

## Distribution des emplois et identification des centres d'affaires : une approche non paramétrique

Aboubacar Amiri, Hakim Hammadou, Moez Kilani

► **To cite this version:**

Aboubacar Amiri, Hakim Hammadou, Moez Kilani. Distribution des emplois et identification des centres d'affaires : une approche non paramétrique. 2017. <hal-01482312>

**HAL Id: hal-01482312**

**<http://hal.univ-lille3.fr/hal-01482312>**

Submitted on 3 Mar 2017

**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.

# Distribution des emplois et identification des centres d'affaires : une approche non paramétrique

Aboubacar Amiri<sup>1</sup> Hakim Hammadou<sup>1</sup> Moez Kilani<sup>\*1</sup>

<sup>1</sup>Université de Lille, LEM – CNRS, UMR 9221

13 février 2017

## Résumé

Nous étudions la distribution des emplois sur la métropole lilloise afin d'identifier les principaux centres d'affaires et caractériser la forme urbaine de cette agglomération. Il ne s'agit pas seulement de repérer les communes avec des niveaux d'emplois élevés, mais celles qui exercent aussi une influence significative sur la distribution des emplois. Nos résultats confirment la structure polycentrique de l'agglomération lilloise et nous trouvons que la commune de Roubaix (parfois accompagnée par Tourcoing) se distingue par une centralité importante. En moyenne, l'éloignement d'un sous-centre d'affaires d'un kilomètre diminue la densité d'emploi de près de 20%. L'identification des sous-centres d'affaires est importante car ces zones d'emploi influencent les décisions de localisation des acteurs économiques, et sont donc à prendre en compte dans l'élaboration des politiques urbaines. La difficulté technique consiste à séparer l'effet du centre ville (le principal centre d'affaires) de l'effet des multiples sous-centres. Comme le révèle notre application empirique, les approches non paramétriques sont plus efficaces que les approches paramétriques.

---

\*Correspondant : moez.kilani@univ-lille3.fr

# 1 Introduction

Le modèle de la ville monocentrique, (cf. Alonso, 1964; Fujita, 1989) est un cadre de référence dans l'analyse de la distribution des emplois. Dans la ville monocentrique, un centre d'affaires principal (CAP) est localisé au centre de la ville, les agents économiques (ménages et entreprises) cherchent à se localiser à proximité du CAP et ces décisions collectives aboutissent à une surenchère de la rente foncière autour de ce centre. Cela entraîne aussi une augmentation de la densité urbaine autour du CAP. Les économistes justifient le regroupement des activités dans un seul centre par des gains générés par les effets d'agglomérations qui se reflètent par une diminution des coûts de production et une augmentation de la productivité. Plusieurs études économétriques ont confirmé la présence de ces forces d'agglomération (Nakamura, 1985; Combes *et al.*, 2010a,b, 2012). Toutefois, le développement d'un CAP est limité par l'espace disponible et par l'augmentation des coûts externes (congestion et pollution, notamment). En conséquence, certaines activités se localisent en dehors du CAP. Ces activités peuvent être dispersées ou se regroupent dans des centres d'affaires secondaires. Ces derniers permettent aux entreprises de bénéficier de certains effets d'agglomération tout en évitant les niveaux élevés de congestion autour du CAP. L'identification des sous-centres nous permet d'évaluer le degré de concentration des emplois et de voir si la ville a une structure monocentrique ou polycentrique. Cette compréhension de la structure de la ville est utile pour les politiques urbaines et les politiques de transport. L'observation des données d'emplois de l'agglomération de Lille Metropole (LM) suggère une structure multacentrique, et on parle couramment de Lille-Roubaix-Tourcoing. Les villes de Roubaix et Tourcoing sont les berceaux de l'industrie textile jusqu'aux années soixante. Le déclin des activités minières et la délocalisation d'une grande partie des industries textiles dans les pays asiatiques et d'Afrique du nord ont conduit à des mutations urbaines importantes au niveau de l'agglomération.

Dans ce papier, nous n'aborderons pas l'évolution historique des changements urbains au niveau de la métropole mais nous utiliserons une approche statistique pour analyser la distribution actuelle de l'emploi au niveau de LM. Plus précisément, nous cherchons à identifier

les centres d'affaires secondaires et leur influence sur la forme urbaine (densité, distribution des emplois et des ménages). Nous accordons une importance particulière à la revue de la littérature car, la définition même d'un sous-centre d'affaires est problématique. L'absence de consensus autour de ce sujet a conduit à l'émergence de techniques statistiques distinctes pour l'identification des sous-centres d'affaires.

La question de l'identification des sous-centres d'affaires à travers une approche statistique remonte, au moins, à McDonald (1987), qui a noté qu'il est plus pertinent d'utiliser les densités de l'emploi, et non pas les valeurs brutes de l'emploi, afin de comparer les sous-centres. À notre connaissance, tous les travaux qui se sont intéressés à l'identification des sous-centres d'affaires se réfèrent au modèle de la ville monocentrique et supposent l'existence d'un centre d'affaires principal. Celui-ci est généralement situé au centre ville.

La première idée que l'on peut se faire d'un sous-centre d'affaires est celle d'un endroit où la densité de l'emploi est jugée élevée. C'est en effet la procédure utilisée par Giuliano et Small (1993) afin d'étudier la distribution des emplois à Los Angeles. La difficulté majeure dans ce cas est de proposer une valeur précise pour le seuil à partir de laquelle on peut juger que le niveau de l'emploi est élevé. En plus, ce seuil, s'il convient à une agglomération ne conviendra pas nécessairement à une autre agglomération. Pour contourner cette difficulté Craig et Ng (2001) ont utilisé une régression quantile de la densité afin d'identifier les sous-centres d'affaires dans la ville de Houston au Texas. La valeur du seuil de la densité d'emploi qui permet d'identifier si une zone d'emploi est un sous-centre d'affaires ou non est déterminée de manière endogène, en fonction des quantiles de la distribution des emplois conditionnellement à la distance du CAP. Cette approche est sophistiquée et présente des avantages techniques au niveau de la méthodologie d'estimation. Néanmoins, elle présente deux inconvénients. Le modèle empirique de base est unidimensionnel et l'analyse statistique ne permet que l'identification des distances au niveau desquelles se trouvent les sous-centres potentiels. La liste définitive des candidats à retenir comme sous-centres nécessite une connaissance de la région d'étude. Des améliorations ont été apportées à cette technique, notamment par Craig *et al.* (No date) qui ont effectué une estimation de surface de quantile en deux dimensions. Le

deuxième problème est principalement soulevé par McMillen (2001, 2004). Ce dernier se base sur des arguments plus forts de l'économie urbaine, notamment le modèle monocentrique, et stipule qu'un centre d'affaires ne doit pas uniquement correspondre à un endroit où la densité des emplois dépasse un seuil donné. Pour qu'un candidat soit retenu définitivement comme centre d'affaires il faut aussi qu'il exerce une déformation significative sur la forme urbaine de la ville en question. Pour mettre en pratique cette idée il développe une procédure en deux étapes. Une présélection de candidats pour des sous-centres est effectuée au cours de la première étape. La deuxième étape consiste à mesurer l'ampleur des effets qu'exerce chaque sous-centre sur son entourage afin de conclure s'il est retenu comme sous-centre ou non. Une approche semi-paramétrique est adoptée au niveau du modèle empirique.

