

# **Distribution des emplois et identification des sous-centres d'affaires dans l'agglomération lilloise**

## **Employment distribution and identification of business subcenters in Lille Metropolitan area**

**Amir ABOUBACAR**

Université de Lille, LEM-CNRS (UMR 9221)  
aboubacar.amiri@univ-lille.fr

**Hakim HAMMADOU**

Université de Lille, LEM-CNRS (UMR9221)  
hakim.hammadou@univ-lille.fr

**Moez KILANI**

Université du Littoral Côte d'Opale (ULCO), TVES  
moez.kilani@univ-littoral.fr  
Auteur correspondant

**Mots-clés** : estimation non-paramétrique, forme urbaine, sous-centres d'affaires

**Keywords** : nonparametric estimation, urban form, business subcenters

**Classification JEL** : C14 ; R12 ; R14

## Résumé

*Nous étudions la distribution des emplois sur la métropole lilloise afin d'identifier les principaux sous-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. L'identification des sous-centres d'affaires est importante car ces zones d'emplois 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. Les méthodes statistiques que nous utilisons permettent de prendre en compte ce problème. 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 %.*

## Abstract

*Understanding the internal structure of cities is an important problem in urban economics. The interaction between major employment locations plays an important role in shaping the city and how households and firms locate. The distribution of employment in urban areas has been studied by several researchers from theoretical and empirical perspectives.*

*The objective of this paper is to study the distribution of employment in the city of Lille in order to identify the most important secondary business districts (SBD). Such identification is an important step in understanding how these centers influence the urban form and how they should be taken into account in transport and urban policies. We do not only consider areas with large employment densities, we also focus on those areas that have an influence on the spatial distribution of employment. The literature review on this question shows the evolution of the concepts and the difficulties in reaching a consensus around the definition of a sub-center. Based on some relevant works, one can define a subcenter as a city with (i) a (relatively) high influence on the global distribution of the density of employment in a metropolitan area and (ii) a significant influence on the density of employment of its neighborhood. The technical difficulty is to isolate the influence of the main business center (the city center), and focus on the effects produced by the multiple business subcenters. We use statistical methods that are suitable for this problem and obtain a set of results confirming the polycentric structure of the city of Lille and showing how the commune of Roubaix (and Tourcoing) plays a particularly important influence. On average, moving away from a business subcenter decreases employment density by about 20%.*

## Points-clés

- La distribution des emplois dans la métropole Lilloise révèle une structure polycentrique.
- La commune de Roubaix a une influence sur la distribution des emplois structurellement comparable à celle de Lille-Centre.
- Une approche semi paramétrique est bien adaptée à l'analyse de la distribution spatiale des emplois
- Nous quantifions l'impact de la proximité à un sous-centre sur la densité des emplois (gradient).

- 1 -

## Introduction

Le modèle de la ville monocentrique (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. En parallèle, cela entraîne 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'agglomération 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.*, 2010 ; Combes *et al.*, 2011 ; Combes *et al.*, 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 regroupées 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. En effet, plusieurs travaux empiriques ont insisté sur les effets que peuvent avoir les sous-centres d'affaires sur l'emploi urbain, la rente foncière et la localisation des ménages (Bender et Hwang, 1985 ; Gordon *et al.*, 1986 ; Griffith, 1981 ; J F McDonald et McMillen, 1990 ; Cervero et Wu, 1997 ; Craig et Kohlhase, 2003).

L'observation des données d'emplois de l'agglomération de Lille Métropole (LM) suggère une structure multicentrique, et on parle d'ailleurs couramment de Lille-Roubaix-Tourcoing. Les villes de Roubaix et Tourcoing ont été 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 deux approches statistiques pour analyser la distribution actuelle de l'emploi au niveau de LM. La première est une méthode non-paramétrique consistant à effectuer une régression quantile de la densité d'emplois sur la distance au CAP (Craig et Ng, 2001). La deuxième approche développée par McMillen (2004), et davantage basée sur des éléments du modèle monocentrique, combine une méthode non-paramétrique et une méthode semi-paramétrique. Notre objectif est d'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 à McDonald (1987). Ce dernier 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, autre que le CAP, où la densité de l'emploi est jugée élevée. C'est en effet cette caractérisation des sous-centres d'affaires qui a été 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 seuil à partir de laquelle on peut juger que le niveau de l'emploi est élevé. En plus, si ce seuil convient à une agglomération, il ne conviendra pas nécessairement à une autre. 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. Bien que cette approche présente des avantages techniques au niveau de la méthodologie d'estimation, elle présente deux inconvénients. Premièrement, 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.* (2016), qui ont effectué une estimation de surface de quantile en deux dimensions. Le deuxième problème est principalement soulevé par McMillen (2001) et McMillen (2004). Ce dernier se base sur des arguments plus forts de l'économie urbaine, notamment le modèle monocentrique, et stipule qu'un sous-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 influence significative sur son propre voisinage. 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.

