UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

TESIS DOCTORAL

TRES ENSAYOS SOBRE UN NUEVO ENFOQUE PARA MEDIR LA CONCENTRACIÓN ESPACIAL DE LAS ACTIVIDADES ECONÓMICAS

THREE ESSAYS ON A NEW APPROACH TO MEASURING SPATIAL CONCENTRATION OF ECONOMIC ACTIVITIES

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

SILIANG WANG

DIRECTORES

FRANCISCO JAVIER VELÁZQUEZ ANGONA
DAVID MARTÍN BARROSO

© Siliang Wang, 2025

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

PROGRAMA DE DOCTORADO EN ECONOMÍA

TESIS DOCTORAL

TRES ENSAYOS SOBRE UN NUEVO ENFOQUE PARA MEDIR LA CONCENTRACIÓN ESPACIAL DE LAS ACTIVIDADES ECONÓMICAS

THREE ESSAYS ON A NEW APPROACH TO MEASURING SPATIAL CONCENTRATION OF ECONOMIC ACTIVITIES

MEMORIA PARA OPTAR AL GRADO DE DOCTORA
PRESENTADA POR

SILIANG WANG

DIRECTORES

FRANCISCO JAVIER VELÁZQUEZ ANGONA
DAVID MARTÍN BARROSO

致我的家人

A mi familia

To my family

AGRADECIMIENTOS Y DEDICATORIAS,

Quiero expresar mi agradecimiento a todas las personas que han hecho posible este trabajo. Este trabajo no habría sido realidad sin el apoyo, la orientación y la confianza de quienes me han acompañado a lo largo de estos años. A todos ellos, les dedico mis más sinceras palabras de gratitud.

En primer lugar, quiero agradecer a la Universidad Complutense de Madrid, donde he tenido el privilegio de formarme a lo largo de más de una década, desde mis estudios de grado hasta la finalización de este doctorado. Durante todos estos años, he tenido la suerte de participar en diversos cursos y seminarios que me han permitido apreciar la riqueza y profundidad del conocimiento en Economía, despertando en mí una verdadera pasión por la disciplina. También quiero extender mi agradecimiento a la Comisión Académica y al grupo de investigación GRIPICO, cuyo respaldo ha sido esencial para la realización de esta tesis.

Mi gratitud más sincera está dirigida a mis padres, quienes han sido mi mayor apoyo, tanto emocional como financiero. Gracias a su esfuerzo incondicional, he podido afrontar los retos de una carrera académica como esta. Su amor, sacrificio y confianza han sido mi mayor fuente de fortaleza, permitiéndome avanzar en los momentos de duda y celebrando conmigo los momentos de éxito. Sin ellos, este logro no habría sido posible.

Por último, quiero expresar mi más profundo agradecimiento a mis directores de tesis, el Dr. Francisco Javier Velázquez Angona y el Dr. David Martín Barroso. Sin su orientación, paciencia y sabiduría, este trabajo no habría sido posible. A lo largo de todo el proceso de investigación, me han ofrecido un apoyo constante, motivándome a superar mis propios límites y a mejorar cada día. Su dedicación y compromiso con mi desarrollo académico han sido invaluables, y por ello, les estoy enormemente agradecido.

RESUMEN

El objetivo de esta tesis doctoral es proponer una nueva familia de métodos basados en la distancia para medir el grado de aglomeración y co-aglomeración. Para ello, se desarrollan tres investigaciones de carácter teórico y empírico que analizan la concentración espacial en diversos contextos urbanos. La investigación combina las dimensiones sectorial y territorial mediante la construcción de indicadores a nivel de establecimiento o punto de producción. Además, se introducen elementos de análisis adicionales y se optimiza, desde una perspectiva computacional, el proceso de contraste de la hipótesis nula de no concentración para los indicadores agregados por sectores y áreas geográficas.

El primer capítulo se centra en un análisis detallado de la aglomeración espacial en las ciudades, tomando como caso de estudio la ciudad de Madrid y utilizando los datos del Censo de Locales y Actividades del Ayuntamiento de Madrid. A través del nuevo método desarrollado para medir esta aglomeración, se comparan los resultados de dos periodos: diciembre de 2014 y diciembre de 2019. El análisis revela una notable concentración en actividades que enfrentan restricciones importantes de localización, como las manufacturas, que hacen un uso intensivo del suelo, tienen un alto valor añadido o están relacionadas con el turismo. Por otro lado, se observa una dispersión significativa en actividades de consumo cotidiano, las cuales requieren proximidad con los consumidores. Desde una perspectiva geográfica, se identifica una persistencia en la alta concentración de actividades en los barrios centrales de la ciudad, así como en algunas zonas periféricas donde se han desarrollado polígonos industriales.

En el segundo capítulo se utiliza la versión ponderada del método para medir la aglomeración sectorial, aplicándolo al sector manufacturero chino con los datos más recientes de 2021. Este enfoque proporciona una visión más precisa de cómo las actividades industriales se agrupan dentro del amplio espacio geográfico. El análisis empírico revela resultados interesantes: la concentración geográfica se produce

principalmente en la Costa Este del país disminuyendo gradualmente hacia el Oeste y el Norte. La aglomeración productiva se concentra en pocos sectores y áreas geográficas específicas (prefecturas). Las prefecturas con un perfil urbano y metropolitano destacado son las que muestran una mayor capacidad para atraer actividad manufacturera en su conjunto. Además, los resultados sugieren que las Zonas de Desarrollo Económico (Economic Development Zones, EDZs) no parecen alterar de forma significativa el patrón general de concentración de las manufacturas chinas.

Finalmente, el tercero capítulo se dedica al estudio de la distribución de la co-aglomeración global de las actividades económicas en la ciudad de Madrid, utilizando el mismo base de datos del primer capítulo, se analiza la interacción entre diferentes sectores y su influencia mutua en la distribución espacial. Los resultados obtenidos muestran una notable estabilidad a nivel individual, sectorial y geográfico entre 2014 y 2019. Desde la perspectiva sectorial, se observa un atractivo significativo de la co-localización en actividades relacionadas con el consumo diario, o el ocio, que requieren proximidad a los consumidores y muestran patrones de co-consumo. Por el contrario, algunas industrias con uso intensivo de suelo, como las manufacturas, la hostelería y las industrias creativas, son menos propensas a presentar un atractivo significativo para la co-localización. Desde una perspectiva geográfica, estas actividades han mostrado un alta global co-aglomeración en los barrios centrales de la ciudad, con gran estabilidad en los dos periodos analizados. Esto sugiere que el fenómeno de la co-aglomeración global es una característica estructural del entorno geográfico, que tiende a cambiar solo a largo plazo. Los resultados de este estudio aportan una nueva perspectiva sobre cómo las actividades económicas se co-agglomeran y coexisten en un espacio urbano, ofreciendo herramientas útiles para la planificación económica y el desarrollo urbano.

ABSTRACT

The objective of this doctoral thesis is to propose a new family of distance-based methods for measuring the degree of agglomeration and co-agglomeration. For this purpose, three theoretical and empirical studies are developed, analyzing spatial concentration in different urban contexts. The research combines both sectorial and territorial dimensions through the construction of indicators at the establishment or point of production level. Additionally, further analytical elements are introduced, and from a computational perspective, the process of testing the null hypothesis of no concentration for aggregated sectoral and geographical area indicators is optimized.

The first chapter focuses on a detailed analysis of spatial agglomeration in cities, using Madrid as a case study and using data from the Census of Premises and Activities of the Madrid City Council. Through the new method developed to measure this agglomeration, the results of two periods -December 2014 and December 2019- are compared. The analysis reveals a significant concentration of activities that face major location restrictions, such as manufacturing, which are land-intensive, have high added value, or are related to tourism. On the other hand, there is notable dispersion in in daily consumption activities, which require proximity to consumers. From a geographical perspective, the analysis identifies a persistence of high concentrations of activities in the central neighborhoods of the city, as well as in certain peripheral areas where industrial parks have been developed.

In the second chapter, the weighted version of the method is used to measure sectoral agglomeration, applying it to the Chinese manufacturing sector with the most recent data from 2021. This approach provides a more accurate view of how industrial activities are clustered within the large geographical space. The empirical analysis reveals interesting results: geographical concentration occurs mainly on the East Coast of the country, gradually decreasing toward the West and North. Productive agglomeration is concentrated in a few sectors and specific geographical areas (prefectures). Prefectures with a prominent urban and metropolitan profile are those that show the higher capacity to attract manufacturing activity. Moreover, the results

suggest that Economic Development Zones (EDZs) do not seem to significantly alter the overall pattern of concentration in Chinese manufacturing.