Enfin, il est à noter qu'en France, et en Europe de manière générale, très peu d'économistes se sont penchés sur les questions de distribution des emplois et l'identification des centres d'affaires. La majorité des travaux existants ont étudié cette question pour des agglomérations situées en Amérique du Nord. Baumont *et al.* (2004) étudient la répartition de l'emploi sur l'agglomération de Dijon en utilisant une approche basée sur les statistiques spatiales. Ils trouvent une faible significativité pour les zones d'emploi en dehors du CAP et confirment que Dijon a plutôt une structure monocentrique. Aguilera *et al.* (2009) observent que dans la région parisienne, le développement de nouveaux quartiers d'affaires à proximité de Paris, comme la Défense, a attiré principalement des Parisiens à haut revenu. Cela a engendré un développement des flux de mobilité de Paris vers les zones périphériques.<sup>1</sup> Aguilera (2005) compare des données entre 1990 et 1999 pour Paris, Lyon et Marseille. Ses résultats montrent que les déplacements du centre ville vers les zones périphériques ont augmenté durant cette période. Aguilera et Mignot (2004) considèrent la même période et étudient la forme urbaine pour sept villes françaises. Leur analyse s'intéresse à la mobilité et distingue les centres urbains proches du centre-ville de ceux situés dans la périphérie.

Au niveau des résultats, nous trouvons que la ville de Roubaix exerce une influence significative sur la distribution de l'emploi. Roubaix est un sous-centre qui se distingue des

---

1. Les auteurs utilisent le terme « reverse commuting » pour désigner ces flux

autres dans le sens où son influence est plus étalée spatialement. En effet, les autres sous-centres comme Villeneuve-d'Ascq ou Armentières exercent une influence limitée à leur proche voisinage, alors que pour Roubaix le modèle statistique laisse apparaître une influence structurellement plus importante. Au niveau théorique chaque approche a ses avantages et ses inconvénients, et il est difficile de décider laquelle est la meilleure. Sur le plan empirique, et par rapport à notre application, les résultats obtenus via l'approche non-paramétrique proposée par McMillen (2001) donne des résultats plus convaincants et plus faciles à interpréter. Sa mise en œuvre pratique reste, toutefois, plus laborieuse étant donnée sa structure itérative. Dans une dernière étape, on utilise le modèle de McMillen (2004) pour mesurer l'impact direct d'un sous-centre d'affaires sur la distribution des emplois. Nous trouvons qu'en s'éloignant du sous-centre le plus proche d'un kilomètre, la densité d'emploi diminue de 20%, approximativement.

Ce papier est organisé de la manière suivante. Dans la deuxième section nous décrivons les principales méthodes statistiques utilisées pour identifier les sous-centres d'affaires. Dans la troisième section nous présentons les données que nous utilisons dans l'analyse empirique (quatrième section). C'est dans cette section que nous caractérisons la distribution des emplois et la forme urbaine de LM. Nous concluons avec quelques perspectives de recherche dans la dernière section.

## 2 Méthodologie

McDonald (1987) évoque cinq critères qui peuvent servir pour identifier un sous centre d'affaires : (1) un niveau élevé de la densité brute d'emploi, (2) de la densité nette d'emploi, (3) un ratio élevé emploi/population, (4) densité brute de la population et (5) densité nette de la population. Il n'en retient finalement que deux : la densité brute d'emploi et le ratio emploi/population qu'il utilise pour identifier les sous-centres d'affaires à Chicago. Les autres critères sont moins convaincants. En particulier, l'augmentation de la densité de la population (brute ou nette) peut bien provenir de la présence d'un cadre (aménité) attractif et pas nécessairement de la présence d'une activité économique importante. Aussi, la densité

nette de l'emploi peut résulter d'une faible valeur du terrain alloué à l'emploi, une possible conséquence de la concurrence avec la demande résidentielle. La densité brute de l'emploi est le rapport entre le niveau d'emploi par l'aire de la région considérée, alors que la densité nette de l'emploi est le rapport entre l'emploi et l'aire du terrain alloué aux emplois. De manière similaire on parle de densités nettes et brutes pour la population. Craig et Ng (2001) utilisent une régression de quantile afin de contourner le choix d'un seuil exogène. Ils estiment une densité d'emplois quantile  $g_\tau(x)$  au niveau  $\tau$  vérifiant  $\tau = \int_{-\infty}^{g_\tau(x)} f_{Y/X}(y/x)dy$ , où  $f_{Y/X}(y/x)$  est la densité de probabilité conditionnelle de  $Y$  (la densité) sachant  $X = x$  (la distance au CAP). L'idée ici est de se focaliser sur les valeurs de  $\tau$  proches de 1 de sorte à repérer les endroits où la densité d'emploi atteint des valeurs significativement élevées. Pour une estimation consistante de la fonction  $g_\tau$ , les auteurs utilisent un lissage non-paramétrique par spline. L'objectif est de trouver une fonction n'ayant pas beaucoup d'ondulations. La fonction objectif consiste à minimiser un coût total qui s'écrit comme :

$$\text{« ajustement »} + \lambda \text{ « ondulations »},$$

où  $\lambda$ , un paramètre positif, reflète le compromis entre « ajustement » et « ondulations ». Plusieurs formes fonctionnelles peuvent être utilisées pour formuler l'ondulation, comme par exemple  $\max |g''(x)|$ . La partie ajustement est une mesure de la distance entre la fonction  $g(x)$  et les observations. Souvent le lissage par spline revient à résoudre un problème de type :

$$\min_g \sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int (g''(x))^2 dx. \quad (1)$$

Si la valeur de  $\lambda$  est proche de zéro, on obtient une courbe  $g$  très proche des observations mais présentant une forme très irrégulière. Si la valeur de  $\lambda$  est élevée, la fonction  $g$  présente peu d'ondulations mais le niveau des erreurs d'ajustement est élevé. Dans la pratique des critères informationnels, de type Schwartz, sont souvent utilisés pour trouver la valeur optimale de  $\lambda$ . Une présentation détaillée de la régression quantile est disponible dans Koenker *et al.* (1994). Il est à remarquer que l'adoption d'un tel critère ne repose pas sur une argumentation très forte, et Koenker *et al.* (1994) rappellent qu'il n'existe pas de choix idéal, et que cela revient en définitive au modélisateur.<sup>2</sup> Pour notre analyse nous avons légèrement modifié cette étape

---

2. Voir la documentation du package « quantreg » de  $R$  proposée par les auteurs.

(choix de  $\lambda$ ) afin de hiérarchiser les sous-centres. Comme nous le détaillons dans la partie empirique, section 4, cela nous permet de mettre en relation la valeur de  $\lambda$  avec le nombre de sous-centres que l'on souhaite faire apparaître.