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 sous-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 : Paris, Lyon, Marseille-Aix, Bordeaux, Grenoble, Dijon et Saint-Étienne. Leur analyse s'intéresse à la mobilité et distingue les centres urbains proches du centre-ville de ceux situés dans la périphérie. Les premiers sont à proximité du centre principal, tandis que les deuxièmes sont situés plus à la périphérie, mais à proximité des principaux axes de transport. Sur la période allant de 1990 à 1999, les auteurs observent un développement des sous-centres. En revanche, ils émettent des réserves vis-à-vis de la poursuite de cette dynamique car ce développement est accompagné par un important rallongement des distances de trajets. Amara *et al.* (2010) utilisent une ESDA<sup>2</sup> pour étudier la distribution spatiale de l'emploi et de la population dans le grand Tunis à partir de données de 1994 et 2004. Ils observent le développement d'une structure polycentrique de Tunis en une dizaine d'années.

La distribution spatiale des emplois a été étudiée par un certain nombre d'auteurs dans un cadre plus général. Dans l'introduction d'un numéro spécial de la RERU consacré à la localisation des activités économiques, Bazillier *et al.* (2014) abordent la question de la localisation des emplois et le développement des territoires en se basant sur des éléments de la nouvelle économie géographique (Krugman, 1993). Ils analysent la littérature et notent la diversité des approches utilisées. Ils expliquent que ces diverses approches sont complémentaires entre elles et distinguent trois principaux axes : une approche centrée sur la localisation du travail et le capital humain, une approche fondée sur la localisation des entreprises et le capital physique et une dernière approche fondée sur la localisation des services et le capital technologique. Pour les auteurs, cette diversité permet de mieux comprendre les questions de spécialisation et de localisation des activités. Wenglenski (2004) étudie la distribution des emplois en Île-de-France et développe un indicateur d'accessibilité à l'emploi qui lui permet d'identifier une ségrégation entre les différentes catégories sociales. Polèse et Shearmur (2007) considèrent la distribution des emplois au Québec entre 1971 et 2001 et identifie une relation entre spécialisation des activités et distance par rapport au CAP. Cette relation semble être stable sur la période étudiée.

Concernant notre analyse dans le présent article, nos résultats confirment la polycentricité de LM et indiquent que la ville de Roubaix exerce une influence particulièrement importante sur la distribution de l'emploi. Roubaix est un sous-centre qui se distingue des 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 déterminer laquelle est la meilleure. Sur le plan empirique, et par rapport à notre application, les résultats obtenus via l'approche proposée par McMillen (2001) donne des résultats plus convaincants et plus faciles à interpréter. D'une part

la méthodologie conduit à une identification du sous-centre et pas seulement sa distance par rapport au CAP. D'autre part, cette méthode se base sur une définition d'un sous-centre qui est plus enracinée dans l'économie urbaine. Ceci se matérialise par la prise en compte dans le modèle économétrique à la fois de la distance à la zone d'emplois et de son inverse. Cette distinction permet de séparer une influence limitée au voisinage immédiat du sous-centre et une influence d'envergure comparable à celle du CAP.<sup>3</sup> La mise en œuvre pratique de cette méthode reste toutefois plus laborieuse étant donné sa structure itérative. Dans la dernière partie de notre analyse, 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. L'originalité de notre contribution réside dans l'utilisation de plusieurs méthodes statistiques que nous adaptons à notre problématique. À titre d'exemple, notre application empirique ne se limite pas à l'identification des sous-centres mais procède aussi à leur classification par ordre d'importance.

L'article 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 la quatrième 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. Tous les Graphiques et les Tableaux ont été produits par les auteurs.

## - 2 -

## Méthodologie

Dans notre analyse on utilisera deux méthodes statistiques développées dans la littérature et présentées ci-dessous. Cela nous permet de les comparer et de tester empiriquement leur efficacité sur la base de données issues de l'agglomération lilloise. Par conséquent, nos conclusions empiriques sont robustes. Comme nous le discutons ci-dessous, dans notre application, nous avons procédé à quelques ajustements qui nous sont apparus utiles.

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) un niveau élevé de la densité nette d'emploi, (3) un ratio élevé emplois/population, (4) un niveau élevé de la densité brute de la population et (5) un niveau élevé de la densité nette de la population.<sup>4</sup> 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.

Craig et Ng (2001) utilisent une régression de quantile afin de contourner le choix d'un seuil exogène. Pour une distance  $x$  du CAP, 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 d'emplois) sachant  $X = x$ , où  $X$  est la variable aléatoire égale à la distance au CAP. L'idée ici est de se focaliser sur les valeurs de  $\tau$  proches de 1 de sorte à repérer les lieux 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. Par exemple,  $\max |g''(x)|$  peut être utilisé. 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 AIC ou BIC<sup>5</sup> 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. Pour notre analyse, en plus de cette méthode, nous adoptons une deuxième approche. Celle-ci consiste à fixer de manière séquentielle le nombre de sous-centres (un sous-centre, deux sous-centres, trois sous-centres, etc.) et à sélectionner la valeur de  $\lambda$  correspondante à chacune de ces situations.

