers

Ont participé à ce rapport :

Pierre-Henri Bono[i], *Quentin David*[ii]

Rodolphe Desbordes[iii], *Loriane Py*[iv]

Assistants de recherche ayant travaillé sur le projet :

Martin Tusl, Julia Fahy, Léna Mardi, Guillaume Roger, Myriam Laadhari

Le Laboratoire Interdisciplinaire d'Évaluation des Politiques Publiques (LIEPP) est un centre d'excellence soutenu par l'Agence Nationale de la Recherche dans le cadre des « Investissements d'Avenir » (ANR-11-LABX-0091, ANR-11-IDEX-0005-02).

i : Sciences Po – LIEPP.
ii : Université de Lille 1.
iii : SKEMA Business School.
iv : Banque de France.

Ce rapport est une version d'étape d'un travail de recherche toujours en cours. Les conclusions générales du rapport ne sont pas remises en cause. Cependant, de nouvelles méthodes ont été mises en œuvre permettant une plus grande robustesse des conclusions obtenues içi.
Les opinions émises dans ce rapport sont propres aux auteurs et n'engagent pas nécessairement la position du LIEPP, de Sciences Po, de SKEMA Business School, de l'université de Lille 1 ou de la Banque de France.

Une version de ce travail a été présenté à l'Université de Lausanne (Workshop Housing and Transport in the City en décembre 2016) et à la conférence des 5 ans du LIEPP à Sciences Po en janvier 2017. Nous remercions tout particulièrement Étienne Wasmer, ainsi que Dominique Bureau, Gilles Duranton, Sylviane Gastaldo, Thiery Mayer, Jean-Claude Prager, Diego Puga, Emile Quinet, Juliette Seban et Jacques Thisse pour leurs relectures, commentaires et suggestions très utiles.

Résumé de l'étude

La Société du Grand Paris (SGP) responsable du développement du Grand Paris Express a chargé le Laboratoire Interdisciplinaire d'Évaluation des Politiques Publiques (LIEPP) de Sciences Po de réaliser une étude de l'impact des infrastructures de transport de type métro sur l'attractivité internationale des villes.

Pour répondre à la question posée par la SGP, le LIEPP a construit une base de données originale d'emprise mondiale à l'échelle des aires urbaines avec les spécificités suivantes. Premièrement, pour mesurer l'attractivité internationale des villes, nous avons géolocalisé à la ville les données des projets d'investissements directs étrangers (*FDI*) de la base *fDiMarkets*, produite par le Financial Times. Deuxièmement, une base de données rassemblant des informations annuelles de 1950 à 2015 sur la taille des infrastructures de type métro dans le monde à l'échelle de la ville a été minutieusement constituée afin de construire nos variables d'intérêt. Troisièmement des informations sur les autres types d'infrastructures de transport disponibles au niveau des villes ont également été collectées. Enfin quatrièmement, ont été rassemblées des informations sur un grand nombre de caractéristiques économiques et géographiques des villes susceptibles aussi d'être facteur d'attractivité pour les multinationales. Ce travail de collecte et d'appariement d'un aussi grand nombre de données au niveau des villes à couverture mondiale est à notre connaissance inédit.

Ce travail nous a permis de mener une analyse économétrique de l'impact des métros sur l'implantation du nombre de projets *FDI* sur la période 2003-2014. Un modèle de Poisson est utilisé pour analyser à la fois l'impact de la présence du métro (marge extensive) et de la taille du réseau (marge intensive) sur l'attractivité d'une ville pour les multinationales.

Nous montrons qu'il existe un lien significatif entre l'attractivité d'une ville et la présence d'un système de transport de type métro. Pour pouvoir interpréter ce lien de manière causale, nous devons traiter l'éventuel biais d'endogénéité lié à l'existence potentielle de variables omises dans le modèle. Pour y parvenir, nous avons mobilisé une méthode innovante mise en place par Lewbel (2012).

Notre analyse nous permet de conclure à un lien positif causal de cette relation entre métro et *FDI* : les infrastructures de transport de type métro attirent l'investissement étranger tant au niveau de la marge extensive que de la marge intensive. Toutes les estimations convergent vers ce résultat. Nos estimations par variables instrumentales pour contrôler l'endogénéité confirment cet effet positif et significatif du métro. Enfin, ces résultats sont robustes à un grand nombre de spécifications différentes et ne sont pas tributaires de nos choix de mesure pour les différents paramètres d'intérêt.

Le Grand Paris Express (GPE) prévoit un doublement de la taille du réseau de métro dans l'aire urbaine parisienne, soit une augmentation de 100 %. En appliquant cette valeur à nos résultats, qui suggèrent une élasticité comprise entre 0,1 et 0,2 pour la marge intensive, nous obtenons un impact attendu pour le GPE de l'ordre de 10 % à 20 % de projets *FDI* supplémentaires. Paris ayant attiré 1 310 projets entre 2003 et 2014, de telles élasticités correspondent à un intervalle de 130 à 262 projets supplémentaires sur une période de 12 ans. De plus, nous constatons dans les données *FdiMarkets* que les projets qui se sont implantés à Paris ont généré, en moyenne sur la période d'étude, 52 emplois par projet. Le Grand Paris Express générerait donc, selon ces valeurs et nos estimations, entre 6 800 et 14 000 emplois directs issus de l'implantation des firmes multinationales.

Table des matières

Résumé de l'étude iii
1 Introduction ..... - 1 -
2 Revue de la littérature ..... - 6 -
2.1 Les déterminants de la localisation des entreprises multinationales : quelle pertinence au niveau des villes ? ..... - 6 -
2.2 Infrastructures de transports et attractivité des villes ..... - 9 -
3 La base de données ..... - 11 -
3.1 La base des aires urbaines (ou villes) ..... - 12 -
3.2 Les projets d'investissements directs étrangers ..... - 14 -
3.3 La base des métros ..... - 17 -
3.4 Les variables de contrôle ..... - 29 -
3.4.1 Population et surface ..... - 29 -
3.4.2 Activité économique ..... - 30 -
3.4.3 Les données G-Econ ..... - 30 -
3.4.4 Les autres infrastructures de transport ..... - 32 -
3.5 Exemples d'intégration ..... - 33 -
4 Modèle économétrique ..... - 36 -
4.1 Modèle de Poisson ..... - 37 -
4.2 Les variables de contrôle au niveau des villes ..... - 38 -
4.3 Variables instrumentales ..... - 43 -
5 Résultats des estimations ..... - 46 -
5.1 Modèle de Poisson ..... - 47 -
5.1.1 Marge extensive ..... - 47 -
5.1.2 Marge intensive ..... - 51 -
5.2 Estimations par variables instrumentales ..... - 59 -
5.3 Résultats complémentaires ..... - 63 -
5.4 Tests de robustesse ..... - 66 -
6 Conclusion ..... - 67 -

Références ..... - 69 -
Annexes ..... - 74 -
A. Statistiques descriptives des systèmes de métro base réduite ..... - 74 -
B. Détail du modèle économétrique ..... - 74 -
C. Test sur les instruments ..... - 76 -
D. Les lumières la nuit comme proxy de l'activité ..... - 76 -
E. Tests de robustesse ..... - 79 -
F. Résultats supplémentaires ..... - 84 -

1 Introduction

Au début du XXIe siècle, les pouvoirs publics français ont entrepris d'étendre de manière importante les réseaux de transport qui desservent la métropole parisienne. Ce choc d'infrastructure se matérialisera par la construction de plus de 200 km de lignes de métro, qui visera soit à prolonger des lignes existantes soit à en créer de nouvelles. En particulier, le projet prévoit la mise en place d'un tracé circulaire autour de Paris, alors que jusqu'à présent le réseau était essentiellement composé de lignes radiales. Ce projet d'extension a pour nom le Grand Paris Express (GPE) et la société en charge de sa mise en œuvre est la Société du Grand Paris (SGP).

Le coût total attendu du GPE s'élève à environ 25,5 milliards d'euros. Le projet est financé dans sa grande majorité par des recettes fiscales, et au vu des montants évoqués, les décideurs attendent de cet investissement qu'il ait un effet transformationnel sur les grands équilibres socioéconomiques et le développement urbain de la métropole. Cet impact multidimensionnel du nouveau métro doit donner naissance à de nouvelles zones de développement économique, mais aussi influencer l'extension de la métropole en reconfigurant la géographie urbaine de la ville, permettant ainsi une densification des zones qu'il impactera. Sur le plan environnemental, le GPE a aussi pour objectif d'améliorer la performance écologique globale de la métropole en diminuant les émissions de gaz à effet de serre (en particulier en réduisant la congestion du trafic routier). Le GPE est un projet ambitieux qui vise à modeler le futur de l'aire urbaine de la capitale française.

D'un point de vue économique, le GPE devrait favoriser les économies d'agglomération en accentuant la concentration des agents économiques sur le territoire. En effet, la proximité des acteurs économiques stimule généralement leur productivité, à condition que les phénomènes de congestion n'excèdent pas les externalités positives. Il est également attendu que le GPE améliore le fonctionnement du marché du travail[1]. Ces effets devraient permettre à la fois à la capitale de rester le moteur de l'économie française, mais aussi de donner à la région parisienne des atouts dans la compétition internationale face aux autres grandes métropoles mondiales telles que Londres, New York, Tokyo, Shanghai ou encore Singapour. Cette liste n'est pas

[1] Voir pour cela le travail de Chapelle, Wasmer, Bono (2017).

exhaustive tant la croissance des grandes métropoles à travers le monde est rapide. En définitive, avec le GPE devrait contribuer à accroître l'attractivité locale et internationale de la métropole parisienne.

Dans un monde où l'aversion du contribuable aux taxes est croissante, la production d'infrastructures publiques est devenue la source d'intenses débats. Pour éclairer les décideurs publics, la mise en œuvre de la plupart de ces projets est maintenant régulièrement soumise à une analyse coût-bénéfice. Pour que cette analyse soit pertinente, les décideurs publics doivent bénéficier d'une d'évaluation fiable et suffisamment robuste de l'impact de ce projet d'infrastructure sur l'activité économique.

De toutes les retombées possibles et attendues qu'un projet comme le GPE peut engendrer, nous nous intéressons dans cette étude à l'impact qu'il peut avoir sur l'attractivité internationale et ses répercussions sur le dynamisme du territoire de la métropole parisienne. Notre mesure de l'attractivité internationale d'une ville repose sur le nombre de projets d'investissements directs étrangers (notés *FDI*, pour *Foreign Direct Investments*, dans la suite de ce rapport) qui s'implantent sur son territoire au cours d'une période donnée. S'intéresser aux firmes internationales n'est pas anodin, car au-delà de l'activité économique qu'elles créent directement, les multinationales tendent aussi à avoir un effet catalyseur sur la croissance du tissu industriel local et sur la création d'emplois. Elles sont donc souvent perçues comme fortement contributrices au dynamisme des villes. En outre, il est probable que l'attractivité internationale reflète aussi l'attractivité de la ville pour l'ensemble des investissements locaux et nationaux qui y sont réalisés.

Bien qu'il existe une littérature abondante sur les déterminants de la localisation des projets *FDI*, ces déterminants restent aujourd'hui encore peu connus à l'échelle qui nous intéresse, à savoir, celle de la ville. En effet, la littérature empirique s'est jusqu'à présent concentrée en très grande majorité sur l'analyse des décisions de localisation des entreprises au niveau des pays ou des régions, principalement en raison du manque de données socioéconomiques à l'échelle des villes, aussi du manque de données sur la localisation précise des FDI. Cependant, en nous appuyant sur cette littérature, nous avons pu identifier quatre grands groupes de déterminants de la localisation des *FDI* : la demande, les bénéfices tirés de l'agglomération des entreprises, les coûts de

production et de transaction ainsi que la fiscalité[2]. Au-delà de ces déterminants traditionnels, l'effet du développement des réseaux de transport de type métro n'a, à notre connaissance, pas encore été analysé. L'absence d'étude sur ce sujet est en grande partie due aux différences d'échelle spatiale. Un métro est une infrastructure d'une aire urbaine alors que la littérature s'est concentrée sur des zones plus étendues comme la région ou le pays.

Dans cette étude, nous évaluons l'impact de la présence d'un système de métro (la marge extensive), et de leur degré de développement en taille (la marge intensive) sur le nombre d'implantations de projets FDI sur la période 2003-2014. La mesure de tels effets sur l'attractivité nécessite d'avoir une quantité d'informations suffisamment complète pour permettre de mesurer le nombre des investissements directs étrangers au niveau des villes, de tenir compte des caractéristiques socioéconomiques et géographiques des villes et, enfin, de déterminer pour toutes les villes du monde la présence et la taille des réseaux de métro. Un travail considérable a été consacré à la collecte de ces informations et à la constitution d'une base de données originale et unique qui couvre l'ensemble de la planète, et dont les unités d'observation sont les aires urbaines.

Pour identifier l'impact du métro, nous mettons en œuvre une stratégie empirique basée sur une modélisation de Poisson. Notre variable dépendante est le nombre de projets qui s'implantent dans une ville au cours d'une période donnée. Il s'agit d'une variable entière positive ou nulle. Le recours au modèle de Poisson permet une meilleure prise en compte à la fois des variables de comptage positives, mais aussi de la proportion importante de zéros, c'est-à-dire de villes où ne s'implante aucun projet.

Pour pouvoir interpréter nos résultats de manière causale, il est nécessaire de tenir compte de l'endogénéité éventuelle du modèle. Des trois sources possibles d'endogénéité : les erreurs de mesure, la causalité circulaire et le biais de variables omises, nous focalisons notre attention sur la dernière. En effet, nous sommes confiants quant à l'absence d'erreur systématique dans notre mesure des systèmes de métro. De plus en estimant notre modèle avec une variable métro provenant d'une source différente de la base que nous avons construite, nous obtenons des résultats similaires. Concernant l'existence de causalité inverse, c'est-à-dire l'implantation de projets *FDI* durant notre période d'analyse qui auraient un impact causal sur la

[2] Voir Crozet, Mayer et Mucchielli (2004) ;Head et Mayer (2004); Basile, Castellani et Zanfei (2008).

présence ou la taille d'un système de métro durant la même période, celle-ci nous paraît peu probable. D'une part, la décision et le temps de construction d'un système de métro s'inscrivent sur de nombreuses années. L'effet du métro que nous capturons dans nos estimations, en particulier pour les métros existant avant 2003, est donc très vraisemblablement la conséquence d'une décision prise bien avant le choix de localisation des multinationales plutôt que l'inverse. Par ailleurs, notre période d'analyse des choix de localisation des multinationales étant relativement courte, il paraît également peu probable que ce soit l'implantation d'une multinationale qui ait entraîné la construction d'un métro dans le cas des métros apparus après 2003. Reste le problème des variables omises qui nous paraît être potentiellement le plus problématique dans notre cadre d'analyse.

Nous tenons compte d'un possible biais de variables omises grâce à une méthode innovante mise en place par Lewbel (2012). Cette méthode consiste à générer des instruments sur la base des données disponibles, sans recours à des informations externes, lorsque le terme d'erreur d'un modèle auxiliaire ayant pour variable dépendante la variable potentiellement endogène est hétéroscédastique. Ces instruments correspondent au produit des résidus de cette régression avec les variables de contrôle, préalablement centrées et présumées exogènes, appartenant au modèle d'intérêt. Intuitivement, la corrélation de ces instruments générés avec la variable endogène tend à augmenter quand la dispersion conditionnelle de cette dernière est fortement fonction des variables de contrôle. Comme le démontre Lewbel (2012) lui-même, le recours à des instruments internes via cette méthode donne des résultats très proches de ceux basés sur des instruments externes (voir aussi Mishra et Smyth, 2015). Cette méthode est à ce titre maintenant utilisée dans de nombreuses études (Emran et Hou, 2013 ; Davies et Desbordes, 2015 ; King, Smith et Wang, 2016 ; Breschi et Lenzi, 2016), et se révèle particulièrement appropriée lorsque des instruments externes valides ne sont pas disponibles comme c'est le cas dans notre analyse.

Les résultats de nos estimations montrent une relation robuste entre la présence d'un métro (marge extensive) ou la taille d'un système de métro (marge intensive) et l'attractivité d'une ville. Toutes choses égales par ailleurs, la présence d'un métro dans une ville augmente de 60 % à 100 % le nombre de projets *FDI* qui s'implantent dans cette ville. En moyenne, la présence d'un métro peut donc permettre le doublement du nombre de projets qui s'implantent dans l'aire urbaine. Pour la marge intensive,

l'élasticité entre la taille du système de métro et le nombre de projets varie de 0,1 à 0,2. Autrement dit, pour une augmentation de 1 % de la taille du système, nous constatons une augmentation du nombre de projets qui se situe entre 0,1 % et 0,2 %.

Le Grand Paris Express prévoit un doublement de la taille du réseau de métro francilien, soit une augmentation de 100 %. En appliquant cette valeur à nos résultats, c'est-à-dire une élasticité comprise entre 0,1 et 0,2 pour la marge intensive, nous obtenons un impact attendu pour le GPE qui varie entre 10 % et 20 % de projets supplémentaires. Paris ayant attiré 1 310 projets entre 2003 et 2014, de telles élasticités correspondent à un intervalle de 130 à 262 projets supplémentaires sur une période de 12 ans. De plus, nous constatons dans les données *FdiMarkets* que les projets qui se sont implantés à Paris génèrent en moyenne 52 emplois par projet[3]. Sous l'hypothèse que les nouveaux projets sont plus nombreux, mais similaires en taille, le Grand Paris Express générerait entre 6 800 et 14 000 emplois directs issus de l'implantation des firmes multinationales.

Dans la partie 2, nous présentons une revue de la littérature sur les choix de localisation des firmes multinationales et le lien entre l'attractivité des villes et les infrastructures de transport. Dans la partie 3 nous détaillons la construction de notre base de données qui est inédite par la richesse des informations que nous avons collectées au niveau des villes (pour la plupart géocodées) et par sa couverture d'envergure mondiale. La partie 4 détaille le modèle économétrique, la construction des variables de contrôle et la génération des instruments. Dans la partie 5, nous présentons les principaux résultats pour l'impact du métro sur les marges extensive et intensive pour le modèle de Poisson (avec et sans variables instrumentales). Enfin, la dernière partie de ce rapport est consacrée à un résumé et à la discussion de la portée de nos résultats.

[3] La base *FdiMarkets* contient des informations sur l'emploi induit par chaque projet d'implantation, mais ces données sont pour la plupart estimées. Il ne nous semblait donc pas rigoureux d'analyser directement l'effet du métro sur l'emploi créé par les multinationales. L'information sur la moyenne des emplois créés par projet s'implantant à Paris peut néanmoins être utilisée pour donner un ordre de grandeur de l'impact potentiel sur l'emploi du GPE via l'implantation de nouveaux projets FDI.

2 Revue de la littérature

Cette étude s'inscrit dans deux courants distincts de la littérature économique : le premier s'intéresse aux déterminants de la localisation des entreprises multinationales, le second à l'impact des infrastructures de transport sur l'attractivité des villes. Notre analyse apporte deux contributions. Tout d'abord, par rapport aux études existantes sur les choix de localisation des multinationales, notre unité géographique d'analyse, la ville, est plus fine que les zonages habituellement considérés. Nous avons constitué une base de données de couverture mondiale qui comprend un grand nombre de caractéristiques des aires urbaines. Nous sommes donc en mesure de tester la pertinence des déterminants traditionnellement identifiés au niveau des pays et/ou des régions pour des choix de localisation au niveau des villes. Ensuite, notre analyse contribue à l'enrichissement de la littérature traitant de l'impact des infrastructures de transport sur l'attractivité des villes en général. Notons d'emblée que l'effet d'une politique d'aménagement aussi importante que celle du développement du réseau de métro sur l'attractivité d'un territoire pour les multinationales n'a fait l'objet que de très peu d'études à notre connaissance.

2.1 Les déterminants de la localisation des entreprises multinationales : quelle pertinence au niveau des villes ?

La littérature existante fait l'hypothèse que les entreprises choisissent la localisation qui maximise leurs profits (Head et Mayer, 2004). Dans ce cadre, la littérature a identifié quatre grands groupes de déterminants : la demande, les bénéfices tirés de l'agglomération des entreprises, les coûts de production, et la fiscalité pesant sur les entreprises. L'originalité de notre étude est qu'elle vise à identifier l'impact des réseaux de métro sur les choix de localisation des entreprises multinationales à l'échelle de la ville. Il est donc nécessaire de discuter de la pertinence des critères identifiés par la littérature en ce qui concerne l'attractivité au niveau de la ville.

De nombreuses entreprises multinationales s'implantent à l'étranger pour servir directement (plutôt que via l'export) les marchés étrangers. Dans ce cadre, la taille du marché de destination est un critère fondamental de localisation. En général, la taille du marché est mesurée par le PIB du pays ou de la région considérée. Néanmoins,

plusieurs études montrent que dans leur stratégie de localisation, les entreprises sont non seulement sensibles à la demande émanant du marché local, mais également à la demande émanant des marchés « proches ». Ainsi, le « potentiel de marché » qui tient compte de ces aspects apparaît comme un déterminant important de l'attractivité des pays (Head et Mayer, 2004 ; Crozet, et al., 2004). Nous pouvons nous attendre à ce que la taille des villes, qui reflète leur poids économique, soit en tant que telle un déterminant important de l'attractivité des villes. Comme nous le verrons, les données de PIB des villes n'étant pas disponibles dans les limites de leurs aires urbaines effectives (qu'il faut distinguer des découpages administratifs), une contribution importante de notre travail a consisté à utiliser les « lumières la nuit » pour mesurer précisément l'activité économique au niveau des aires urbaines. Cette technique a déjà été adoptée par la littérature récente en économie urbaine[4], mais pas dans le contexte qui nous intéresse. Le potentiel de marché d'une ville est possiblement différent du potentiel de marché d'un pays. Nous l'avons séparé en deux composantes : un potentiel de marché local autour de la ville et un potentiel de marché national mesuré principalement par le PIB du pays.

La littérature existante (Duranton et Puga, 2004; Fujita et Thisse, 2013) suggère par ailleurs que les bénéfices tirés de l'agglomération des entreprises l'emportent sur les coûts en termes de concurrence ou de congestion qui peuvent émaner de la concentration des entreprises en un même lieu. Ainsi, les entreprises ont tendance à s'implanter dans les pays ou régions qui attirent déjà de nombreuses entreprises. Cela leur permet de bénéficier des externalités positives liées à la présence de fournisseurs à proximité, de bénéficier d'un bassin de main-d'œuvre aux compétences et qualifications adaptées, mais également de tirer profit des partenariats potentiels pouvant être créés. En général, ces effets sont pris en compte par le décompte du nombre d'entreprises (domestiques et/ou étrangères) appartenant au même secteur que l'entreprise qui investit dans le pays et/ou la région considérés. Des travaux récents appartenant au domaine de l'économie urbaine ont par ailleurs mis en évidence l'existence d'un phénomène de spécialisation « fonctionnelle ». Certaines villes se spécialiseraient dans les activités d'innovation et de décision alors que d'autres se spécialiseraient dans des « fonctions » de production (Duranton et Puga, 2005). Ainsi, la présence de « *headquarters* » (quartiers généraux et/ou centres de décision) dans

[4] Voir l'annexe D pour une revue de littérature sur le sujet plus détaillée et une analyse de la pertinence de cette mesure dans le cadre qui est le nôtre.

une ville contribuerait à attirer l'implantation d'autres *headquarters* (Davis et Henderson, 2008, Strauss-Kahn et Vives, 2009).

En ce qui concerne l'impact des coûts de production sur la localisation des multinationales, les conclusions de la littérature empirique sont relativement contrastées. Plusieurs études concluent à un impact non significatif des salaires sur les choix de localisation (Devereux et Griffith, 1998 ; Head, Ries et Swenson, 1999 ; Head et Mayer, 2004). Différents éléments sont avancés pour expliquer ce résultat. D'abord, les régions au potentiel de marché élevé sont aussi celles où les salaires sont les plus élevés (Head et Mayer, 2006). Ensuite, un salaire élevé peut aussi refléter un important degré de qualification de la main-d'œuvre. Ainsi, Liu, Wei, Wang et Liu (2006) montrent dans le cas de la Chine, qu'en raison de biais d'endogénéité, l'impact des salaires sur la localisation des multinationales a pu être sous-estimé. À notre connaissance, il n'existe pas de données disponibles sur les salaires ni sur les degrés de qualification de la main-d'œuvre au niveau des villes à l'échelle mondiale. Comme nous contrôlons pour la taille de la population au niveau de la ville, les lumières la nuit sont là aussi une variable proxy de la productivité, plus imprécise que la taille du marché local, mais c'est, à notre connaissance, la seule mesure disponible (voir l'annexe D) au niveau des villes.

Enfin, la fiscalité peut également avoir un impact sur l'attractivité. De nombreuses études empiriques se sont attachées à essayer de quantifier l'impact de la fiscalité sur les choix de localisation des entreprises. C'est généralement le choix de localisation des entreprises au niveau des pays qui est étudié. Dans l'ensemble, l'effet des taxes est plutôt négatif et significatif (Devereux et Griffith, 1998 ; Bénassy-Quéré, Fontagné et Lahrèche-Révil, 2003 ; Head et Mayer, 2004). Au niveau infranational (choix de localisation à l'intérieur des pays), les résultats sont plus contrastés. Dans le cas du Royaume-Uni, Duranton, Gobillon et Overman (2011) montrent que le niveau de taxe foncière n'affecte pas l'entrée d'entreprises dans une localité, mais a un effet négatif sur l'emploi. Dans le cas de la France, Rathelot et Sillard (2008) concluent qu'une fiscalité plus importante pesant sur les entreprises décourage la localisation de ces dernières. Mayer, Mayneris et Py (2015) analysent l'impact en France de la politique des zones franches urbaines et montrent que des exonérations géographiquement ciblées affectent positivement la localisation des entreprises, incitant les entreprises à se (re)localiser de la partie de la ville ne bénéficiant pas d'exonérations vers la partie

de la ville en bénéficiant. Là encore, nous ne sommes pas en mesure d'obtenir des informations aussi détaillées pour l'ensemble des villes de notre échantillon. Néanmoins, l'utilisation d'effets fixes au niveau des pays (des états ou des provinces) permet de tenir compte (en partie) de ces effets et de savoir si les différences de fiscalité auxquelles sont exposées les multinationales sont davantage le reflet d'une variation entre pays que de variations de fiscalité entre villes au sein d'un même pays.

2.2 Infrastructures de transports et attractivité des villes

Certaines études tentent d'analyser directement l'impact des réseaux de transport sur l'activité locale. D'un point de vue théorique, la nouvelle économie géographique (Fujita, Krugman, Venables et Fujita, 1999, Krugman, 1991a, Krugman, 1991b) donne un cadre de travail pour analyser le rôle des infrastructures dans le développement économique d'une zone. Les modèles de la nouvelle économie géographique prennent en compte, dans le cadre d'un modèle d'équilibre général avec des hypothèses microéconomiques robustes, les effets des rendements d'échelle. Une des idées fortes de cette littérature est que les infrastructures de transport ont un effet sur les grandeurs économiques locales (Ottaviano et Thisse, 2004).