Finally, the third chapter is dedicated to studying the global co-agglomeration distribution of economic activities in the city of Madrid. Using the same database as in the first chapter, the interaction between different sectors and their mutual influence on spatial distribution is analyzed. The results obtained show a notable stability at the individual, sectoral, and geographical levels between 2014 and 2019. From a sectoral perspective, there is a significant attractiveness for co-location in activities related to daily consumption or leisure, which require proximity to consumers and exhibit patterns of co-consumption. In contrast, some land-intensive industries, such as manufacturing, hospitality, and creative industries, are less likely to present a significant attractiveness for co-location. From a geographical perspective, these activities have shown high global co-agglomeration in the central neighborhoods of the city, with a high stability in both analyzed periods. This suggests that the global co-agglomeration phenomenon is a structural feature of the geographical environment, which tends to change only in the long term. The results of this study offer a new perspective on how economic activities co-agglomerate and coexist in an urban space, providing useful tools for economic planning and urban development.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

Marshall (1890) is often cited as the seminal contribution to the study of productive agglomeration, although Weber (1909) is generally credited as the first to use this specific term. The extensive body of literature that has since analysed the determinants and effects of spatial concentration of productive activities underscores the importance of territory and cities in shaping the location decisions of firms (Krugman, 1991, 1992; Tabuchi, 1998; Porter, 2000; Fujita & Thisse, 2002; Behrens et al., 2014, among many others). Consequently, part of the economic literature has focused on identifying the determinants of this concentration process, introducing the term "agglomeration economies" with its various subcategories, i.e., internal, localisation, and urbanisation economies, or typologies such as intra-industry agglomeration, inter-industry agglomeration, or co-agglomeration (Goldstein & Gronberg, 1984; Helsley & Strange, 1990; Audretsch & Feldman, 1996; Glaeser, 1999; Rosenthal & Strange, 2001; Dumais et al., 2002, among others[1]). On the other hand, another strand of the literature has sought to evaluate the effects of this concentration process, both on the behaviour of industries or firms and on territories. This research addresses both the productive perspective, such as the interrelation with productivity (Ciccone & Hall, 1996; Brülhart & Mathys, 2008; Andersson & Lööf, 2011), and the implications for urban planning (Eberts & McMillen, 1999; Giuliano et al., 2019).

Many of these studies share a common requirement: the need for measures to approximate the "vague" concept of productive agglomeration. This has also been a significant topic of interest in the literature on agglomeration, with its evolution shaped by the availability and nature of the data. Numerous authors, based on the data utilised, classify these measures into two typologies (discrete and continuous) and three generations of indicators (Combes & Overman, 2004; De Dominicis et al., 2013; Chain et al., 2019; Gómez-Antonio & Alañón-Pardo, 2020). The first and second generations rely on data aggregated by geographical areas. In the first generation, existing

[1] See Quigley (1998) for a summary of these early contributions, which are primarily theoretical, focusing on the analysis of the determinants of agglomeration economies.

inequality indicators are adapted to measure the concentration of various activities or sectors, resulting in sectoral agglomeration indicators. Additionally, productive density or specialisation indicators are used to assess the level of concentration in specific territories. In this context, agglomeration becomes both a technological characteristic of sectors and a structural feature of territories (and activities).

Second-generation indicators, also based on discrete data, i.e., geographically aggregated, represent a significant advancement by controlling for industrial concentration within sectors and being derived from a theoretical model. Among these, the indicators proposed by Ellison & Glaeser (1997) and Maurel & Sédillot (1999) stand out. These indicators, the first specifically designed to measure spatial agglomeration, mark a crucial step forward in understanding the elements that need to be controlled when measuring productive agglomeration. They have undoubtedly seen widespread empirical application across various geographic and sectoral contexts. However, these indicators suffer from issues related to the Modifiable Area Unit Problem (MAUP), meaning their sensitivity to the level of aggregation of the geographic data available. Furthermore, they do not account for interactions between production points, such as proximity, and fail to provide measures for the geographic dimension of agglomeration processes, focusing solely on sectoral analysis. As a result, first-generation indicators based on density continue to be used for geographic measurement.

It is for this reason that, almost simultaneously, distance-based indicators emerged, leveraging the availability of georeferenced microdata of production points. These indicators meet the five desirability criteria for agglomeration indicators outlined by Duranton & Overman (2002, 2005):

1. Comparability across industries.
2. Control for overall agglomeration trends across industries within the territory.
3. Clear separation between sectoral and geographic concentration.
4. Unbiased with respect to the degree of spatial aggregation.
5. Development of statistical significance tests for the indicators.

Additionally, Kominers (2008) introduced two new criteria:

6. Justification by an appropriate theoretical model, which distance-based indicators fulfill through their foundation in Ripley's K-function density framework.
7. Feasibility of measurement with the available information.

This final criterion is only partially met, as assumptions are sometimes necessary to address data limitations, e.g., by not controlling for the size of production points or by applying restrictive assumptions for edge-effect correction.

However, Kopczewska et al. (2019) incorporated three new criteria that some of the earlier indicators fail to meet or only partially satisfy:

8. Sensitivity to different spatial distributions of productive activity. This criterion is only partially fulfilled by previous indicators, as it depends on specific decisions made during their construction, such as the number of buffers, the distance intervals between them, and other methodological choices.
9. Weighting by the size of production points. This is achieved by earlier weighted indicators, which adjust for differences in production point sizes.
10. Providing unique numerical indicators of concentration levels, independent of buffer size. This criterion addresses the need for a single, consistent measure that does not vary with the buffer's dimensions, something previous indicators often struggled to deliver.

In any case, the theoretical literature and empirical contributions in this field of study have been scarce in recent years since the publication of the excellent surveys by Marcon & Puech (2017) and Gómez-Antonio & Alañón-Pardo (2020), which describe, analyse, and compare all agglomeration indicators available up to that point. Specifically, the first article represents a significant contribution to the theoretical literature, as it streamlines the procedure for deriving all agglomeration indicators into five steps that these indicators must or generally follow. It seems appropriate to take this contribution and its terminology as a starting point to advance, as the authors themselves suggest, with more proposals that either provide new theoretical interpretations of agglomeration measures, deepen, modify, or eliminate some of the restrictive assumptions (explicit or implicit), and offer new analytical tools for studying this phenomenon.

Specifically, in the literature on the measurement of agglomeration indicators, some of which has already been cited in relation to their properties or their application in empirical studies, six shortcomings are identified. These serve as the basis for the main theoretical contribution of this doctoral thesis:

1. Absence of distance-based indicators at the point level: While the dominant interpretation of agglomeration as a technological characteristic of sectors is consistent with this gap, applications using microdata often require agglomeration indicators. In their absence, alternative measures, such as potential markets or accessibility levels, are used instead.
2. Lack of geographical agglomeration indicators in the third generation: Similar to the first generation, the dominance of sectoral interpretation (whether intra-industry or inter-industry) has limited the development of indicators with a geographic perspective.
3. Absence of single, synthetic numerical indicators for agglomeration: Distance-based indicators have not yet provided solutions for summarising agglomeration levels, whether sectoral or geographical, into a single numerical value.
4. Use of extreme assumptions in measurement or correction processes: Examples include the edge-effect correction when the buffer used extends beyond the study area.
5. Complex null hypothesis testing methods: The hypothesis of no concentration (or dispersion) is often tested using computationally intensive methods, which can pose obstacles due to the time and computational resources required.
6. Limited adaptability of available indicators: Many current indicators lack flexibility to adapt to data characteristics, different aspects of agglomeration, various contexts (especially urban), and specific empirical requirements. They do not provide tools for adjusting their application to the needs of diverse studies.

To address these aspects, even partially, this thesis is structured into three chapters and a final conclusions section. The first chapter builds upon the typology and steps developed by Marcon & Puech (2017), aiming to generalise their approach. Based on this theoretical development, a new family of indicators is proposed, starting with the construction of point-level agglomeration indicators. This approach involves modifying the order in which the steps outlined by Marcon & Puech (2017) are executed. Additionally, greater emphasis is placed on controlling for random spatial concentration, which should be accounted for (referred to as first-order concentration or join-location). Agglomeration indicators should exclusively capture spatial location derived from supply or demand interactions among production points (referred to as second-order concentration or co-location)[2]. Specifically, first-order concentration is divided into two aspects:

1. **Local-global join-location**: This refers to the average spatial configuration around each production point, irrespective of its activity. This aspect has not been controlled in available indicators to date.
2. **Local join-location**: This accounts for the expectation of a higher density of production points in specific local areas due to the overall concentration of firms from all sectors in that area.

The proposed indicator addresses both aspects. The local-global join-location is controlled at the point-level indicators, while the local join-location is controlled in the aggregated sectoral indicators. This separation eliminates geographical biases when calculating sectoral agglomeration indicators.

Similarly, from the point-level indicators, it is possible to derive geographical agglomeration indicators or indicators corresponding to specific subareas, while controlling for sectoral composition bias. This represents a significant contribution of the doctoral thesis.