Une fois la densité de quantile estimée, on peut en déduire le gradient de la densité d'emploi. Si cette densité a une forme exponentielle, alors la courbe du gradient est une ligne horizontale correspondant à un niveau négatif. L'existence de sous-centres d'affaires conduit à des irrégularités dans la courbe du gradient. Plus précisément le gradient augmente autour des sous-centres. À ce niveau le gradient passe à un niveau négatif après qu'il ait pris des valeurs positives. Craig et Ng (2001) ont appliqué leur méthode sur l'agglomération de Houston (Texas) pour laquelle ils disposaient de 578 observations sur la densité des emplois. Ils identifient l'existence de centres d'affaires sur trois rayons autour du centre ville.

La dernière étape est l'identification des sous-centres proprement dits. Ceux-ci nécessitent l'examen des zones d'emplois se situant au niveau des distances trouvées par la procédure que nous venons de décrire. La méthode proposée par Craig et Ng (2001) repose sur des outils statistiques solides mais présente l'inconvénient de cette dernière étape où un examen subjectif de la distribution des emplois est nécessaire pour l'identification des sous-centres d'affaires. Ce problème provient de la structure unidimensionnelle du modèle empirique utilisé. En effet, chaque endroit de la ville est identifié uniquement par la distance qui le sépare du CAP. Une manière de dépasser ce problème est d'étendre le modèle à un cadre bidimensionnel. Cette ligne a été explorée par Craig *et al.* (No date) et Craig *et al.* (2016). L'idée est d'estimer une fonction de régression quantile de niveau  $\tau$  qui dépend de deux paramètres  $x$  et  $y$  qui correspondent à la latitude et la longitude. En notant  $z$  la log-densité des emplois, l'objectif est de minimiser la fonction de coût donnée par :

$$\sum_{i=1}^n \left\{ \tau \left( z_i - g(x_i, y_i) \right)^+ + (1 - \tau) \left( g(x_i, y_i) - z_i \right)^+ \right\} + \lambda J(g), \quad (2)$$

où  $n$  est le nombre d'observations, et pour un réel  $I$  donné, on note  $I^+ = I$  si  $I > 0$  et  $I^+ = 0$  sinon. Le deuxième terme dans l'équation (2) mesure l'ondulation  $J(g)$  de la fonction de lissage  $g$  pondérée par le paramètre  $\lambda$  qui joue le même rôle que dans le cas univarié discuté ci-dessus. Dans l'expression (2), on donne un poids  $\tau$  aux observations se trouvant



au dessus de la surface de régression quantile et un poids  $1 - \tau$  aux observations se trouvant au dessous de la même surface. Le lissage d'une fonction bivariable permet de repérer les endroits où la densité de l'emploi est élevée mais aussi de mesurer l'influence de chaque zone sur son voisinage. L'avantage du triogramme bivarié, par rapport à l'approche univariée, est de permettre l'identification des sous-centres, directement en cherchant les endroits où la surface de quantile atteint un maximum local. Souvent le graphique des courbes de niveaux de la surface de quantile est suffisant pour trouver ces maxima. Un algorithme d'estimation est proposé par Koenker et Ng (2005).

McMillen (2001) propose une approche non-paramétrique différente. Il part d'une définition pertinente, au sens économique, d'un sous-centre d'affaires pour développer sa modélisation statistique. Il considère qu'une zone particulière ne peut être considérée comme sous-centre pour la simple raison d'une densité d'emploi élevée. Dans une ville multicentrique, on doit observer des zones avec des densités d'emplois élevées mais qui, comme le CAP, exercent une influence sur la densité d'emploi des autres régions. Pour identifier de tels sous-centres, McMillen développe une approche à deux étapes. Dans la première étape une régression à coefficients localement pondérés permet d'estimer spatialement la distribution du logarithme de la densité des emplois. Avec cette procédure, pour estimer la densité au niveau d'un point donné, on accorde des pondérations élevées aux points les plus proches et des pondérations plus faibles pour les observations éloignées. En développant sa méthode, McMillen part du constat qu'une procédure idéale doit vérifier cinq propriétés : (1) invariabilité vis à vis des unités de mesure (aire d'un polygone, centroïdes de zones, etc.), (2) aptitude à repérer des augmentations de la densité statistiquement significatifs, (3) prise en compte des effets du CAP, (4) prise en compte des variations de densité provenant des spécificités locales (nature du terrain, par exemple), et (5) quantification de l'influence géographique du sous-centre. Si l'on note  $y_i$  le logarithme de la densité des emplois sur le site  $i$ , alors on désigne par  $\hat{y}_i$  l'ajustement par régression localement pondérée de  $y_i$  et par  $\hat{\sigma}_i$  l'écart-type de l'erreur d'ajustement. Les candidats pour un sous-centre sont les sites où la densité est significativement élevée par rapport à sa valeur ajustée. Pour un niveau de risque  $\alpha$  fixé,

on présélectionne alors les sites avec une densité  $y_i$  telle que  $(y_i - \hat{y}_i)/\hat{\sigma}_i > t_\alpha$ , où  $t_\alpha$  est le fractile d'ordre  $1 - \alpha/2$  d'une loi normale centrée réduite. Remarquons que le sous-centre est déterminé de manière endogène. Lorsque plusieurs sites contigus sont candidats, seul le site ayant la densité la plus élevée est retenu pour la deuxième étape. Cela permet d'éviter de traiter un grand nombre de candidats et d'éviter aussi des problèmes de multicollinéarité dans les distances.

Dans la deuxième étape on cherche à évaluer l'influence qu'exerce chaque sous-centre présélectionné sur toutes les observations (autres que les candidats). Soient  $D_i$  la distance entre le CAP et le site  $i$ , et  $d_{ij}$  la distance entre le site  $i$  et le candidat  $j$ . Une régression semi-paramétrique ayant la forme

$$y_i = g(D_i) + \sum_{j=1}^S (\delta_{1j} d_{ij} + \delta_{2j} d_{ij}^{-1}) + u_i \quad (3)$$

est alors effectuée. La fonction  $g(D_i)$  peut être estimée par une régression localement pondérée ou par noyau. Une approximation par spline est aussi possible. McMillen (2001) et nous-même utilisons un développement de Fourier. Dans l'expression (3), la distance entre le site  $i$  et le candidat  $j$  rentre de deux manières :  $d_{ij}$  et  $d_{ij}^{-1}$ . L'inverse de la distance diminue rapidement lorsque  $d_{ij}$  augmente et reflète donc une influence très localisée du candidat  $j$ , alors que  $d_{ij}$  reflète une influence plus étendue du même candidat. Une fois l'estimation effectuée, seul le paramètre statistiquement le plus significatif, parmi  $d_{ij}$  et  $d_{ij}^{-1}$  est retenu, et la procédure est répétée jusqu'à ce que tous les coefficients des distances soient significatifs à 20%.