Des études analysent empiriquement l'impact des politiques publiques visant à améliorer l'offre d'infrastructure dans les pays et/ou les régions considérés sur la localisation des multinationales. Elles tendent à mettre en évidence l'existence d'un impact indirect assez limité des réseaux de transport. Crozet, et al. (2004) concluent à un effet très faible de la prime d'aménagement du territoire sur la localisation des investissements directs étrangers dans les régions françaises. Basile, et al. (2008) obtiennent des effets positifs pour les fonds structurels et de cohésion européens dans le choix de localisation des multinationales en Europe, mais cet impact est faible en comparaison des autres déterminants. Ces études suggèrent, indirectement, un impact potentiellement faible des infrastructures sur la localisation des multinationales. Néanmoins, il faut considérer les résultats avec précaution : d'abord, ces politiques ne concernent pas uniquement les infrastructures de transport. Par ailleurs, elles sont attribuées aux régions les plus défavorisées et donc les moins attractives pour les investissements étrangers, quelles que soient les infrastructures présentes.

{, #1@@hidden}

D'autres études s'intéressent directement à l'effet des réseaux de transport sur la localisation des entreprises (domestiques ou étrangères). Coughlin et Segev (2000) trouvent que les autoroutes ont un impact sur la localisation de firmes étrangères dans différents comtés aux États-Unis. Holl (2004a) analyse l'effet de la construction de routes sur la localisation d'entreprises manufacturières en Espagne et montre que la probabilité de localisation des entreprises augmente dans les villes situées à moins de 10 km d'une autoroute. Holl (2004b) conclut également à un impact positif du réseau routier dans le cas du Portugal. Gibbons, Lyytikäinen, Overman et Sanchis-Guarner (2016) analysent dans le cas de la Grande-Bretagne, l'impact de la construction de nouvelles infrastructures routières et montrent que celles-ci permettent d'attirer de nouvelles entreprises.

Ces études concluent dans l'ensemble à un impact positif des infrastructures de transport sur la localisation des entreprises, mais ne concernent en général qu'un seul pays ou région[5] et se focalisent sur l'impact des réseaux routiers plutôt que des réseaux ferrés. Or dans une synthèse récente de la littérature, Redding et Turner (2015) rappellent que les différents types de transport ne sont pas interchangeables. La structure spatiale de la population est plus sensible à la route et aux transports intra-urbains alors que les activités de production sont plus sensibles au rail. Les résultats des études citées précédemment ne sont donc pas nécessairement applicables dans le cas dans du métro. Particulièrement proche de notre propos, l'étude de Mayer et Trévien (2015) montre que l'ouverture de lignes de RER en région parisienne a eu un impact sur l'emploi et la localisation des entreprises à l'échelle de la ville, celui-ci étant plus fort (de l'ordre de 32%) pour la localisation des entreprises étrangères.

Identifier l'impact des réseaux de transport et, en particulier, des métros sur l'attractivité des entreprises, soulève en réalité un certain nombre de défis, dont le premier s'articule autour de la relation de causalité. La construction ou l'extension de réseaux de transport n'est en général pas exogène à la localisation de l'activité économique. En termes méthodologiques, ces études se distinguent par le recours à des estimations par variables instrumentales relativement originales souvent basées sur des données historiques et géolocalisées très précises (telles que des plans initiaux

[5] Il existe des études se focalisant directement sur l'impact des infrastructures de transport sur la localisation des multinationales entre différentes régions et/ou pays. Ainsi Khadaroo et Seetanah (2009) trouvent que les infrastructures de transport, mesurées par la longueur de routes pavées par kilomètre carré, influencent l'entrée d'investissements directs étrangers dans 33 économies africaines. Dans une autre étude, Azemar et Desbordes (2009) ne trouvent pas d'effet des infrastructures dans les pays en voie de développement.

de routes ou de voies ferrées ayant été abandonnés pour des raisons exogènes). Notre contribution consistera à vérifier si ces effets positifs des infrastructures de transport sur la localisation des entreprises mis en évidence à l'échelle locale (d'un pays ou d'une région) ont également un impact sur la localisation des multinationales dans les villes à l'échelle mondiale. Néanmoins, en l'absence de données aussi détaillées pour toutes les villes du monde, un important défi de notre travail consiste à résoudre ce problème potentiel d'endogénéité.

3 La base de données

La première étape de notre travail a consisté à construire une base de données de couverture mondiale dont le niveau d'observation est l'aire urbaine ou la ville (leurs contours sont définis dans la sous-partie suivante). Compte tenu du nombre trop restreint de systèmes de métro (moins de 200 systèmes), l'échelle mondiale a été retenue comme référence, rejetant ainsi une analyse à l'échelle nationale ou continentale afin de disposer d'un nombre important d'observations. Cependant travailler à l'échelle mondiale soulève bon nombre de difficultés. En effet, il faut pouvoir disposer de données fiables et uniformisées sur les caractéristiques socioéconomiques des villes pour l'ensemble de la planète. Or, apparier de telles données est difficile soit en raison du manque d'homogénéisation des noms des villes entre différentes bases, soit parce que les données disponibles n'ont pas la même emprise spatiale, ou encore parce que les découpages administratifs reflètent des réalités différentes selon les pays. Pour contourner ces problèmes, nous avons fait le choix de recourir à des données géolocalisées, issues des données de satellite, dont la précision ne cesse d'augmenter. Notre parti pris a été de construire une base entièrement géolocalisée en faisant un usage intensif des systèmes d'information géographique (SIG)[6]. Ce choix présente l'avantage majeur de faciliter les appariements entre des bases géolocalisées en utilisant la position géographique, sans passer par des identifiants communs qui, la plupart du temps, n'existent pas. Cette étape indispensable a nécessité un travail conséquent de préparation pour collecter des données issues des différentes sources, et les intégrer au niveau de la ville de manière géolocalisée.

[6] L'ensemble des traitements a été effectué sous le logiciel ARCGIS de la société ESRI.

La base de données que nous avons constituée, inédite par la richesse des informations qu'elle contient au niveau des villes et par sa couverture d'envergure mondiale, peut se décomposer en 4 grandes sous-parties : la base des aires urbaines, la base des *FDI*, la base des métros et la base des variables de contrôle incluant des caractéristiques économiques, géographiques, climatiques et des informations sur les infrastructures de transport autres que les métros. Nous détaillons ci-dessous les sources et méthodes mobilisées pour chacune de ces parties afin d'aboutir à la base finale sur laquelle repose l'analyse empirique.

3.1 La base des aires urbaines (ou villes)

La définition des aires urbaines

Pour créer une base au niveau des villes, la première étape a consisté à trouver une liste mondiale identifiant le plus précisément possible le contour et nom de chaque agglomération pour servir de support au reste des données géolocalisées. Potere, Schneider, Angel et Civco (2009) ont étudié 10 bases issues de 8 groupes gouvernementaux ou académiques représentant le recouvrement de la Terre par de la surface urbanisée. À partir de ces 10 cartes, ils ont estimé que la base MODIS[7] 500 (Schneider, 2009) était la plus fiable en termes de représentation de l'espace urbanisé étant donné sa précision. Cette base utilise une résolution de 500 mètres pour évaluer l'ensemble du bâti de la planète pour l'année 2000. Nous avons décidé d'utiliser cette base pour construire notre base des aires urbaines. Cependant nous n'avons pas directement utilisé les données brutes issues des données MODIS, mais les données mises à disposition par le *Lincoln Institute*[8]. En effet, en utilisant les données MODIS 500, le *Lincoln Institute* (Angel, 2010) a créé une base de données géolocalisées comprenant 3 649 aires urbaines identifiées et nommées dont la population était supérieure à 100 000 habitants en 2000. Ce premier travail d'intégration a servi de socle pour notre base des aires urbaines.

En comparant la carte géographique des aires urbaines *Lincoln Institute* avec les cartes sur les lumières la nuit (dont nous détaillons plus loin les sources) et la base des aires urbaines de la société de consulting DeLorme, (base qui est livrée avec le logiciel Arcgis), nous avons identifié des villes manquantes. Nous avons ainsi retravaillé

[7] MODIS : Moderate Resolution Imaging Spectroradiometer.
[8] http://www.lincolninst.edu/subcenters/atlas-urban-expansion/google-earth-data.aspx.

l'univers des aires urbaines pour finalement obtenir une base de 3 729 unités. Notons que les nouvelles aires sont, soit des aires qui manquaient dans la base du *Lincoln Institute*, soit des aires que nous avons divisées, car, même s'il existe un bâti continu, nous avons constaté qu'il s'agissait d'aires urbaines différentes. Le principal critère pour réaliser cette division d'une aire urbaine en plusieurs a été de vérifier s'il existait un système de transport urbain intégré ou séparé.

Notre base des aires urbaines identifie et géolocalise ainsi la surface des principales villes du monde. Pour chaque ville, nous disposons également de la valeur de la population pour l'année 2000. Cette base sert de fondation pour construire et apparier les données nécessaires à nos estimations. Pour chaque aire urbaine de notre base, nous avons construit trois zones « tampons » : de un, cinq et dix kilomètres. Ces tampons facilitent les jointures spatiales, c'est-à-dire l'appariement entre deux ou plusieurs bases de données reposant sur des critères spatiaux différents.

Tableau 1 : Statistiques descriptives des aires urbaines

	Moyenne	Écart-type	Médiane	Somme
Population 2000	546 939	1 369 819	212 640	2 039 535 937
Surface (en km²)	90,4	252,7	33,66	337 119

Notre base urbaine représente près de 2,04 milliards d'habitants en 2000. La taille moyenne des villes est de 546 939 habitants avec un écart-type de 1 369 819 et une médiane se situant à 212 640 habitants. La surface moyenne des villes est de 90 km² pour une superficie totale de 340 000 km². À titre de comparaison[9], selon la Banque Mondiale, 2,835 milliards de personnes vivaient dans les aires urbaines en 2000. Cette étude couvre donc 70 % de la population mondiale.

Les données de population des Nations Unies

Afin d'obtenir des données avec une profondeur temporelle pour la population, nous avons appareillé à notre base des aires urbaines, la base de population des villes des Nations Unies[10]. Cette base comprend la population de toutes les agglomérations de plus de 300 000 habitants en 2014, par année depuis 1950 avec des prévisions de

[9] La surface de la Terre est d'environ 510 millions de km² dont 149,4 millions de terres émergées.
[10] http://esa.un.org/unpd/wup/CD-ROM/

population jusqu'en 2030. Une grande partie des valeurs de cette base sont issues d'interpolations entre deux recensements.

La base des Nations Unies est fournie avec la longitude et la latitude du centre de la ville que nous avons utilisées pour appareiller les données avec notre base des aires urbaines. Sur les 3 729 villes de notre base, nous avons pu apparier les données de population des Nations Unies pour 1 602 villes. Si une métropole de notre base, qui est représentée par un polygone, recouvrait plusieurs points de la base population des Nations Unies, nous avons additionné les différentes valeurs pour obtenir la population de notre aire urbaine.

Bien que les données de la base des Nations Unies soient un sous-ensemble de notre base des aires urbaines, sa profondeur temporelle, unique à notre connaissance, en fait la source principale de la population des villes dans cette étude. Nos résultats se limitent donc à ce sous-ensemble des aires urbaines. D'autres sources de données sur la population existent, en particulier basées sur des grilles surfaciques : elles ont l'avantage de couvrir l'ensemble de la planète, mais ne sont pas disponibles avec une telle profondeur temporelle.

3.2 Les projets d'investissements directs étrangers

Pour mener à bien notre analyse, il a été nécessaire de définir une mesure objective de l'attractivité internationale de chacune de nos aires urbaines. Notre choix s'est porté sur le nombre de projets d'investissements directs étrangers ou *FDI (pour Foreign Direct Investment)*. Dans notre étude, nous ne tenons pas compte des fusions et acquisitions. En effet, les projets *FDI* nouveaux semblent plus pertinents, car ils sont créateurs d'emplois et de nouvelles activités, ce qui n'est pas le cas des fusions et acquisitions.

Pour construire les données *FDI,* nous nous sommes appuyés sur la base *fDiMarkets* créée par le *Financial Times* pour valoriser l'information économique que le journal détient sur les investissements. *fDiMarkets* recense tous les projets d'investissements annoncés et/ou réalisés par les firmes multinationales et abonde une base dont chaque observation représente un investissement étranger « pur » (investissement dit « *greenfield* », les joint-ventures pouvant parfois être présentes, mais restant très rares). Pour chaque projet, des informations portant sur la société qui investit, le pays source de la multinationale, le secteur et le type d'activité de

l'investissement, la localisation dans le pays de destination au niveau de la ville, le capital investi et le nombre d'emplois créés sont disponibles. Concernant les deux dernières variables, les valeurs sont souvent estimées par *fDiMarkets,* elles sont donc très délicates à utiliser pour une analyse économétrique. À notre connaissance, *fDiMarkets* est la seule base de données d'emprise mondiale, disponible pour les chercheurs, qui permette une localisation à l'échelle de la ville d'un projet *FDI greenfield*. Des données de ce type ont été utilisées par Defever (2006) et Py et Hatem (2009), mais uniquement pour des choix de localisation réalisés par les multinationales du monde entier se localisant au sein de l'Europe. Les données dont nous disposons couvrent une période de 12 ans, s'étalonnant de 2003 à 2014. La base complète comporte 143 395 observations pour autant de projets d'investissements.

Un long travail a été nécessaire pour géolocaliser précisément les investissements, c'est-à-dire affecter une longitude et une latitude à chaque projet à partir du nom de la ville. Cette localisation a été réalisée à l'aide de la base *Geonames*[11] qui contient environ 8 millions de noms de localités avec la longitude et la latitude correspondante. L'appariement entre la base *FDI* et la base *Geonames* nous a demandé un effort conséquent d'appariement. En effet, pour un grand nombre de pays, il est impossible d'identifier de manière unique une ville, car il existe plusieurs villes portant le même nom. Par exemple, presque tous les états des États-Unis possèdent une ville qui s'appelle Springfield. Pour identifier de façon unique une ville dans un pays, il faut donc une division administrative supplémentaire. Pour chaque pays, cette division prend des formes différentes : états, régions, préfectures, villes, etc. Il existe deux tentatives de normalisation des régions administratives de niveau 1 (le niveau 0 étant les pays). L'une découle de la standardisation ISO (Standard ISO 3166-2) et l'autre de l'institut de standardisation américain FIPS (FIPS 10-4). Si ces deux standardisations sont similaires pour certains pays, elles sont incompatibles pour d'autres. La base *fDiMarkets* utilise les standards ISO alors que la base GEONAME utilise le standard FIPS. Avant de commencer tout appariement, il a d'abord fallu réaliser une table de correspondance entre ces deux standards. Une fois cette étape effectuée, nous avons standardisé un grand nombre de chaînes de caractères manuellement pour pouvoir faire l'appariement entre les bases *fDiMarkets* et *Geonames*. Cette étape fastidieuse consiste essentiellement à harmoniser les noms des pays, des régions et des villes. La

[11] http://www.geonames.org/

base géolocalisée des *FDI* issue de ce processus comporte 112 190 observations. Les 32 069 observations restantes n'ont pas pu être localisées, soit parce qu'aucune concordance n'a pu être trouvée entre les bases, soit parce que l'information de localisation était manquante dans la base *fDiMarkets*.

Une fois les projets *FDI* précisément géocodés, il est facile de les affecter à une aire urbaine. Néanmoins, un des enjeux a été de définir l'échelle géographique pertinente pour la localisation des projets *FDI*. Il est en effet très probable que certains projets d'investissements se localisent à la frontière des contours de la ville (pour des raisons de disponibilité des terrains et/ou de prix de l'immobilier par exemple) tout en restant proches de leur périphérie pour bénéficier de toutes les aménités et du marché potentiel offert par ces villes. Ainsi, nous avons pris le parti d'utiliser le tampon de 5 km autour de chaque aire urbaine pour tenir compte d'éventuels effets de bord. Un projet situé dans une aire urbaine ou dans le tampon qui l'entoure est donc considéré comme faisant partie de l'aire urbaine. Les 17 136 projets qui sont en dehors des aires urbaines que nous avons définies sont considérés comme des investissements soit à la campagne, soit dans de très petites villes qui n'appartiennent pas à notre base des aires urbaines.

La Figure 1 représente le volume total de *FDI* sur la période 2003-2014 pour les principales métropoles[12]. Notons que le contraste entre les villes européennes qui tendent à attirer de nombreux *FDI*, qu'elles soient proches de la mer ou non, et les autres villes du monde, dont la quasi-totalité des villes attractives se trouve en bord de mer, est saisissant.

Dans les Figure 2 et Figure 3, nous nous intéressons aux 20 villes les plus attractives pour la période 2003-2014. Shanghai est la métropole qui arrive en tête de ce classement avec plus de 3 500 projets d'investissements directs étrangers en 12 ans. Le podium est complété par Singapour et Londres. Paris se classe à la 7e place avec 1 461 projets, juste devant New York. C'est la seconde ville européenne du classement, derrière Londres et ses 3 149 projets. Dans ce classement, 6 villes se situent en Europe, 3 en Chine, 3 en Inde et 2 aux États-Unis. L'hémisphère nord comporte ainsi 18 villes en tête de classement contre seulement 2 dans l'hémisphère sud (Sao Paulo et Sydney).

[12] Les villes qui totalisent moins de 10 projets *FDI* sur la période ne sont pas représentées sur la carte.

La Figure 3 représente, pour ces 20 premières villes, la part des *FDI* qu'elles concentrent par rapport aux autres grandes villes et aux zones rurales du pays auquel elles appartiennent. Sur l'ensemble des *FDI* localisés en France entre 2003 et 2014, 27,3 % l'ont été dans l'agglomération parisienne, 42,8 % dans une autre aire urbaine française, 10,6 % en dehors des grandes aires urbaines et enfin, 19,2 % n'ont pu être géolocalisés . Shanghai représente à elle seule 22 % des *FDI* chinois et un projet d'investissement étranger sur deux au Japon se localise dans la mégapole tokyoïte. À l'opposé, en Allemagne ou aux États-Unis, les investissements directs étrangers ont tendance à être plus répartis entre les différentes villes du pays.

Dans les Figure 4 et Figure 6, nous présentons respectivement l'évolution annuelle du nombre total de projets dans notre base, du nombre d'emplois induits, et enfin de la somme du capital investi. Nous constatons sans peine que la crise de 2008 a eu un impact fort sur l'évolution des *FDI*. Cet impact concerne à la fois le volume, qui diminue considérablement en 2009 comparé à 2008, mais aussi la structure des *FDI*. Après 2008, les projets FDI sont en moyenne bien moins importants comme en témoignent les graphiques des Figure 5 et Figure 6.

3.3 La base des métros

Depuis l'ouverture en 1863 du métro de Londres avec ses locomotives à vapeur, les systèmes de transport urbain de type métro se sont répandus à travers le monde. À notre connaissance, il n'existe pas de base mondiale capable d'identifier de manière précise l'ensemble des systèmes de métro. De fait, il a donc fallu créer notre propre base. Pour autant, produire une base qui recense l'ensemble des systèmes de métro dans le monde soulève en premier lieu le problème de la définition d'un système de transport de type métro. En effet, il existe de nombreux types de transports rapides intra-urbains. Conformément aux standards de la littérature sur le sujet, les caractéristiques suivantes ont été retenues pour définir un système de métro :

- La voie que le système emprunte ne doit pas être partagée avec d'autres moyens de transport, en particulier l'automobile.
- Les fréquences de passage doivent être élevées.
- Les stations doivent être spatialement proches.

Cette définition, assez large, a le mérite de faire la différence entre les moyens de transport de type tramway (qui souvent partagent leurs voies avec les automobiles) et

les métros. Elle n'est en revanche pas suffisamment précise pour faire la différence avec certains types de train, notamment les trains légers. Notre stratégie a été de construire une base métro la plus large possible en identifiant des systèmes hybrides ou des systèmes simplement très différents de la représentation classique que nous avons du métro, mais qui restent des systèmes de transports rapides intra-urbains. De cette base avec une définition élargie, nous avons extrait une base avec une définition plus restreinte des systèmes de métro. Par exemple, Sydney figure dans la base élargie, mais est absente de la base restreinte, car son système de transport urbain est assez proche d'un système métro dans le centre de l'agglomération et ressemble à un train de banlieue lorsque l'on s'éloigne du centre. De la même manière, les *U-Bahn* allemands de Francfort ou encore Düsseldorf seront dans la base élargie et non dans la base restreinte.

Cette stratégie nous permettra de faire des tests de robustesse pour nous assurer que notre choix de définition et de délimitation des réseaux de métro n'affecte pas qualitativement nos résultats. Notons que lorsqu'une ville se trouve seulement dans la base élargie, elle est considérée comme une ville sans métro dans la base restreinte.

Les métros suscitent de l'intérêt et produisent des communautés de passionnés sur le sujet. Nous nous sommes appuyés sur cette passion pour construire notre base des métros en consultant les sites Internet communautaires sur le sujet. Les plus notables sont sans aucun doute Metrobits[13] et UrbanRail.Net[14]. À ces deux sites, il faut ajouter les pages Wikipédia[15] qui traitent du sujet. C'est principalement à partir de ces sources que nous avons construit notre base.

[13] http://mic-ro.com/metro/index.html.
[14] http://www.urbanrail.net/.
[15] https://fr.wikipe ia.org/wiki/Liste_des_m%C3%A9tros_du_monde.

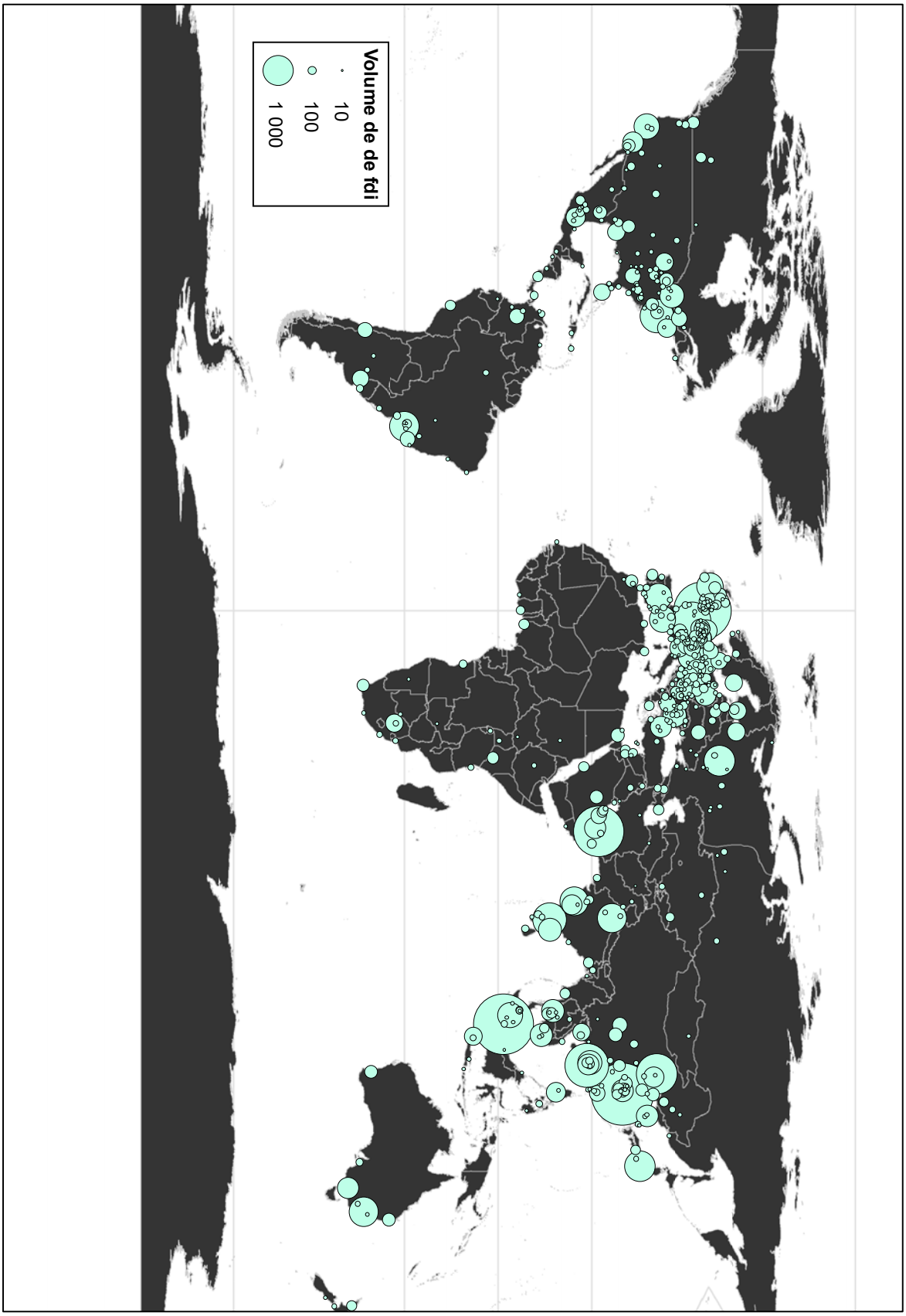
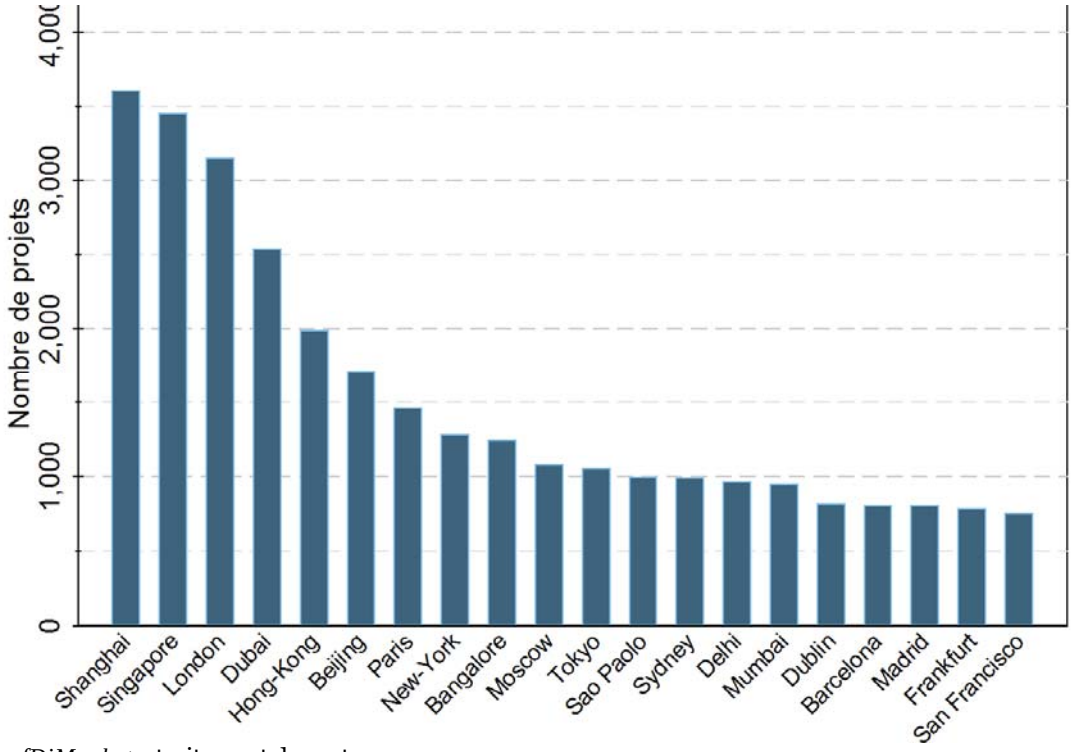
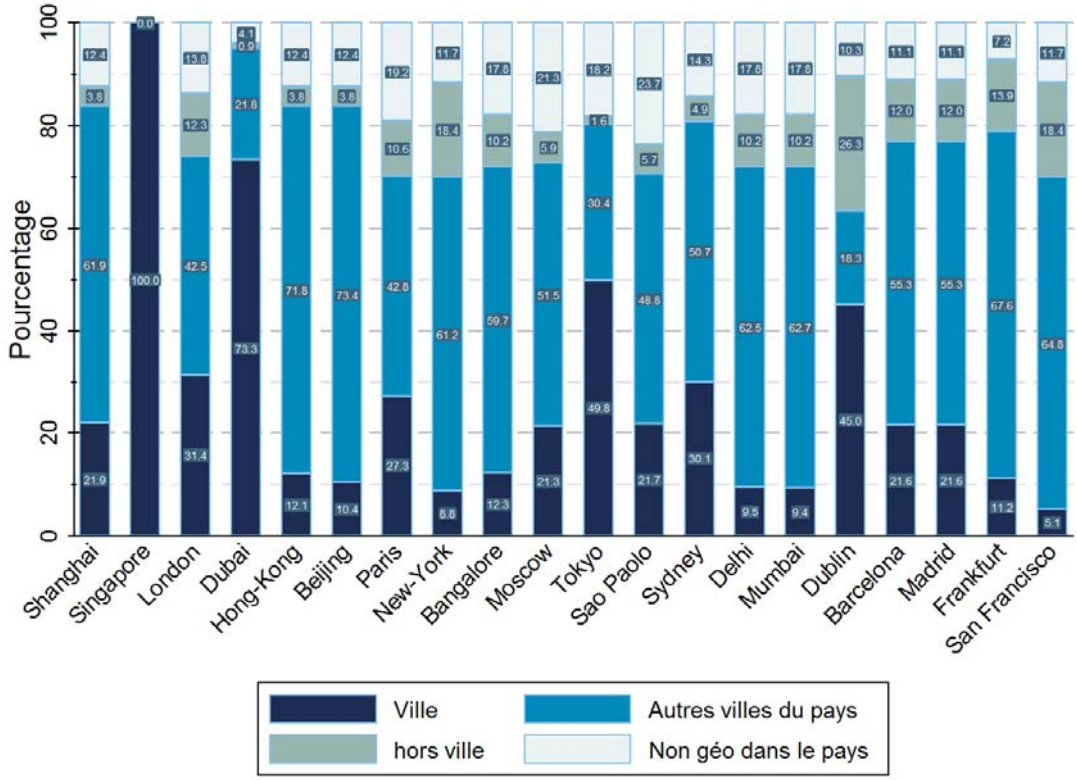


Figure 1 : Volume de *FDI* par ville 2003-2014



Source : *fDiMarkets*, traitement des auteurs.