The proposed methodology generates a wide range of tools for measuring

[2] A detailed discussion differentiating joint-location and co-location can be found in Pablo-Marti and Arauzo-Carod (2020). In their work, an indicator is developed using raster analysis, which provides a methodological framework to distinguish between these two concepts in the context of spatial concentration.

agglomeration: individual indicators for specific buffers and synthetic indicators; aggregated agglomeration indicators dependent on the buffers considered, as well as synthetic ones. It also produces agglomeration curves at individual, sectoral, and geographical levels, which reflect how the measurement changes depending on the distance considered. Finally, it provides statistical synthetic agglomeration indicators for each sector and specific subareas.

Additionally, a data-driven correction for the edge-effect is proposed. This correction addresses situations where the measurement buffer extends beyond the administrative area for which data is available, avoiding the imposition of specific assumptions about the density of production points in adjacent areas. However, a particularly significant contribution lies in the approach to testing the null hypothesis of no concentration. In previous literature, two assumptions regarding the distribution of establishments have been used to define the absence of concentration (or dispersion). The first assumes that production points for a given sector could be located anywhere within the analysed administrative area. This is an extreme assumption, rarely implemented, as it suggests the possibility of establishing production points even in geographic obstacles such as rivers, mountains, or the sea. For this reason, the more common assumption is that establishments of one activity could locate anywhere there are production points from any other activity. However, in urban environments, characterised by spatial limitations, restrictions due to pollution (e.g., smoke, noise, light), potential congestion, and the division of land use between residential, commercial, and industrial zones, this assumption also appears unrealistic.

Therefore, a possibly restrictive assumption is introduced: the current distribution of production points is limited to their existing locations. This allows for testing the null hypothesis by creating confidence intervals derived from sufficiently large random subsamples that are representative of each activity and generated in sufficient numbers. This approach eliminates the need to recalculate individual indicators and instead focuses on constructing new aggregates for the subsamples, which significantly simplifies the process of building these intervals and, most importantly, substantially reduces computation times.

In this first chapter, an indicator unweighted by the size of the production point is used due to a lack of available information, to measure productive agglomeration based on the Census of Premises and Activities of the Madrid City Council, covering approximately 100,000 production points in 2014 and 2019. The selection of Madrid as a case study stems from the availability of freely accessible, periodically updated, georeferenced data, providing detailed information at the production point level. Similar data are available for other major European cities, facilitating the immediate extrapolation of this study.

The second and third chapters of this doctoral thesis aim to demonstrate the adaptability and computational efficiency of the proposed family of indicators. In the second chapter, the indicator introduced in the first chapter is adapted to account for the importance of the weight of each production point, measured by employment. While this adaptation is straightforward, ensuring it rigorously meets all the properties of the unweighted indicator is not trivial. Since the data used in the first chapter for Madrid lack economic information for each production point, the analysis shifts to Chinese manufacturing firms, utilising the Orbis database for 2021. This dataset includes 1.7 million manufacturing firms, providing a robust test of the improvement in computational efficiency. For context, no previous studies using geographically referenced microdata for China have conducted this type of analysis. Prior research has typically been limited to specific regions and/or sectors, especially in the eastern coastal zone, and has focused on large enterprises or subsets that never exceed 100,000 firms. It is worth noting that the necessary calculations involve computing bilateral distances between all firms. This empirical analysis is particularly relevant as it allows for the refinement or even correction of biases present in previous studies on agglomeration measurement in the "World's Factory". Furthermore, the chapter concludes with a small experiment examining whether synthetic agglomeration indicators change when considering the presence of economic development zones throughout China. The results may provide insights into the significance of this industrial policy instrument in shaping the spatial distribution of activities in the country.

Finally, the third chapter adapts the indicator for measuring co-agglomeration, also referred to in some previous works as inter-industry agglomeration (Ellison & Glaeser, 1997; Duranton & Overman, 2005). The literature on this specific measure has primarily focused on bilateral co-agglomeration, that is, the co-location of two specific activities. To address this, the indicator is developed for the case of global co-agglomeration for each sector. This refers to the joint location of production points in a sector with those of all others, which can be interpreted as the attraction (or repulsion) capacity of establishments in one sector to those of similar sectors across all activities. This is a measure that existing indicators cannot achieve due to the way the relative reference is defined in their calculation. This measure is particularly relevant for development policy and urban planning, as it considers the intersectoral synergies of a given sector in relation to all others. It likely holds greater significance, and differentiation, from a geographical perspective than from a sectoral one. From the geographical perspective, it indicates, after controlling for appropriate sectoral biases, the ability of firms in a given area to attract economic activity. From the sectoral perspective, the measure, by construction, will show greater similarity across sectors. As in the first chapter, these indicators are calculated using the same database for Madrid.

In conclusion, this doctoral thesis aims to combine theoretical and empirical contributions on a topic that was considered settled, adapting the indicators, expanding the analytical tools, and introducing new interpretations. At the same time, it addresses challenges in data processing and computational efficiency that this new family of indicators enables.

Chapter 1. A new approach of measuring and interpreting spatial agglomeration in the cities: The case of Madrid

1.1. Introduction

The spatial concentration of economic activities and the analysis of the location of establishments have been widely studied by many economists over the years. It has been proven that economic activity tends to concentrate (Krugman, 1992; Porter, 2000; Henderson & Thisse, 2004). The interest in studying spatial concentration patterns can be traced back to Marshall (1890). In the 1990s, the theory of "new economic geography" revitalized the study of agglomeration economies, primarily inspired by the work of Krugman (1991, 1992), Venables (1996), Fujita et al. (1999), and Ottaviano and Thisse (2004), among others. This research rediscovered space as a central element in economics, aiming to explain why economic activity is geographically concentrated. Economic activities tend to agglomerate, largely due to the gathering of many people, companies, and institutions.

In this context, cities are increasingly relevant actors, concentrating new physical, human, and technological capital. Additionally, proximity and ease of interaction favour the matching between the needs of some firms in terms of human capital and inputs, and on the one hand, those of other firms seeking markets, and on the other hand, workers with their specific skills. Moreover, this closeness facilitates technological spillovers, leading to benefits such as lower costs and productivity improvements (Fujita and Thisse, 2002).

This area of research has made substantial progress in identifying the processes and determinants of firm location and the associated agglomeration patterns of economic activities[3]. This has led to the development of measures assessing the relative importance of these agglomeration processes in each activity. However, despite the

[3] For a comprehensive overview and analysis of the literature, see Gómez-Antonio & Alañón-Pardo (2020).

aforementioned contributions, there remain reinterpretations and the need for new measures (Marcon & Puech, 2017). Specifically, there is a need for measurements at the production point level that can be used in concrete applications -such as evaluating the effects on productivity-, a single value that is not dependent on the distances used for measurement, and above all, measures that provide a better understanding of the concentration phenomenon not only from a sectoral perspective, as is currently the case, but also from a geographical one.

These challenges become more pronounced when measuring in cities for three main reasons: the greater location constraints for many activities, the appropriate distance measure that should be applied, and finally, the exacerbation of the two previous issues due to the city's geography, which often involves a high concentration of activity in small areas[4]. This requires the use of smaller measurement areas or distances, given that many activities are concentrated in close proximities to, for example, meet the everyday needs of citizens. As a result, traditional measures developed for broader contexts, such as countries or regions, require adaptation.

In this regard, it is surprising that despite the fact that the latest proposed measures are based on the availability of microdata, the construction of aggregated indicators at the activity level or according to certain characteristics of firms, such as size, is still being considered, rather than making the leap toward concentration or dispersion indicators at the firm level or, more precisely, around each individual firm. Having basic indicators at this level represents a substantial improvement in the efficiency of testing the null hypothesis regarding the values of the spatial concentration indicator for a specific sector or group of establishments, compared to the usual Monte Carlo simulations.

Moreover, based on these indicators, it is possible to interpret them not only from a sectoral perspective, as has traditionally been the case, but also from a geographical one

[4] Albert et al., (2012) also highlight the distortion caused by urban environments when using a distance-based agglomeration indicator applied to Spanish manufacturing. This distortion arises from the installation restrictions imposed on certain productive activities within cities.

(such as neighbourhoods, districts, etc.), and also provide data-driven solutions to the problems related to geographical boundaries, rather than relying on assumptions about the density of establishments outside the geographic area for which information is originally unavailable (Duranton & Overman, 2005; Marcon & Puech, 2003, 2010, 2017; Arbia et al., 2012; Mori & Smith, 2014; Bonneu & Thomas-Agnan, 2015, among others).

Therefore, the objective of this chapter is twofold. From a theoretical perspective, it introduces a novel methodology based on distance-based indicators for spatial concentration, defining the construction of agglomeration measures of economic activity around each production point. From these, different tools are proposed to obtain sectoral concentration averages and by geographical units, such as districts, neighbourhoods, etc. From an empirical perspective, the aforementioned measures are applied to the Census of Establishments from the Madrid City Council for two years: 2014, the first year available, and 2019, before the COVID-19 pandemic.