Cette démarche permet en même temps de hiérarchiser les sous-centres par ordre d'importance. 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.* (2016).<sup>6</sup> 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 :

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

où  $n$  est le nombre d'observations, et où 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 bivariée 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). Néamoins, cette méthode conduit à des erreurs d'interpolation plus importantes que la régression unidimensionnelle.

McMillen (2001) propose une approche 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, qui est non-paramétrique, 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 multicolinéarité dans les distances. Dans la deuxième étape, qui est semi-paramétrique, 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 (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'influence du sous-centre le plus proche sur la distribution des emplois.

## - 3 - Les données

La zone d'étude est Lille Métropole (LM), administrativement appelée la « Métropole Européenne de Lille ». La population totale dans cette zone est supérieure à un million d'habitants (1 141 440 en 2015) sur une superficie de 671,71 km<sup>2</sup>, donc une densité de 1 762 hab/km<sup>2</sup>. Historiquement on parle de Lille-Roubaix-Tourcoing qui constituait un grand pôle pour l'industrie textile. LM est située au nord du bassin minier du Nord-Pas-de-Calais. Depuis des dizaines d'années la délocalisation de l'industrie du textile et la cessation des activités minières ont conduit à des mutations urbaines importantes dans LM.

Les données sont à l'échelle de l'IRIS.<sup>7</sup> Ces données de l'année 2010 proviennent du recensement démographique réalisé sur l'agglomération Lilloise. Nous ne disposons pas de données plus récentes à cette échelle. De plus, la distribution des emplois n'est pas très volatile à travers le temps et l'utilisation de données datant de quelques années ne devrait pas poser de problèmes. 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 CEREMA Nord-Picardie<sup>8</sup> qui avait géolocalisé l'emploi au niveau de l'IRIS selon le secteur d'activité. La base

de données SIRENE qui décrit la structure et la démographie des établissements a été utilisée pour aboutir à ces données. 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.

**Tableau 1 – Statistiques descriptives (données exprimées mètres)**

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

LM une agglomération multicentrique avec trois villes historiques : Lille, Roubaix et Tourcoing. Cette structure est visible sur la Figure 1(b) 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 1(a) 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 -

## Résultats empiriques

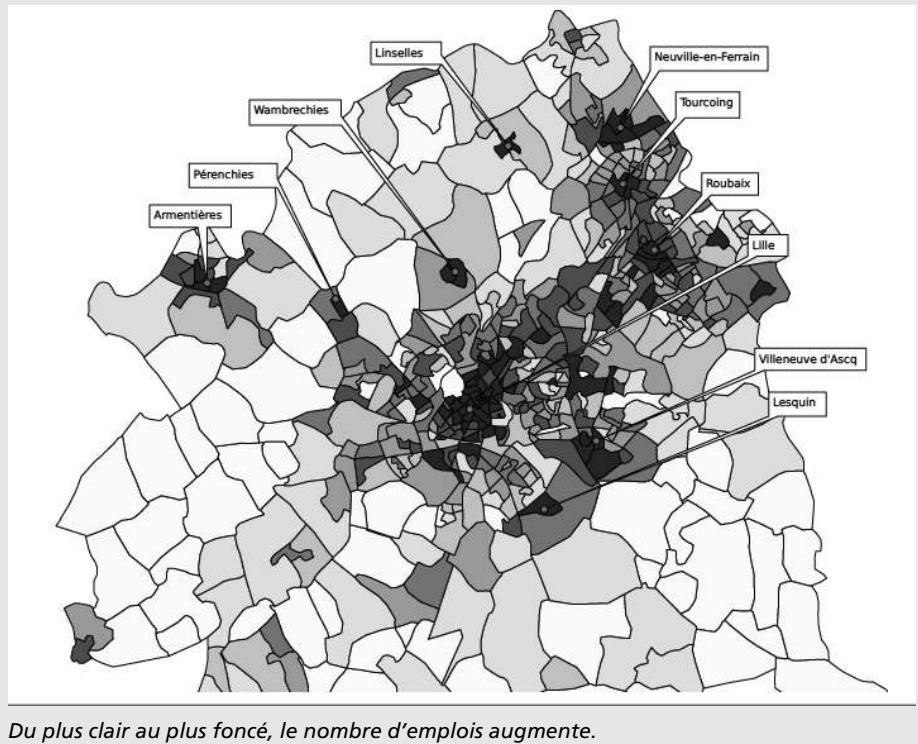
Dans les deux sous-sections suivantes nous appliquons les deux méthodes décrites dans la section 2 afin de caractériser la forme urbaine de LM.