Figure 2 : Les 20 villes les plus attractives

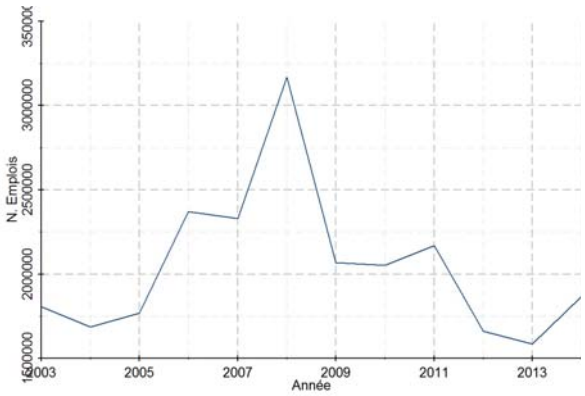


Source : *fDiMarkets*, traitement des auteurs.

Figure 3 : Les 20 villes les plus attractives dans leur pays

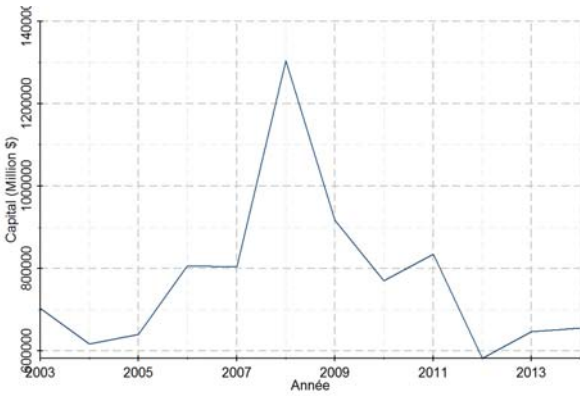
Source : *fDiMarkets*, traitement des auteurs.
Note de lecture : Ce graphique représente le nombre total de projets *FDI* par année présents dans la base brute de *fDiMarkets*.

Figure 4 : Nombre de projets FDI par année



Source : *fDiMarkets*, traitement des auteurs.
Note de lecture : Ce graphique représente le nombre total d'emplois induits par les projets FDI. La crise de 2008 a un fort impact sur les emplois.

Figure 5 : Nombre d'emplois induits par année



Source : *fDiMarkets*, traitement des auteurs.
Note de lecture : Ce graphique représente la somme des capitaux investis pour les projets FDI. La crise de 2008 a un fort impact sur le volume de capitaux.

Figure 6 : Capital investi

Pour la majeure partie des systèmes de métro que nous avons identifiés, nous avons ainsi réussi à obtenir, pour chaque année, depuis la mise en service du système, la taille en kilomètres du réseau. Le fait de travailler sur des données annuelles nous permet d'accéder à l'évolution du système au fil des années. Toutefois, outre la taille du réseau, les données relatives au nombre de stations et de lignes de métro ne sont disponibles que pour l'ensemble de l'année 2015. Les données sur le nombre annuel de voyageurs existent également, mais celles-ci restent largement plus parcellaires, avec des années de collecte différentes d'un système à l'autre.

Pour illustrer la construction de notre base, prenons l'exemple de l'agglomération parisienne. Comme indiqué précédemment, nous avons cherché à identifier le système de métro et à le différencier des autres modes de transports existants, ce qui nous a conduits à faire un certain nombre de choix. Premièrement, seules les lignes de métro ont été comptabilisées tandis que les lignes de tramway et celles de RER ont été exclues. Deuxièmement, les stations desservies par plusieurs lignes ont été comptées autant de fois qu'elles sont traversées par une ligne dans la base élargie, et de manière unique dans la base réduite. Nous obtenons ainsi pour le métro parisien 220 km de lignes de métro pour 303 stations dans la base restreinte et 383 stations dans la base élargie. Le fait d'avoir comme cela deux définitions pour les systèmes de métro nous permet de tester la robustesse de nos résultats quant à notre choix concernant la typologie de ces infrastructures de transport.

Ce travail minutieux nous a ainsi permis d'identifier 187 villes avec un système métro dans le monde pour la base élargie et 142 villes pour la base réduite. En plus de ces métros existants, nous avons identifié 34 systèmes en construction, c'est-à-dire des villes qui n'ont pas de système métro en 2015, mais qui ont lancé des travaux pour en construire un.

La Figure 7 présente la localisation des métros en 2015 en séparant la base restreinte de la base élargie et les systèmes en construction. La base élargie comprend tous les systèmes de la base restreinte auxquels s'ajoutent les systèmes moins standards dont les caractéristiques sont plus éloignées de celles des métros traditionnels comme ceux de Londres ou Paris. Notons que la majorité des futurs systèmes de métro se situent en Inde, en Chine ou au Moyen-Orient. Nous n'avons identifié qu'un seul nouveau projet en Europe, à Thessalonique en Grèce. Ce phénomène peut s'expliquer par le fait que les grandes villes européennes possèdent

déjà toutes un système de métro, tandis que les villes moins importantes n'en font pas une priorité dans leur politique de transport. En dépit du faible nombre de métros présents sur le continent américain, il n'y a qu'un seul projet en développement à Quito en Équateur et aucun aux États-Unis. Enfin, exception faite à Alger et Le Caire, aucune autre ville d'Afrique n'a investi dans un système de métro.

L'analyse de la distribution des années de mise en service des systèmes de métro nous renseigne sur les différentes périodes de construction de ces systèmes. Ainsi, si certaines villes pionnières ont construit leur réseau de métro avant la Seconde Guerre mondiale, il y a bien quelques systèmes qui sont apparus dans les années 50 à 70, mais c'est à partir des années 80 que le nombre de mises en service a vraiment explosé (Figure 8). Enfin, la popularité des systèmes de métro ne saurait s'affaiblir puisqu'entre 2003 et 2015, 50 systèmes ont été mis en service dont 22 depuis 2010 (principalement en Asie).

La taille cumulée des systèmes de métro (Figure 9) est un processus en croissance constante. Jusqu'au milieu des années 60, la progression était assez lente. Après cette période, la progression n'a au contraire cessé de s'accroître. Notons une sensible accélération à partir de 2010, portée en particulier par 22 nouveaux systèmes.

La Figure 10 fait la synthèse des deux précédentes figures. Elle propose l'évolution cumulée de la taille des systèmes en fonction de la période de mise en service. Cette figure nous permet de constater que les systèmes ne sont pas figés dans le temps et que même longtemps après leur année de mise en service, les systèmes continuent de se développer.

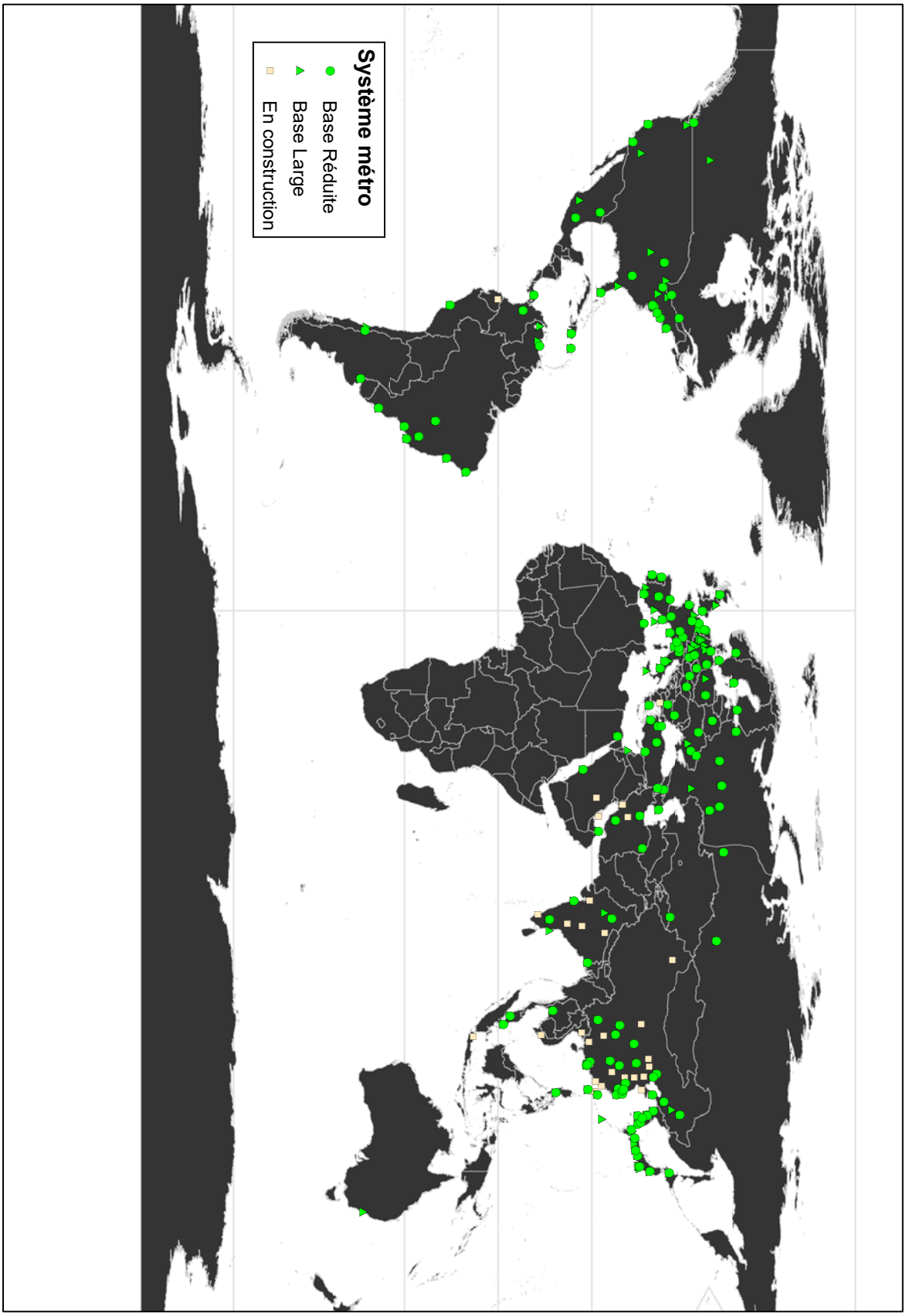


Figure 7 : Localisation des métros, base large, base réduite et systèmes en construction en 2015

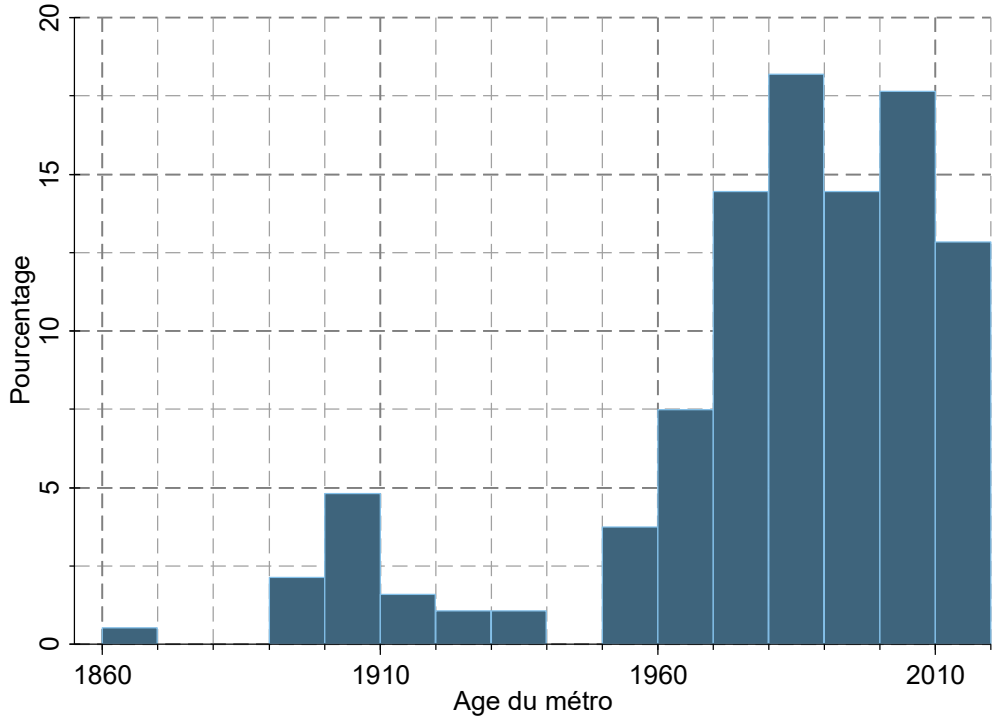
Tableau 2 : Statistiques descriptives des systèmes de métro (base élargie)

	Longueur moyenne (km)	Nombre de lignes	Nombre de stations	Âge
Nombre	187			
Moyenne	67,91	3,50	58,05	34,35
Écart-type	92,12	3,89	75,34	31,02
Médiane	37,3	2	31	29
Somme	12 700	654	10 855	

Au total, les villes de la base élargie comprennent 12 700 km de voies de métro, 654 lignes différentes et 10 855 stations. En 2015, le système de métro moyen possède une longueur de 68 km, 3,5 lignes, 58 stations et est âgé de 35 ans. La moitié des réseaux mesure moins de 37 km. Pour les statistiques descriptives, il y a peu de différence avec la base restreinte (en annexe A) et la base élargie. Néanmoins, les systèmes de la base restreinte sont en moyenne un peu plus grands et plus âgés.

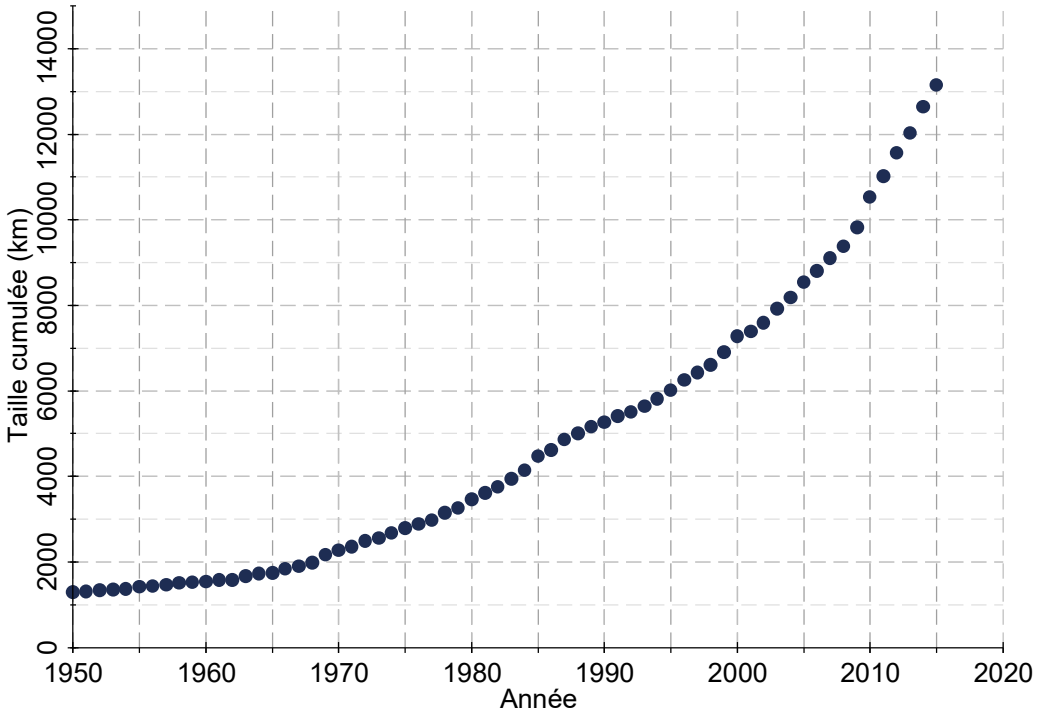
Dans la Figure 10, nous représentons pour chaque ville le nombre de *FDI* en fonction de la position de la ville dans le classement du nombre de *FDI* pour la période 2003-2014. Il est intéressant d'observer que toutes les villes du top 20 en termes de nombre de *FDI* ont un métro, à l'exception de Dublin. Autour de la 50e place, le nombre de *FDI* n'est plus que de 425 environ. À la 250e place, le nombre de projets est inférieur à 60. Pour les villes avec moins de 500 *FDI*, ce qui correspond à la 40eplace, la répartition entre villes « avec » et « sans » système de métro est bien équilibrée.

Dans la Figure 12, nous représentons le volume de *FDI* d'une ville avec métro de la base large en fonction de la taille de son système. Nous pouvons observer une corrélation entre taille de métro et volume de *FDI*. Enfin, si l'on s'intéresse à l'évolution annuelle de la part du nombre de *FDI* en fonction des villes avec et sans métro en 2015 (Figure 13), nous observons que même si les villes possédant un métro sont très peu nombreuses (moins de 200), elles représentent une part importante du volume total de *FDI*. Ce chiffre croît avant 2009 et est stable après cette date, alors que pour les villes sans métro, le volume est plus volatile après 2009.



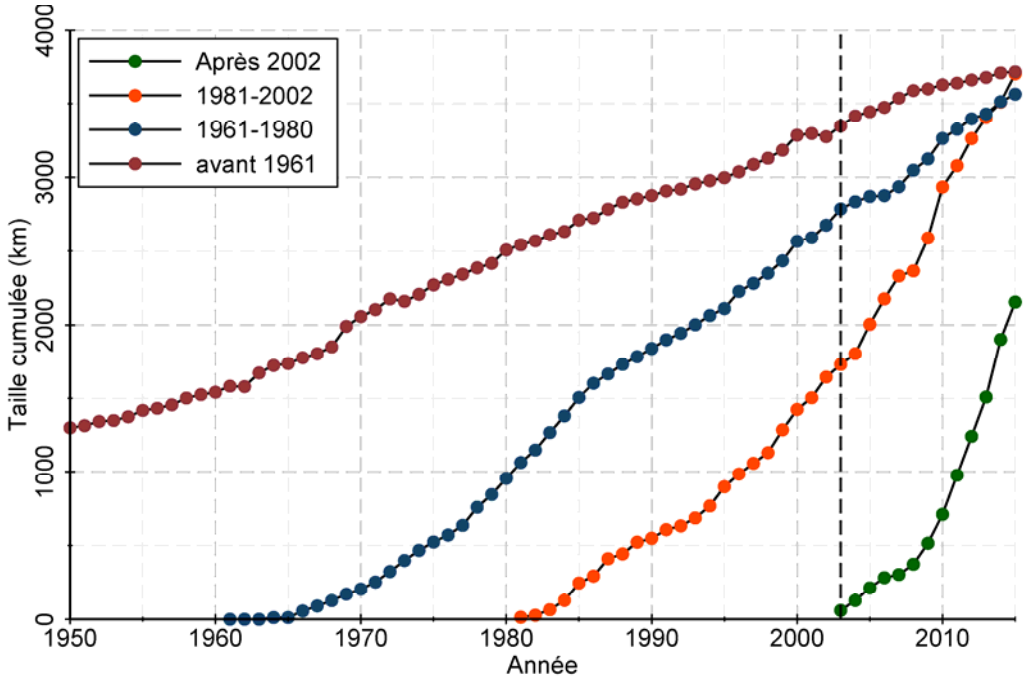
Source : Auteurs.
Note de lecture : Dans ce graphique, nous reportons le pourcentage de mises en service d'un système de métro pour une période donnée (dans notre cas des périodes de 10 ans) pour l'ensemble de la base large en 2015. Ainsi entre 1960 et 1970, 7,5 % de la totalité des 187 métros ont été mis en service.

Figure 8 : Distribution de la date de mise en service des métros (base large)



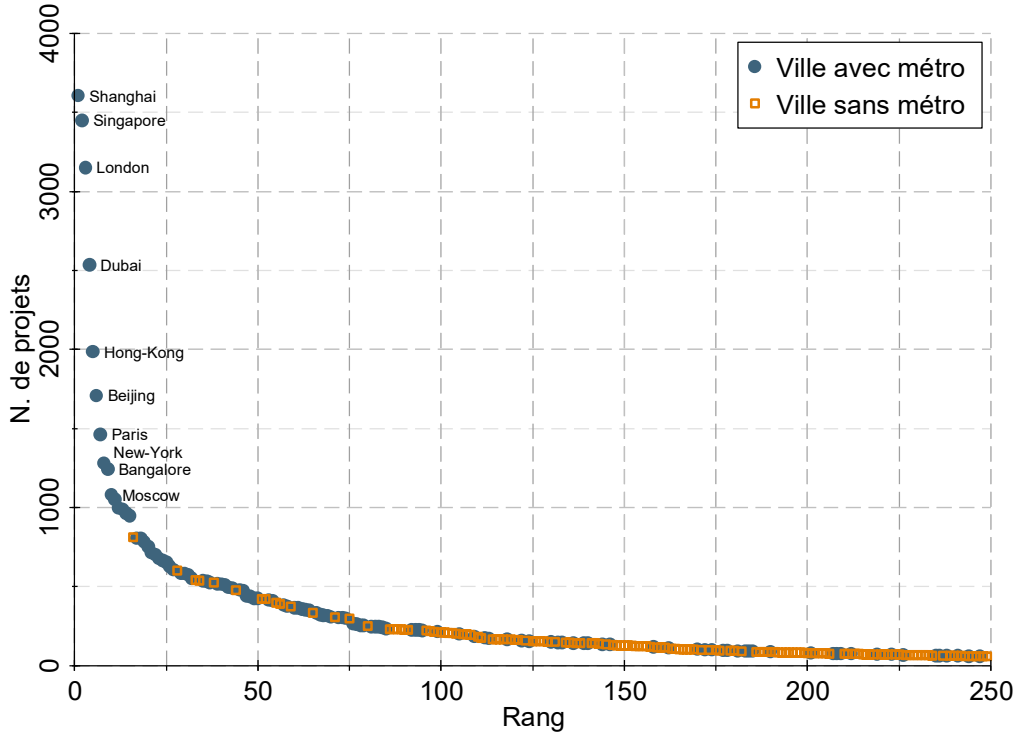
Source : Auteurs.
Note de lecture : Évolution pour les métros de la base large.

Figure 9 : Évolution de la taille agrégée des systèmes de métro depuis 1950



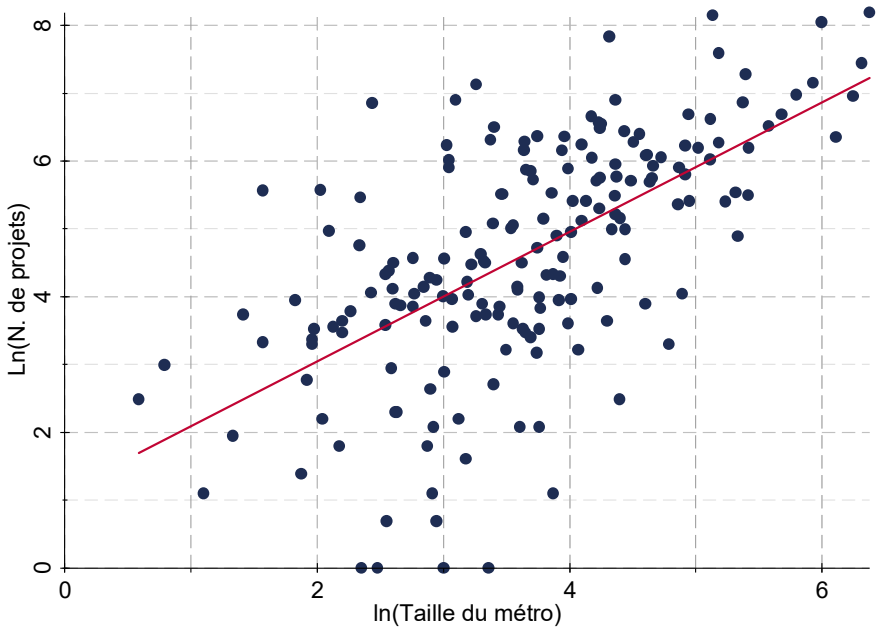
Source : *fDiMarkets*, traitement des auteurs.