This work contributes to the literature on the subject by introducing a new methodology better suited to measuring in the context of cities, with a measure of attraction or repulsion for each production point. Additionally, it discusses certain aspects of previous measures, suggesting alternative corrections to raw measures based on data-driven procedures, such as addressing the edge-effect problem. Furthermore, based on point-level measures, synthetic agglomeration indicators are obtained for each activity and geographical area, independent of the selected buffer, which was a limitation in the interpretative scope of previously available measures. Finally, it proposes a different approach to testing the significance of the proposed indicators, which is particularly simple and significantly reduces computation time.

This chapter has six sections. Following this brief introduction, the second section provides an overview of the various concentration indicators offered in the literature, while analysing the most widely used methods for measuring the geographic concentration of activities, classifying them, and highlighting their similarities and differences. The third section outlines the methods we employed. The fourth section presents the data used for the empirical application to the city of Madrid. In the fifth section, we discuss the results. The main conclusions of this chapter are summarised in the final section.

1.2. Literature review

The measurement of the spatial concentration of economic activity has been a recurring topic in regional economics and economic geography. These measures have advanced by incorporating theoretical aspects and have evolved towards the use of disaggregated data as the necessary databases became available for implementation.

Duranton & Overman (2005) establish a set of five desirable properties that these indicators should ideally meet. (i) The results must be comparable across industries. (ii) They should control for overall agglomeration trends across industries. (iii) They should separate spatial concentration from industrial concentration. (iv) They should be unbiased with respect to the degree of spatial aggregation, and (v) they should indicate the statistical significance of the results. Kominers (2008) introduces two more relevant criteria: (vi) they must be measurable using accessible information and (vii) they must be justified by an appropriate theoretical model. Additionally, Kopczewska et al. (2019) propose three additional criteria for continuous spatial indicators: (viii) sensitivity to different spatial distributions; (ix) inclusion of the area of the territory, as well as firm size; and (x) results that are easy to understand and interpret. Meeting this set of criteria is highly complex, and, for this reason, some authors, such as Marcon & Puech (2017), suggest that contrary to what might seem the case, measuring geographic concentration processes is a topic with further room for development, as both the theoretical foundation and its implementation require significant advances.

The various measures proposed in the literature can be classified into two categories depending on whether the data used in their construction are aggregated (discrete measures) or microdata (continuous measures). This distinction not only determines the method of calculation for the indicators but also their potential properties. However, three generations of measures are commonly identified (Combes & Overman, 2004; Duranton & Overman, 2005; Albert et al., 2012; De Dominicis et al., 2013; Chain et al., 2019): two constructed using aggregated data and a third using microdata. Additionally, the different spatial concentration indicators can be classified as sectoral or geographical, depending on whether they analyse the concentration of certain productive activities across space, or whether certain geographic areas concentrate more productive activities than others. The former are typically referred to as agglomeration measures, while the latter are referred to as specialisation or concentration in geographical terms (see Table 1.1).

Table 1.1 Sectoral and geographical indicators for spatial concentration of production

Discrete indicators	**First generation**	Industrial and geographical indicators - Hirschman-Herfindahl spatial index - Location Quotient - Dissimilarity indices - Gini spatial index
	Second generation	Industrial indicators - Ellison & Glaeser (1997) - Maurel & Sédillot (1999) - Rysman & Greenstein (2005) - Mori et al., (2005)
Continuous indicators	**Third generation**	Industrial indicators - g and K function (Ripley, 1976, 1977) - K_{mn} function (Penttinen et al., 1992; Penttinen, 2006) - D function (Diggle & Chetwynd, 1991) - g_{inhom} and K_{inhom} functions (Baddeley et al., 2000) - K_d function (Duranton & Overman, 2005) - M function (Marcon & Puech, 2010) - m function (Lang et al., 2020)

The first generation of measures adapts pre-existing concentration, inequality, dispersion, or specialisation indicators to measure the spatial concentration of productive activity. In this sense, with geographically aggregated data, the Hirschman-Herfindahl index, the location quotient, entropy indicators, dissimilarity indicators (such as the inequality measure proposed by Audretsch & Feldman, 1996), and, ultimately, geographic indicators like Krugman's (1992) productive specialisation index or geographic Gini indices have been used.

This first generation of indices provides spatial concentration measures for specific productive activities, as well as for geographic areas. These indices aim to assess the extent to which the activities of a sector (or in a geographic area) are more geographically concentrated than those of other sectors (or areas). In general, these indicators only satisfy the first of the desirable properties, with some, such as the Gini index, also fulfilling the second (Bertinelli & Decrop, 2005). At the time of their development, they were suitable for the available data. However, the literature has mainly highlighted the failure to meet the second and third conditions, which led to the emergence of the second generation of measures.

Ellison & Glaeser (1997) launched the “second generation” method by introducing a new index that establishes a benchmark for comparing the distribution of all activities in space and controls for market agglomeration. Maurel & Sédillot (1999) redefined the previous index based on the correlation between the location of any pair of production points within an industry, thereby transforming Ellison & Glaeser’s (1997) proposal to adapt it to a model derived from the “dartboard model”. In their case, the proposed agglomeration indicator measures the difference between the distribution of an industry’s production points and random chance.

Two little-used contributions have been designed based on a statistical test, although they lack an explicit underlying model. Rysman & Greenstein (2005) developed a combinatoric test for agglomeration called the Multinomial Test for Agglomeration and Dispersion (MTAD). This measure tests the agglomeration behaviour of production points, assuming that firms concentrate if and only if they exhibit similar location choice behaviours. Additionally, the measure provides information on actual levels of dispersion, unlike previous tests that only indicate whether dispersion exists. The two main criticisms of this indicator are its applicability and the lack of an explicit reference model, although the idea seems to be based on the “dartboard” approach. The second proposal was made by Mori et al. (2005), who suggest measuring agglomeration through a comparison between the D-index introduced by Kullback & Leibler (1951) and a reference model of complete spatial dispersion (e.g., the Gini index). This proposal has numerous statistical and practical properties, but it shares the issue of relying on aggregated discrete spatial data.

In general, these second-generation indicators present two types of problems. The first is sensitivity to the degree of spatial aggregation, the structure, and the relative position of territorial units, known as the Modifiable Area Unit Problem (MAUP) as outlined by Openshaw & Taylor (1979)[5]. The second is that they ignore geographical relationships between locations that are close to each other but fall under different administrative units.

Precisely, this is the main issue addressed by the third generation of measures, generally referred to as distance-based measures of spatial concentration. Initially based on Ripley's K-function (1976, 1977), they were developed following the seminal works of Duranton & Overman (2002) and Marcon & Puech (2003). These indicators had previously seen significant development in the field of ecology (Pélissier & Goreaud, 2001).

Marcon & Puech (2017) establish a typology of distance-based measures of spatial concentration, all constructed from georeferenced microdata of production points[6]. This typology identifies five steps (though not in all cases) in the construction of these indicators, following the spatial point pattern approach (Cressie, 1993). The review that follows, based on this typology from Marcon & Puech (2017), generalises and classifies the available indicators. This framework serves as the foundation for presenting and comparing the proposal in the next section. The specific formulas used by the authors, along with discussions of their applications, advantages, and limitations, are detailed in the aforementioned article and in Gómez-Antonio & Alañón-Pardo (2020). The latter also offers a list of applications for each indicator. The five-step typology is presented below.

---

[5] A discussion on this issue can be found in Arbia (2001), Bonneu & Thomas-Agnan (2015), and Kopczewska (2018).

[6] Some studies that use these measures employ data from firms, assuming that they are located at their headquarters.

The first step is to count the number of production points[7]around each reference point, considering a certain distance. Based on certain differential elements used in this counting, the following distinctions can be made: a) density and cumulative measures; b) weighted or unweighted by the density of the point process around the reference point; c) weighted or unweighted by the size of the production point; d) weighted or unweighted by the distance to the reference point; and finally, e) with or without border-effect correction.

In summary, it could be said that the counting of the number of neighbours takes the following form:

- For cumulative measures:

$$n(x_i, r) = \sum_{\substack{j \neq i \\ j,i \in t}} k\left(\left\|x_i - x_j\right\| \leq r\right) w\left(x_j\right) z\left(x_j\right) c(i,j) \qquad [1.a]$$

- For density measures:

$$n(x_i, r) = \sum_{\substack{j \neq i \\ j,i \in t}} k\left(\left\|x_i - x_j\right\|, r\right) w\left(x_j\right) z\left(x_j\right) c(i,j) \qquad [1.b]$$

Where $n(x_i, r)$ is the number of neighbours x_j around the reference point x_i of a specific type (same sector, same size, etc.) at a distance r (at that exact distance for density measures and up to that distance for cumulative measures); $w\left(x_j\right)$ represents the weights calculated based on the density of the process around x_j (therefore equal to one if the measures are unweighted by the density around the neighbour); $z\left(x_j\right)$ is an indicator of the size of the productive activity at x_j (being equal to one for unweighted measures by the size of the neighbour); k is a Kernel function inversely related to distance (in the case of unweighted measures by distance, it would be equal to one); and finally, $c(i,j)$ is the correction for edge effects (also equal to one if the measure is without edge-effect correction).