### 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<sup>9</sup>. La courbe du 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, à mesure que l'on s'éloigne du CAP. En fixant la valeur du paramètre

$\lambda$  de sorte à minimiser le critère d'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 10 km, c'est Roubaix qui constitue un sous-centre d'affaires important. À 17,5 km

**Figure 1 – La distribution des emplois sur l'agglomération lilloise**



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, qui sont 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. 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

**Figure 2 – La distribution des densités d'emplois sur l'agglomération lilloise**

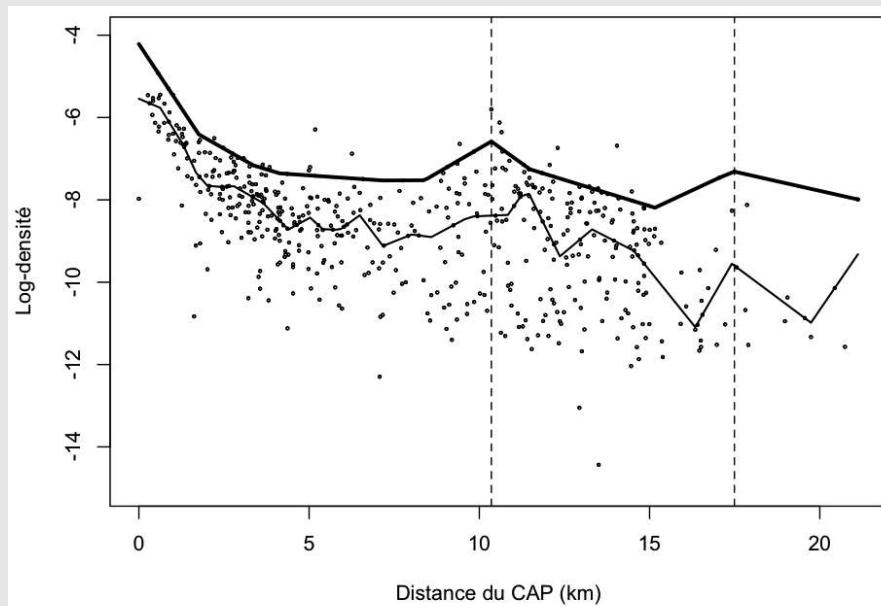


*Du plus clair au plus foncé, la densité augmente.*

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, 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  $\lambda$  de l'équation (1), est discutable. En particulier, dans l'application empirique on doit choisir une grille de valeurs dans laquelle on optimise le critère AIC. Plus les valeurs de  $\lambda$  sont élevées, plus la sélection de sous-centres est stricte (moins de zones seront considérées comme des sous-centres), et plus les valeurs de  $\lambda$  sont faibles, et moins la sélection de sous-centres est stricte. Dans cette application, ce critère est optimisé sur une grille allant de 100 à 1 000 m avec un pas de 50 m. Ainsi on a pu identifier deux sous-centres. Afin de contourner ce choix arbitraire de grille, nous proposons de modifier la procédure initiale et d'adopter 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 sous-centre le plus influent sur la forme urbaine est celui qui apparaît en premier, le deuxième est

**Figure 3 – Estimation de quantiles des emplois.** Dans ce graphique, les valeurs de  $\lambda$  qui minimisent le critère AIC sont 383.52 pour la médiane et 767.4 pour le quantile d'ordre 0.95