Figure 10 : Évolution de la taille cumulée en fonction de la période de mise en service



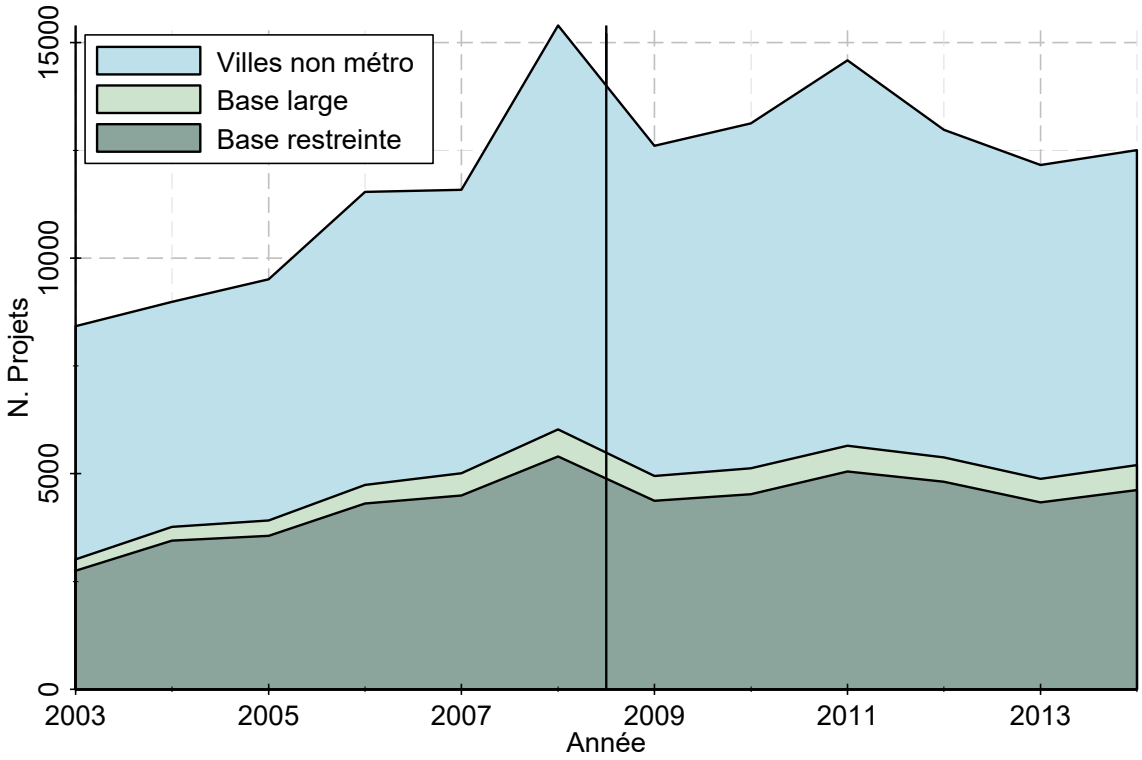
Source : *fDiMarkets*, traitement des auteurs.

Figure 11 : Évolution du nombre de FDI en fonction du rang (2003-2014)



Source : *fDiMarkets*, traitement des auteurs.

Figure 12 : Nombre de projets et taille du métro en logarithme



Source : *fDiMarkets*, traitement des auteurs.
Note de lecture : pour les villes ayant un métro en 2015 dans la base large.

Figure 13 : Répartition agrégée des projets FDI en fonction des villes avec et sans métro

3.4 Les variables de contrôle

Nous contrôlons pour toutes les caractéristiques des villes autres que le métro qui pourraient influencer les choix de localisation des firmes multinationales. Pour répondre à cette exigence, nous avons réuni en plus des données sur les *FDI* et les métros, tout un ensemble de données au niveau des aires urbaines.

Notre choix des variables de contrôle est basé à la fois sur la disponibilité des données, mais aussi sur la littérature existante sur le sujet. Selon la littérature sur les déterminants de la localisation des multinationales au niveau des pays ou des régions sont principalement : la demande (approchée par la taille du marché), les coûts de production, les effets d'agglomération et la fiscalité (Head et Mayer, 2004; Crozet, et al., 2004; Basile, et al., 2008). L'originalité de notre étude est double puisqu'il s'agit d'identifier, d'une part, la pertinence de ces critères standards dans l'analyse des choix de localisation des entreprises à une échelle géographique plus fine, c'est-à-dire au niveau des villes, et, d'autre part, au-delà de ces critères, de quantifier l'impact d'autres caractéristiques plus locales avec comme variables d'intérêt les systèmes de métro sur l'attractivité d'une ville.

Dans les sous-parties suivantes, nous détaillons les sources et les constructions de l'ensemble des variables de contrôle que nous utilisons dans nos estimations.

3.4.1 Population et surface

Comme nous l'avons déjà mentionné, les données sur la population proviennent des travaux du *Lincoln Institute* pour l'année 2000 (nous utilisons les données des agences statistiques nationales pour les villes que nous avons séparées ou rajoutées) et des Nations Unies pour le sous-ensemble des villes de plus 300 000 habitants en 2014. Pour la surface géographique, nous utilisons l'aire en kilomètres carrés de chaque métropole en 2000, calculée à partir des données MODIS500 dans le logiciel Arcgis. La conjonction de ces deux variables nous permet aussi de contrôler pour la densité de population dans les estimations.

3.4.2 Activité économique

Il nous fallait une mesure de l'activité économique des villes. À notre connaissance, une telle base n'existe pas pour un grand nombre de pays. Nous avons porté notre choix sur une variable proxy, la mesure de la luminosité la nuit, qui est de plus en plus utilisée pour mesurer l'activité humaine et sa croissance au fil du temps. Cette mesure, issue des satellites météorologiques de défense américaine, présente l'avantage d'être mondiale (à l'exception des pôles Nord et Sud) et de proposer une mesure uniforme pour l'ensemble de la planète, évitant ainsi les problèmes de différence de comptabilité entre les pays ou encore d'absence de données au niveau local pour bon nombre de pays.

Les données brutes sont disponibles sur le site du NOAA,[16] une agence américaine en charge de la recherche sur les changements climatiques (entre autres). Il existe deux types de données disponibles pour les lumières la nuit. Le premier concerne les lumières dites saturées avec une valeur allant de 0 à 63. Ces données sont disponibles entre 1992 et 2013. Le second type correspond aux lumières dites en radiance. Elles ne sont disponibles que pour certaines années souvent couplées (1996-97, 1999, 2000, 2002-03, 2004, 2005-06, 2010, 2010-11). Si les données des lumières saturées ont l'avantage d'être disponibles pour toutes les années, elles sont moins corrélées avec les grandeurs économiques. Ainsi dans le cadre de notre analyse, nous avons privilégié l'utilisation des données de radiance, car elles représentent plus fidèlement l'activité économique agrégée locale. L'annexe D présente un peu plus en détail les données de lumières la nuit et le niveau de corrélation avec l'activité économique pour un sous-ensemble de villes.

3.4.3 Les données G-Econ

Au-delà des variables de population, il n'existe pas d'autre base de données à couverture mondiale renseignant sur les caractéristiques socioéconomiques à l'échelle des villes. Cependant, il existe de nombreuses bases de données géographiques qui peuvent se révéler extrêmement importantes dans les choix de localisation. La base *G-Econ* est l'une d'entre elles. L'Université de Yale a développé le projet *G-Econ*[17] (Nordhaus, Azam, Corderi, Hood *et al.*, 2006) qui, en plus de nombreuses caractéristiques géographiques, permet d'avoir pour des cellules de 1 degré de

[16] http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.
[17] La version que nous utilisons est la version 4.0. http://gecon.yale.edu/.

longitude et de 1 degré de latitude (ce qui correspond à des cellules d'environ 100 km de côté à l'équateur), des données sur l'activité économique de ces cellules. Contrairement aux bases de données traditionnelles qui reposent sur des découpages administratifs (pays, états, régions, etc.), la base *G-Econ* est une base géophysique qui s'appuie sur une grille surfacique de plus de 27 000 cellules. La variable retenue pour l'activité économique est la valeur ajoutée, un concept similaire au produit intérieur brut. Les auteurs de la base ont défini le *produit brut de la cellule* de la même manière qu'est défini le PIB au niveau d'un pays, c'est-à-dire la valeur ajoutée d'une région mesurée comme la production totale des marchés des biens et services moins les achats des autres secteurs. Cette mesure est équivalente aux revenus des facteurs de production de la région. Afin d'obtenir la valeur pour une cellule, les auteurs estiment la valeur ajoutée de la région et pondèrent ensuite par la population de la cellule. La production et la précision des données économiques dépendent fortement des données nationales sous-jacentes. Les données à la cellule s'appuient sur les données régionales quand elles sont disponibles, sur les revenus industriels par secteur, sur l'emploi régional et, lorsque toutes ces mesures ne sont pas disponibles, sur des mesures de la population. Le passage des données régionales aux données cellulaires a été effectué par des procédures de repositionnement spatial (« *Spatial Rescaling* »). Pour intégrer les données de grille à nos données métropoles, pour chaque aire urbaine, nous calculons la somme pondérée par la proportion de la surface de la ville qui se trouve sur la cellule. Malgré ce travail conséquent pour obtenir une valeur de la production locale, nos investigations ont montré que les lumières la nuit étaient des mesures plus fiables de la mesure de l'activité économique.

Nous utilisons également les données de distance entre le centre des cellules et l'Océan, et entre le centre des cellules et les principales voies navigables à l'intérieur des terres issues de la base *G-Econ* comme contrôle dans nos estimations avec la même procédure de somme pondérée pour passer des données en grille/surface aux données des aires urbaines. Concernant les distances, nous avons aussi calculé la distance minimale entre une ville et la frontière terrestre la plus proche. Ensuite, cette base nous fournit deux autres indicateurs géographiques : l'élévation et la rugosité du terrain. Ce dernier est un indicateur qui représente la nature accidentée du relief d'une ville. Enfin, la base *G-Econ* nous fournit aussi des données météorologiques. Nous utilisons les moyennes et écarts-types des précipitations et températures entre 1980 et 2008, car il s'agit également de facteurs potentiellement importants de l'attractivité d'une ville.

3.4.4 Les autres infrastructures de transport

L'accessibilité d'une ville est une composante importante de son attractivité. Une première variable, la distance à une voie navigable majeure (issue des données *G-Econ*), a déjà été évoquée. Mais le transport fluvial n'est pas le seul moyen de transport. Ainsi, nous avons décidé d'ajouter à nos données de transport, les ports, les aéroports, les voies de chemin de fer et les routes. Enfin, nous avons ajouté une autre variable capable de mesurer les transports intra-urbains : la taille du réseau de tramway.

Pour construire nos variables d'infrastructure, nous avons utilisé le tampon de 10 km[18] autour de chaque aire urbaine et nous mesurons dans ce tampon la longueur ou le nombre des infrastructures de transport. Pour certaines de ces données, plusieurs sources (d'emprise mondiale) coexistent. Nous avons fait le choix de n'en exclure aucune et de tester si nos estimations étaient sensibles à la source de données utilisée. Aucun changement notable n'est apparu lorsque nous changions la source des données, c'est pourquoi nous présentons les résultats avec la source qui nous est apparue la plus fiable.

Les données relatives aux routes, aux voies ferrées et aux aéroports proviennent du projet VMAP[19], issu des systèmes d'information géographique du gouvernement américain qui a décidé de les mettre dans le domaine public. L'échelle des données que nous utilisons est au 1:1 000 000. Pour les ports, nous disposons d'une information extrêmement détaillée sur l'ensemble des ports de la planète issue du Wold Port Index[20]. Là encore, les données sont issues d'une agence gouvernementale américaine.

D'autres bases de données géographiques existent pour ces infrastructures. Nous en avons identifié deux : le projet communautaire Natural Earth et les données fournies avec le logiciel ESRI qui sont issues de la société Delorme. Moins complètes que les données VMAP, ces deux autres sources de données présentent à certains moments l'avantage de proposer des typologies différentes. Par exemple, VMAP décompose le réseau routier en trois catégories : routes primaires, routes secondaires et chemins alors que les données issues de Delorme proposent une typologie basée sur un découpage du réseau routier en 3 catégories : autoroutes, routes majeures et routes

[18] Changer la taille du tampon (5 km, 10 km ou même 15 km) ne change pas nos résultats. Nous avons choisi de travailler avec une zone tampon pour simuler une certaine croissance entre 2000 et 2015.
[19] http://geoengine.nga.mil/geospatial/SW_TOOLS/NIMAMUSE/webinter/rast_roam.html.
[20] http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_62&pubCode=0015

locales. Enfin, les données communautaires possèdent l'avantage d'être mises à jour plus régulièrement, ce qui ne semble plus être le cas pour les données VMAP.

Pour les infrastructures intra-urbaines, nous avons dans un premier temps mobilisé les données issues du site communautaire OpenStreet[21]. Ce site nous permet d'extraire différentes données d'infrastructure intra-urbaines. Nous avons décidé de retenir les métros (cela nous donnera une seconde source), les tramways, les trains légers et les monorails. Une fois ces données extraites, nous les avons transformées dans un format utilisable dans Arcgis. Pour construire un indicateur de la taille de ces infrastructures, nous avons dû retravailler énormément ces données en utilisant une procédure qui fusionne des lignes proches les unes des autres. En effet, le propre d'OpenStreet est de proposer une alternative libre aux différents systèmes de cartographie Internet existants qui se prête davantage à la représentation visuelle qu'à l'analyse statistique.

Malgré une quantité de données impressionnantes, les données issues du projet OpenStreet ne sont pas exhaustives et finalement assez peu appropriées à l'analyse statistique. Par exemple, nous ne retrouvons pas l'ensemble de nos systèmes de métro. Un autre inconvénient, inhérent au projet communautaire, est un certain manque d'uniformisation. Dans certaines villes, les contributeurs vont jusqu'à identifier les attractions des parcs récréatifs comme système de transport. Cependant, l'utilisation d'OpenStreet nous a permis d'identifier les tramways, comme étant le type de transport le plus important pour les transports intra-urbains, après le métro.

Pour augmenter la précision de notre variable de contrôle tramway, nous avons mis en place une procédure d'aspiration de données sur Wikipédia. Cette procédure nous a permis de récupérer la taille en 2016 et l'année de mise en service de l'ensemble des systèmes de tramway dans le monde. C'est cette variable que nous privilégions.

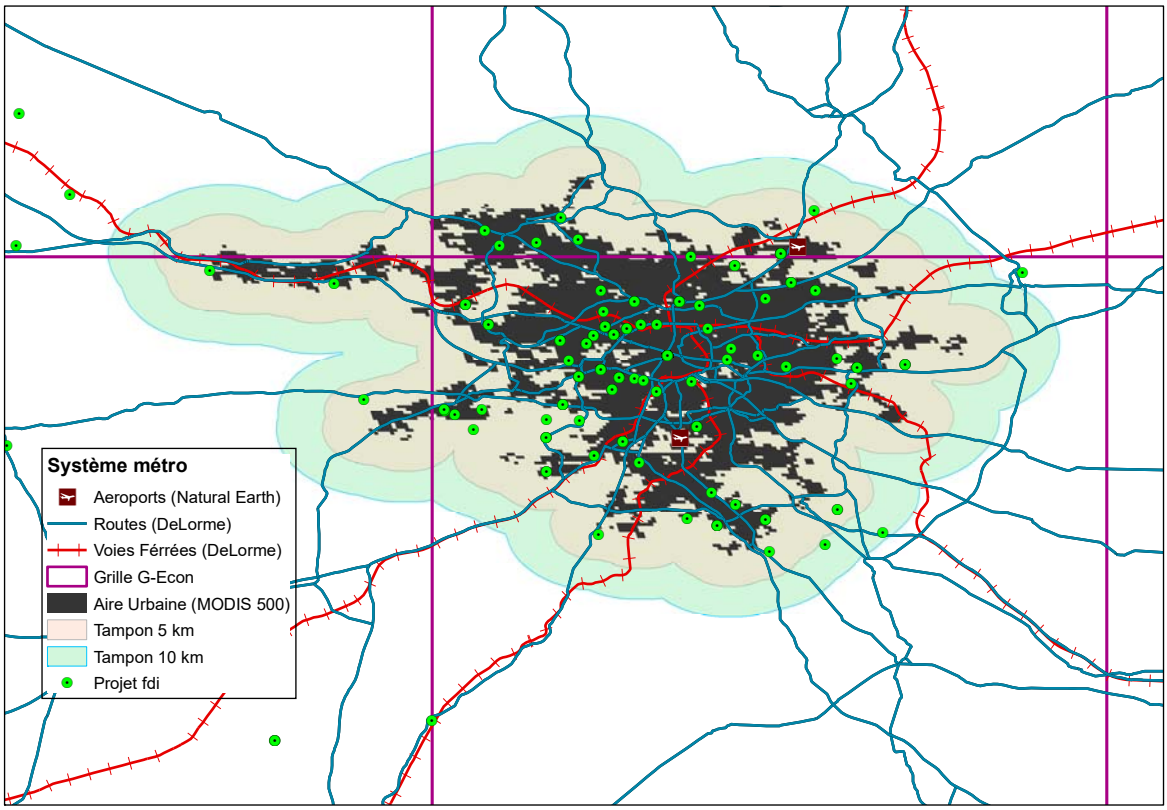
3.5 Exemples d'intégration

Les cartes 3 et 4 nous donnent un aperçu du travail d'intégration des données pour deux aires urbaines : Paris et New York. Les points rouges correspondent aux localisations des projets *FDI*. Pour un même point, il peut y avoir plusieurs projets puisque l'unité de localisation est la ville. Les *FDI* sont localisés dans l'aire urbaine s'ils

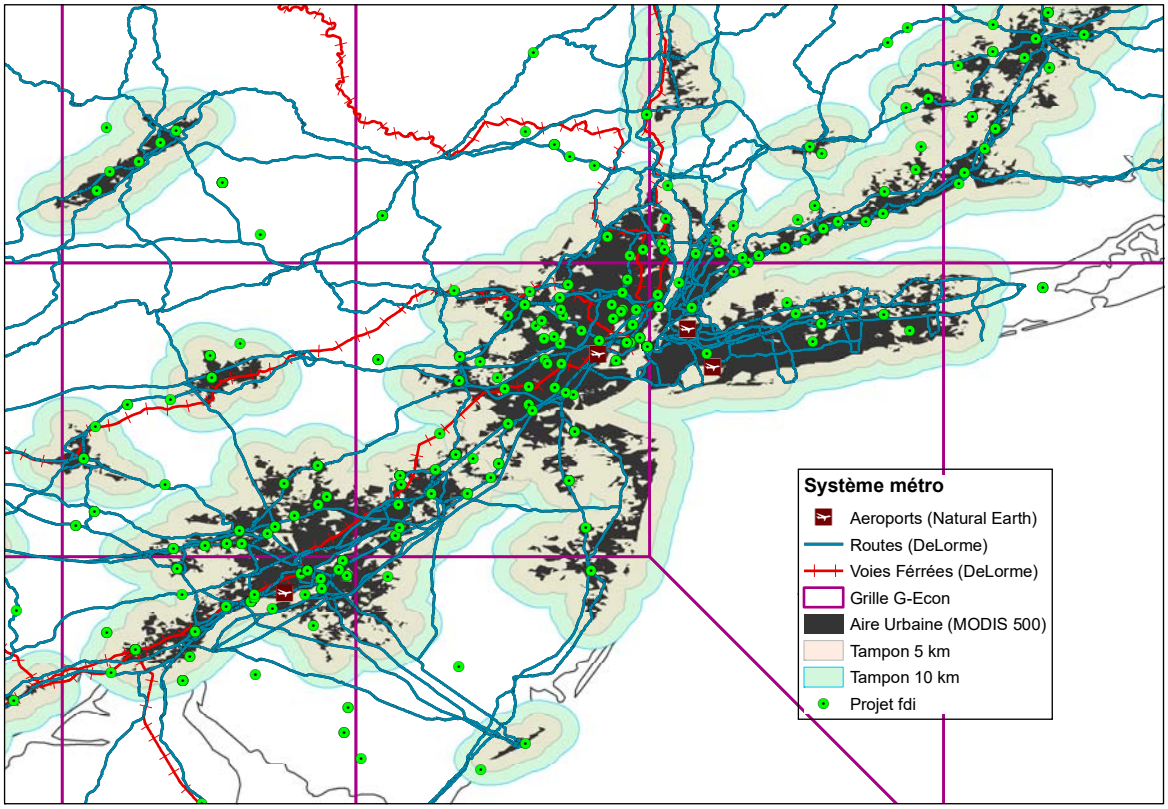
[21] https://www.openstreetmap.org. Données extraites le 1er février 2016.

se trouvent à l'intérieur du tampon de 5 km. Nous pouvons noter l'existence de projets localisés en dehors des aires urbaines. Pour les routes, les voies ferrées, les aéroports et les ports, c'est le tampon à 10 km. Ces cartes nous permettent aussi de nous rendre compte du maillage des données *G-Econ* et de voir que les grandes agglomérations se situent sur plusieurs cellules.

Le travail de collecte de l'information et son intégration dans une base statistiquement exploitable nous permettent de disposer d'une base de données inédite rassemblant des informations sur la localisation des investissements étrangers directs au niveau des villes, sur le développement des réseaux de métro et sur un grand nombre de caractéristiques propres aux villes. Le fait de disposer d'un grand nombre de variables de contrôle nous assure une plus grande confiance quant à l'absence d'un possible biais de variables omises. Maintenant que nous avons détaillé la base de données et sa construction, la partie suivante est consacrée à la présentation du modèle économétrique sous-jacent à notre analyse.



Carte 3 : Carte de l'aire urbaine parisienne



Carte 4 : Carte de l'aire urbaine new-yorkaise

4 Modèle économétrique

Le but de cette étude est de produire une estimation de l'impact d'un système de métro sur l'attractivité d'une ville, mesurée comme le nombre de projets *FDI* qui se localisent dans cette ville durant une période donnée. Notre variable expliquée est donc une variable de comptage qui peut être positive avec des valeurs assez élevées pour les villes championnes de l'attractivité, mais comportant également une part non négligeable de valeurs nulles. En effet, de nombreuses villes ne reçoivent aucun projet durant la période d'étude. Il faut alors utiliser une forme fonctionnelle qui soit en mesure de traiter, à la fois, le nombre élevé de valeurs nulles, mais aussi de prendre en compte des valeurs moyennes de la variable dépendante assez élevées. Pour l'analyse, nous avons opté pour le modèle de Poisson en coupe transversale qui permet de prendre en compte de manière très satisfaisante les difficultés auxquelles nous sommes confrontés.

Notre préférence pour un modèle en coupe transversale au détriment d'un modèle en panel repose sur plusieurs arguments. Le premier est que les estimations transversales donnent des résultats d'équilibre général de long terme alors que des estimations en panel, du fait de notre faible profondeur temporelle sur les données *FDI* (12 ans), nous donneraient des estimations de l'impact consécutif à un choc sur le système de métro de très court terme. Ces chocs peuvent provenir de la mise en service d'un système ou d'une augmentation de la taille du système préexistant. Un autre argument nous ayant conduits à préférer le modèle en coupe transversale est la forte volatilité du nombre d'implantations de projets d'une année sur l'autre. Cette caractéristique est mise en évidence en Figure 14, annexe B, pour les 11 premières villes en termes d'attractivité de projets *FDI*. Cette volatilité annuelle rend l'utilisation d'un modèle en données de panel peu appropriée.

Le modèle en coupe transversale n'est cependant pas sans écueils et pose un certain nombre de défis. Le premier d'entre eux est relatif à l'apparition de nouveaux systèmes de métro durant notre période de référence. Comme nous le notions précédemment, 50 nouveaux systèmes de métro ont été mis en service entre 2003 et 2014. Ainsi, afin d'isoler l'impact de l'effet des métros préexistants, c'est-à-dire l'impact des métros mis en service avant 2003, nous utilisons deux variables pour représenter le paramètre d'intérêt : une variable pour les métros préexistants et une autre pour les nouveaux

métros. Néanmoins, pour la marge intensive et pour certaines spécifications, nous avons pu réaliser des estimations avec une seule variable capable de capturer ces différences.

Le second défi de notre stratégie concerne la possible présence d'endogénéité dans nos estimations. En particulier, il est possible que des variables omises (au niveau des villes) puissent biaiser la valeur de nos coefficients estimés. Notre première réponse à cette difficulté a consisté à construire une base de données la plus complète possible, et ainsi utiliser certaines variables comme des mesures proxy de variables auxquelles nous n'avons pas accès. L'exemple emblématique de notre travail est l'utilisation des lumières la nuit en lieu et place de l'activité économique locale. Malgré la grande qualité de notre base de données, il reste difficile d'exclure la possibilité qu'il existe des variables omises, systématiquement corrélées avec les systèmes de métro. Pour remédier à cela, nous avons réalisé des estimations par variables instrumentales. Pour les raisons qui seront exposées ci-dessous, nous avons opté pour la méthode proposée par Lewbel, (1997, 2012), qui vient compléter nos estimations traditionnelles du modèle de Poisson.

Dans les sous-parties suivantes, nous détaillons le modèle empirique que nous avons mis en œuvre pour ensuite détailler la procédure par variables instrumentales.

4.1 Modèle de Poisson

Supposons que nous soyons en présence d'un échantillon $y_1, y_2, \cdots, y_n$ de taille n de notre variable dépendante qui peut être traitée comme des réalisations indépendantes de variables aléatoires suivant une loi de Poisson avec $Y_i \sim P(\mu_i)$ et supposons que la moyenne de chaque variable aléatoire dépende d'un vecteur de variables explicatives X_i. L'espérance, μ_i, du nombre d'occurrences pour chaque variable aléatoire doit être positive[22]. Pour obtenir une valeur positive, on modélise le logarithme de l'espérance plutôt que l'espérance. On obtient le modèle linéaire suivant :

$$\ln(\mu_i) = X_i'\beta$$

En prenant l'exponentielle de l'équation ci-dessus on obtient :

$$\mu_i = e^{X_i'\beta}$$

[22] Pour une présentation plus détaillée du modèle de Poisson voir Cameron et Trivedi (2013).

Dans ce modèle, l'exponentiel du coefficient e^{β_j} représente l'effet multiplicatif de la $j^{\text{ième}}$ explicative sur la moyenne.

Les coefficients estimés du modèle de Poisson s'interprètent de la manière suivante. Augmenter X_j d'une unité, multiplie la moyenne par un facteur e^{β_j}. Pour une variable explicative transformée en logarithme, la valeur du coefficient s'interprète donc comme une élasticité.

Le modèle de Poisson s'estime généralement par maximum de vraisemblance. Dans notre cas nous utiliserons la procédure « ppml » dans Stata écrite par Silva et Tenreyro (2011). Cette procédure, qui utilise le pseudo maximum de vraisemblance, présente l'avantage de tester l'existence des paramètres estimés avant la procédure de convergence, évitant ainsi de nombreux problèmes numériques. Les écarts-types tiennent compte des effets de grappe au niveau des pays.

4.2 Les variables de contrôle au niveau des villes

Les variables explicatives que nous utilisons dans nos estimations peuvent être réparties en 4 grandes catégories : économiques, pays, infrastructures (routières et ferroviaires) et géographiques (et météorologiques). Nous décrivons ces variables et leur construction, dans les paragraphes suivants.