McMillen (2004) reprend une approche similaire avec deux changements majeurs. D'abord, seul le sous-centre candidat le plus proche du site  $i$  est pris en compte, ensuite seul l'inverse de la distance est pris en compte dans le modèle empirique. Le modèle semi-paramétrique a la forme suivante :

$$y_i = g(D_i) + \delta d_i^{-1} + u_i, \quad (4)$$

où  $d_i$  désigne la distance entre le site  $i$  et le sous-centre candidat le plus proche (remarquons que l'indice  $j$  n'est plus utilisé dans ce cas). McMillen compare le modèle semi-paramétrique

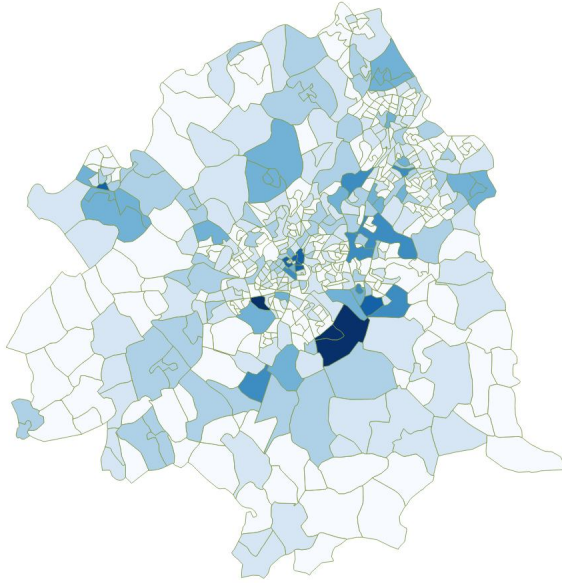
	Min	Moyenne	Max	Ecart-type	Obs.
Distance du CAP (mètre)	0	8 295,73	21 120,98	5 014,97	547
Densité d'emploi (emploi/km <sup>2</sup> )	0	412,31	7 298,52	709,17	547

TABLE 1 – Statistiques descriptives

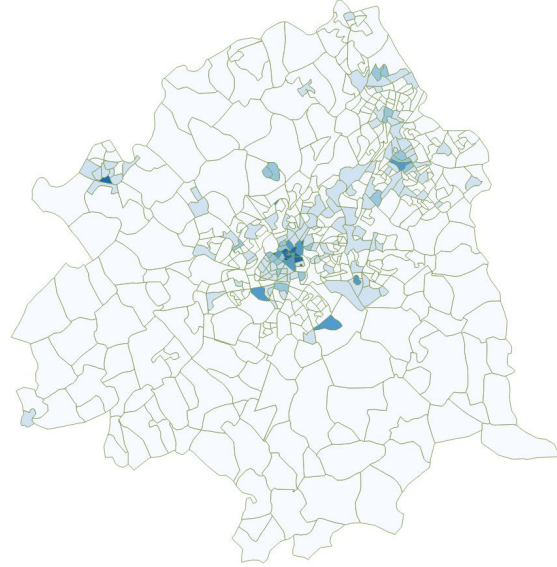
(4) avec un modèle linéaire en mesurant l'autocorrélation spatiale dans chaque cas. Il obtient des résultats meilleurs à travers le modèle semi-paramétrique et en déduit qu'une mauvaise spécification du modèle conduit à des problèmes d'autocorrélation spatiale. L'avantage du modèle (4) est que le paramètre  $\delta$  mesure l'importance de la distance du sous-centre le plus proche.

### 3 Les données

Les données sont à l'échelle de l'IRIS : « Ilots Regroupés pour l'Information Statistique ». Selon l'INSEE, les communes qui ont plus de 10 000 habitants sont découpées en IRIS. Pour les autres communes on conserve l'échelle communale. Ces données proviennent du recensement démographique réalisé sur l'agglomération Lilloise. Cette base de données décrit la population active sur le lieu de résidence selon différentes catégories (actifs occupés, chômeurs, etc.). Nous avons complété cette base avec des données du « Centre d'Études Techniques de l'Équipement (CETE) » de Lille, qui avait géolocalisé l'emploi au niveau de l'IRIS selon le secteur d'activité. Le CETE a utilisé pour cela la base de données SIRENE qui décrit la structure et la démographie des établissements. Ces données sont relatives à l'année 2010. Le tableau 1 décrit les statistiques descriptives pour les 547 observations de notre base de données. La distance par rapport au CAP est calculée en considérant les centroids des IRIS. La densité de l'emploi est calculée comme le rapport du nombre d'emplois sur la superficie.



(a) Les emplois (effectifs)



(b) La densité de l'emploi

FIGURE 1 – La distribution des emplois su l'agglomération de Lille.

## 4 Résultats empiriques

Lille métropole est une agglomération multicentrique avec trois villes historiques : Lille, Roubaix et Tourcoing. Cette structure est visible sur la figure (1b) où, en plus des principaux centres quelques autres sites avec des niveaux de densité importants sont visibles (Villeneuve-d'Ascq, Armentières, Lesquin, etc.). Plus la couleur est foncée, plus le niveau de la variable est élevé. La figure (1a) où sont reportés les effectifs des emplois ne fait pas clairement apparaître les zones d'emploi. Il est facile de voir qu'en considérant les effectifs au lieu des densités, on va surestimer l'importance des zones ayant une grande taille géographique. C'est la densité qui est utilisée dans l'analyse de la distribution des emplois.

### 4.1 Régression quantile

Nous appliquons la méthode proposée par Craig et Ng (2001) afin d'identifier les sous-centres d'affaires. Notre analyse empirique nous permet de détecter deux groupes de sous-centres d'affaires. Chaque groupe est situé à une distance donnée par rapport au CAP. Ce

résultat est visible sur la figure 2. La courbe en haut correspond aux quantiles de niveau 95% et celle d'en bas au niveau 50%, la médiane. Les points correspondent aux observations et les lignes verticales aux emplacements des sous-centres d'affaires. Le nuage de points montre que la densité d'emploi diminue, en médiane, au fur et à mesure que l'on s'éloigne du CAP. En fixant la valeur du paramètre  $\lambda$  de sorte à minimiser l'information AIC, on identifie l'existence de sous-centres à deux distances du CAP. Les premiers sous-centres se trouvent à 10 km et les autres à 17,5 km. En regardant de près au niveau de ces distances, on trouve qu'à dix kilomètres, c'est Roubaix qui constitue un sous-centre d'affaires important. À 17,5km c'est la commune de Halluin qui apparaît comme un sous-centre. Remarquons que par rapport à ce dernier point, le nombre d'observations est relativement faible à plus de 15 km, et le résultat obtenu est moins robuste.