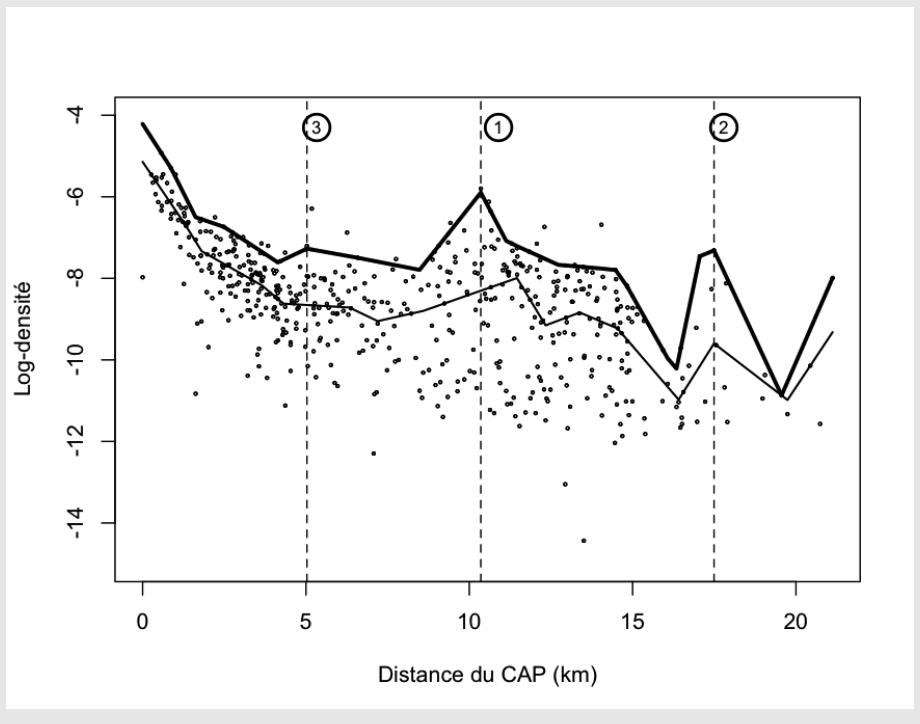


Les points correspondent aux valeurs observées ; la ligne épaisse correspond à la régression quantile d'ordre 0,95 ; la ligne fine correspond à la régression médiane.

celui qui apparaît en deuxième lieu et ainsi de suite. Nous nous arrêtons lorsque la recherche ne fait plus apparaître de zones d'emplois 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. Avec cette procédure et en fixant le nombre de sous centres à sélectionner à trois, nous aboutissons aux courbes de quantiles représentées dans 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 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

**Figure 4 – Estimation de quantiles des emplois. Les valeurs de  $\lambda$  obtenues lorsque trois sous-centres d'affaires sont à sélectionner valent 650 pour la médiane et 250 pour le quantile d'ordre 0.95.**

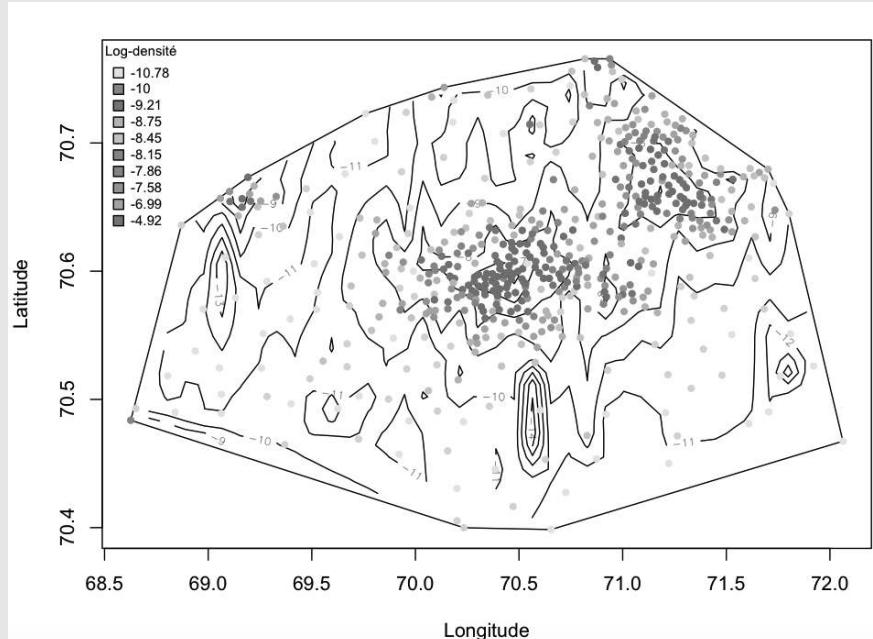


*les points indiquent les valeurs observées; la ligne épaisse correspond à la régression quantile d'ordre 0,95 ; la ligne fine correspond à la régression médiane.*

de 15 km est trop 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'Est de Lille-centre. Notons que la déformation de la courbe de quantiles à 95 % est faible, 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. On explore les zones d'emplois qui se trouvent à un rayon égal à une distance fournie par la régression afin de repérer celles ayant les densités les plus élevées (voir section 2). 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. Comme indiqué dans la section 2, il est possible de généraliser cette méthode unidimensionnelle à une régression de quantile bivariée dans laquelle chaque zone est repérée à la fois par sa latitude et sa longitude. Le résultat de cette estimation est donné sur la Figure 4.

**Figure 5 – Estimation bivariée de la densité d'emplois**



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. Juste autour de cette zone d'emplois on identifie 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. Sur les bords de la zone d'étude on observe l'existence de certains pics qui correspondent aux communes d'Halluin et 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.

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, *a 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 régression de la log-densité sur la distance par rapport au centre. Le modèle linéaire est

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

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 qu'en moyenne, en s'éloignant d'un kilomètre du centre-ville la densité d'emplois diminue de près de 16 %.