Variables économiques

Notre premier groupe de variables de contrôle représente les caractéristiques économiques des villes. Voici les variables que nous avons retenues :

- Le logarithme de la taille de la population en 2002 (données des Nations Unies). Cette variable prend à la fois en compte l'effet taille du marché local, mais aussi la taille du marché de l'emploi.
- La croissance de la population entre 2002 et 2014 et la croissance de la population entre 1990 et 2002 (données des Nations Unies). Ces deux variables permettent de tenir compte du dynamisme démographique contemporain et passé de la ville. Par souci de cohérence, nous découpons nos périodes de croissance en périodes de 12 années pour qu'elles correspondent au découpage de notre mesure des *FDI*.

- Le logarithme de la somme des lumières la nuit en radiance en 2003 (données NASA). Cette variable est un proxy de l'activité agrégée de la ville, le PIB de celle-ci.
- Le logarithme de la surface de l'aire urbaine brute en km² en 2000 (données issues du traitement géomatique des auteurs à partir des données MODIS). Combinée avec les données de population, la surface permet de contrôler pour la densité de population.
- Une indicatrice valant 1 si la ville est la capitale de son pays et 0 sinon. En effet les capitales sont souvent plus attractives.
- Le nombre de villes se situant à moins de 500 km de notre observation. Il s'agit d'une mesure du potentiel de marché local qui est issue du traitement géomatique des auteurs.

En ce qui concerne la population, les lumières la nuit ainsi que la surface, nous introduisons également le carré et le cube du logarithme de ces variables dans nos estimations. Cela permet de capturer d'éventuels effets non linéaires.

Nous devons noter que les variables de population, de surface et de lumière sont sensiblement corrélées comme le montre le tableau de corrélations présenté en annexe (Tableau 14). Cependant, comme ces variables ne sont pas au centre de notre analyse et que le niveau de corrélation avec la taille des métros est relativement contenu, nous avons décidé de garder l'ensemble des variables économiques.

Variables pays

La deuxième catégorie de variables de contrôle comprend des variables mesurées au niveau des pays. Nous avons retenu :

- Le logarithme du PIB du pays en 2003 en million de dollars[23] (Données issues des *PENN World tables*). Cette variable sert de contrôle pour le potentiel de marché au niveau du pays et non plus local comme précédemment.
- Le rang du pays dans le classement pour la facilité à faire des affaires de la Banque Mondiale. Cette variable sert de proxy pour l'ensemble des caractéristiques relatives à la stabilité du pays, le respect des lois, la facilité d'entreprendre, etc.

[23] Real GDP at constant 2011 national prices (in mil. 2011US$).

- Un ensemble de 4 variables indicatrices séparant le niveau de revenu du pays : revenu élevé, revenu moyen haut (référence), revenu moyen bas et revenu bas. Là encore, cette classification provient de la Banque Mondiale. Ces variables permettent de prendre en compte d'une manière non linéaire les différents facteurs associes au niveau de vie des habitants du pays.

Infrastructures de transport

Nous disposons de variables de contrôle pour tous les principaux modes de transport inter et intra-urbains disponibles dans/ou à proximité des aires urbaines.

Transport aérien

L'offre de transport aérien est capturée par les 5 variables suivantes :

- Le nombre de grands aéroports civils se trouvant dans le tampon de 10 km de l'aire urbaine.
- Le nombre de grands aéroports militaires se trouvant dans le tampon de 10 km de l'aire urbaine.
- Le nombre d'aéroports mixtes (civils et militaires) se trouvant dans le tampon de 10 km de l'aire urbaine.
- Le nombre d'aéroports de taille moyenne se trouvant dans le tampon de 10 km de l'aire urbaine.
- Le nombre de grands aéroports se trouvant dans un rayon de 250 km de l'aire urbaine.

Ce découpage fin permet de différencier l'impact en fonction à la fois du type d'aéroport, civil ou militaire, de sa taille, mais aussi en fonction du maillage d'aéroports dans le voisinage proche.

Transport maritime ou fluvial

- La présence d'un grand port.
- La présence d'un port de taille moyenne.

Là encore, la motivation est de voir s'il existe des effets différenciés non linéaires en fonction de la taille du port.

Transport routier

Dans cette catégorie, nous avons différentes mesures de l'accessibilité routière des aires urbaines. Les données utilisées proviennent du projet VMAP, et semblent être à notre connaissance les plus complètes au niveau mondial. Nous avons construit 4 variables continues en utilisant la typologie de VMAP :

- Le logarithme du nombre de kilomètres de routes principales dans le tampon de 10 km. Cette variable permet de contrôler pour l'accessibilité interne de l'aire urbaine. Ainsi, plus le réseau principal est dense plus il est facile de circuler. Mais cette variable représente aussi l'accessibilité externe de la ville, c'est-à-dire la capacité de pouvoir s'y rendre à partir d'une autre ville.
- Le logarithme du nombre de kilomètres de routes secondaires dans le tampon de 10 km. Cette variable représente surtout l'accessibilité interne de l'aire urbaine.
- Le logarithme du nombre de kilomètres de chemins dans le tampon de 10 km. Cette variable est surtout présente en Afrique et permet de contrôler pour les réseaux de routes non goudronnées.

Transport ferroviaire

- Le logarithme du nombre de kilomètres de voies de chemin de fer dans le tampon de 10 km. Une fois de plus, cette variable permet de contrôler l'accessibilité interne en donnant une mesure de la qualité du réseau de trains de banlieue, mais aussi l'accessibilité externe via la capacité à rejoindre la ville en train.
- Enfin, la dernière variable d'infrastructure mesure le logarithme du nombre de kilomètres de tramway en 2016. Il nous a semblé très important de pouvoir contrôler pour l'autre grand type de transport intra-urbain. Notons que la distribution spatiale de ces deux types de systèmes n'est pas identique comme nous pouvons le constater en analysant la Figure 15 en annexe. Les systèmes de tramway sont extrêmement présents dans les pays du nord alors que les pays émergents ont plutôt choisi les systèmes de type métro pour leur transport de masse.

Géographie et météorologie

Nous regroupons dans cette catégorie toutes les variables qui ont un lien soit à la géographie et au positionnement dans l'espace de l'aire urbaine, soit aux caractéristiques météorologiques des zones. Nous débutons cette catégorie avec 3 variables de distance :

- Le logarithme du nombre de kilomètres à une voie de navigation fluviale majeure. Cette variable nous permet de contrôler pour une autre voie de communication principalement externe à la ville et sa capacité à exporter et importer.
- Le logarithme du nombre de kilomètres à l'océan. La proximité à l'océan est une caractéristique importante de l'attractivité économique d'une ville.
- Le logarithme du nombre de kilomètres à la frontière terrestre la plus proche. Là encore, être à proximité d'un autre pays peut se révéler une caractéristique importante en termes d'attractivité.

Viennent ensuite 3 variables géographiques :

- Le logarithme de la valeur absolue de la latitude. Cette variable permet de contrôler pour l'effet négatif d'être à proximité de l'équateur.
- Le logarithme de l'élévation moyenne de l'aire urbaine.
- Le logarithme de la rugosité du terrain de l'aire urbaine. Cette variable permet de prendre en compte la facilité avec laquelle on pourra développer de l'activité dans l'aire urbaine.

Les caractéristiques météorologiques sont prises en compte par 4 variables :

- Le logarithme des précipitations moyennes de l'aire urbaine.
- Le logarithme de la température moyenne de l'aire urbaine.
- Le logarithme de l'écart-type des précipitations moyennes de l'aire urbaine.
- Le logarithme de l'écart-type des températures moyennes de l'aire urbaine.

Notons que ces 4 dernières variables sont fortement corrélées entre elles. Enfin nous construisons 6 variables indicatrices qui représentent le continent où se situe l'aire urbaine. La modalité de référence est l'Amérique du Sud.

Après ce descriptif détaillé de nos variables de contrôle, nous présentons notre stratégie par variables instrumentales.

4.3 Variables instrumentales

Dans la stratégie traditionnelle d'estimation par variables instrumentales, il est nécessaire de trouver un bon instrument, c'est-à-dire une variable qui soit corrélée avec la variable explicative endogène, mais qui soit exclue de la liste des explicatives de la variable dépendante. On parle dans ce dernier cas de la condition d'exclusion. Dans le contexte d'analyse qui est le nôtre, cela signifie qu'il faut trouver des variables qui soient corrélées avec le métro à la fois dans ses dimensions extensive et intensive, mais qui ne soient pas reliées de manière directe au nombre d'implantations de projets *FDI* dans la ville. Pour la marge extensive, il nous faudrait trouver une variable qui explique la présence d'un métro dans une ville sans pour autant être corrélée à son attractivité internationale sauf à via son effet par le métro. Pour la marge intensive, le défi est encore plus grand, car il faut pouvoir expliquer tous les zéros (les villes sans métro) et en plus pouvoir expliquer la taille du système.

Une voie envisageable concerne les caractéristiques géologiques des villes et principalement celle de la qualité des sols. Des sols plus durs pourraient empêcher ou limiter la taille des systèmes. Cependant tous les tests que nous avons pu conduire montrent que même si l'on suppose la condition d'exclusion vérifiée, que les variables géologiques n'ont pas d'impact sur l'attractivité, ces variables ne sont pas des instruments valides pour estimer de manière causale l'impact du métro, car pas assez corrélé avec les paramètres d'intérêt métro. Notre expérience sur l'histoire et la construction des systèmes de métro montrent que le génie humain est capable de surmonter de nombreuses difficultés techniques : il existe des métros dans des zones où le sol est dur ou bien humide. Dès lors, si certains types de sols peuvent rendre la construction d'un métro plus difficile et plus onéreuse, ces différences géologiques retardent probablement la construction de métros davantage qu'elles ne l'empêchent. Or dans le cadre d'analyse en coupe transversale qui est le nôtre, nous ne sommes pas en mesure d'exploiter cette possible source de variation temporelle.

De manière plus générale, dans notre recherche d'instruments, toutes les variables susceptibles d'être corrélées avec la présence d'un métro sont potentiellement liées à l'activité économique de la ville et donc à son attractivité mesurée par le nombre de projets *FDI*. Or, quand la condition d'exclusion n'est pas vérifiée, les instruments ne sont plus valides et l'estimation causale de l'impact de la variable endogène est biaisée. Cependant Lewbel (2012) a démontré que même en l'absence d'instruments, il est

possible d'identifier le coefficient de la variable endogène. En particulier, dans le cadre d'un système triangulaire qui correspond à de l'endogénéité causée par une variable omise, Lewbel démontre que l'on peut identifier de manière causale l'impact de la variable endogène en utilisant des restrictions sur les moments d'ordre supérieur par rapport à l'estimation avec des variables instrumentales traditionnelles. Ces conditions sur les moments s'appuient sur la présence d'hétéroscédasticité lors d'une première étape dans une équation liant la variable endogène et les variables exogènes. En définitive avec cette procédure, l'identification et l'estimation de l'impact causal du métro sur l'attractivité de la ville sont obtenues sans la nécessité de recourir à de l'information externe au modèle.

Cette méthode récente et novatrice a été mise en œuvre dans plusieurs branches de l'économie et son utilisation tend à se démocratiser. Emran et Hou (2013) analysent l'amélioration de l'accessibilité au marché en Chine, ils utilisent la méthode de Lewel en plus d'instruments traditionnels pour montrer qu'un meilleur accès aux marchés locaux et internationaux entraîne un effet positif sur les PIB par habitant. Dans un cadre plus urbain, Breschi et Lenzi (2016) montrent que plus les réseaux sociaux entre les inventeurs d'une ville sont denses, plus cette ville sera créative. Cette méthode a également été mobilisée dans les domaines du marché du travail, de l'éducation et la mobilité sociale. Caliendo, Lee et Mahlstedt (2017) exploitent l'hétéroscédasticité pour traiter l'endogénéité dans une analyse du salaire de réserve dans les différences de salaire entre les genres. Cheng, King, Smyth et Denny et Oppedisano (2013) s'appuient sur cette méthode pour analyser la taille des classes. Wang (2016) met en œuvre cette méthode pour analyser les relations entre la propriété et le bienêtre. Elle a également été utilisée par Magnani et Zhu (2015) dans le cadre d'une étude sur mobilité sociale en Chine. Enfin, cette méthode a aussi été utilisée en économie de la santé (Hollard et Sene, 2016 ; Craig et Howard 2014), en finance et économie internationale (Arcand, Berkes et Panizza, 2015 ; Davies et Desbordes, 2015 ; Chowdhury, Bhattacharya, Mallick et Ulubaşoğlu, 2014) ou encore en économie de la croissance (Mallick, 2012) sur l'élasticité de substitution ; Huang et Xie (2013) sur le lien entre population et croissance.

Afin de mieux comprendre les implications de la méthode, nous la présentons de manière plus détaillée dans ce qui suit en adoptant les notations de Lewbel (2012). Le

problème de variables omises peut être envisagé comme l'absence d'identification dans le système triangulaire suivant :

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \epsilon_1$$

$$Y_2 = X'\beta_2 + \epsilon_2$$

Où Y_1 est notre variable expliquée (le nombre de projets *FDI*), X est un vecteur de variables de contrôle exogène (l'ensemble de nos variables de contrôle décrites dans la partie précédente) et Y_2 est la variable endogène (notre variable d'intérêt relative aux métros). La présence d'endogénéité se traduit par l'existence d'une variable omise non observée que l'on retrouve dans les deux termes d'erreur :

$$\epsilon_1 = \alpha U + V_1$$
$$\epsilon_2 = \alpha_2 U + V_2,$$

où U est la variable omise qui peut influencer à la fois Y_1 et Y_2. V_1 et V_2 sont des erreurs idiosyncratiques. Lewbel propose de construire des instruments de manière interne au modèle. Ces instruments peuvent être construits du moment qu'existe de l'hétéroscédasticité d'échelle avec Z, un sous-ensemble de X, ou X lui-même, c'est-à-dire que $cov(Z, \epsilon_2^2) \neq 0$ et que les conditions sur les moments :

$$E(X\epsilon_1) = 0, E(X\epsilon_2) = 0 \text{ et } cov(Z, \epsilon_1\epsilon_2) = 0$$

soient valides. Sous ces conditions, qui ne sont pas plus contraignantes que les conditions classiques des estimations par IV si ce n'est par la présence d'hétéroscédasticité, des instruments adéquats peuvent être générés. Ils sont construits comme le produit des résidus $\hat{\epsilon}_2$ de la régression de la variable endogène sur toutes les variables exogènes avec les Z instruments centrés c'est-à-dire $(Z - \bar{Z})\hat{\epsilon}_2$. C'est un bon instrument, car l'hypothèse $cov(Z, \epsilon_1\epsilon_2) = 0$ garantit que l'instrument n'est pas corrélé avec ϵ_1 et la présence d'hétéroscédasticité, $cov(Z, \epsilon_2^2) \neq 0$, accroit la possibilité que l'instrument généré est corrélé avec Y_2 ϵ_2.

Concernant les hypothèses d'identification, la seule vraie différence est l'existence l'hétéroscédasticité dans ϵ_2 dont Lewbel déduit qu'elle doit être de la forme : $cov(Z, \epsilon_2^2) \neq 0$. Les résidus $\hat{\epsilon}_2$ sont par définition non corrélés avec les régresseurs exogènes utilisés pour les construire. L'existence d'hétéroscédasticité implique que leur produit avec les régresseurs exogènes centrés sera différent de zéro. Plus l'hétéroscédasticité sera prononcée plus l'instrument généré sera fort.

Pour construire nos instruments générés, et n'ayant pas d'a priori sur le niveau de l'hétéroscédasticité, nous prenons pour Z, l'ensemble des variables de X. Une fois que nous avons créé nos instruments à partir de la procédure ci-dessus, nous les utilisons dans une estimation du modèle de Poisson par IV. Quelles que soient les raisons de la présence d'hétéroscédasticité, la pertinence de nos instruments peut être testée formellement par des tests d'identification faible.

Nous venons de présenter l'ensemble de nos stratégies économétriques d'estimation. Dans la partie suivante, nous présentons les résultats des estimations en ventilant nos résultats pour la marge extensive et la marge intensive.

5 Résultats des estimations

Dans cette partie, nous présentons les principaux résultats issus des estimations des modèles de Poisson pour différentes mesures des systèmes de métro à la fois pour la marge extensive et pour la marge intensive. Pour l'ensemble des spécifications, la variable dépendante est le nombre de projets *FDI* nouveaux qui se sont implantés dans l'aire urbaine entre 2003 et 2014. L'unité d'observation est celle des aires urbaines associées aux différentes villes. L'échantillon de travail est constitué des villes avec les données de population des Nations Unies.

Les caractéristiques régionales (le pays, l'état ou encore la province) de l'aire urbaine représentent une part importante dans le choix de localisation des firmes multinationales. Ces caractéristiques regroupent par exemple : le niveau de formation, le poids administratif, le degré de protection de la propriété, la qualité générale du gouvernement ou encore des institutions, etc. Pour tenir compte de ces effets régionaux, nous mettons en œuvre deux stratégies. D'une part nous estimons des modèles avec des variables mesurées au niveau des pays (PIB, climat des affaires, niveaux de vie). D'autre part nous remplaçons ces variables mesurées au niveau des pays par des effets fixes régionaux. Les effets fixes régionaux sont les pays, sauf pour certaines villes des États-Unis, d'Inde, du Brésil et de la Chine où les effets fixes sont les états ou les provinces. Cette distinction intra pays intervient seulement dans les états ou provinces où il y a au moins deux villes. En dehors de ces états ou provinces, l'effet fixe reste celui du pays. Lors de nos estimations, nous utilisons soit directement un modèle de Poisson avec effets fixes soit des modèles avec des variables indicatrices

régionales (principalement pour les estimations par variables instrumentales avec effets fixes). Quelle que soit la méthode, cette stratégie à effets fixes impose qu'il y ait au moins deux villes dans l'échantillon pour chaque région et nous force donc à travailler avec un sous-ensemble de l'échantillon principal. Pour chacune de nos modélisations, nous présentons les résultats avec ces deux stratégies, variables de contrôle mesurées au niveau pays et effets fixes. Cette démarche permet de vérifier qu'il n'y a pas de biais d'échantillon en supprimant les pays avec une seule ville. Nous utilisons le même découpage en régions pour les estimations des écarts-types par grappe (« *cluster* ») quelle que soit la stratégie adoptée.

Dans la section suivante, nous présentons les résultats des estimations des modèles de Poisson pour la marge intensive et la marge extensive. La sous-partie suivante présente nos estimations par variables instrumentales pour tenir compte d'une possible endogénéité. Enfin, nous concluons cette partie avec des résultats complémentaires en estimant notre modèle grâce à l'utilisation de différents sous-échantillons comme celui des villes de plus de 1 million d'habitants, ou encore, les sous-échantillons excluant les deux grands pays en développement que sont l'Inde et la Chine.

5.1 Modèle de Poisson

5.1.1 Marge extensive

Nous commençons l'analyse de nos résultats par l'estimation de l'impact de la présence d'un métro sur l'attractivité de la ville. Deux spécifications alternatives sont mises en œuvre pour mesurer l'impact de la marge extensive (présence d'un métro) :

- Une spécification avec une seule variable indicatrice qui vaut 1 si la ville possède un métro en 2015 et 0 sinon.
- Une spécification avec deux variables indicatrices. La première vaut 1 en cas de présence d'un métro mis en service avant 2003 et zéro sinon et la seconde variable indicatrice vaut 1 en cas de présence d'un métro mis en service en 2003 ou après et 0 sinon.

Notre première spécification possède l'avantage de ne nécessiter l'estimation que d'un seul coefficient pour le paramètre d'intérêt, mais possède le défaut de donner le même poids à l'impact du métro pour les villes dont le système est présent sur

l'ensemble de la période et les villes où celui-ci est présent seulement sur une fraction de la période. Pour la seconde spécification, nous retenons deux variables indicatrices afin de traiter ce problème de différence de présence temporelle. Nous souhaitons pouvoir évaluer s'il existe l'effet varie pour des métros présents avant la période d'implantation des projets *FDI* (avant 2003) avec ceux mis en place pendant la période (entre 2003 et 2014). Afin de tester cette hypothèse, nous mettons en œuvre un test de Wald pour vérifier s'il existe des différences significatives entre les estimations des coefficients associés aux indicatrices métros nouveaux et métros anciens.

Nous commençons l'analyse de nos résultats par le Tableau 3 où sont présentées les estimations du modèle de Poisson pour différentes spécifications des variables de contrôle. La marge extensive est mesurée par la variable indicatrice de la présence d'un métro en 2015. Les spécifications (1) à (5) sont sans effets fixes. La spécification (1) comporte uniquement les contrôles pour les caractéristiques socioéconomiques des villes. Cette spécification est ensuite augmentée d'un certain nombre de contrôles détaillés qui sont progressivement ajoutés : la colonne (2) autorise la possibilité d'effets non linéaires de la taille des villes, la colonne (3) présente les résultats d'une estimation avec contrôles économiques au niveau des pays, la colonne (4) inclut des contrôles pour d'autres infrastructures de transport (aéroports, port, routes, tramways), enfin la spécification présentée en colonne (5) comporte aussi des contrôles relatifs à la géographie, au climat et des indicatrices continents. La spécification (6) comporte des effets fixes pays/états/province en plus de tous ces contrôles ce qui permet de contrôler aussi pour les facteurs constants dans le temps inobservés qui peuvent affecter l'attractivité d'un pays, d'une province ou d'une région, par exemple la qualité des institutions, ou la qualité de vie perçue.

Les estimations du coefficient associé à la présence d'un métro (marge extensive) sont toutes positives et significatives, quelle que soit la spécification retenue. Les résultats présentés dans la colonne (1) suggèrent que la présence d'un métro en 2015 a un effet positif et significatif sur le nombre de projets FDI s'implantant au niveau des villes entre 2003 et 2014. L'effet de la taille de la ville en termes de population et son poids économique (mesuré par les lumières la nuit) et le fait d'être une capitale sont également des facteurs importants de l'attractivité d'une ville pour les multinationales. L'ajout des variables de contrôle diminue sensiblement l'impact des indicatrices métro. Les variables mesurées au niveau du pays sont celles qui font diminuer l'impact de la

présence d'un système de métro le plus fortement. L'autre groupe de variables, celui des variables d'infrastructure, modifie la valeur du coefficient associé à la marge extensive, mais ceci, dans une moindre mesure. Les variables géographiques, météorologiques et les indicatrices continent ne modifient pas la valeur de l'impact du métro une fois contrôlé pour les autres variables.

Les spécifications (5) et (6) sont celles que nous privilégierons pour l'interprétation des résultats. Nous parlerons respectivement de la spécification complète sans effets fixes et de la spécification complète avec effets fixes. Il est à noter que si le nombre d'observations diffère entre les modèles (1) à (5) et le (6), cela s'explique par la suppression d'observations là où il n'y a qu'une seule ville dans la région[24]. Pour la mise en œuvre du modèle à effets fixes, nous sommes obligés de supprimer 60 villes, dont 8 avec un métro en 2015. Enfin, nous obtenons 189 régions différentes pour le modèle sans effets fixes et 131 pour le modèle à effets fixes. La qualité de notre ajustement est donnée de manière informative pour les modèles sans effets fixes par un pseudo R^2 calculé comme le carré du coefficient de corrélation entre les valeurs prédites et les valeurs observées.

Le modèle complet (5) sans effet fixe donne une estimation de la marge extensive de l'ordre de 80 %. Le modèle à effets fixes (6) est celui qui donne l'estimation de l'effet du métro la plus faible. Toutes choses égales par ailleurs, la présence d'un métro en 2015 augmente le nombre de projets FDI de $100 \times (\exp(0{,}386) - 1) = 47{,}11\ \%$. En d'autres termes, la présence d'un métro dans une ville permettrait d'obtenir 47,11 % d'implantations de projets *FDI* supplémentaires entre 2003 et 2014, que si cette même ville ne disposait pas de métro.

Dans le Tableau 4, nous séparons l'effet du métro en deux variables indicatrices identifiant les métros présents avant notre période de référence (2003-2014) et les systèmes mis en service pendant la période. Les résultats, qui sont présentés en suivant la même structure que dans le Tableau 3, sont cohérents avec ceux que nous avons obtenus avec une seule variable indicatrice. Pour l'ensemble des spécifications, les coefficients associés aux deux variables indicatrices sont statistiquement significatifs au seuil 5 % et voire de 1 % le plus souvent. Lorsque l'on s'intéresse aux résultats de l'estimation (6) du Tableau 4, nous constatons que l'impact des métros présents sur

[24] La procédure `xtpoisson` de Stata que nous avons utilisé supprime automatiquement ces observations.

l'ensemble de la période est plus fort que pour les métros mis en service pendant la période. Toutes choses égales par ailleurs, la présence d'un métro avant 2002 augmente le nombre de projets *FDI* de $100 \times (\exp(0{,}434) - 1) = 54{,}34\ \%$ et la présence d'un métro uniquement à partir de 2003 augmente le nombre de projets *FDI* de $100 \times (\exp(0{,}274) - 1) = 31{,}52\ \%$. Cependant, le test de Wald sur l'égalité des coefficients entre métro « ancien » et métro « nouveau » ne permet pas de rejeter l'hypothèse nulle d'égalité et cela pour toutes les spécifications. Il n'y a donc pas de différence statistiquement significative entre les deux coefficients.

De ces deux premiers ensembles de résultats, nous pouvons conclure qu'en moyenne sur notre période de référence, la présence d'un système de métro est un déterminant important de l'attractivité pour les multinationales. Ce résultat est robuste pour l'ensemble des spécifications. Toutes choses égales par ailleurs, la présence d'un métro en 2015 augmente de 47 % le nombre de projets FDI qui se sont implantés dans la ville entre 2003 et 2014. Néanmoins, la simple présence d'un métro peut refléter des réalités très diverses en termes de développement du réseau de métro. Après nous être intéressés à l'effet de la seule marge extensive (présence ou absence de métro), la partie suivante est consacrée à l'estimation de l'impact la taille du réseau de métro sur l'attractivité des villes pour les multinationales.