The first distinction concerns whether the counting of production points in the sector around the reference point is done up to a maximum distance (cumulative) or at a

[7] In this section, the word "sector" is used, but it could refer to a typology of production points: by size, technological intensity, etc.

specific distance or buffer (density). The second distinction, represented by $w(x_j)$, adjusts for the higher density of production points around each neighbour by considering the greater productive intensity in the areas where the counting takes place (weighted or unweighted by the density around each neighbour). This is important because in areas with more productive activity, there is a higher likelihood of finding production points from the sector being analysed. In a way, the indicators that introduce this type of adjustment in neighbour counting anticipate the introduction of a local reference for the entire area by applying it to each neighbour individually.

The third distinction, represented by $z(x_j)$, considers whether neighboring points are counted equally (unweighted by the size of the neighbour) or based on their size in terms of employment or production (weighted by the size of the neighbour). Sometimes this distinction depends on the availability of data, and transforming one measure into another is relatively simple, although it affects various steps in the calculation process. The fourth distinction, represented by k, addresses whether more distant neighbours from the reference point should have less weight in the accounting process. In this case, some authors propose introducing a Kernel weight (weighted by the distance), while others do not apply this weighting (unweighted by the distance).

Finally, the buffer used to count the number of neighbours sometimes extends beyond the geographic area for which data is available. This can happen either into a neighbouring area that may contain production points from the sector being analysed, or into geographical features like seas, lakes, rivers, or mountains where no such points exist. This creates two types of measures: those that do not apply any correction and those that do, represented by $c(i,j)$. The most common correction assumes that in the buffer area not overlapping with the administrative area where data is available, the density of establishments is the same as in the overlapping area. As a result, the number of neighbours is adjusted by multiplying by the inverse of the proportion of the buffer area that overlaps with the known area. In the third section of this chapter, an alternative correction is proposed, applied after the initial count, and based on the spatial distribution of the data.

In the second step, the point-level analysis is abandoned, and the construction of the sectoral indicator begins. For this, an average is calculated for all the points in the sector based on the previous count (number of neighbours around each point), including the necessary corrections. Once again, two types of averages emerge: an unweighted average, where all points have equal relevance, and a weighted average, where the average is adjusted according to the intensity of the process, i.e. density, around each point, or its productive importance.

In this way,

$$\bar{n}_s(r) = \sum_{i \in s} \delta(x_i) n(x_i, r) \quad [2]$$

Where $\bar{n}_s(r)$ is the average number of neighbours at a distance r (cumulative or density) for the n production points in the sector; $\delta(x_i)$ is the relative weight of each of the analysed points. In the case of unweighted measures $\delta(x_i) = \frac{1}{n}$.

In the third step, a local reference is established to account for how the rest of the activities (or all of them) are distributed, considering the same buffer size. As previously mentioned, this relativisation aims to create a reference for comparison with other activities. Thus, three types of measures arise, as noted by Brülhart & Traeger (2005): a) topographical, which simply considers the buffer size to normalise the measure, since a larger buffer, *ceteris paribus*, will contain a greater number of production points; b) relative measures, which use the same method of measurement applied to the production points of other (or all) sectors around the analysed point[8]. Finally, absolute measures do not introduce any local reference. It should be noted that in some measures there is a certain degree of prior local normalisation, both in the counting and in the averaging, based on the weights of either the neighbours' density or the reference point itself. Therefore,

[8] In general, neighbours from the same sector are also included, although some authors believe it is preferable to exclude them (Diggle et al., 2007 and Marcon et al., 2012).

$$\bar{n}_l(r) = \frac{\bar{n}_s(r)}{z(r)} \qquad [3]$$

Where $\bar{n}_l(r)$ is the average number of neighbours in the sector, relative to the total number found at a distance r, and (r) is a local reference that, depending on the type of measure, can take the following forms:

$z(r) = \pi r^2$ for topographic cumulative measures,

$z(r) = 2\pi r$ for topographic density measures,

$z(r) = 1$ for absolute measures,

$z(r) = \sum_i \delta(x_i) n(x_i, r)$ for relative measures. Basically, in this case, steps [1] and [2] are followed, but the number of production points from all activities around each point in the analysed sector is counted, and their average is calculated. $\delta(x_i) = \frac{1}{n}$ in the case of unweighted measures. This step is carried out with the goal of removing first-order concentration (the joint location) from the indicator and leaving only second-order concentration (colocation) of the point process.

In the fourth step, a global reference is established with the goal of obtaining a reference value to compare with that of the analysed sector. This reference indicates the absence of a relative process of concentration or dispersion of productive activity. Specifically, the final sectoral spatial concentration indicator would be:

$$SSCI(r) = \frac{\bar{n}_l(r)}{g(r)} \qquad [4]$$

Where the spatial concentration indicator at a distance r is obtained from the previous step and is globally normalised by $g(r) = \frac{n-1}{A}$, where A is the area (cumulative) or the circumference length (density) of the buffer in the case of topographic measures. Alternatives forms for $g(r)$ include:

$g(r) = n - 1$ for absolute measures,

$g(r) = \frac{W_t}{W}$ for relative measures. W_t is the total weight of all the points in the sector, and W is the total weight of all production points.

Finally, in the fifth step, the aim is to test the significance of the constructed measure. In this regard, two types of analyses have been conducted. The most common approach has been to construct confidence intervals for the null hypothesis of reference value equality using Monte Carlo simulations, either by distributing the points as a Poisson process (homogeneous or inhomogeneous) or by redistributing the points across the actual locations of all activities. In this way, spatial concentration would be considered absent if the estimated value of the $SSCI(r)$ indicator falls within the confidence interval for the null hypothesis of no concentration or dispersion structure. Conversely, if it falls outside, the null hypothesis can be rejected, indicating either concentration or dispersion depending on whether the value is greater or smaller than the reference.

An alternative and less common approach is explored by Lagache et al. (2013), where confidence intervals are created for the calculated indicator value, similar to how confidence intervals are constructed in conventional inference. In this case, if the value corresponding to the null hypothesis is included within the confidence interval, the absence of concentration or dispersion cannot be rejected.

In both procedures, the values are tested locally, which is referred to as pointwise. This aspect has been criticized as restrictive by Duranton & Overman (2005), Loosmore & Ford (2006), and Loop & McClure (2015). For this reason, Duranton & Overman (2005) propose a global test, known as the rank envelope test (Myllymäki et al., 2017). However, this type of test becomes very restrictive when the number of simulations is small relative to the number of points. Table 1.2 presents the main indicators proposed in the literature and their typology in relation to the one previously presented.

Table 1.2 Typology of distance-based spatial concentration indicators (based on Marcon & Puech, 2017)

Measure	First step					Second step	Third step	Fourth step	Fifth step
	Type counting the neighbours	Weighted by the density	Weighted by activity size	Weighted by the distance	Border-effect correction	Average measure	Local reference	Global reference	Null hypothesis (*)
	(a)	(b)	(c)	(d)	(e)				
g	Density	NO	NO	YES	YES	Non-weighted	Topographic	Point density	HPP
K	Cumulative	NO	NO	NO	YES	Non-weighted	Topographic	Point density	HPP
K_{mm}	Cumulative	NO	YES	NO	YES	Weighted	Topographic	Point density	HPP with random labelling
D	Cumulative	NO	NO	NO	YES	Non-weighted	Topographic	Point density	HPP
g_{inhom}	Density	YES	NO	YES	YES	Weighted	Topographic	Point density	IPP
K_{inhom}	Cumulative	YES	NO	NO	YES	Weighted	Topographic	Point density	IPP
K_d	Density	NO	NO	YES	NO	Non-weighted	Absolute	Number of points	PRAAL
K_d^{emp}	Density	NO	YES	YES	NO	Weighted	Absolute	Number of points	PRAAL
M	Cumulative	NO	YES	NO	NO	Non-weighted	Relative	Global adjustment	PRAAL
m	Density	NO	YES	YES	NO	Non-weighted	Relative	Global adjustment	PRAAL

(*) HPP: Homogeneous Poisson Process; IPP: Inhomogeneous Poisson Process; PRAAL: Points are redistributed across actual locations.