Il y a plusieurs points importants révélés par cette estimation et visibles sur la figure 2. Premièrement, toutes les zones ayant un niveau de densité d'emplois élevé ne sont pas des sous-centres, parce qu'elles n'exercent pas une influence significative sur la densité d'emploi des sites avoisinants. Certaines conclusions à ce niveau sont relativement surprenantes. C'est le cas notamment pour Tourcoing, Neuville et Wambrechies, des communes dont les densités d'emploi correspondantes se trouvent au dessus de la courbe de quantile à 95%, mais pour lesquelles cette courbe de quantiles n'atteint pas un maximum local. Même si la densité d'emploi est élevée, elle ne produit pas une déformation significative sur la courbe de quantiles. Deuxièmement, et comme dans le cas de Craig et Ng (2001), la courbe de quantile a une forme assez irrégulière, ce qui montre qu'une estimation paramétrique de la distribution n'est pas convenable pour ce problème. Ce problème est aussi souligné par McMillen (2004). Troisièmement, le quantile de niveau 50%, la médiane, n'est pas très informatif pour l'identification des sous-centres. Ainsi, l'observation de la tendance centrale de la distribution des emplois n'est pas particulièrement utile dans ce cas. Enfin, on peut observer l'hétérogénéité des termes d'erreurs. Si ceux-ci étaient indépendants et identiquement distribués on aurait des courbes équidistantes. Toutefois, les deux courbes sur la figure 2 ne révèlent pas une telle structure même aux environs du CAP où le nombre d'observations est élevé. En plus,

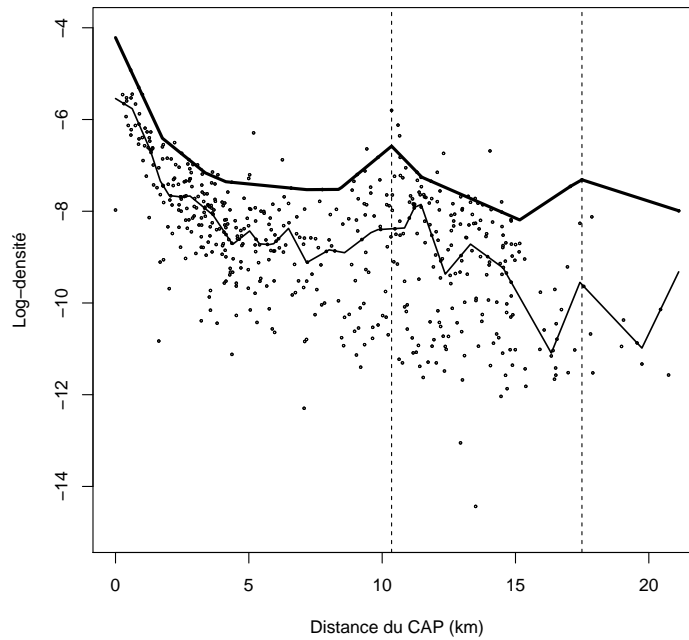


FIGURE 2 – Estimation de quantiles des emplois ( $\lambda$  qui minimise le critère AIC)

la distance entre les deux courbes de quantile ne semble pas varier selon une structure particulière. Il n'est donc pas facile de traiter la question de l'hétérogénéité des erreurs dans ce cas.

Le critère informationnel sur lequel repose le choix de la valeur du paramètre  $\lambda$ , le paramètre dans l'équation (1), est discutable. C'est pour cette raison que nous essayons de modifier la procédure initiale en adoptant une procédure itérative : on fixe le nombre de sous-centres d'affaires à identifier et on cherche la valeur de  $\lambda$  qui lui correspond. Cela nous permet de hiérarchiser les différents sous-centres. Le plus important est celui qui apparaît en premier, le deuxième, qui apparaît en deuxième lieu et ainsi de suite. Nous nous arrêtons lorsque la recherche ne fait plus apparaître des zones d'emploi significatives. Par rapport à la cohérence de nos conclusions, nous vérifions que les rayons sélectionnés restent invariants lorsque la valeur de  $\lambda$  est mise à jour. C'est à dire que la distance qui apparaît au début reste sélectionnée avec deux distances, et ainsi de suite.

Avec cette procédure nous aboutissons à des courbes de quantiles données sur la figure 3. Les nombres à côté des lignes verticales correspondent à l'ordre d'apparition des distances. Les premiers sous-centres sont à 10 km du CAP, les deuxièmes à 17,5 km et un troisième rayon, qui n'apparaissait pas initialement, est situé à 5 km du CAP. Ce dernier correspond à des zones d'emplois dans la commune de Villeneuve-d'Ascq. Nous avons décidé de nous arrêter à ce niveau car il n'existe aucune valeur de  $\lambda$  donnant quatre sous-centres, et en fixant cinq sous-centres ou plus on ne trouve que des zones d'emploi de faible importance. Les trois lignes verticales discontinues indiquent que la courbe de quantile d'ordre 95% atteint un maximum local. Roubaix apparaît comme un sous-centre de grande importance. Cette zone cause une déformation significative de la courbe de quantile, et influence significativement la distribution des emplois sur ses alentours. La commune d'Halluin, même si sa densité n'est pas très élevée, exerce une influence sur la distribution des emplois. Toutefois, comme nous l'avons indiqué ci-dessus, le nombre d'observations au delà de 15km est relativement faible pour que l'on puisse affirmer cette conclusion avec assurance. Le troisième ensemble de sous-centres d'affaires se trouve à une distance de 5 km de Lille-centre. À cette distance il y a les zones d'emplois de Villeneuve-d'Ascq, une commune située à l'ouest de Lille-centre. Notons que la déformation de la courbe de quantiles à 95% n'est pas très annoncée, mais qu'à ce niveau il y a un grand nombre d'observations avec des densités relativement élevées. C'est donc une zone qui se distingue parmi d'autres zones où la densité de l'emploi est déjà élevée.

Avec cette approche univariée, l'identification des sous-centres nécessite une connaissance de la région d'étude car l'estimation statistique ne donne que la distance des sous centres par rapport au CAP. On explore les zones d'emplois qui se trouvent à un rayon égal à cette distance afin de repérer celles ayant les densités les plus élevées. C'est sur ce point que cette démarche est principalement critiquée, car le modélisateur doit, en quelque sorte, utiliser sa propre connaissance de l'agglomération dans cette dernière étape. Craig *et al.* (No date) ont développé une régression de quantile bivariée, où chaque zone est repérée à la fois par sa latitude et sa longitude. L'idée est de trouver une surface de quantile correspondant à l'équa-