Tableau 3 : Résultats marge extensive avec un paramètre

	(1)	(2)	(3)	(4)	(5)	(6)
Metro en 2015	0,935***	0,925***	0,681***	0,576***	0,584***	0,386***
	(0,171)	(0,208)	(0,117)	(0,128)	(0,114)	(0,094)
Pop. 02	0,352*	-6,202	-8,357**	-8,627**	-1,227	-2,449
	(0,204)	(5,841)	(4,076)	(3,422)	(3,419)	(5,556)
Lumière 03	0,591***	-3,691**	-4,182***	-3,448**	-3,561**	-1,841
	(0,154)	(1,714)	(1,259)	(1,668)	(1,510)	(1,753)
Surface 00	-0,188	1,344	0,971	1,282	0,804	-0,125
	(0,252)	(1,811)	(1,281)	(1,440)	(1,475)	(0,989)
Villes <500km	0,003*	0,003*	0,007***	0,005***	0,003**	0,002
	(0,002)	(0,002)	(0,001)	(0,002)	(0,001)	(0,003)
Capitale	0,667***	0,655***	0,044	0,162	0,082	0,242**
	(0,222)	(0,232)	(0,164)	(0,155)	(0,136)	(0,108)
Puissance	Non	Oui	Oui	Oui	Oui	Oui
Pays	Non	Non	Oui	Oui	Oui	Non
Infrastructures	Non	Non	Non	Oui	Oui	Oui
Géographie	Non	Non	Non	Non	Oui	Oui
Météo	Non	Non	Non	Non	Oui	Oui
Continent	Non	Non	Non	Non	Oui	Non
Observations	1574	1574	1574	1574	1574	1514
Paramètres	7	15	20	32	47	36
Effets fixes	Non	Non	Non	Non	Non	Oui
R-carré	0,389	0,471	0,620	0,726	0,851	-
N. <2003 (1)	134	134	134	134	134	127
N. >=2003 (1)	48	48	48	48	48	47
N. grappes	189	189	189	189	189	131

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes de métro. Les effets fixes sont au niveau pays/états/provinces.
Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays/états/provinces.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.1.2 Marge intensive

Pour évaluer l'impact de la marge intensive, notre variable d'intérêt est maintenant la taille (longueur) du système sur l'attractivité d'une ville. Tout comme pour la marge extensive, plusieurs possibilités s'offrent à nous pour mesurer l'impact de la taille d'un système. La spécification de base que nous avons retenue contient deux variables : le logarithme plus 1[25] de la taille du réseau en 2002 pour les métros mis en service avant 2003 (la variable vaut donc 0 pour toutes les villes sans métro en 2002) et pour prendre en compte les systèmes mis en service en 2003 ou après, le logarithme plus 1 de la taille du métro lors de l'inauguration du système (cette variable vaut 0 pour toutes les autres

[25] Nous faisons dans l'annexe E, une analyse de robustesse de ce choix de modélisation. Notons que la construction de nos paramètres d'intérêt pour la taille du métro génère un grand nombre de zéros dans les variables pour les villes sans métro. Les estimations effectuées en utilisant le sous-échantillon des villes avec métro, et donc sans les 0, permettent de tenir compte de ce problème.

villes). Là encore, nous souhaitons différencier l'impact des métros présents avant notre période de référence de celui des métros mis en service durant la période. Nous ajoutons aussi 3 variables qui pour tenir compte de la dynamique passée de la croissance des systèmes pour notre période de référence et pour des périodes antérieures :

- La croissance du système entre 2002 et 2014,
- La croissance du système entre 1990 et 2002,
- La croissance du système entre 1978 et 1990.

Si la ville n'a pas de métro durant l'une de ces périodes, cette variable prend la valeur de 0. Les résultats des estimations que nous reportons dans le Tableau 5 nous indiquent que la taille des systèmes a un impact sur l'attractivité à la fois pour les métros anciens, mais aussi pour les systèmes plus récents. Tout comme pour la marge extensive, le test d'égalité de Wald ne permet pas de rejeter l'hypothèse nulle d'égalité des coefficients. Il n'y a donc pas de différence statistiquement significative entre les deux valeurs que nous obtenons. L'ajout des variables de contrôle diminue sensiblement la valeur de l'impact et c'est encore une fois dans la spécification complète avec effets fixes (colonne (6)) que cet impact est le plus faible.

Pour le modèle complet sans effets fixes, colonne (5), une augmentation de 1 % de la taille du métro se traduit par une implantation de projets supérieure de 0,16 % pour les métros les plus anciens et de 0,22 % pour les plus récents. Pour le modèle complet avec effets fixes, colonne (6), une augmentation de 1 % de la taille d'un métro augmenterait de 0,12 % le nombre d'implantations de projets *FDI* durant notre période de 12 ans. Pour les nouveaux métros (mis en service après 2003), une augmentation de 1 % de la taille du métro correspond à une augmentation de 0,10 % du nombre de *FDI*. À la lecture de ces résultats, ne pas tenir compte des effets régionaux (pays/états/provinces) biaiserait positivement nos résultats. Sans les effets régionaux, il semble que les métros capitalisent une certaine dynamique économique locale. C'est donc une fois de plus la spécification présentée en colonne (6), modèle complet avec effets fixes, que nous allons privilégier pour procéder aux exercices de quantification.

Tableau 4 : Résultats marge extensive avec deux paramètres

	(1)	(2)	(3)	(4)	(5)	(6)
Metro < 2003	0,894***	1,020***	0,747***	0,601***	0,527***	0,434***
	(0,220)	(0,229)	(0,173)	(0,155)	(0,118)	(0,102)
Métro >= 2003	1,033***	0,772**	0,563**	0,541**	0,698***	0,274**
	(0,367)	(0,355)	(0,259)	(0,228)	(0,180)	(0,120)
Pop. 02	0,346+	-5,873	-8,095**	-8,503**	-1,462	-2,438
	(0,219)	(5,804)	(4,066)	(3,485)	(3,456)	(5,589)
Lumière 03	0,591***	-3,795**	-4,226***	-3,472**	-3,454**	-1,966
	(0,155)	(1,901)	(1,324)	(1,716)	(1,527)	(1,745)
Surface 00	-0,173	1,795	1,241	1,376	0,697	-0,110
	(0,265)	(2,113)	(1,372)	(1,528)	(1,434)	(0,999)
Villes <500km	0,003*	0,003*	0,007***	0,005***	0,003**	0,002
	(0,002)	(0,002)	(0,001)	(0,002)	(0,001)	(0,003)
Capitale	0,692***	0,610**	0,010	0,152	0,110	0,223**
	(0,258)	(0,243)	(0,185)	(0,161)	(0,131)	(0,109)
Puissance	Non	Oui	Oui	Oui	Oui	Oui
Pays	Non	Non	Oui	Oui	Oui	Non
Infrastructures	Non	Non	Non	Oui	Oui	Oui
Géographie	Non	Non	Non	Non	Oui	Oui
Météo	Non	Non	Non	Non	Oui	Oui
Continent	Non	Non	Non	Non	Oui	Non
Observations	1574	1574	1574	1574	1574	1514
Paramètres	8	16	21	33	48	37
Effets fixes	Non	Non	Non	Non	Non	Oui
R-carré	0,388	0,472	0,622	0,727	0,849	.
N <2003 (1)	134	134	134	134	134	127
N >=2003 (1)	48	48	48	48	48	47
N. grappe	189	189	189	189	189	131
Test égalité (2)	0,0884	0,470	0,277	0,0472	0,864	1,818
P-valeur (2)	0,766	0,493	0,599	0,828	0,353	0,178

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation. (2) Test de Wald pour l'égalité des coefficients des métros.
Estimations effectuées pour la définition élargie des métros.
Les effets fixes sont au niveau pays/états/provinces. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays/états/provinces.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dans le Tableau 6, nous présentons des estimations du modèle complet sans et avec effets fixes avec deux mesures alternatives de la marge intensive. Dans la première (estimations 1 et 2), le paramètre d'intérêt est calculé comme le logarithme plus 1 de la moyenne annuelle de la taille du réseau entre 2003 et 2014. Pour la seconde spécification (estimations 3 et 4), nous utilisons le logarithme de la taille du métro en 2015, c'est-à-dire en fin de période. Ces deux spécifications alternatives ont l'avantage de mesurer le paramètre d'intérêt avec une seule variable au lieu de deux, le premier pondérant la taille du métro en fonction du nombre d'années de la période.

Les résultats que nous obtenons avec ces mesures alternatives sont du même ordre de grandeur que précédemment. Une nouvelle fois, les estimations des coefficients des paramètres d'intérêt dans la spécification avec effets fixes sont les plus faibles. Le

Tableau 7 résume le degré de similarité des différentes mesures en présentant la matrice de corrélation entre les variables pour les sous-échantillons des villes avec métro en 2015. Certes les niveaux de corrélations sont élevés, mais ne représentent pas pour autant avec exactitude les mêmes réalités. À titre d'exemple, la corrélation entre la taille des métros en 2002 et 2015 des villes qui ont un métro en fin de période n'est que de 56 %.

Le Tableau 8 présente les estimations lorsqu'on limite l'échantillon aux seules villes avec métro de la base élargie mis en service avant 2003[26]. Limiter notre échantillon aux villes avec métro avant la période permet de focaliser notre analyse spécifiquement sur la marge intensive. Nous constatons que les variables qui représentent le stock de métros restent significatives, et cela pour toutes les mesures. Toutes choses égales par ailleurs, avoir une taille de métro en 2002 qui soit 1 % plus grande augmente le nombre de projets de 0,3 %. Ce résultat est plus élevé que ceux que nous avons obtenus pour l'ensemble de notre l'échantillon. Le modèle complet et celui avec effets fixes donnent des résultats extrêmement similaires. Les variables de contrôle au niveau pays ainsi que les indicatrices « continent » sont équivalentes aux effets fixes pays. Néanmoins, pour ces spécifications, les effets fixes sont uniquement au niveau des pays et non au niveau pays province ou état, car il n'y a pas assez de villes avec un métro dans les états ou provinces. Ces spécifications peuvent donc souffrir d'un biais de variables omises provenant du manque d'informations sur la situation régionale de la ville.

[26] Les résultats sur la base restreinte sont présentés en annexe dans le Tableau 21.

Tableau 5 : Résultats marge intensive I

	(1)	(2)	(3)	(4)	(5)	(6)
Taille métro 02	0,254***	0,380***	0,284***	0,231***	0,164***	0,122***
	(0,070)	(0,081)	(0,077)	(0,054)	(0,037)	(0,030)
Taille nouveau	0,386***	0,285**	0,202***	0,162***	0,222***	0,096**
	(0,140)	(0,116)	(0,067)	(0,058)	(0,050)	(0,042)
C. métro 02-14	0,031*	0,004	0,016	0,024+	0,023*	0,009
	(0,017)	(0,013)	(0,015)	(0,016)	(0,012)	(0,011)
C. métro 90-02	0,031	-0,263+	-0,326**	-0,539***	-0,208+	0,160+
	(0,151)	(0,161)	(0,140)	(0,143)	(0,128)	(0,109)
C. métro 78-90	-0,122***	-0,043	-0,058*	0,015	-0,026	-0,020
	(0,042)	(0,057)	(0,030)	(0,032)	(0,029)	(0,043)
Pop. 02	0,300	-4,605	-6,522*	-6,788**	-0,944	-2,748
	(0,233)	(5,041)	(3,666)	(3,290)	(3,396)	(5,586)
Lumière 03	0,581***	-4,293**	-3,857**	-4,153***	-3,306*	-2,271
	(0,134)	(1,924)	(1,542)	(1,606)	(1,737)	(1,708)
Surface 00	-0,206	1,613	0,803	1,298	0,770	0,035
	(0,269)	(1,880)	(1,185)	(1,539)	(1,465)	(1,091)
Villes <500km	0,002	0,002	0,006***	0,004***	0,002*	0,001
	(0,002)	(0,002)	(0,001)	(0,001)	(0,001)	(0,003)
Capitale	0,662**	0,427*	-0,137	-0,035	0,051	0,162
	(0,266)	(0,250)	(0,193)	(0,142)	(0,139)	(0,115)
Puissance	Non	Oui	Oui	Oui	Oui	Oui
Pays	Non	Non	Oui	Oui	Oui	Non
Infrastructures	Non	Non	Non	Oui	Oui	Oui
Géographie	Non	Non	Non	Non	Oui	Oui
Météo	Non	Non	Non	Non	Oui	Oui
Continent	Non	Non	Non	Non	Oui	Non
Observations	1574	1574	1574	1574	1574	1514
Paramètres	11	19	24	36	51	40
Effets fixes	Non	Non	Non	Non	Non	Oui
R-carré	0,420	0,571	0,697	0,794	0,862	.
N <2003 (1)	134	134	134	134	134	127
N >=2003 (1)	48	48	48	48	48	47
N. Pays	189	189	189	189	189	131
Test égalité (2)	0,713	0,497	0,532	0,755	1,175	0,467
P-valeur (2)	0,398	0,481	0,466	0,385	0,278	0,495

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros. Les effets fixes sont au niveau pays/états/provinces. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays/états/provinces.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dans l'ensemble, les spécifications alternatives testées dans cette partie nous permettent de conclure que nos résultats ne sont pas dépendants de notre choix de mesure de la marge intensive. Les résultats du modèle complet avec effet fixes que nous préférons pour l'exercice de quantification indiquent qu'une augmentation de 1 % de la taille d'un métro accroît de 0,12 % le nombre d'implantations de projets FDI durant notre période d'analyse.

Tableau 6 : Résultats marge intensive II

	(1)	(2)	(3)	(4)
Moy. métro 03-14	0,167***	0,105***		
	(0,035)	(0,031)		
Taille métro 15*			0,181***	0,106***
			(0,033)	(0,028)
C. métro 02-14	0,001	-0,002	-0,003	-0,002
	(0,014)	(0,011)	(0,014)	(0,011)
C. métro 90-02	-0,258*	0,171+	-0,263**	0,173+
	(0,134)	(0,115)	(0,130)	(0,110)
C. métro 78-90	-0,016	-0,019	-0,016	-0,015
	(0,030)	(0,042)	(0,030)	(0,040)
Var éco.	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non
Observations	1574	1514	1574	1514
Paramètres	50	39	50	39
Effets-fixes	Non	Oui	Non	Oui
R-carré	0,864	.	0,870	.
N <2003 (1)	134	127	134	127
N >=2003 (1)	48	47	48	47
N. Pays	189	131	189	131

C. : croissance ;
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Les effets-fixes sont au niveau pays/états/provinces. Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tableau 7 : Matrice de corrélations entre les différentes mesures de la marge intensive du métro

	Taille 2015	Taille 2002	Taille moyenne	Croissance 2002-2014
Taille 2015	1,0000			
Taille 2002	0,5575	1,0000		
Taille moyenne	0,8689	0,8076	1,0000	
Croissance 2002-2014	0,2444	0,0467	0,1898	1,0000

Pour l'échantillon des villes avec métro de la base élargie (n=187)

Tableau 8 : Résultats marge intensive III, ville métro uniquement base élargie

	(1)	(2)	(3)	(4)	(5)	(6)
Taille métro 02	0,274**	0,295**				
	(0,110)	(0,115)				
Moy. métro 03-14			0,278**	0,278**		
			(0,111)	(0,116)		
Taille métro 15*					0,305***	0,301***
					(0,099)	(0,104)
C. métro 02-14	0,023	0,055**	0,001	0,031	-0,005	0,027
	(0,023)	(0,026)	(0,018)	(0,027)	(0,017)	(0,027)
C. métro 90-02	0,034	0,279*	0,048	0,285*	0,030	0,295*
	(0,114)	(0,167)	(0,115)	(0,172)	(0,109)	(0,168)
C. métro 78-90	-0,026	-0,058	-0,023	-0,052	-0,013	-0,040
	(0,040)	(0,048)	(0,040)	(0,047)	(0,037)	(0,045)
Var éco.	Oui	Oui	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non
Observations	134	106	134	106	134	106
Paramètres	48	38	48	38	48	38
Effets fixes Pays	Non	Oui	Non	Oui	Non	Oui
R-carré	0,942	.	0,943	.	0,944	.
N. Pays	48	20	48	20	48	20

C. : croissance ; (1) Nombre de systèmes dans l'estimation.
Les effets fixes sont au niveau pays.
Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays. $^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Afin d'interpréter les résultats de nos estimations de manière causale, notre principal problème est de gérer le problème d'endogénéité, c'est-à-dire qu'il n'y ait pas de corrélation entre nos paramètres d'intérêt sur les mesures du métro et le terme d'erreur. Les sources d'endogénéité peuvent être multiples. Nous pouvons avoir des erreurs de mesure, une causalité circulaire ou encore un biais de variables omises.

Nous sommes assez confiants quant aux éventuelles erreurs de mesure sur la longueur des systèmes de métro. D'une part, nous avons collecté nos données de la manière la plus minutieuse possible. D'autre part nous avons fait des tests avec une source de données complètement différente pour la mesure de la taille des systèmes : le site communautaire OpenStreet. Nous avons extrait de la base la taille des systèmes de transport identifiés comme des métros en 2015 par les contributeurs. L'utilisation de cette variable pour la marge intensive, construite de manière totalement différente de la nôtre, nous a donné des résultats similaires à ceux obtenus avec les données que

nous avions construites[27]. C'est pourquoi nous pensons que de telles erreurs ne peuvent pas biaiser nos estimations.

Le second problème générant potentiellement de l'endogénéité est celui de la causalité inverse : l'implantation de projets durant notre période d'analyse pourrait avoir un impact causal sur la taille ou la présence d'un système de métro durant la même période. Néanmoins, ce problème nous paraît peu probable dans le cadre de notre analyse pour deux raisons. D'une part, la décision et le temps de construction d'un système de métro s'inscrivent sur de nombreuses années. Il est donc très probable que l'effet du métro que nous capturons dans nos estimations, en particulier pour les métros existant avant 2003, soit la conséquence d'une décision prise bien avant le choix de localisation des multinationales et non l'inverse. Par ailleurs, notre période d'analyse des choix de localisation des multinationales étant relativement courte, il paraît également peu probable que ce soit l'implantation d'une multinationale qui ait entraîné la construction d'un métro dans le cas des métros apparus après 2003.

Le problème de variable omise constitue la source principale de l'endogénéité qui puisse biaiser nos estimations. Bien que nous ayons prêté une attention particulière à la construction de notre base de données, nous ne pouvons affirmer sans équivoque que nous sommes capables de tenir compte de l'ensemble des variables de contrôle pertinentes. Pour surmonter ce problème, nous nous sommes appuyés sur la méthode d'estimation par variables instrumentales. Nous en présentons les résultats dans la partie suivante.

[27] Voir le Tableau 22 en annexe F.

5.2 Estimations par variables instrumentales

Nos estimations par variables instrumentales font appel à la méthode de Lewbel (2012) qui permet de générer des instruments valides à partir des données du modèle et ainsi de contrôler le biais de variables omises. Pour les estimations avec effets fixes, nous utilisons maintenant un ensemble de variables indicatrices de la zone avec au minimum deux villes

Le Tableau 9 présente les résultats pour la marge extensive. Les estimations par variables instrumentales confirment non seulement les résultats obtenus précédemment, mais donnent également des élasticités plus élevées. La spécification avec des effets fixes donne une nouvelle fois les résultats les moins élevés. Toutes choses égales par ailleurs, la présence d'un métro en 2015 impacte le nombre de projets par $100 \times (\exp(0{,}524) - 1) = 68{,}88\,\%$. En moyenne, la présence d'un système de métro entraîne une implantation de projets *FDI* supérieure de 68,88 % par rapport à la même ville sans système de métro.

Si l'on s'intéresse à la spécification avec deux variables indicatrices, les résultats sont encore plus prononcés. La présence d'un métro avant 2003 impacte le nombre de projets de $100 \times (\exp(0{,}709) - 1) = 103{,}20\,\%$. En moyenne, la présence d'un système avant notre période de référence double le nombre d'implantations de projets. L'impact des métros récents, mis en service après 2002, est extrêmement différent entre le modèle avec effets fixes et celui avec les variables pays. On ne peut rejeter l'hypothèse d'égalité des coefficients pour le modèle sans effets fixes. En revanche, pour le modèle avec effets fixes, il existe une différence statistiquement significative entre les nouveaux métros et les métros anciens.

Pour la marge intensive (Tableau 10), les valeurs des estimations par variables instrumentales sont aussi légèrement plus grandes que les estimations du modèle de Poisson classique. Encore une fois, les estimations avec effets fixes sont systématiquement plus faibles que les estimations sans effets fixes avec les variables de contrôle au niveau du pays. Nous pouvons aussi constater que c'est surtout pour la variable mesurant la taille du système en 2012 que les estimations par variables instrumentales sont différentes des estimations classiques. En effet, pour une estimation classique l'élasticité est de 0,122 alors qu'elle est de 0,195 pour l'estimation par variables instrumentales.

Tableau 9 : Estimations par variables instrumentales marge extensive

	(1)	(2)	(3)	(4)
Métro en 2015[(i)]	0,624***	0,524***		
	(0,132)	(0,124)		
Métro < 2003[(i)]			0,741***	0,709***
			(0,135)	(0,141)
Métro >= 2003[(i)]			0,820***	0,285**
			(0,218)	(0,142)
Pop. 02	0,721	6,925	2,280	6,191
	(5,855)	(6,798)	(5,854)	(6,458)
Lumière 03	-5,742+	11,743	-7,298**	9,562
	(3,850)	(10,590)	(3,062)	(10,087)
Surface 00	-3,043	-5,056	-0,028	-4,288
	(3,336)	(3,715)	(4,976)	(3,573)
Villes <500km	0,007***	0,001	0,006**	0,001
	(0,002)	(0,003)	(0,002)	(0,003)
Capitale	0,271	0,277**	0,124	0,261*
	(0,191)	(0,135)	(0,207)	(0,134)
Puissance	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non
Indicatrices	Non	Oui	Non	Oui
Observations	1574	1514	1574	1514
Paramètres	47	168	48	169
Métros <2003	134	127	134	127
Métros >=2003	48	47	48	47
N. Pays	189	131	189	131
Test égalité (2)			0,152	12,95
P-valeur (2)			0,696	0,000321

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros.
Les indicatrices sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[(i)] Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir de l'ensemble des autres variables de contrôle à l'exception des indicatrices pays.

Le Tableau 11 regroupe les résultats des estimations par variables instrumentales pour le sous-échantillon des villes avec un métro en 2002. Les estimations obtenues sont bien supérieures aux estimations obtenues sans les variables instrumentales. Pour la taille du métro en 2002 et pour le modèle avec effets fixes, une augmentation de 1 % de la taille du métro correspond à une augmentation de 0,75 % du nombre d'implantations de projets. Pour ces estimations sur le sous-ensemble des villes avec métro, nous pouvons aussi constater un impact plus significatif des variables de croissance. Cette particularité n'était jamais présente pour les autres spécifications.

Tableau 10 : Estimations par variables instrumentales marge intensive

	(1)	(2)	(3)	(4)	(5)	(6)
Taille métro 02[(i)]	0,242***	0,195***				
	(0,043)	(0,041)				
Taille nouveau[(i)]	0,218***	0,107***				
	(0,076)	(0,039)				
Moy. métro 03-14[(i)]			0,172***	0,125***		
			(0,049)	(0,043)		
Taille métro 15[(i)]					0,207***	0,119***
					(0,045)	(0,041)
C. métro 02-14[(i)]	0,001	0,023	-0,027*	0,028	-0,024*	0,028
	(0,014)	(0,016)	(0,015)	(0,022)	(0,013)	(0,022)
C. métro 90-02	-0,248+	0,140	-0,185	0,214	-0,168	0,231+
	(0,167)	(0,143)	(0,162)	(0,156)	(0,149)	(0,160)
C. métro 78-90	0,046	-0,063	0,043	-0,082+	0,028	-0,080+
	(0,053)	(0,048)	(0,049)	(0,050)	(0,042)	(0,050)
Eco	Oui	Oui	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui	Oui	Oui
Pays	Non	Non	Non	Non	Non	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non
Indicatrices	Non	Oui	Non	Oui	Non	Oui
Observations	1574	1514	1574	1514	1574	1574
Paramètres	51	172	50	171	50	229
N. métros	134	127	134	127	134	134
N. Nouv. métros	48	47	48	47	48	48
N. Pays	189	131	189	131	189	189
Wald	0,119	5,288				
P-valeur	0,731	0,0215				

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros.
Les indicatrices sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[(i)] Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir de l'ensemble des autres variables de contrôle à l'exception des indicatrices pays.

Cependant, ces résultats restent délicats à interpréter, car comme cela est expliqué précédemment, nous ne sommes pas en mesure de contrôler précisément les effets régionaux.

Les résultats que nous obtenons avec les estimations par variables instrumentales confirment dans l'ensemble nos résultats précédents. La présence et la taille d'un système de transport de type métro sont bien une source de croissance du nombre de *FDI* que nous interprétons comme un indicateur d'une hausse de l'attractivité d'une ville. En ce qui concerne la marge extensive, la présence d'un métro peut aller jusqu'à doubler le nombre de projets *FDI* se localisant dans cette ville par rapport à une ville.

Tableau 11 : Estimations IV de la marge intensive pour les villes avec métro uniquement, base large

	(1)	(2)	(3)	(4)	(5)	(6)
Taille métro 02[i]	0,750*	0,808+				
	(0,406)	(0,500)				
Moy. métro 03-14[i]			0,683***	0,418***		
			(0,132)	(0,142)		
Taille métro 15[i]					0,663***	0,504***
					(0,128)	(0,146)
C. métro 02-14[i]	0,031	0,048+	0,048**	0,050**	0,029+	0,047**
	(0,023)	(0,031)	(0,019)	(0,020)	(0,019)	(0,022)
C. métro 90-02	0,434**	0,441***	0,434**	0,590***	0,333**	0,589***
	(0,175)	(0,158)	(0,170)	(0,162)	(0,170)	(0,141)
C. métro 78-90	-0,143***	-0,096**	-0,148***	-0,050	-0,110***	-0,041
	(0,042)	(0,042)	(0,032)	(0,040)	(0,032)	(0,039)
Eco	Oui	Oui	Oui	Oui	Oui	Oui
Pays	Non	Non	Non	Non	Non	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non
Indicatrices	Non	Oui	Non	Oui	Non	Oui
Observations	134	106	134	106	134	106
Paramètres	43	66	42	65	42	65
N. Pays	48	20	48	20	48	20

C. : croissance
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros.
Les effets-fixes sont au niveau pays.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[i] Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir de l'ensemble des autres variables de contrôle à l'exception des indicatrices pays.

sans métro. En ce qui concerne la marge intensive, les élasticités entre la taille du système et le nombre de projets s'établissent dans une fourchette allant de 0,1 à 0,2 pour les spécifications que nous privilégions.

Dans la partie suivante, nous présentons des résultats complémentaires obtenus sur différents sous-échantillons.

5.3 Résultats complémentaires

Dans les sous-parties précédentes, nous avons établi que si l'on se restreignait aux seules villes avec métro, l'estimation de la marge intensive reste positive et significative. Il est intéressant d'analyser la variation des estimations de nos élasticités en fonction d'autres sous-échantillons.

Pour notre premier sous-échantillon, nous ne considérons que les villes de plus d'un million d'habitants en 2003 selon les données des Nations Unies. Le Tableau 12 présente les résultats pour la marge extensive et la marge intensive. Les résultats obtenus sont tous plus élevés que leurs équivalents estimés sur l'ensemble de l'échantillon. En particulier pour la présence d'un métro dans le modèle de Poisson estimé par variables instrumentales avec effets fixes région, le coefficient est de 1,29 pour un impact de 263 %. Les résultats sur la marge intensive restent plus contenus, variant de 0,18 % à 0,26 % pour une augmentation de 1 % de la taille du métro en 2002. Compte tenu du faible nombre de nouveaux métros dans les villes de plus d'un million d'habitants (40 pour les estimations sans effets fixes et 29 pour les estimations avec effets fixes), les résultats sont beaucoup moins robustes et sont parfois non significatifs. D'ailleurs, nous rejetons bien plus souvent l'hypothèse nulle d'égalité des coefficients entre les métros anciens et récents.

Il nous est impossible de mener la même analyse pour les villes de moins de 1 million d'habitants, car s'il y a de nombreuses villes de cette taille attirant des projets FDI, peu d'entre elles possèdent un métro. Dans l'ensemble, la comparaison de ces résultats avec ceux obtenus à partir de notre échantillon global suggère que l'absence d'une infrastructure lourde comme le métro pénalise davantage les grandes villes que les petites.