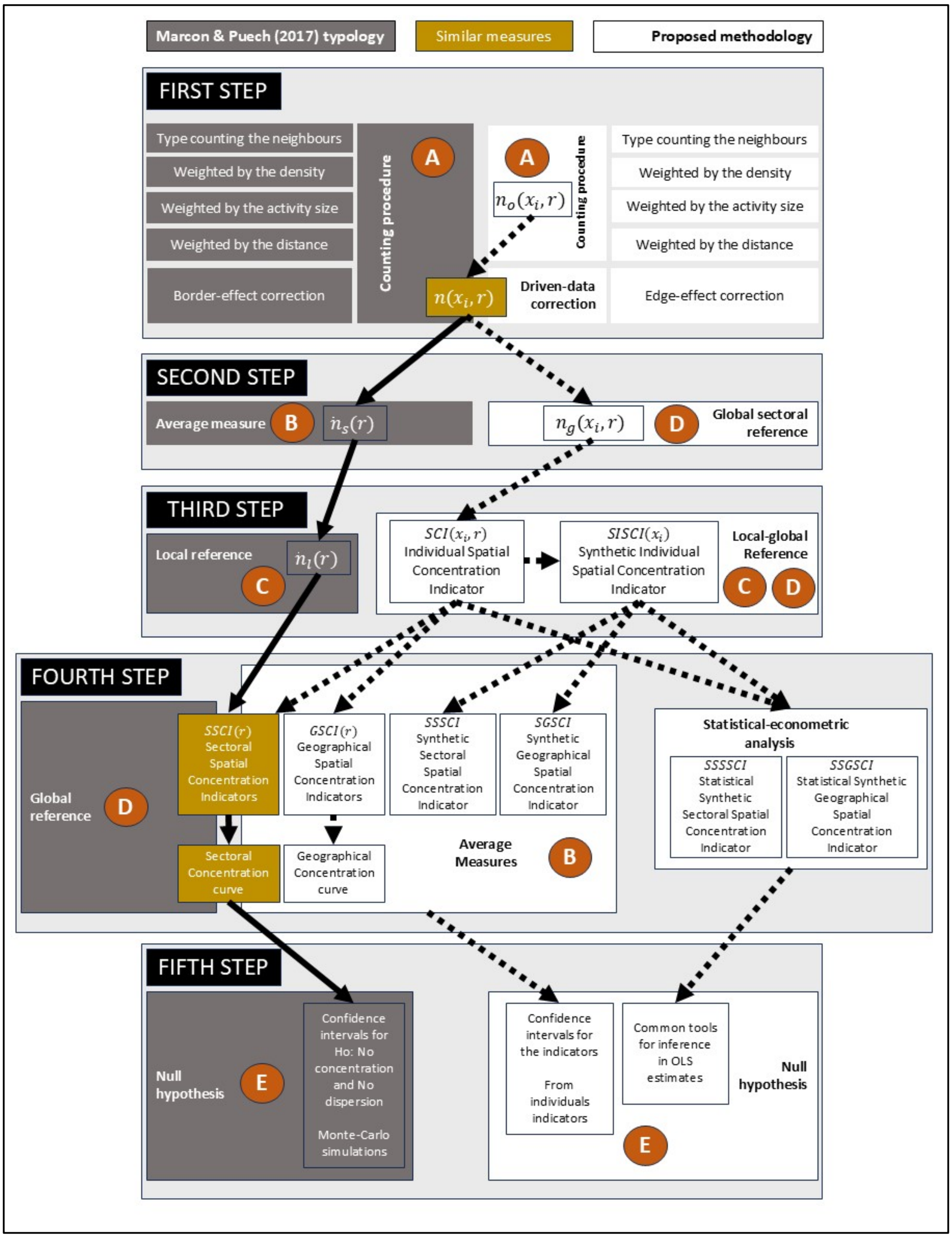
Own elaboration based on Marcon & Puech (2017) and the author's generalisation.

1.3. A new distance-based point-level and derived aggregate indicators for sectoral agglomeration

To present the procedure for obtaining the sectoral and geographical agglomeration indicators proposed here, the process follows and compares with the stylised procedure outlined in the previous section, which in turn follows the typology proposed by Marcon & Puech (2017). However, the order of the steps is adjusted to fit the proposed approach. Substantial changes are introduced in the calculation of the indicator. Its meaning is reinterpreted to include a geographical perspective, and new analytical tools are developed (see the diagram in Figure 1.1).

The main difference and contribution of this chapter is that, while previous proposals aim for a sectoral indicator, all the concentration indicators proposed here revolve around the construction of distance-based indicators for sectoral agglomeration at the point level. From these, it is possible to derive aggregate indicators at the activity level, for specific characteristics of the production points, or geographically based on the location of these points (e.g., neighbourhoods or rasters). The creation of these latter indicators offers an alternative proposal, based on this methodology, to the geographic indicators of the first generation, which disappear in the second and third generations of spatial concentration measures for productive activity. Their interpretation is simple, aiming to approximate the attraction or repulsion capacity of production points located in a specific geographic area. Additionally, agglomeration curves can be derived from the path described by the indicators as the buffer size for which they are calculated increases, and synthetic indicators can be obtained using various techniques for calculating aggregated averages for each sector (or geographic area). These synthetic indicators can also be obtained, and likely more robustly, by applying simple statistical techniques.

Figure 1.1 Comparison between the Marcon & Puech (2017) typology and the proposed methodology



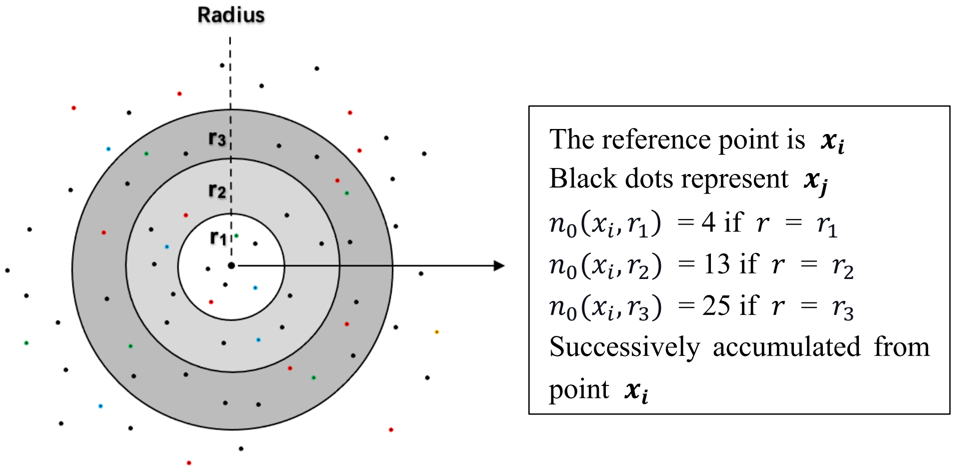
1.3.1. Distance-based point-level indicators for sectoral agglomeration

Let us assume we have an activity with n_S geographic production points (stores, establishments, local units of activity, etc.) in a specific administrative area (city, region, country, etc.) for which information is available for all sectors, and where, in total, there are n production points across all activities. The first step in defining a sectoral agglomeration indicator is to count the neighbours of the same sector or activity around each reference point at a maximum distance *r*.

$$n_0(x_i, r) = \sum_{\substack{j \neq i \\ i,j \in s}} 1\left(\left\|x_j - x_i\right\| \leq r\right) \qquad [5]$$

We denote the reference point as $\boldsymbol{x}_{\mathrm{i}}$, which represents establishment *i* of sector *s*, while its neighbours of the same sector are $\boldsymbol{x}_{\mathrm{j}}$; $n_0(x_i, r)$ refers to the number of neighbours, also from sector *s*, within a maximum distance *r* around reference establishment i. Therefore, in this case, it is a cumulative function. Expression [5] is equivalent to [1.a] presented in the previous section as the standard way to measure neighbours. However, for simplicity and due to the data availability that will follow, the measure is developed in an unweighted form: unweighted by density, i.e. $w(x_j) = 1$; unweighted by the activity size of the neighbour, $z(x_j) = 1$; and unweighted by the distance between x_j and x_i, $k = 1$. However, the indicator could easily include these types of weightings or, alternatively, be defined as a density measure. Consequently, the agglomeration measures will depend on the distance considered for the buffer (see Figure 1.2).

Figure 1.2 Different buffers and neighbour counting around the reference point



The number of neighbours is counted without edge-effect correction, at least in the conventional way established in expression [1.a]. However, when the buffers extend beyond the administrative geographic limits of the database's coverage, the number of points that would fall inside the buffer but outside the study area is unknown. Therefore, a correction for the border effect is usually introduced. This is typically an *ad-hoc* correction. For example, it is common to assume that the external area has the same density as the observed density inside the study area. In other words, the correction is applied by using the inverse of the buffer overlap with the administrative area for which information is available.

However, this *ad-hoc* correction could be considered a maximum correction. In fact, there is no argument to arbitrarily assume a higher density outside the study area but within the buffer. On the contrary, it could be reasonable to assume a lower density outside the administrative boundaries, especially when analysing areas of concentrated activity such as large cities, as in our case. This is particularly relevant when the administrative boundary is a geographic barrier that prevents the continuity of the production area. In such cases, as indeed happens, there may be a higher production density within the administrative area, and therefore, using the previous correction could overestimate the number of neighbours within the buffer. Nevertheless, in this scenario, analysing agglomeration within the area for which information is available, and particularly in the buffer in relation to other parts of the study area, provides valuable insights into this potential response process by firms.