**Tableau 2 – Estimation par moindre carrées ordinaires**

	Estimation	Erreur	t-value	p-value
<b>Constante</b>	-7,40	0,11	-68,91	$<2e-16$
<b>Gradient</b>	-0,16	0,01	-13,95	$<2e-16$
$R^2$	0,27			
$R^2$ ajusté	0,27			

Les résultats de cette estimation sont résumés dans le Tableau 2. Ils 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 la régression quantile (Figure 2), la distribution des densités ne semble pas respecter une structure fonctionnelle apparente. Ceci explique l'inadéquation de la modélisation linéaire dans ce contexte. De plus, le modèle linéaire ne prend pas en compte l'autocorrélation spatiale des erreurs. À travers plusieurs exemples, McMillen (2004) a conclu qu'une spécification non (ou semi)-paramétrique est préférable à un modèle linéaire pour modéliser la relation entre la densité d'emplois et la distance au CAP. Dans la section suivante nous utiliserons cette approche.

## 4.2. Approche de McMillen (2001)

Dans cette section nous abordons le même problème avec une approche différente. Nous appliquons la méthode proposée par McMillen (2001) présentée dans la section 2. Notre objectif est de vérifier si cette approche conduit aux mêmes conclusions que celles de la section précédente. Le point central dans cette démarche est l'estimation du modèle donné par l'équation (3). Le processus est itératif. On commence par sélectionner des candidats à travers une estimation non-paramétrique de la distribution des densités d'emplois. Ensuite on estime (3) plusieurs fois en éliminant à chaque itération les coefficients présentant de mauvais signes ou une faible

significativité. L'étape de présélection donne un total de 31 zones d'emplois comme candidats pour un sous-centre d'affaires avec un risque  $\alpha = 20\%$ <sup>10</sup>. Nous avons ensuite regroupé les candidats se situant à proximité l'un de l'autre et nous avons retenu une seule zone, celle ayant la densité la plus élevée. Cela nous permet de réduire la multiplicité des candidats se situant dans une petite région géographique. À partir de ce groupe on sélectionne ainsi 16 candidats.

Dans le Tableau 3 figurent les zones d'emplois présélectionnées. Les résultats des estimations du modèle (3) sont donnés pour la première itération (label « Initiales ») et pour la dernière itération (label « Finales »). 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, ce qui est incompatible avec la définition d'un sous-centre. Lorsque le modèle ne présente que des coefficients négatifs, nous gardons les zones correspondant à deux coefficients  $\delta_1$  et  $\delta_2$  significatifs. Dans la dernière étape on élimine les sous-centres ayant un coefficient avec une  $p$ -value inférieure à 5 %<sup>11</sup>. 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 Roubaix, commune pour laquelle nous avons délibérément gardé les deux coefficients. En effet, c'est le seul sous-centre pour lequel la distance, et non l'inverse de la distance, est significative. 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 sous-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 visibles par rapport à l'estimation unidimensionnelle. Halluin ne figure plus dans l'ensemble des sous-centres sélectionnés, alors qu'Armentières est présente dans cette liste. Même la régression quantile bidimensionnelle ne faisait pas apparaître Armentières comme particulièrement importante 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 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. Sur la base

d'un argument qualitatif et de la connaissance *a priori* de la zone d'étude, ces différentes remarques nous laissent penser que l'approche de McMillen (2001) est plus appropriée pour l'étude de la structure urbaine de la métropole lilloise.

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. 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 puisqu'ils sont simultanément estimés, alors que dans le modèle actuel, seul le sous-centre le plus proche est pris en considération.

## - 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 urbaines 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 d'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 sur 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 semi-paramétriques 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.

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

Commune	IRIS	Initiales		Estimations	
		$\delta_1$	$\delta_2$	$\delta_1$	$\delta_2$
	590090207	-169 (0,67)	4,23 (0,67)		
	590090305	-304 (0,47)	18,71 (0,27)		
	590090504	519 (0,20	12,69 (0,35)		
Villeneuve-d'Ascq	590090704	634 (0,12)	25,79 (0,06)		0,25 (0,04)
Armentières	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)
Lannoy	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)		
Roubaix	595120502	-940 (0,01)	31,72 (0,00)	0,004 (0,00)	0,72 (0,00)
Tourcoing	595990902	988 (0,00)	70,99 (0,00)		0,30 (0,00)
	596480201	862 (0,82)	16,71 (0,15)		

**Tableau 4 – Estimation du modèle semi-paramétrique de (McMillen, 2004)**

	Estimation	Erreur	t-value	p-value
Constante	-9,37	1,57	-5,98	0,00
Gradient	-0,21	0,04	5,82	0,00

Nous avons appliqué plusieurs méthodes statistiques 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 manuellement les zones situées dans ces rayons pour décider de la localisation exacte des sous-centres. L'estimation de quantile bivarié 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. Au vu de notre analyse nous concluons 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'emplois, 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 est liée au fait que les modèles de la ville polycentrique sont peu développés, et les quelques modèles existants sont difficilement transposables à un cadre empirique. Il est, par ailleurs, utile de prendre en compte les différents secteurs d'emploi Polèse et Shearmur (2007) et ne pas se limiter à une mesure agrégée pour tous les secteurs.