Dans le Tableau 13, nous présentons les résultats des estimations sans la Chine et l'Inde, deux pays qui ont tendance à être surreprésentés dans l'échantillon en raison de leur démographie. En effet, près d'un tiers des villes de notre échantillon de travail appartient à la Chine ou à l'Inde et les supprimer diminue d'autant la taille de l'échantillon. La mise en œuvre de ces estimations montre néanmoins que nos résultats ne dépendent pas des villes de Chine et d'Inde, car les coefficients associés au métro sont tous positifs et significatifs et du même ordre de grandeur que nos résultats précédents

Tableau 12 : Estimations pour les villes de plus d'un million d'habitants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Poisson	Poisson-FE	IV Poisson	IV Poisson FE	Poisson	Poisson FE	IV Poisson	IV Poisson FE
Métro < 03(i)	0,633***	0,794***	0,864***	1,290***				
	(0,151)	(0,138)	(0,176)	(0,186)				
Métro>= 03(i)	0,771***	0,215	1,156***	0,073				
	(0,208)	(0,161)	(0,244)	(0,159)				
Taille mét. 02(i)					0,179***	0,223***	0,261***	0,259***
					(0,043)	(0,037)	(0,046)	(0,051)
Taille >=03(i)					0,233***	0,109**	0,301***	0,031
					(0,056)	(0,054)	(0,075)	(0,058)
C. mét. 02-14(i)					0,021*	0,022	0,017	0,063***
					(0,012)	(0,020)	(0,012)	(0,017)
C. métro 90-02					-0,249+	0,075	-0,303*	0,310**
					(0,154)	(0,105)	(0,156)	(0,138)
C. métro 78-90					-0,022	0,022	0,028	0,151*
					(0,035)	(0,037)	(0,044)	(0,083)
Pop. 02	-13,335	15,764	-17,410	47,316***	-19,375	16,622	-36,406	1,005***
	(17,606)	(17,871)	(21,260)	(9,036)	(17,356)	(18,107)	(30,709)	(0,164)
Lumière 03	-11,18***	-2,89	-38,37***	9,85*	-10,21**	-4,07	-39,98***	0,50***
	(3,727)	(6,442)	(8,939)	(5,748)	(4,160)	(7,128)	(10,914)	(0,177)
Surface 00	4,887	-1,782	11,261*	-8,110+	5,445	-1,320	20,069***	0,307
	(4,807)	(3,106)	(6,625)	(5,364)	(5,117)	(3,529)	(6,501)	(0,233)
Villes <500km	0,003**	0,001	0,004*	0,004	0,002+	0,000	0,000	-0,004
	(0,001)	(0,003)	(0,002)	(0,003)	(0,001)	(0,004)	(0,002)	(0,004)
Capitale	0,203	0,222+	0,276	0,446*	0,151	0,088	0,262	0,109
	(0,161)	(0,151)	(0,223)	(0,263)	(0,170)	(0,152)	(0,210)	(0,217)
Puissance	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non	Oui	Non
Indicatrice pays	Non	Non	Non	Oui	Non	Non	Non	Oui
Observations	375	298	375	298	375	298	375	298
Effets fixes	Non	Oui	Non	Oui	Non	Oui	Non	Oui
Paramètres	48	37	48	109	51	40	51	104
Métros<2003[1]	104	82	104	82	104	82	104	82
Métros>=2003[1]	40	29	40	29	40	29	40	29
N. Pays	148	71	148	71	148	71	148	71
Wald[2]	0,417	14,08	1,899	45,70	0,754	4,983	0,300	20,57
P-valeur[2]	0,519	0,000	0,168	0,000	0,385	0,026	0,584	0,000

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
Estimation 1 à 4 les paramètres d'intérêt sont la taille du métro en 2002 et la taille du métro lors de la mise en service pour un système construit après 2002. Pour les estimations 5 à 8, c'est la taille du métro en 2015. Les estimations 3,4,7 et 8 sont estimées par variables instrumentales avec les instruments générés par la méthode de Lewbel.
(1) Nombre de systèmes dans l'estimation. (2) Test de Wald pour l'égalité des coefficients métro.
Estimations effectuées pour la définition élargie des métros.
Les effets fixes sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays. $^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
(i) Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir des autres variables de contrôle à l'exception des indicatrices pays.

Tableau 13 : Estimations sans les villes de Chine et d'Inde

	(1) Poisson	(2) Poisson-FE	(3) IV Poisson	(4) IV Poisson FE	(5) Poisson	(6) Poisson FE	(7) IV Poisson	(8) IV Poisson FE
Métro<03[(i)]	0,329***	0,317***	0,591***	0,681***				
	(0,126)	(0,113)	(0,166)	(0,116)				
Métro>=03[(i)]	0,208	-0,081	0,086	0,120				
	(0,162)	(0,160)	(0,279)	(0,200)				
Taille mét. 02[(i)]					0,132***	0,106***	0,154***	0,154***
					(0,035)	(0,032)	(0,034)	(0,039)
Taille nouveau[(i)]					0,080+	-0,025	-0,073	-0,047
					(0,055)	(0,050)	(0,117)	(0,071)
C. métro 02-14[(i)]					-0,183**	-0,319*	-0,115	-0,104
					(0,088)	(0,169)	(0,082)	(0,159)
C. métro 90-02					0,022	0,248**	0,158	0,592***
					(0,109)	(0,104)	(0,132)	(0,174)
C. métro 78-90					-0,043*	-0,029	-0,08***	-0,085*
					(0,025)	(0,042)	(0,030)	(0,047)
Pop. 02	1,893	1,392	2,157	10,778*	2,605	1,734	6,074+	0,309+
	(3,823)	(5,846)	(4,107)	(6,213)	(3,692)	(5,374)	(4,120)	(0,204)
Lumière 03	-2,208*	-1,509	-2,442	0,029	-2,420*	-1,958+	-3,977**	0,632**
	(1,322)	(1,478)	(2,392)	(9,149)	(1,322)	(1,295)	(1,951)	(0,255)
Surface 00	-1,017	-1,116	-5,781**	-1,404	-0,820	-1,005	-2,952	0,378
	(1,367)	(1,087)	(2,673)	(3,231)	(1,372)	(1,184)	(3,625)	(0,320)
Villes <500km	0,003	0,006**	0,004	0,004	0,002	0,004	0,004	-0,002
	(0,002)	(0,003)	(0,003)	(0,004)	(0,002)	(0,004)	(0,003)	(0,005)
Capitale	0,183	0,205*	0,060	0,136	0,144	0,208*	-0,088	0,262+
	(0,131)	(0,107)	(0,137)	(0,141)	(0,135)	(0,107)	(0,157)	(0,166)
Puissance	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non	Oui	Non
Indicatrice	Non	Non	Non	Oui	Non	Non	Non	Oui
Observations	1045	985	1045	985	1045	985	1045	985
Effets fixes	Non	Oui	Non	Oui	Non	Oui	Non	Oui
Paramètres	48	37	48	133	51	40	51	128
Métros <03[(1)]	124	117	124	117	124	117	124	117
Métros >=03[(1)]	28	27	28	27	28	27	28	27
N. Pays	153	95	153	95	153	95	153	95
Wald[(2)]	0,409	8,714	3,671	10,87	0,683	11,02	3,975	7,919
P-valeur[(2)]	0,522	0,003	0,055	0,001	0,409	0,001	0,046	0,005

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
Estimation 1 à 4 les paramètres d'intérêt sont la taille du métro en 2002 et la taille du métro lors de la mise en service pour un système construit après 2002. Pour les estimations 5 à 8, c'est la taille du métro en 2015. Les estimations 3,4,7 et 8 sont estimées par variables instrumentales avec les instruments générés par la méthode de Lewbel.
[(1)] Nombre de systèmes dans l'estimation.
[(2)] Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros.
Les effets fixes sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[(i)] Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir de l'ensemble des autres variables de contrôle à l'exception des indicatrices pays.

5.4 Tests de robustesse

En plus des résultats complémentaires, nous avons mis en œuvre un certain nombre de tests de robustesse afin d'analyser l'impact de nos choix de modélisation sur les valeurs et la significativité des estimations. Le détail de ces tests se trouve dans l'annexe E de ce document.

Le premier de ces tests de robustesse consiste à analyser l'impact que peut avoir la transformation de notre variable dépendante par le logarithme plus 1. Nous testons donc l'effet de cette transformation contre 3 alternatives concurrentes et aussi en faisant varier la constante ajoutée. Nous constatons à la lecture du Tableau 17 ;Tableau 18 et Tableau 19 que le choix de la transformation n'a que très peu d'influence sur la significativité de la marge intensive.

Pour le second test, nous voulons savoir, si nos estimations sont sensibles au choix du modèle. Nous avons choisi d'estimer un modèle de Poisson, car celui-ci a la particularité de bien prendre en compte les variables de comptage avec une certaine fraction de zéros. Pour autant, d'autres possibilités existent. En particulier, le modèle binomial négatif est proposé comme une alternative du modèle de Poisson pour mieux modéliser le problème de *surdispersion*. Il y a *surdispersion* quand la variance pour une variable modélisée par une loi de Poisson est supérieure à la moyenne. Les résultats sont présentés dans le Tableau 20. Encore une fois, nos estimations des variables de métro restent significatives et très proches de celles obtenues avec le modèle de Poisson.

Enfin, nous avons aussi mis en œuvre un ensemble d'estimations avec des sources de données différentes pour les routes, les aéroports et les ports sans que cela ne change nos résultats de manière significative.

6 Conclusion

La Société du Grand Paris a chargé le LIEPP-Sciences Po d'évaluer l'impact potentiel que pourrait avoir la construction du Grand Paris Express sur l'attractivité de la région Île-de-France pour les entreprises multinationales. Pour répondre à cette question, nous avons construit une base de données inédite au niveau des villes et ce, à l'échelle mondiale, pour évaluer le lien entre système de métro et attractivité d'une ville pour la localisation des multinationales. Cette base unique rassemble d'importantes informations sur les caractéristiques socioéconomiques et sur les systèmes de transport au niveau des villes. Nous avons montré dans ce rapport qu'il existe un lien significatif entre l'attractivité d'une ville (mesurée par le nombre de projets *FDI*) et la présence d'un système de transport de type métro.

Les résultats suggèrent que les villes qui ont un métro attirent davantage d'investissements directs étrangers, qu'il s'agisse de la marge extensive (présence d'un métro) ou de la marge intensive (longueur du métro). Notre analyse suggère que cette relation est causale.

Nos résultats sont robustes à un grand nombre de spécifications et ne sont pas tributaires de nos choix de mesure des systèmes de métro. Les résultats ne semblent pas non plus tirés par la présence des villes de Chine ou d'Inde (particulièrement représentées en raison de leur démographie), ni tributaires de notre choix de modélisation.

Le Grand Paris Express prévoit un doublement de la taille du métro, soit une augmentation de 100 % de la taille de l'infrastructure de métro. En appliquant cette valeur aux résultats du modèle avec effets fixes que nous privilégions, c'est-à-dire une élasticité comprise entre 0,1 et 0,2, nous obtenons un impact pour le GPE qui varie entre 10 % et 20 % de projets en plus. Paris ayant attiré 1 310 projets entre 2003 et 2014, de telles élasticités correspondent à un intervalle de 130 à 262 projets supplémentaires. De plus, nous constatons dans les données *FdiMarkets* que les *FDI* qui se sont implantés à Paris génèrent 52 emplois en moyenne. Le GPE pourrait donc générer, selon ces valeurs et nos estimations, entre 6 800 et 14 000 emplois directs supplémentaires issus de l'implantation de firmes multinationales.

Ce rapport apporte une importante contribution à la littérature pour deux raisons. Tout d'abord, contrairement à la plupart des études existantes sur les choix de

localisation des multinationales, notre unité géographique d'analyse est la ville. Pour parvenir à cette analyse, nous avons constitué une base de données de couverture mondiale qui comprend un grand nombre de caractéristiques des aires urbaines. Nous avons donc été en mesure de tester la pertinence des déterminants traditionnellement identifiés au niveau des pays et/ou des régions et de montrer que pour des choix de localisation au niveau des villes, les déterminants locaux ont un impact significatif. Par ailleurs, notre analyse contribue à la littérature relative à l'impact des infrastructures de transport sur l'attractivité des villes en général en montrant que l'effet d'une politique d'aménagement aussi importante que celle du développement du réseau de métro peut contribuer à attirer les entreprises multinationales.

Nos résultats soulèvent néanmoins un certain nombre de questions qui vont au-delà de ce rapport, mais qui pourraient être importantes pour les décideurs publics. Si les caractéristiques propres d'une ville sont importantes pour son attractivité, une importante littérature a montré que les déterminants au niveau des pays peuvent compter et que les choix de localisation peuvent suivre une structure hiérarchisée (par exemple un pays puis une région à l'intérieur d'un pays). Dès lors, la question se pose de savoir si la concurrence pour l'attractivité des multinationales entre territoires se fait davantage au niveau des villes ou des pays.

Si notre analyse permet de répondre à la question de l'existence d'un lien causal entre infrastructure de type métro et attractivité, nous sommes toutefois dans l'incapacité d'en déterminer précisément le(s) mécanisme(s) sous-jacent(s). D'abord, la présence d'un système de transport de type métro pourrait améliorer le marché du travail en améliorant les appariements entre emplois et travailleurs et en favorisant la concentration des activités. Il se pourrait également que le métro constitue simplement un « effet signal », et que la présence (ou la construction) d'un système de métro soit associée à un signe de modernité qui contribue à attirer les multinationales. Si la question des raisons de cet effet positif du métro pour les multinationales va au-delà de ce rapport, c'est pourtant une question fondamentale pour quantifier les effets indirects de la construction de telles infrastructures.

En définitive, bien que le nombre de villes disposant d'un système de métro ne soit pas si élevé, ce nombre est en augmentation depuis quelques années et cette tendance semble ne pas ralentir. Les villes en font d'ailleurs un argument de communication, à la fois pour justifier ces investissements, mais aussi pour attirer de nouveaux

investissements. Dans ce contexte, ce rapport a permis de mieux appréhender l'impact attendu de la construction ou de l'extension d'un système de métro sur les investissements directs étrangers et de préciser les effets que l'on peut espérer du Grand Paris Express sur l'attractivité pour les multinationales.

Références

Angel, Shlomo, Jason Parent, Daniel Civco, Alexander Blei, and David Potere. 2010. "Urban Land Cover Estimates and Projections for All Countries, 2000-2050." *Lincoln Institute of Land Policy*.

Arcand, Jean Louis; Enrico Berkes et Ugo Panizza. 2015. "Too Much Finance?" *Journal of Economic Growth*, 20(2), 105-48.

Azemar, Celine et Rodolphe Desbordes. 2009. "Public Governance, Health and Foreign Direct Investment in Sub-Saharan Africa." *Journal of African Economies*.

Basile, Roberto; Davide Castellani et Antonello Zanfei. 2008. "Location Choices of Multinational Firms in Europe: The Role of Eu Cohesion Policy." *Journal of International Economics*, 74(2), 328-40.

Bénassy-Quéré, Agnès; Lionel Fontagné et Amina Lahrèche-Révil. 2003. *Tax Competition and Foreign Direct Investment*. CEPII Paris.

Caliendo, Marco; Wang-Sheng Lee et Robert Mahlstedt. 2017. "The Gender Wage Gap and the Role of Reservation Wages: New Evidence for Unemployed Workers." *Journal of Economic Behavior & Organization*, 136, 161-73.

Cameron, A Colin et Pravin K Trivedi. 2013. *Regression Analysis of Count Data*. Cambridge university press.

Chowdhury, Mohammad Tarequl H; Prasad Sankar Bhattacharya; Debdulal Mallick et Mehmet Ali Ulubaşoğlu. 2014. "An Empirical Inquiry into the Role of Sectoral Diversification in Exchange Rate Regime Choice." *European Economic Review*, 67, 210-27.

Coughlin, Cletus C et Eran Segev. 2000. "Location Determinants of New Foreign Owned Manufacturing Plants." *Journal of regional Science*, 40(2), 323-51.

Crozet, Matthieu; Thierry Mayer et Jean-Louis Mucchielli. 2004. "How Do Firms Agglomerate? A Study of Fdi in France." *Regional Science and Urban Economics*, 34(1), 27-54.

Davies, Ronald B et Rodolphe Desbordes. 2015. "Greenfield Fdi and Skill Upgrading: A Polarized Issue." *Canadian Journal of Economics/Revue canadienne d'économique*, 48(1), 207-44.

Davis, James C et J Vernon Henderson. 2008. "The Agglomeration of Headquarters." *Regional Science and Urban Economics*, 38(5), 445-60.

Defever, Fabrice. 2006. "Functional Fragmentation and the Location of Multinational Firms in the Enlarged Europe." *Regional Science and Urban Economics*, 36(5), 658-77.

Devereux, Michael P et Rachel Griffith. 1998. "Taxes and the Location of Production: Evidence from a Panel of Us Multinationals." *Journal of public Economics*, 68(3), 335-67.

Duranton, Gilles; Laurent Gobillon et Henry G Overman. 2011. "Assessing the Effects of Local Taxation Using Microgeographic Data." *The Economic Journal*, 121(555), 1017-46.

Duranton, Gilles et Diego Puga. 2005. "From Sectoral to Functional Urban Specialisation." *Journal of urban Economics*, 57(2), 343-70.

Duranton, Gilles et Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." *Handbook of regional and urban economics*, 4, 2063-117.

Elvidge, Christopher D; Kimberley E Baugh; Eric A Kihn; Herbert W Kroehl; Ethan R Davis et Chris W Davis. 1997. "Relation between Satellite Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing*, 18(6), 1373-79.

Emran, M Shahe et Zhaoyang Hou. 2013. "Access to Markets and Rural Poverty: Evidence from Household Consumption in China." *Review of economics and statistics*, 95(2), 682-97.

Fujita, Masahisa; Paul R Krugman; Anthony J Venables et Massahisa Fujita. 1999. *The Spatial Economy: Cities, Regions and International Trade*. Wiley Online Library.

Fujita, Masahisa et Jacques-François Thisse. 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge university press.

Gaudry, Marc et Emile Quinet. 2013. "Track Wear-and-Tear Cost by Traffic Class: Functional Form, Zero Output Levels and Marginal Cost Pricing Recovery on the French Rail Network."

Gibbons, Stephen; Teemu Lyytikäinen; Henry G Overman et Rosa Sanchis-Guarner. 2016. "New Road Infrastructure: The Effects on Firms."

Head, C Keith; John C Ries et Deborah L Swenson. 1999. "Attracting Foreign Manufacturing: Investment Promotion and Agglomeration." *Regional Science and Urban Economics*, 29(2), 197-218.

Head, Keith et Thierry Mayer. 2004. "Market Potential and the Location of Japanese Investment in the European Union." *Review of economics and statistics*, 86(4), 959-72.

Head, Keith et Thierry Mayer. 2006. "Regional Wage and Employment Responses to Market Potential in the Eu." *Regional Science and Urban Economics*, 36(5), 573-94.

Holl, Adelheid. 2004a. "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain." *Regional Science and Urban Economics*, 34(3), 341-63.

Holl, Adelheid. 2004b. "Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal." *Journal of regional Science*, 44(4), 693-712.

Hollard, Guillaume et Omar Sene. 2016. "Social Capital and Access to Primary Health Care in Developing Countries: Evidence from Sub-Saharan Africa." *Journal of health economics*, 45, 1-11.

Huang, Tai-Hsin et Zixiong Xie. 2013. "Population and Economic Growth: A Simultaneous Equation Perspective." *Applied Economics*, 45(27), 3820-26.

Khadaroo, Jameel et Boopen Seetanah. 2009. "The Role of Transport Infrastructure in Fdi: Evidence from Africa Using Gmm Estimates." *Journal of Transport Economics and Policy (JTEP)*, 43(3), 365-84.

Krugman, Paul. 1991a. "Increasing Returns and Economic Geography." *Journal of political economy*, 99(3), 483-99.

Krugman, Paul R. 1991b. *Geography and Trade*. MIT press.

Lewbel, Arthur. 1997. "Constructing Instruments for Regressions with Measurement Error When No Additional Data Are Available, with an Application to Patents and R&D." *Econometrica*, 65(5), 1201-13.

Lewbel, Arthur. 2012. "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business & Economic Statistics*, 30(1), 67-80.

Liu, Xiaming; Yingqi Wei; Chengang Wang et Bo Liu. 2006. "Foreign Direct Investment, Transactional Linkages and Productivity Spillovers in Chinese Manufacturing." *Journal of International Business Studies*.

Mallick, Debdulal. 2012. "The Role of the Elasticity of Substitution in Economic Growth: A Cross-Country Investigation." *Labour Economics*, 19(5), 682-94.

Mayer, Thierry; Florian Mayneris et Loriane Py. 2015. "The Impact of Urban Enterprise Zones on Establishment Location Decisions and Labor Market Outcomes: Evidence from France." *Journal of Economic Geography*, lbv035.

Mayer, Thierry et Corentin Trévien. 2015. "*The Impacts of Urban Public Transportation. Evidence from the Paris Metropolitan Area*," Working Paper g2015/03, INSEE.

Mellander, Charlotta; José Lobo; Kevin Stolarick et Zara Matheson. 2015. "Night-Time Light Data: A Good Proxy Measure for Economic Activity?" *PLoS ONE*, 10(10), e0139779.

Nordhaus, William; Quazi Azam; David Corderi; Kyle Hood; Nadejda Makarova Victor; Mukhtar Mohammed; Alexandra Miltner et Jyldyz Weiss. 2006. "The G-Econ Database on Gridded Output: Methods and Data." *Yale University, New Haven*.

Ottaviano, Gianmarco et Jacques-François Thisse. 2004. "Agglomeration and Economic Geography." *Handbook of regional and urban economics*, 4, 2563-608.

Potere, David; Annemarie Schneider; Shlomo Angel et Daniel L Civco. 2009. "Mapping Urban Areas on a Global Scale: Which of the Eight Maps Now Available Is More Accurate?" *International Journal of Remote Sensing*, 30(24), 6531-58.

Py, Loriane et Fabrice Hatem. 2009. "Internationalisation Et Localisation Des Services: Une Analyse Sectorielle Et Fonctionnelle Appliquée Aux Firmes Multinationales En Europe." *Economie et statistique*, 426(1), 67-95.

Rathelot, Roland et Patrick Sillard. 2008. "The Importance of Local Corporate Taxes in Business Location Decisions: Evidence from French Micro Data." *The Economic Journal*, 118(527), 499-514.

Sanderson, Eleanor et Frank Windmeijer. 2016. "A Weak Instrument F-Test in Linear Iv Models with Multiple Endogenous Variables." *Journal of Econometrics*, 190(2), 212-21.

Schneider, A., M. A. Friedl and D. Potere 2009. "A New Map of Global Urban Extent from Modis Data." *Environmental Research Letters*, 4.

Silva, JMCS et Silvana Tenreyro. 2011. "Poisson: Some Convergence Issues." *Stata journal*, 11(2), 207-12.

Strauss-Kahn, Vanessa et Xavier Vives. 2009. "Why and Where Do Headquarters Move?" *Regional Science and Urban Economics*, 39(2), 168-86.

Sutton, Paul C et Robert Costanza. 2002. "Global Estimates of Market and Non-Market Values Derived from Nighttime Satellite Imagery, Land Cover, and Ecosystem Service Valuation." *Ecological Economics*, 41(3), 509-27.

Zhang, Mingliang; John C. Fortney; John M. Tilford et Kathryn M. Rost. 2000. "An Application of the Inverse Hyperbolic Sine Transformation—a Note." *Health Services and Outcomes Research Methodology*, 1(2), 165-71.