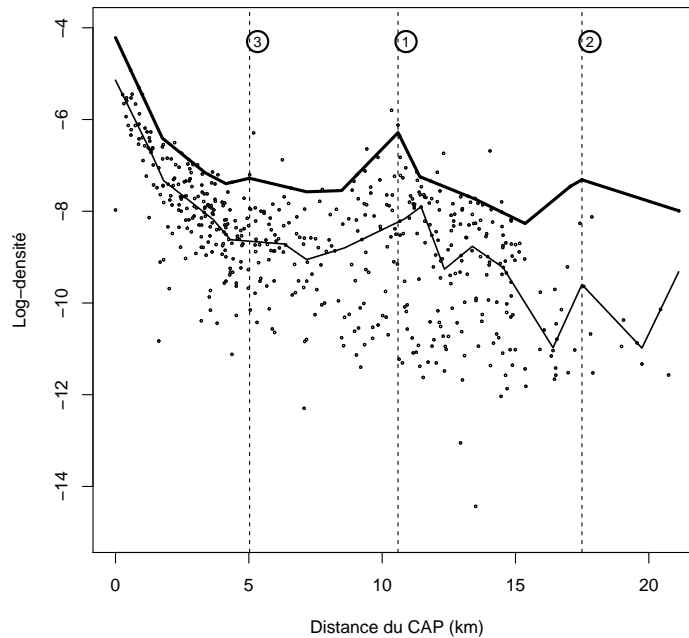


FIGURE 3 – Estimation de quantiles des emplois ( $\lambda$  obtenu avec itération)

tion (2). Le résultat de cette estimation est donné sur la figure 4. Les coordonnées sur les axes correspondent aux longitudes et latitudes, divisées par 10 000 et 100 000, respectivement. Les courbes de niveau correspondent à la surface de quantile de niveau 95%. La surface de quantile atteint un maximum là où la couleur devient foncée et c'est en ces points que nous identifions les sous-centres. La densité la plus élevée correspond à Lille-centre, et autour on observe les sous-centres de La Madeleine et Marquette. Cette estimation confirme encore l'importance de Roubaix. Nous retrouvons aussi deux sous-centres au niveau de Villeneuve-d'Ascq (ISRIS : 590090704). Au niveau des limites de l'agglomération on voit que sur certains endroits on tend vers un maximum, c'est le cas des communes d'Halluin mais aussi d'Armentières. Dans le cas univarié, seule la commune d'Halluin était identifiée comme sous-centre. On remarque aussi une augmentation de la densité des emplois à la frontière sud-ouest de l'agglomération. C'est la commune de la Bassée qui provoque cette augmentation de la densité.



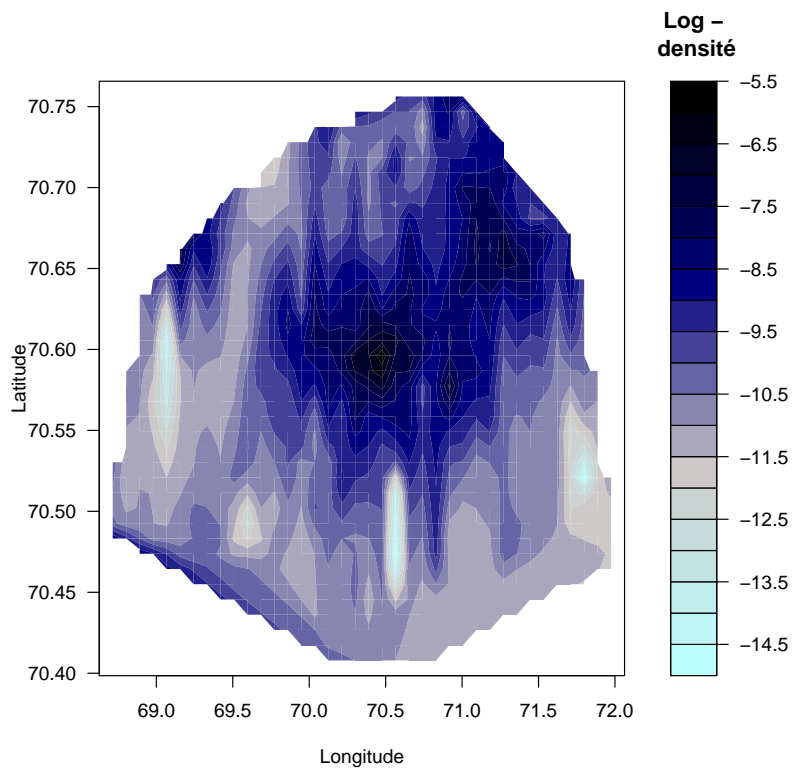


FIGURE 4 – Estimation bivariée de la densité d’emplois.

Avec l'estimation bivariée la relation de proximité est plus réaliste car c'est la distance entre deux points qui est considérée. Dans le cas univarié, deux points peuvent se trouver à proximité alors que la distance qui les sépare est élevée. C'est le cas lorsque les deux points sont diamétralement opposés par rapport au centre ville. Le modèle univarié reprend la structure du modèle monocentrique dans le fait que la forme urbaine (la densité) en un point donné dépend de la distance qui sépare ce point du centre ville. L'agglomération lilloise ne semble pas, à priori monocentrique. C'est en particulier la présence des villes de Roubaix et de Tourcoing, centres historiques des industries du textile qui favorise cette thèse.

Pour vérifier cet argument, on peut effectuer une estimation avec la distance par rapport au centre comme variable explicative et la log-densité comme variable expliquée. Le modèle linéaire est

$$y_i = \alpha + \beta D_i + \epsilon_i. \quad (5)$$

Comme  $y_i$  est la log-densité de l'emploi, le paramètre  $\beta$  est le gradient de la densité de l'emploi. Notre estimation donne une valeur de -0.157 pour ce paramètre, ce qui veut dire que, en moyenne, en s'éloignant d'un kilomètre du centre ville la densité d'emploi diminue de près de 16%. Les résultats de cette estimation sont résumés dans le tableau 2 et montrent que même si les coefficients du modèle sont significatifs, le modèle lui-même n'est pas approprié aux données que nous analysons. En effet, la faible valeur de la statistique  $R^2$  est liée à la spécification linéaire dans l'équation (5). Comme nous l'avons observé avec l'estimation de quantile la distribution des densités ne semble pas respecter une structure fonctionnelle apparente et la modélisation linéaire conduit à de grosses restrictions. Ensuite, le modèle linéaire ne traite pas convenablement les problèmes liés à l'autocorrélation spatiale des erreurs. Comme nous l'indiquons ci-dessous, McMillen (2004) a étudié plusieurs jeux de données et a trouvé qu'il est plus convenable d'aborder la modélisation spatiale des densités d'emploi à travers une spécification non paramétrique. Dans la section suivante nous utiliserons cette approche de McMillen.

	Estimation	Erreur	$t$ -value	$p$ -value
Constante	-7,39534	0,10732	-68,91	$<2e - 16$
Gradient	-0,15698	0,01126	-13,95	$<2e - 16$
$R^2$	0,267			
$R^2$ ajusté	0,2656			

TABLE 2 – Estimation de moindre carrées ordinaires.