Precisely for this reason, in this case, the issue will be addressed using the information suggested by the data. The correction proposed is to use the inverse of the buffer area that overlaps with the administrative study area (λ_{ir}), but raising this ratio to an exponent that ranges between 0 and 1. That is, in expression [1.a], $c(i,j) = \frac{1}{\lambda_{ir}^{\gamma_{sr}}}$. Consequently, the number of establishments or production points from the same sector around the reference point is calculated according to expression [6].

$$n(x_i, r) = \frac{n_0(x_i, r)}{\lambda_{ir}^{\gamma_{sr}}} \quad [6]$$

Where γ_{sr} is the exponent that may vary between sectors and also depending on the buffer size. It differs between sectors because concentration processes have specific characteristics for each activity, and therefore the border effect is expected to differ as well. It also varies based on the buffer size because when smaller buffer sizes are considered, the border effect is likely to be more significant (hence, the exponent will be closer to one), whereas when the buffers are relatively larger, the exponent tends to zero. If the exponent γ_{sr} takes the lower value, i.e. zero, then no correction would be necessary, while a value of 1 would result in the commonly used correction. Intermediate values would produce a correction somewhere between these two situations.

The value of γ_{sr} for each sector and buffer, it is estimated based on a relationship between the corrected number of points and qualitative characteristics of the production points z_i that might affect the measurement of concentration in their surroundings, such as whether the production point belongs to a shopping centre, with fixed effects for the sector (d_s) and the buffer (d_r), expression [7.a].

$$\ln n(x_i, r) = \alpha + \beta z_i + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \varepsilon_{ir} \quad [7.a]$$

As $n(x_i, r)$ is not observed, it is replaced by expression [6], yielding expression [7.b].

$$ln\, n_0(x_i, r) = \alpha + \beta_1 z_i + \gamma_{sr} ln\, (\lambda_{ir}) + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \varepsilon_{ir} \quad [7.b]$$

The coefficient γ_{sr} is approximated based on fixed effects for sector and buffer, the two dimensions in which it varies, expression [7.c].

$$\gamma_{sr}\, ln(\lambda_{ir}) = \gamma\, ln(\lambda_{ir}) + \sum_s \vartheta_s\, ln(\lambda_{ir})\, d_s + \sum_r \nu_r ln\, (\lambda_{ir}) d_r \qquad [7.c]$$

Substituting [7.c] into [7.b] results in the final expression [7.d] to be estimated.

$$ln\, n_0(x_i, r) = \alpha + \beta_1 z_i + \gamma\, ln(\lambda_{ir}) + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \sum_s \vartheta_s\, ln(\lambda_{ir})\, d_s + \sum_r \nu_r ln\, (\lambda_{ir}) d_r + \varepsilon_{ir} \qquad [7.d]$$

Thus, the different values of γ_{sr} are obtained as $\gamma_{sr} = \gamma + \vartheta_s + \nu_r$.

The second step we follow to construct the agglomeration indicator involves establishing a sectoral reference, allowing for comparison between firms within the same sector and, to some extent, across sectors, since larger sectors would have a higher number of neighbours and smaller sectors would have fewer. In other words, the size of the sector in terms of the number of production points is taken into account, indirectly controlling for sectoral concentration within the overall study area. This step corresponds to the fourth step identified by Marcon & Puech (2017). The change in the order of the steps comes from the goal of obtaining an indicator of relative attraction or repulsion for each production point, rather than just an aggregate one. To do this, $n(x_i, r)$ is divided by the number of establishments in the sector, which is also corrected (as was done with the numerator) by the mean of the overlaps, similarly to expression [6][9], so that the corrections do not only affect the numerator, as this would result in indicators outside the feasible theoretical range (since it is a probability). In this way, the proportion of production points from the same sector found around the reference firm within a radius *r* is obtained from expression [8].

$$p(x_i, r) = \frac{n(x_i, r)}{n_s \big/ \left(\frac{1}{n_s} \sum_{j \in s} \lambda_{jr}^{\gamma_{sr}}\right)} \qquad [8]$$

If the previous step involves taking a global reference at the sector level, in the third step, a measure of agglomeration around each establishment is obtained by taking a

[9] If this correction was not applied, the reference of $p(x_i, r)$ when *r* reaches the maximum distance between the establishments in the sector would not be equal to one, which is the goal of this normalisation.

global reference based on the local count for all activities. In other words, a Spatial Concentration Indicator (*SCI*) at the point level for a buffer *r* is obtained by normalising expression [8] by the average for all points in all sectors of the previous proportion, always considering a fixed buffer.

$$SCI(x_i, r) = \frac{p(x_i, r)}{\overline{p(r)}} \qquad [9]$$

Where $\overline{p(r)} = \frac{1}{n}\sum_i p(x_i, r)$ represents the average of the relative frequency with which establishments of the same sector are found within a maximum distance of *r* around each activity point. In other words, instead of counting the production points from other activities around each point at a distance *r*, as is done in the relative measures developed so far, the reference is constructed like the numerator but calculated for all activities. This has a direct implication on the result. In the case of the *M* measure proposed by Marcon & Puech (2003), with local normalisation, the goal is to assess whether, in areas with a higher density of production points, each activity has a greater or lesser presence, so the local reference aims to control for the varying densities of productive locations throughout the studied area. This normalisation prevents the interpretation of the indicators from a geographic perspective and can yield seemingly contradictory results for a relative measure[10]. Therefore, if $SCI(x_i, r)$ has a value greater than one, it would imply a higher geographic concentration than the average around that activity point at a certain distance. Consequently, as a relative indicator of agglomeration, the reference value is set at one, representing a state of neither agglomeration nor dispersion of productive activity. Values above one indicate spatial concentration, whereas values below one reflect spatial dispersion. Notably, the range of dispersion values is bounded between 0 and 1, whereas agglomeration values

[10] For example, let us assume a city has only two activities that are geographically disjoint but symmetrically placed, and in one of the analysis buffers, no establishments from the other activity are present. In this case, the *M* indicator could not be calculated unless the establishments of the same sector were considered. However, if the buffer is expanded, the two activities would appear concentrated, even though they are the only two activities in the city, which seems inconsistent with the calculation of a relative indicator, where one activity is concentrated relative to the average. In contrast, with the proposed measure, the indicator can always be calculated, since at least the denominator will have a value for establishments in the same sector. Additionally, not all sectors can be either agglomerated or dispersed.

exceeding one are unbounded and have no upper limit[11]. Compared to the typology of Marcon & Puech (2017), this step partially corresponds to the third and fourth, since the proposed reference is both local, i.e. measuring the behaviour of all activities at a certain distance, and global, i.e. referring to the entire study area.

However, the previous indicator does not provide a unique answer regarding the concentration level of a specific activity around the reference establishment, as its value depends on the size of the buffer used, similar to standard agglomeration measures. For this reason, it is also proposed to construct the Synthetic Individual Spatial Concentration Indicator (*SISCI*) at the point level, which does not depend on distance, by creating a weighted aggregation of the indicators for different buffers as follows:

$$SISCI(x_i) = \frac{\sum_r SCI(x_i,r)f(r)}{\sum_r f(r)} \qquad [10]$$

Where $f(r) = \left(\frac{r_0}{r}\right)^g$ is a function of distance. The exponent g is greater than zero and may take on different values such as 0.5, 1, 2, etc. r_0 refers to the radius of the smallest buffer considered, and r refers to the full range of buffers for which the previous indicators are calculated. In this way, greater weight is given to the level of concentration around each point in the buffers closest to the point being analysed. Note that in this aggregation, the points contained within the smaller buffers are included in all the larger ones, so their weighting is much higher than what is reflected in the evolution of the function $f(r)$ between buffers. In any case, both the collection of indicators $SCI(x_i, r)$ and $SISCI(x_i)$ are measures of agglomeration at the point or establishment level.

1.3.2. Aggregated agglomeration indicators and agglomeration curves

Once establishment or point-level indicators are available, the fourth step is to obtain sectoral agglomeration indicators (or by unit or geographic area, i.e., raster, neighbourhood, census section, etc.) as an average of the production point indicators

[11] Some authors propose addressing this issue of asymmetry by taking the logarithms of the indicator. This transformation sets the reference value to zero, eliminating asymmetry in its distribution.

for the sector. Therefore, the Sectoral Spatial Concentration Indicator[12], $SSCI(r)$, will be constructed as follows:

$$SSCI(r) = \frac{1}{n_s}\sum_{i\in s} SCI(x_i, r) \qquad [11]$$

In other words, it is an unweighted average[13], with $\delta(x_i) = 1$. Therefore, a relative indicator of spatial concentration of productive activity will be available for each sector (or geographic area) and buffer size. By ordering the $SSCI(r)$ indicators according to the values of buffer size *r*, an agglomeration curve can be constructed, which tends toward a value of one as *r* approaches the maximum distance between establishments in that sector within the studied area.

Each sector (or geographic area) will follow a different trajectory of attracting production points from the same sector (agglomeration) as the buffer size where activity points are counted increases. Classifying these behaviours also helps identify sectoral (or geographic) agglomeration patterns based on distance. Two criteria are used for this (Table 1.3): whether the described curve crosses the unit (the reference value) and the shape of the curve, distinguishing between decreasing, U or V-shaped, inverse U or V, double U or V in opposite directions, and, finally, increasing.