En effet, cela nous aidera à faire le lien entre la forme urbaine et les économies d'agglomération (Ellison *et al.*, 2010).

Enfin soulignons que la compréhension de la structure de la ville et de sa forme est importante pour identifier les politiques urbaines les plus pertinentes. Par exemple, au niveau du transport public, lorsqu'il n'existe pas de sous-centres d'affaires, le réseau de transport à développer est souvent de type « étoile », avec des lignes radiales reliant les zones résidentielles situées à la périphérie et le centre-ville où se situe le CAP. À l'opposé, lorsqu'il y a une présence de plusieurs de zones d'activités économiques importantes à la périphérie, le réseau de transport doit répondre à la demande de mobilité entre ces sous-centres. À titre d'illustration, le développement des activités économiques dans les banlieues autour de Paris est à l'origine du projet « Grand Paris Express » dont l'objectif est de relier directement ces zones d'emplois en évitant des passages par le centre de Paris. Dans un registre similaire, la politique de logement social est dépendante des niveaux de prix (loyers et ventes) sur le marché immobilier. L'équilibre sur le marché immobilier est lui-même dépendant de la présence de sous-centres. Mais plus important encore, les décisions de localisation des ménages dépendent des lieux de travail et des lieux où des opportunités futures de travail peuvent apparaître. Il y a donc une dépendance à la présence de sous-centres. Plusieurs arguments dans ce sens peuvent être développés afin de souligner l'importance de la caractérisation de la forme urbaine et donc de l'identification des sous-centres d'affaires. Dans une revue de la littérature se focalisant sur la recherche en France, et en rapport avec des questions d'emploi et de métropolisation, Perrin et Rousier (2002) parlent du travail urbain, ses multiples dimensions et les différentes approches utilisées pour l'étudier. Ils évoquent, en particulier la relation entre la taille de la ville et le niveau de chômage dans celle-ci, une relation initialement étudiée par Maillard (1997). Ce dernier a observé que le niveau de chômage augmente avec la taille de la ville, mais que cette relation est inversée pour les quelques métropoles ayant une population dépassant un million d'habitants. Les travaux portant sur la relation entre le marché de l'emploi et la forme urbaine sont encore en développement et les applications empiriques restent encore limitées. Étant donné l'importance des politiques de l'emploi, cette voie de recherche devrait se développer significativement durant les prochaines années. L'identification des sous-centres d'affaires et la caractérisation de la forme urbaine s'inscrivent en grande partie dans cette perspective.

## Remerciements

Ce travail a bénéficié d'un soutien financier de l'Université de Lille (Projet CUAS). Il a été présenté et discuté au colloque annuel de l'Association de Science Régionale de Langue Française (ASRLDF, Caen 2018). Les retours de la part des participants nous ont permis de clarifier la problématique considérée dans ce travail. La révision de cet article a bénéficié des commentaires et remarques détaillés de deux rapporteurs anonymes et de l'éditrice Camélia Turcu. Nous remercions Marion Romo et Amandine Ghintran pour la relecture du texte ainsi que Ngagne Demba

Diop pour son aide à la mise en forme finale du papier. Les éventuelles erreurs sont de la responsabilité des auteurs.