Annexes

A. Statistiques descriptives des systèmes de métro base réduite

Tableau 4 : Statistiques descriptives des systèmes de métro (base réduite)

	Longueur moyenne	Nombre de lignes	Nombre de stations	Âge
Nombre	142			
Moyenne	75,18	3,51	59,53	34,25
Écart-type	89,89	3,47	65,75	31,62
Médiane	42,95	2	37,5	28,5
Somme	10 675,85	498	8453	

B. Détail du modèle économétrique

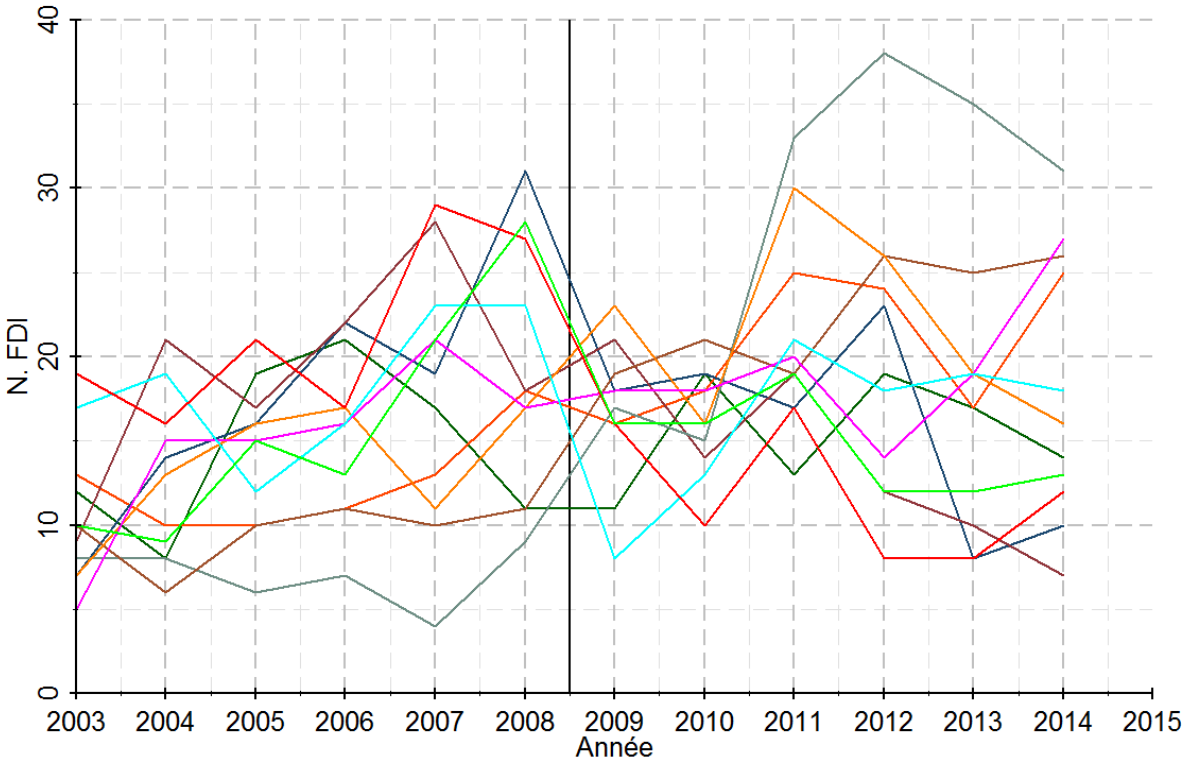


Figure 14 : Évolution annuelle du nombre de projets FDI pour les 11 premières villes

Tableau 14 : Table des corrélations entre les différentes variables économiques et la taille des métros

	Population	Lumière	Surface	Potentiel marché	Capitale	Métro ancien	Métro nouveau
Population	1						
Lumière	0,6752	1					
Surface	0,7379	0,8905	1				
Potentiel de marché	-0,0568	-0,158	-0,2149	1			
Capitale	0,3347	0,2007	0,2417	-0,178	1		
Métro ancien	0,5081	0,444	0,477	-0,0082	0,2904	1	
Métro nouveau	0,2475	0,188	0,1671	0,0451	-0,0182	-0,0491	1

Pour les 1594 villes avec les données des Nation Unies

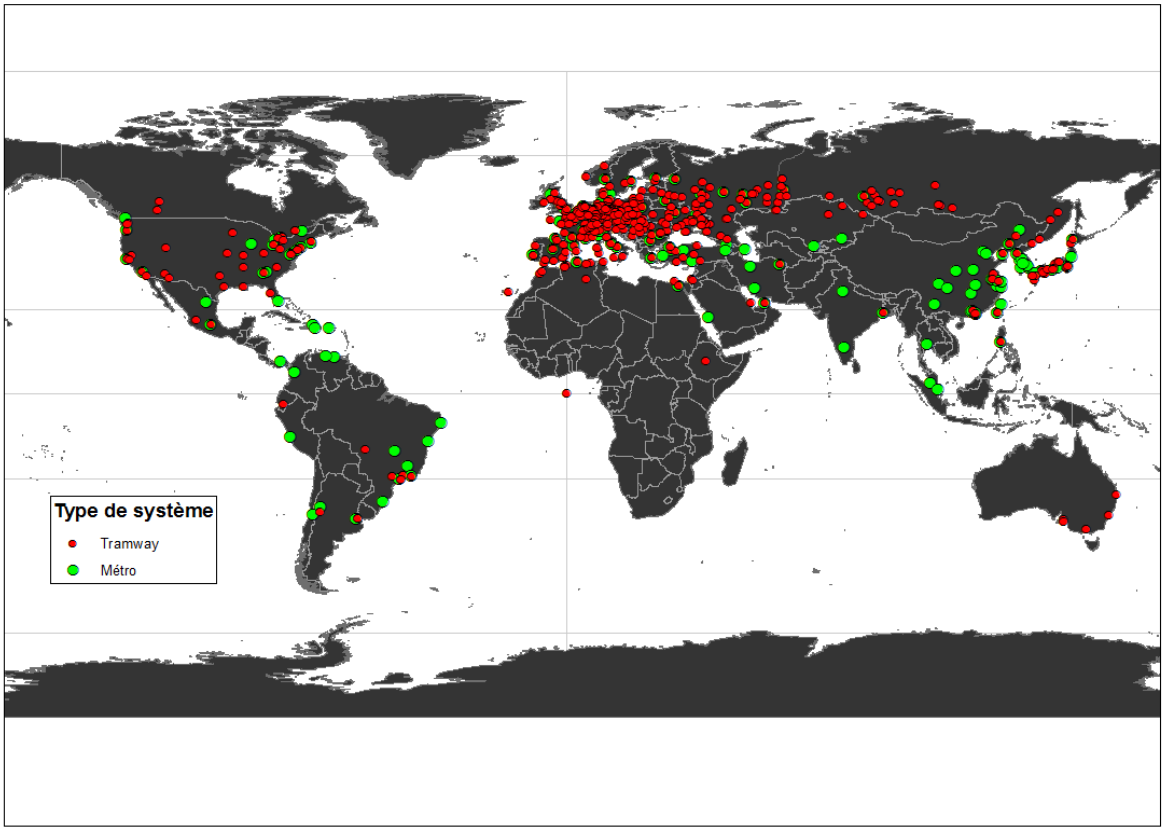


Figure 15 : Distribution spatiale des systèmes de métro et de tramway

C. Test sur les instruments

Dans cette partie de l'annexe, nous mettons en œuvre des tests afin d'évaluer la qualité des instruments générés par la méthode de Lewbel. Dans le Tableau 15 nous testons, pour l'ensemble de l'échantillon, la qualité des instruments au travers du test de Sanderson et Windmeijer (2016). Nous constatons la grande qualité de nos instruments générés, chaque statistique *F* étant largement plus grande que le seuil de 10 habituellement utilisé. Notre modèle génère beaucoup d'hétéroscédasticité. Ce résultat est tout à fait compréhensible dans la mesure où nous sommes en présence de villes à caractéristiques très différentes, telles que la taille. Cette spécificité prend tout son sens, lorsque nous savons que les grandes villes sont celles qui possèdent de grands métros.

Tableau 15 : Test d'identification faible pour l'ensemble de l'échantillon

	F test	Sanderson-Windmeijer F
Indicatrice métro 2015	119.95	7495.43
Indicatrice métro ancien (<2003)	130.48	4522.56
Indicatrice métro nouveau (<=2003)	194.92	4174.42
Taille moyenne	107.35	645.95
Taille nouveau	230.46	525.11
Taille en 2015	106.44	394.69
Taille en 2002	107.06	566.91
Croissance 2002-2014	1102.94	1168.44
Tramway	49.04	52.08

D. Les lumières la nuit comme proxy de l'activité

Depuis maintenant plus d'une vingtaine d'années, les chercheurs dans différents domaines utilisent les lumières la nuit comme variable de remplacement de l'activité économique ou de la démographie. Les lumières la nuit présentent de nombreux avantages. Elles sont disponibles dans une version exploitable pour les chercheurs depuis 1992. Elles couvrent la quasi-totalité de la terre, à l'exception des pôles, avec une granularité extrêmement fine. Les données disponibles ont une résolution de 30 secondes d'arc, ce qui correspond à des carrés de 1 km à l'équateur. Des travaux antérieurs ont déjà montré que les lumières la nuit étaient de très bons proxys pour le produit intérieur brut au niveau des pays. Ainsi, Elvidge, Baugh, Kihn, Kroehl *et al.*

(1997) ont montré qu'il existait pour un groupe de 21 pays une corrélation de 97 % entre le logarithme du PIB et celui de la lumière. Plus récemment, Sutton et Costanza (2002) ont montré un niveau de corrélation de 72 % entre le logarithme du PIB des États américains et le logarithme des lumières la nuit de ces états. Enfin, une étude de Mellander, Lobo, Stolarick et Matheson (2015) a montré grâce à des données suédoises que même pour des niveaux de zonage extrêmement fin (des carrés de 1 km de côté en campagne et de 250 mètres de côté en zone urbaine), le niveau de corrélation des lumières la nuit avec l'activité économique locale est extrêmement fort. Ainsi, selon leur étude, le niveau de corrélation entre les lumières la nuit et les salaires (qui peut être considéré comme une mesure ultra locale de PIB) est d'un peu moins de 70 % pour les lumières saturées et de 76 % pour les lumières en radiance. Toujours selon ces auteurs, les lumières la nuit sont aussi largement corrélées avec le nombre d'établissements et le nombre d'habitants. Enfin, les auteurs notent que les lumières la nuit reflètent mieux les densités que les niveaux des grandeurs économiques dans ce contexte très local.

Au regard de ces résultats, il semblait important de vérifier si les lumières la nuit étaient un bon prédicteur de l'activité pour un zonage intermédiaire : l'aire urbaine. Cependant, à notre connaissance, aucune analyse de corrélation n'a encore été effectuée au niveau des aires urbaines. Pour pallier l'absence d'étude, nous utilisons les données des aires métropolitaines de l'OCDE[28] pour construire notre propre analyse. Ces données donnent accès à une variable du PIB des villes et de la productivité du travail pour 281 aires urbaines des pays de l'OCDE. En utilisant les fichiers géographiques des aires urbaines mis à disposition par l'OCDE, nous avons mesuré différents indicateurs des lumières la nuit pour chacune des aires urbaines. Le Tableau 16 ci-dessous résume le niveau de corrélation entre l'activité des villes mesurée par l'OCDE et celle obtenue à partir des lumières la nuit en 2004. Toutes les valeurs sont en logarithme. Pour les variables en niveau, PIB, population et la somme des lumières la nuit en radiance sont fortement corrélées. Concernant la densité de population, c'est la valeur moyenne qui donne la plus forte corrélation. Enfin, la corrélation entre le PIB par habitant et le niveau de lumière par habitant est de presque 60 %.

À la vue des résultats de ce tableau, nous avons choisi la somme des lumières la nuit en radiance comme un proxy de l'activité d'une ville.

[28] https://stats.oecd.org/Index.aspx?DataSetCode=CITIES

Tableau 16 : Corrélations entre les données économiques OCDE et les lumières la nuit en 2004

	Somme lumière radiance par habitant	Écart-type lumière radiance	Moyenne Lumière radiance	Maximum lumière radiance	Somme lumière radiance	Densité de population	PIB par habitant	Productivité du travail	Population	PIB
PIB	0,3054	0,4349	0,3726	0,5296	0,876	0,0878	0,6728	0,6608	0,8897	1
Population	0,0331	0,4882	0,3575	0,5661	0,7973	0,1963	0,2608	0,262	1	
Productivité du travail	0,5585	0,1004	0,2009	0,1755	0,5409	-0,1085	0,9727	1		
Pib par habitant	0,5921	0,1285	0,2087	0,2025	0,5603	-0,1325	1			
Densité de population	-0,5654	0,2598	0,4979	-0,042	-0,1888	1				
Somme lumière radiance	0,6297	0,4959	0,2704	0,6785	1					
Maximum lumière radiance	0,3948	0,7718	0,3694	1						
Moyenne Lumière radiance	-0,0124	0,8164	1							
Écart-type lumière radiance	0,1927	1								
Somme lumière radiance par habitant	1									

E. Tests de robustesse

Dans cette annexe, nous confrontons nos choix de modélisation à d'autres alternatives afin d'étudier la robustesse de nos résultats face à nos décisions.

Autres transformations du paramètre d'intérêt

Le premier choix que nous avons fait est de transformer notre paramètre d'intérêt pour la marge intensive par la fonction logarithme. Toutefois pour éviter le problème des zéros, nous ajoutons 1 à notre transformée par le logarithme, c'est-à-dire que nous transformons x en $\ln(x+1)$. Ce choix de modélisation n'a d'impact que pour la marge intensive sur l'ensemble de l'échantillon. Pour la marge extensive, le paramètre d'intérêt est représenté par des variables indicatrices et pour le sous-échantillon des villes avec un métro, il n'y pas de problème d'indétermination du logarithme puisque l'ensemble des valeurs associé au métro est positif. Pour mesurer l'impact de ce choix sur nos estimations, nous estimons nos modèles avec d'autres types de transformations. Nous proposons trois alternatives de spécification.

Nous notons t_0 la transformation que nous avons utilisée :

$$t_0(x) = \ln(x+1).$$

Pour la première alternative, nous utilisons la transformée suivante :

$$t_1(x) = \begin{cases} \ln(x) \text{ si } x > 0 \\ 0 \text{ si } x = 0. \end{cases}$$

Dans cette spécification, seules les valeurs strictement positives sont transformées par le logarithme. Les valeurs égales à zéro restent zéro.

Pour la deuxième alternative, nous suivons les recommandations de Gaudry et Quinet (2013) et rajoutons une indicatrice qui vaut un si la taille du métro vaut 0 et zéro sinon. La troisième alternative transforme la variable en utilisant la racine cubique,

$$t_3(x) = x^{\frac{1}{3}}.$$

Enfin, pour la quatrième et dernière alternative nous utilisons la transformation inverse du sinus hyperbolique (voir Zhang, Fortney, Tilford et Rost, 2000, pour des applications) définie comme :

$$t_4(x) = \ln\left(x + \sqrt{x^2 + 1}\right),$$

Cette dernière transformation possède le double avantage de prendre à la fois en compte les valeurs nulles et négatives, mais en plus, son interprétation est identique à une transformée logarithme classique (excepté pour de petites valeurs, ce qui n'est pas notre cas ici).

Dans la Figure 16 ci-dessous, nous avons représenté les trois transformations pour l'intervalle entre 0 et la taille du plus grand métro. Nous constatons que c'est le logarithme qui resserre le mieux les données.

Nous estimons la marge intensive pour chacune des transformations pour deux spécifications du paramètre d'intérêt : en premier avec la taille en 2002 et la taille des métros nouveaux lors de la mise en service tandis que la seconde est la taille des métros en fin de période en 2015. Ces deux spécifications résument parfaitement les différentes possibilités pour modéliser le paramètre d'intérêt pour la marge intensive. Les résultats sont présentés respectivement dans le Tableau 17 et le Tableau 18.

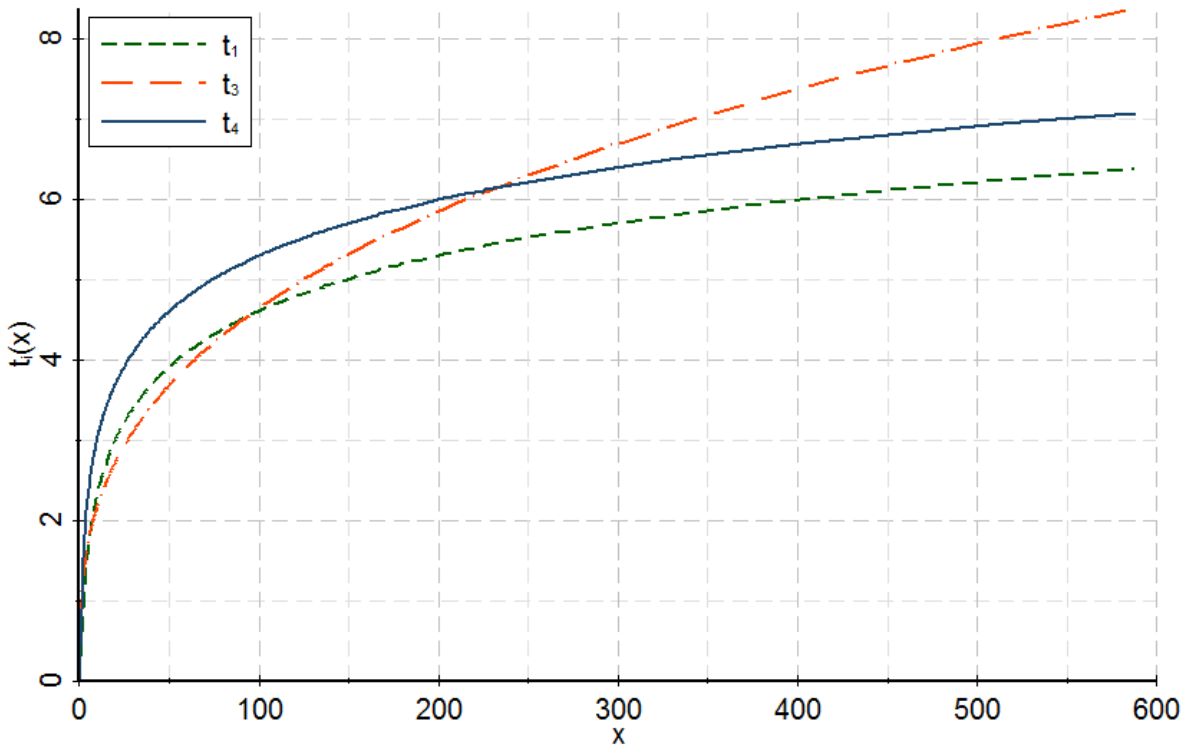


Figure 16 : Transformations alternatives

Enfin dans le Tableau 19, nous analysons l'impact sur nos résultats d'un changement de la constante. En d'autres mots nous transformons par $ln(x + c)$ pour différentes valeurs de c.

Nous constatons que nos résultats ne dépendent que très peu du choix de la transformée. Nous obtenons strictement les mêmes ordres de grandeur pour les coefficients et pour les écarts-types. Pour l'estimation 3 du Tableau 17, rajouter l'indicatrice nouveau métro et la taille de celui-ci dans le même modèle rend l'estimation du paramètre négatif et non significatif. Cela peut se comprendre car, comparativement à la taille de l'échantillon, il n'y a que 48 nouveaux métros et rajouter une indicatrice, annihile le pouvoir explicatif de la variable continue. Dans notre situation, l'inverse du cosinus hyperbolique donne des résultats inférieurs aux autres transformations. Néanmoins, chaque transformation se trouve dans l'intervalle de confiance des autres.

Si nous avons choisi de conserver la transformée $\ln(x + 1)$ c'est parce qu'elle offrait à la fois une souplesse de mise en œuvre et une facilité d'interprétation.

Tableau 17 : Estimations de la marge intensive avec d'autres transformations du paramètre d'intérêt

	(1) $\ln(x+1)$	(2) $\ln(x)$	(3) $\ln(x)$ + indicatrice	(4) $x^{\frac{1}{3}}$	(5) $\ln\left(x+\sqrt{x^2+1}\right)$
Métro 02	0,164***	0,165***	0,152*	0,146***	0,142***
	(0,037)	(0,037)	(0,081)	(0,036)	(0,031)
Métro nouveau	0,222***	0,223***	-0,092	0,233***	0,187***
	(0,050)	(0,050)	(0,275)	(0,056)	(0,042)
Eco	Oui	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui	Oui
Infrastructures	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Oui	Oui	Oui	Oui
Observations	1574	1574	1574	1574	1574
Effets-fixes	Non	Non	Non	Non	Non
Paramètres	51	51	53	51	51
N. métros	134	134	134	134	134
N. Nouv. métros	48	48	48	48	48
N. Pays	189	189	189	189	189
Wald	1,001	1,198	0,790	2,095	1,001
P-valeur	0,317	0,274	0,374	0,148	0,317

(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros.
Les effets-fixes sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tableau 18 : Autres transformations sur la taille en 2015

	(1)	(2)	(3)	(4)	(5)
Métro 15	0,181***	0,182***	0,187**	0,180***	0,155***
	(0,033)	(0,034)	(0,083)	(0,032)	(0,029)
Eco	Oui	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui	Oui
Infrastructures	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Oui	Oui	Oui	Oui
Observations	1574	1574	1574	1574	1574
Effets-fixes	Non	Non	Non	Non	Non
Paramètres	50	50	51	50	50
N. métros	134	134	134	134	134
N. Nouv. métros	48	48	48	48	48
N. Pays	189	189	189	189	189

Estimations effectuées pour la définition élargie des métros.
Les effets-fixes sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tableau 19 : Variation transformation $ln(x + c)$

	(1) Log(x+1)	(2) Log(x+0.1)	(3) Log(x+0.5)	(4) Log(x+10)
Métro 02	0,164***	0,105***	0,142***	0,287***
	(0,037)	(0,023)	(0,031)	(0,076)
Métro nouveau	0,222***	0,133***	0,186***	0,497***
	(0,050)	(0,031)	(0,042)	(0,118)
Eco	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui
Infrastructures	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui
Continent	Oui	Oui	Oui	Oui
Observations	1574	1574	1574	1574
Effets-fixes région	Non	Non	Non	Non
Paramètres	51	51	51	51
N. métros	134	134	134	134
N. Nouv. métros	48	48	48	48
N. Pays	189	189	189	189

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Les effets-fixes sont au niveau pays/états/provinces.
Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays/états/provinces.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Modèle binomial négatif

Le modèle binomial négatif permet de prendre en compte un excès de dispersion du modèle de Poisson. Le Tableau 20 présente les résultats d'estimation du modèle binomial négatif pour la marge extensive et la marge intensive. Nous observons des résultats toujours significatifs, mais d'une ampleur plus faible. Néanmoins, les ordres de grandeur restent les mêmes. Nos estimations peuvent souffrir d'un excès de dispersion, mais sans que cela ne biaise manière conséquente nos résultats.

Tableau 20 : Résultats du modèle binomial négatif

	(1)	(2)	(3)	(4)	(5)	(6)
Métro < 2003	0,390***	0,314***				
	(0,118)	(0,075)				
Métro >= 2003	0,535**	0,223**				
	(0,245)	(0,097)				
Taille métro 02			0,161***	0,106***		
			(0,045)	(0,022)		
Taille nouveau			0,212***	0,090***		
			(0,072)	(0,035)		
Taille métro 15*					0,176***	0,099***
					(0,039)	(0,020)
C. métro 02-14			0,042***	0,019*	0,018	0,009
			(0,013)	(0,011)	(0,013)	(0,011)
C. métro 90-02			-0,178	0,208***	-0,218*	0,208***
			(0,127)	(0,065)	(0,123)	(0,063)
C. métro 78-90			-0,031	-0,038*	-0,022	-0,033+
			(0,026)	(0,021)	(0,026)	(0,021)
Puissance	Oui	Oui	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non
Observations	1574	1574	1574	1514	1574	1514
Effets fixes	Non	Non	Non	Oui	Non	Oui
Paramètres	49	40	52	41	51	40
Métros <2003	134	134	134	127	134	127
Métros >=2003	48	48	48	47	48	47
N. Pays	189	189	189	131	189	131
Wald	0,308	0,802	0,400	0,215		
P-valeur	0,579	0,370	0,527	0,643		

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Les effets-fixes sont au niveau pays/états/provinces.
Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F. Résultats supplémentaires

Tableau 21 : Résultats marge intensive III, base restreinte

	(1)	(2)	(3)	(4)	(5)	(6)
Taille métro 02	0,287**	0,501***				
	(0,133)	(0,169)				
Moy. métro 03-14			0,300**	0,399***		
			(0,144)	(0,142)		
Taille métro 15*					0,333**	0,399**
					(0,139)	(0,176)
C. métro 02-14	0,039	0,090***	0,017	0,041+	0,012	0,036
	(0,030)	(0,029)	(0,024)	(0,025)	(0,022)	(0,026)
C. métro 90-02	0,093	0,307+	0,117	0,283+	0,105	0,297+
	(0,148)	(0,189)	(0,149)	(0,185)	(0,147)	(0,182)
C. métro 78-90	-0,028	-0,033	-0,026	-0,011	-0,016	0,004
	(0,039)	(0,033)	(0,040)	(0,030)	(0,039)	(0,029)
Var éco.	Oui	Oui	Oui	Oui	Oui	Oui
Puissance	Oui	Oui	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non	Oui	Non
Observations	103	76	103	76	103	76
Paramètres	47	38	47	38	47	38
Effets-fixes pays	Non	Oui	Non	Oui	Non	Oui
R-carré	0,945	-	0,945	-	0,946	-
N. Pays	45	18	45	18	45	18

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Les effets-fixes sont au niveau pays.
Estimations effectuées pour la définition restreinte des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tableau 22 : Marge intensive avec les données issues de la base OpenStreet

	(1)	(2)	(3)	(4)	(5)	(6)
Metro OS 2015	0,226***	0,232***	0,178***	0,153***	0,131***	0,067**
	(0,063)	(0,059)	(0,041)	(0,039)	(0,033)	(0,028)
Pop. 02	0,286	-5,179	-7,571**	-7,973**	-1,169	-2,536
	(0,207)	(5,748)	(3,854)	(3,248)	(3,478)	(5,694)
Lumière 03	0,545***	-4,517***	-4,654***	-4,075***	-4,232***	-2,579*
	(0,132)	(1,262)	(1,115)	(1,448)	(1,413)	(1,530)
Surface 00	-0,143	1,490	0,967	1,708	1,237	-0,021
	(0,250)	(1,505)	(1,187)	(1,489)	(1,686)	(1,112)
Villes <500km	0,002	0,002	0,006***	0,004***	0,003**	0,002
	(0,002)	(0,002)	(0,001)	(0,002)	(0,001)	(0,003)
Capitale	0,508**	0,479**	-0,103	0,022	-0,019	0,188*
	(0,240)	(0,243)	(0,170)	(0,155)	(0,139)	(0,113)
Puissance	Non	Oui	Oui	Oui	Oui	Oui
Infrastructures	Non	Non	Non	Oui	Oui	Oui
Géographie	Non	Non	Non	Non	Oui	Oui
Météo	Non	Non	Non	Non	Oui	Oui
Continent	Non	Non	Non	Non	Oui	Non
Observations	1574	1574	1574	1574	1574	1514
Effets fixes	Oui	Oui	Oui	Oui	Oui	Oui
Paramètres	7	15	20	32	47	36
N. métros 2015[1]	191	191	191	191	191	181
N. régions	189	189	189	189	189	131

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
[1] Nombre de systèmes dans l'estimation.
Estimations effectuées pour la définition élargie des métros.
Les effets-fixes sont au niveau pays/états/provinces.
Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays/états/provinces.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Le Tableau 22 présente les résultats pour l'estimation de la marge intensive en utilisant les données du site communautaire OpenStreet. Nous constatons que les contributeurs du site ont une vision plus large des systèmes de métro. En effet, pour notre échantillon de travail nous identifions 183 systèmes alors que pour le site, le nombre est de 191. Néanmoins, les estimations utilisant les données OpenStreet donnent des résultats similaires aux estimations avec nos propres données sur les métros.

Si nous n'avons pas utilisé les données OpenStreet, c'est d'abord parce qu'elles ne possédaient pas de profondeur temporelle, mais aussi, et surtout, parce qu'elles ne nous semblaient pas suffisamment précises. Cette absence de précisions provient d'une part du manque d'uniformisation de la part des contributeurs, mais également du fait que le site a pour but la simple représentation cartographique des éléments et non leur utilisation à des fins statistiques. Cette particularité entraîne l'apparition de nombreux doublons dans les lignes, par exemple, nous voyons apparaître deux lignes pour figurer les deux sens de circulation

Tableau 23 : Estimations marge extensive pour les villes de plus d'un million d'habitants

	(1) Poisson	(2) Poisson	(3) IV-Poisson	(4) IV-Poisson
Métro en 2015[(i)]	0,654***	0,604***	0,892***	1,141***
	(0,158)	(0,106)	(0,182)	(0,145)
Pop. 02	-17,387	5,057	-10,377	44,690**
	(17,276)	(16,285)	(17,625)	(20,676)
Lumière 03	-11,096**	0,645	-32,273***	64,846***
	(4,476)	(7,003)	(8,615)	(18,492)
Surface 00	4,867	2,590	8,527+	-18,267***
	(5,380)	(3,961)	(5,777)	(6,526)
Villes <500km	0,003**	0,001	0,005**	0,006+
	(0,001)	(0,001)	(0,002)	(0,004)
Capitale	0,178	0,084	0,266	-0,578*
	(0,136)	(0,215)	(0,187)	(0,316)
Puissance	Oui	Oui	Oui	Oui
Pays	Oui	Non	Oui	Non
Infrastructures	Oui	Oui	Oui	Oui
Géographie	Oui	Oui	Oui	Oui
Météo	Oui	Oui	Oui	Oui
Continent	Oui	Non	Oui	Non
Indicatrice pays	Non	Non	Non	Oui
Observations	370	311	370	311
Effets fixes pays	Non	Oui	Non	Non
Paramètres	47	36	47	76
N métro 2015	144	117	144	117
N. Pays	100	39	100	39

C. : croissance ; Pop. : population de la ville (x1000) ; surf. : surface de la ville (km²).
(1) Nombre de systèmes dans l'estimation.
(2) Test de Wald pour l'égalité des coefficients des variables métro.
Estimations effectuées pour la définition élargie des métros. Entre parenthèses les écarts-types avec prise en compte des grappes au niveau des pays.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[(i)] Variables instrumentées, pour les variables d'infrastructure seule la taille des tramways est instrumentée. Les instruments sont générés à partir de l'ensemble des autres variables de contrôle à l'exception des indicatrices pays.