## 4.2 Approche non paramétrique de McMillen

Nous appliquons la méthode non paramétrique à deux étapes proposée par McMillen (2001), afin de comparer avec les résultats de la section précédente. Le point central dans cette démarche est l'estimation du modèle donné par l'équation (3). Le processus est itératif car il faut commencer par la sélection des candidats à travers une estimation non paramétrique de la distribution des densités d'emploi, et ensuite estimer plusieurs fois ce modèle en éliminant les coefficients présentant de mauvais signes ou une faible significativité. La première étape nous donne un total de 31 zones d'emplois comme candidats pour un sous-centre d'affaires avec un risque  $\alpha = 20\%$ . Nous avons ensuite regroupé les candidats se situant à proximité l'une de l'autre et nous avons retenu la seule zone ayant la densité la plus élevée. Cela nous permet de réduire la multiplicité des candidats se situant dans une petite région géographique. A partir de ce groupe on sélectionne ainsi 16 candidats.

Sur le tableau 3 figurent les zones d'emplois qui sont sélectionnées au terme de la première étape et qui sont au nombre de 16. Les résultats des estimations sont donnés pour la première itération (Initiale) et pour la dernière itération (Finale). Au niveau de la première itération certains coefficients  $\delta_1$  présentent des signes positifs, alors qu'il n'y a pas de problème de signe pour les coefficients  $\delta_2$ . Nous commençons par supprimer le coefficient  $\delta_1$  du modèle lorsqu'il est positif. Un tel signe indique que la densité de l'emploi augmente lorsqu'on s'éloigne du sous-centre en question, ceci étant incompatible avec la définition d'un sous-centre. Une fois tous les coefficients négatifs, nous commençons par ne retenir, pour les sous-centres ayant

encore les deux coefficients  $\delta_1$  et  $\delta_2$ , que les coefficients les plus significatifs. Dans la dernière étape on élimine les sous-centres ayant un coefficient avec une  $p$ -valeur inférieure à 20%. Avec nos données, cinq itérations ont été suffisantes pour aboutir au résultat du tableau 3. Notons que c'est la variable  $\delta_2$  qui est retenue dans quasiment tous les cas. L'inverse de la distance varie plus rapidement que la distance. Ainsi, la distance mesure une influence de grande envergure du sous-centre alors que l'inverse de la distance mesure une influence du sous-centre sur les zones d'emplois qui sont à sa proximité. Donc, la majorité des sous-centres identifiés par cette procédure exercent plutôt une influence limitée dans l'espace. La seule exception ici est par rapport à Roubaix pour lequel nous avons délibérément gardé les deux coefficients, car c'est le seul sous-centre pour lequel la distance, et non pas l'inverse de la distance, est significatif. Ce sous-centre se distingue clairement par une influence importante et structurellement comparable à celle du centre d'affaire principal (centre de Lille), même si la valeur de  $\delta_1$  est relativement proche de zéro.

On peut voir qu'à travers cette approche on identifie finalement cinq centres d'affaires. Des similarités et des différences par rapport aux régressions quantile, présentées dans la section précédente, sont à noter. Les différences sont particulièrement annoncées par rapport à l'estimation unidimensionnelle. Halluin ne figure plus dans l'ensemble des sous-centres sélectionnés, alors qu'Armentières est présent dans cette liste. Même la régression quantile bidimensionnelle ne faisait pas apparaître Armentières comme particulièrement important par rapport à Halluin. Tourcoing n'apparaissait pas comme sous-centre alors que cette commune est présente dans la liste. Il est à noter que Lambersart est la commune la plus proche du centre ville et identifiée comme sous-centre. Sur la figure 3, correspondant à l'estimation de quantile bidimensionnelle, on peut remarquer la présence d'un étirement de la densité vers le nord-ouest, là où se trouve Lambersart, mais on n'identifie pas un maximum local de la densité d'emploi à ce niveau. Ces différentes remarques nous laissent penser que l'approche de McMillen (2001) est plus appropriée à la structure urbaine de Lille.

Nous avons effectué une analyse de sensibilité en introduisant une perturbation au niveau de la matrice des distances. Cet exercice nous permet de tester la robustesse de nos résultats.

Candidats		Estimations			
Commune	Iris	Initiales		Finales	
		$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$
	590090207	-169 (0.67)	4.23 (0.67)		
	590090305	-304 (0.47)	18.71 (0.27)		
	590090504	519 (0.20)	12.69 (0.35)		
<b>Villeneuve-d'Ascq</b>	590090704	634 (0.12)	25.79 (0.06)		0.25 (0.04)
<b>Armentières</b>	590170402	855 (0.97)	76.30 (0.00)		0.80 (0.00)
Lambersart	593280501	174 (0.67)	16.58 (0.18)		0.18 (0.15)
Lonnoy	593320000	302 (0.38)	25.12 (0.07)		0.20 (0.14)
	593500702	610 (0.12)	13.82 (0.12)		
	593501001	285 (0.48)	5.75 (0.40)		
	593600104	-652 (0.09)	8.49 (0.44)		
	593680202	-330 (0.43)	6.36 (0.50)		
	594100302	-410 (0.34)	7.86 (0.30)		
	595070303	-197 (0.60)	6.32 (0.59)		
<b>Roubaix</b>	595120502	-940 (0.01)	31.72 (0.00)	0.004 (0.00)	0.72 (0.00)
<b>Tourcoing</b>	595990902	988 (0.00)	70.99 (0.00)		0.30 (0.00)
	596480201	862 (0.82)	16.71 (0.15)		

TABLE 3 – Résultats de la procédure à deux étapes (entre parenthèses la  $p$ -value ; pour  $\delta_2$  les chiffres sont à multiplier par  $10^{-2}$ )

	Estimation	Erreur	$t$ -value	$p$ -value
Constante	-9,3680	1,56700	-5,977	0
Gradient	-0,2079	0,03574	5,818	0

TABLE 4 – Estimation du modèle semi-paramétrique de McMillen (2004)

Nous avons ajouté pour chaque élément de cette matrice un choc tiré d’une variable aléatoire normale de moyenne nulle et de faible variance (de l’ordre de  $10^{-3}$ ). En répétant plusieurs fois ces simulations, tous les sous-centres, à l’exception d’Armentières sont sélectionnés dans tous les cas. Dans certains cas, Lannoy est sélectionné alors qu’Armentières ne l’est pas. La deuxième remarque concerne le coefficient  $\delta_1$  qui apparaît parfois pour Tourcoing au lieu de Roubaix. C’est clairement la proximité de ces deux sous-centres qui génère ce résultat.

Pour mesurer l’impact de la proximité à un sous-centre d’affaires on utilise le modèle de McMillen (2001) donné par l’équation (4). Le résultat de cette estimation est donné dans le tableau 4. Nous utilisons les résultats du modèle précédent et nous partons de l’ensemble des cinq sous-centres que nous avons déjà identifié. La valeur du gradient suggère que lorsqu’on s’éloigne du sous-centre d’affaires le plus proche d’un kilomètre, la densité de l’emploi diminue d’un peu plus de 20%. Dans les estimations données sur le tableau 3 les effets des sous-centres sont cumulés puisque qu’ils sont simultanément estimés, alors que dans le modèle actuel, seul le sous-centre le plus proche est pris en considération. McMillen (2004) montre, à travers plusieurs exemples, que l’estimation non paramétrique basée sur l’équation (4) est structurellement plus appropriée qu’un modèle linéaire et qu’elle permet de réduire les problèmes liés à l’autocorrélation spatiale des erreurs.