## Références bibliographiques

- Aguilera A (2005) Growth in commuting distances in French polycentric metropolitan areas: Paris, Lyon and Marseille. *Urban studies* 42(9) : 1537-1547.
- Aguilera A, Mignot D (2004) Urban sprawl, polycentrism and commuting. A comparison of seven French urban areas. *Urban Public Economics Review* 1(1): 93-114.
- Aguilera A, Wenglenski S, Proulhac L (2009) 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.
- Alonso W (1964) Location and land use. Toward a general theory of land rent. *Harvard University Press*.
- Amara M, Kriaa M, Montacer M (2010) Centralité des activités dans la région urbaine de Tunis (1994-2004). *Revue d'Économie Régionale et Urbaine* (3) : 473-509.
- Baumont C, Ertur C, Le Gallo J (2004) Spatial analysis of employment and population density: the case of the agglomeration of Dijon 1999. *Geographical analysis* 36(2): 146-176.
- Bazillier R, Rabaud I, Turcu C (2014) Compétitivité territoriale et localisation du travail et des entreprises: une introduction. *Revue d'Économie Régionale et Urbaine* (2) : 197-217.
- Bender B, Hwang H-S (1985) Hedonic housing price indices and secondary employment centers. *Journal of Urban Economics* 17(1): 90-107.
- Cervero R, Wu K-L (1997) Polycentrism, commuting, and residential location in the San Francisco Bay area. *Environment and Planning A: Economy and Space* 29(5): 865-886.
- Combes P-P, Duranton G, Gobillon L, Roux S (2010) Estimating agglomeration economies with history, geology, and worker effects. In Glaeser E L (ed) *Agglomeration Economics*. University of Chicago Press: 15-66.
- Combes P-P, Duranton G, Gobillon L (2011) The identification of agglomeration economies. *Journal of Economic Geography* 11(2): 253-266.
- Combes P-P, Duranton G, Gobillon L, Puga D, Roux S (2012) The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica* 80(6): 2543-2594.
- Craig S G, Kohlhase J E (2003) Employment subcenters and the distribution of economic activity. *unpublished, Department of Economics, University of Houston*.
- Craig S G, Kohlhase J E, Perdue A W (2016) Empirical polycentricty: the complex relationship between employment centers. *Journal of Regional Science* 56(1): 25-52.
- Craig S G, Ng P T (2001) Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *Journal of Urban Economics* 49(1): 100-120.
- Ellison G, Glaeser E L, Kerr W R (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review* 100(3): 1195-1213.
- Fujita M (1989) *Urban Economic Theory: Land Use and City Size*. Cambridge University Press.
- Giuliano G, Small K A (1993) Is the journey to work explained by urban structure? *Urban Studies* 30(9): 1485-1500.
- Gordon P, Richardson H W, Wong H L (1986) The distribution of population and employment in a polycentric city: the case of Los Angeles. *Environment and Planning A: Economy and Space* 18(2): 161-173.
- Griffith D A (1981) Modelling urban population density in a multi-centered city. *Journal of Urban Economics* 9(3): 298-310.
- Koenker R, Ng P T (2005) A Frisch-Newton algorithm for sparse quantile regression. *Acta Mathematicae Applicatae Sinica* 21(2): 225-236.

## Distribution des emplois et identification des sous-centres d'affaires

- Koenker R, Ng P T, Portnoy S (1994) Quantile smoothing splines. *Biometrika* 81(4): 673-680.
- Krugman P R (1993) *Geography and trade*. MIT Press.
- Maillard M (1997) La taille de la ville influence-t-elle son taux de chômage? Le cas français. *Revue d'Économie Régionale et Urbaine* : 307-320.
- McDonald J F, McMillen D P (1990) Employment subcenters and land values in a polycentric urban area: the case of Chicago. *Environment and Planning A: Economy and Space* 22(12): 1561-1574.
- McDonald J F (1987) The identification of urban employment subcenters. *Journal of Urban Economics* 21(2): 242-258.
- McMillen D P (2001) Nonparametric employment subcenter identification. *Journal of Urban Economics* 50(3): 448-473.
- McMillen D P (2004) Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas. *Journal of Regional Science* 44(2): 225-244.
- Nakamura R (1985) Agglomeration economies in urban manufacturing industries: a case of Japanese cities. *Journal of Urban Economics* 17(1): 108-124.
- Perrin É, Rousier N (2002) Métropolisation, emploi et politiques locales une relecture de quinze années de recherche en France. *Revue d'Économie Régionale et Urbaine* (5) : 707-728.
- Polèse M, Shearmur R (2007) La stabilité des modèles de localisation industrielle . *Revue d'Économie Régionale et Urbaine* (4) : 677-706.
- Wenglenski S (2004) Une mesure des disparités sociales d'accessibilité au marché de l'emploi en île-de-France. *Revue d'Économie Régionale et Urbaine* (4) : 539-550.

## Notes

- 1 - Les auteurs utilisent le terme « reverse commuting » pour désigner ces flux.
- 2 - Exploratory Spatial Data Analysis.
- 3 - Nous développons ce point dans la section 2.
- 4 - Pour une zone donnée, on définit la densité « brute » (resp. « nette ») d'emplois comme le ratio du nombre d'emplois par la superficie de cette zone (resp. la superficie de la partie dédiée à des activités professionnelles dans cette zone). Les densités brute et nette de population sont définies de manière similaire.
- 5 - AIC pour « Akaike Information Criterion » et BIC pour « Bayesian Information Criterion ».
- 6 - Cette idée a été initialement discutée dans un manuscrit non daté : « Craig, Steven G, Janet E Kohlhase, et Pin Ng. «Statistical identification of employment subcenters.» mimeo.
- 7 - IRIS : « Îlots 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.
- 8 - Anciennement le Centre d'Études Techniques de l'Équipement (CETE) de Lille.
- 9 - Notons que dans l'analyse statistique, la densité d'emplois est exprimée en nombre d'emplois par mètre carré. Ce changement d'échelle conduit à des log-densités négatives mais n'a évidemment aucun effet sur les résultats.
- 10 - En dessous de ce seuil  $\alpha$  le nombre de candidats présélectionnés est très limité.
- 11 - Dans le Tableau 3 nous avons également conservé des sous-centres avec des p-values comprises entre 5 % et 20 % afin d'enrichir la procédure de sélection.