## 5 Conclusion

Il est important de comprendre la structure interne des villes. Cette structure interne implique de complexes interactions entre les zones d’emploi. Cette question est à la fois utile sur le plan théorique, mais aussi sur le plan pratique afin d’affiner les politiques ur-

baines d'emploi et de transport, notamment. L'objectif de ce travail est de comprendre la distribution des emplois sur l'aire urbaine de Lille et identifier les zones qui exercent une influence significative au niveau de la localisation des activités économiques. La revue de la littérature montre une évolution des concepts utilisés et des difficultés pour aboutir à un consensus autour de la définition d'un sous-centre. Les travaux les plus aboutis considèrent qu'un sous-centre est un endroit avec une densité (relativement) élevée, mais qui en plus exerce une influence significative de la densité d'emploi de son voisinage. Le sens exact de « niveau élevé de densité » doit être déduit de la distribution des emplois, et dans ce cadre les méthodes non paramétriques et semiparamétriques semblent plus robustes que les méthodes paramétriques. Ainsi, le modèle exponentiel, très utilisé dans la pratique, n'est pas approprié dans une agglomération avec une structure polycentrique.

Nous avons appliqué plusieurs méthodes non paramétriques afin d'identifier les sous-centres d'affaires. Ces sous-centres permettent, d'une part, un regroupement des emplois qui échappent au CAP et, d'autre part, à certaines entreprises de bénéficier de certaines forces d'agglomération tout en subissant moins de coûts externes (congestion, pollution). Notre analyse révèle une importance de premier ordre pour Roubaix. L'analyse de sensibilité montre que cet effet semble bien lié à la proximité de la ville de Tourcoing. Nous avons identifié l'existence d'autres sous-centres comme Villeneuve-d'Ascq et Armentières. D'autres sous-centres sont présents mais leur significativité statistique est plus faible. C'est le cas, par exemple, d'Halluin. Cette commune à la périphérie de l'agglomération est identifiée seulement par une estimation de quantile univarié, mais semble être dominée par Armentières à travers les autres méthodes. Les communes de « Lannoy » et de « Lambersart » présentent les mêmes propriétés. L'estimation de quantile est assez rapide à mettre en œuvre et présente des propriétés statistiques intéressantes, mais présente l'inconvénient de trop simplifier la géométrie de la ville. En effet, on se limite à l'identification des rayons au niveau desquels les sous-centres existent. Il faut ensuite explorer ces rayons manuellement pour conclure pour les sous-centres. L'estimation de quantile bivarié de Craig *et al.* (No date) cherche à éviter cet inconvénient, mais elle semble moins efficace dans les zones périphériques où

le nombre d'observations est faible. Pour la structure des emplois dans Lille Métropole, l'approche de McMillen (2004) présente de bonnes propriétés. La possibilité de faire une analyse de sensibilité est assez naturelle puisqu'elle revient à introduire une perturbation dans la matrice des distances, ce qui nous permet de tester la robustesse de nos résultats.

Plusieurs extensions de ce travail sont possibles. Il y a d'abord l'utilité de multiplier les applications pour d'autres agglomérations afin d'étudier les qualités des différents estimateurs sur plusieurs jeux de données. Nous sommes arrivés à la conclusion que l'approche de McMillen est plus adaptée à la structure urbaine de Lille Métropole, mais il sera utile de voir si c'est le cas pour les villes européennes de manière générale. Notons que Lille Métropole est assez particulière, dans le sens où elle est composée d'importantes agglomérations et zones d'emploi, autres que Lille-centre. Ce travail repose sur des hypothèses théoriques, dont notamment la définition de sous-centres d'affaires. Ce point est évoqué par Craig *et al.* (2016), cette question importante peut aussi être discutée dans un cadre analytique d'économie urbaine, mais la difficulté étant que les modèles de la ville polycentrique sont peu développés, et les quelques modèles existants sont difficilement transposables à un cadre empirique.

## Références

- Anne AGUILERA : Growth in commuting distances in French polycentric metropolitan areas : Paris, lyon and marseille. *Urban studies*, 42(9):1537–1547, 2005.
- Anne AGUILERA et Dominique MIGNOT : Urban sprawl, polycentrism and commuting. a comparison of seven French urban areas. *Urban Public Economics Review*, 1(1):93–114, 2004.
- Anne AGUILERA, Sandrine WENGLANSKI et Laurent PROULHAC : Employment suburbanisation, reverse commuting and travel behaviour by residents of the central city in the Paris metropolitan area. *Transportation Research Part A : Policy and Practice*, 43(7):685–691, 2009.
- William ALONSO : *Location and Land Use*. MA : Harvard University Press, 1964.



- Catherine BAUMONT, Cem ERTUR et Julie GALLO : Spatial analysis of employment and population density : the case of the agglomeration of Dijon 1999. *Geographical Analysis*, 36(2):146–176, 2004.
- Pierre-Philippe COMBES, Gilles DURANTON et Laurent GOBILLON : The identification of agglomeration economies. *Journal of Economic Geography*, page lbq038, 2010a.
- Pierre-Philippe COMBES, Gilles DURANTON, Laurent GOBILLON, Diego PUGA et Sébastien ROUX : The productivity advantages of large cities : Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594, 2012.
- Pierre-Philippe COMBES, Gilles DURANTON, Laurent GOBILLON et Sébastien ROUX : Estimating agglomeration economies with history, geology, and worker effects. *In Agglomeration Economics*, pages 15–66. University of Chicago Press, 2010b.
- Steven G CRAIG, Janet E KOHLHASE et Pin NG : statistical identification of employment subcenters. mimeo, No date.
- Steven G CRAIG, Janet E KOHLHASE et Adam W PERDUE : Empirical polycentricity : The complex relationship between employment centers. *Journal of Regional Science*, 56(1):25–52, 2016.
- Steven G CRAIG et Pin T NG : Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *Journal of Urban Economics*, 49(1):100–120, 2001.
- Masahisa FUJITA : *Urban economic theory*. Cambridge University Press, 1989.
- Genevieve GIULIANO et Kenneth A SMALL : Is the journey to work explained by urban structure? *Urban studies*, 30(9):1485–1500, 1993.
- Roger KOENKER et Pin NG : A frisch-newton algorithm for sparse quantile regression. *Acta Mathematicae Applicatae Sinica*, 21(2):225–236, 2005.
- Roger KOENKER, Pin NG et Stephen PORTNOY : Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

John F MCDONALD : The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2):242–258, 1987.

Daniel P MCMILLEN : Nonparametric employment subcenter identification. *Journal of Urban economics*, 50(3):448–473, 2001.

Daniel P MCMILLEN : Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas. *Journal of regional science*, 44(2):225–244, 2004.

Ryohei NAKAMURA : Agglomeration economies in urban manufacturing industries : a case of japanese cities. *Journal of Urban Economics*, 17(1):108–124, 1985.