Curves that always lie above (or below) the unit value indicate that, regardless of the buffer size considered, this activity (or geographic area) has greater (or lesser) attraction power than the average of activities, irrespective of the curve's shape. On the other hand, curves that cross the unit value suggest that the diagnosis of concentration depends on the buffer radius considered for the measurement. Therefore, the assessment of a sector's (or area's) ability to attract firms should take into account different

---

[12] The construction of Geographical Spatial Concentration Indicators, $SSCI(r)$, in this case, would be obtained by aggregating the individual indicators for all production points within the geographic area for which the indicator is calculated (neighbourhood, district, etc.):

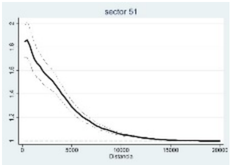
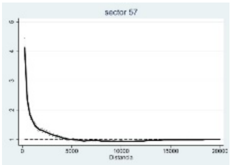
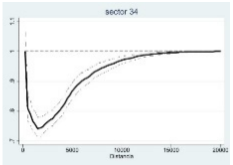
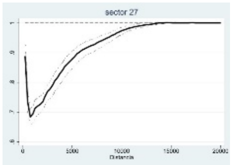
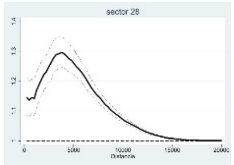
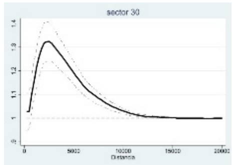
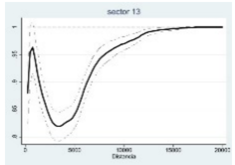
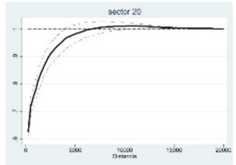
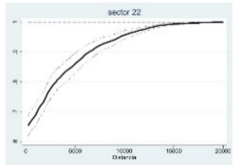
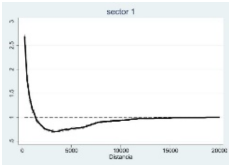
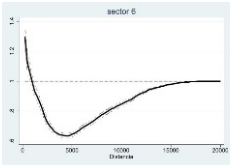
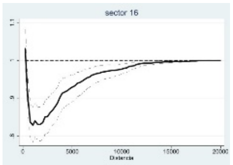
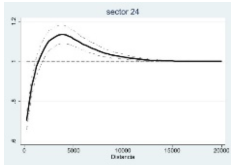
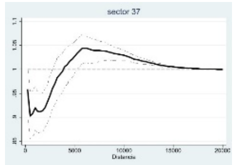
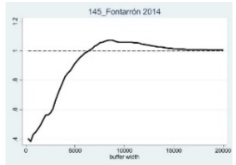
$$GSCI(r) = \frac{1}{n_g}\sum_{i\in g} SCI(x_i, r)$$

Where g refers to the geographic area for which the indicator is calculated and n_g is the number of companies from all activities located in g. It should also be noted that a specific geographic indicator could be constructed for each individual activity.

[13] However, if information at the point level is available for the activity, it is possible to construct weighted measures.

distances. However, agglomeration curves for sectors (or geographic areas) where smaller buffers have high values above (or below) the unit will tend to indicate relative concentration (or dispersion) behaviours. In contrast, if the unit is crossed from higher (or lower) values to significantly lower (or higher) ones, the sign can be reversed. Thus, sectors (or geographic areas) with type 6 agglomeration curves are expected to show general sectoral agglomeration patterns, while types 7 and 10 could indicate dispersion. Type 8 and 9 curves, in general, will have behaviours that are harder to classify and therefore more variable.

Table 1.3 Typology of spatial concentration curves (*)

	Decreasing or L	U or V	Inverse U or V	Double U and V	Increasing or inverse L
Do not cross the unit	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**
	 	 	 		 
Do cross the unit	**Type 6**	**Type 7**	**Type 8**	**Type 9**	**Type 10**
		 			

(*) Real examples obtained for the case of Madrid in 2014 and 2019.

Consequently, the previous sectoral indicators, similar to those available in the literature, do not provide a single answer regarding the level of agglomeration of each sector (or geographic area) if the curve has both positive and negative values. Additionally, comparing sectors (or geographic areas) may be complicated. For this reason, the calculation of Synthetic Sectoral Spatial Concentration Indicators (*SSSCI*) is proposed, based on the unweighted average of the individual synthetic indicators for each establishment[14].

$$SSSCI = \frac{1}{n_s}\sum_{i \in s} SISCI(x_i) \qquad [12]$$

Finally, both for each of the sectoral (or geographical) agglomeration indicators related to a specific buffer and for the synthetic indicator that does not depend on the calculation distance, their statistical significance must be evaluated. The absence of an agglomeration or dispersion process at a specific distance would be indicated by indicators that are not significantly different from one. Typically, to perform this test, confidence intervals are constructed around the null hypothesis (in this case, around the value 1). For this, a sufficiently large number of Monte Carlo simulations of the indicator are usually estimated, randomly selecting n_s points (or n_g in the case of geographical indicators), either following a Poisson process or using the actual locations of all production points from the sector and the rest of the activities. From these simulations, the confidence interval for the null hypothesis of random location is obtained, that is, the absence of an agglomeration or dispersion process. These simulations allow testing the null hypothesis of no agglomeration at each point on the curve. When the estimated agglomeration curve, or the points corresponding to a specific buffer, falls within the confidence interval of the null hypothesis, that part of the curve is considered to show no statistically significant deviation from randomness, hence, it is neither more nor less concentrated than the average.

However, this approach has two main problems, one theoretical and one with practical

[14] The Synthetic Geographical Spatial Concentration Indicators (*SGSCI*) are obtained as follows:

$$SGSCI = \frac{1}{n_g}\sum_{i \in g} SISCI(x_i)$$

consequences. The theoretical issue is that it assumes the location of establishments is random. However, in many potential locations, especially in cities, the location may be administratively restricted. As a result, using Monte Carlo simulations to construct confidence intervals can lead to rejecting the null hypothesis of no concentration, even if it is true for the subset of feasible locations for the sector, thus inflating the Type I error. In this sense, the assumption of randomness is acceptable in the context of analysing productive concentration at the regional or national level, since large distances are involved and the difference between being located in the city centre or the periphery is irrelevant. But in an urban environment, this assumption is less acceptable. Therefore, in the urban context, it seems reasonable to discard both the hypothesis of total randomness, given the division of urban space between residential and productive activity, and the idea that establishments could be located in any other activity point within the city.

This chapter proposes a different approach, which consists of creating confidence intervals for each aggregated agglomeration indicator based on those obtained for each individual production point, rather than from its null hypothesis, and evaluating whether the value for no agglomeration (in this case, one) falls within the interval, as is typically done in classical inference. Additionally, the computational cost is substantially reduced, which is especially important in the case of large sectors or numerous alternative locations.

In this case, it would be sufficient to assume that the means of the indicators calculated in [11] or [12] follow a normal distribution, allowing the confidence interval to be constructed based on the standard deviation of the mean derived from the individual indicators. However, if normality is not assumed, confidence intervals must be computed from a sufficiently large number of replications of the average indicator for each sector, calculated from representative subsamples of the point-level indicators. In other words, k_s subsamples of point-level indicators are created, and from these, an equivalent number of average spatial concentration indicators are calculated (expressions [13.a] and [13.b]). While this example uses sectoral indicators, the same process can be applied to geographical indicators, providing alternatives to those calculated with all the establishments.

$$SSCI(r)^j = \frac{1}{k_s}\sum_{\substack{i=1 \\ i\in s}}^{k_s} SCI(x_i, r), \text{ where } k_s < n_s \qquad [13.a]$$

$$SSSCI^j = \frac{1}{k_s}\sum_{\substack{i=1 \\ i\in s}}^{k_s} SISCI(x_i) \qquad [13.b]$$

To determine the number of replications k_s and the number of point-level indicators (*j*) considered in each of these "alternative" indicators, the minimum sample size required for random sampling in finite populations is calculated according to expression [13.c]:

$$k_s = \frac{N_s z_\alpha^2 pq}{e^2(N_s - 1) + z_\alpha^2 pq} \qquad [13.c]$$

Where z_α is the confidence level according to the normal distribution; *p* is the expected proportion and *q* = 1 - *p*. Since an expected proportion is absent in this case, the sample size is maximized when *p* = *q* = 0.5; *e* refers to the admissible error or precision. In this way, a number of indicators are calculated, and by eliminating the upper and lower 2.5% of the extreme values, the confidence interval for each sectoral (or geographic) agglomeration indicator is obtained. Additionally, similar to how the agglomeration curve was constructed, the confidence interval for this curve is obtained based on the previous values for each of the buffers considered. Thus, if the unit lies within this interval, the null hypothesis of no agglomeration for that sector (or geographic area) in that buffer cannot be rejected. The same approach is applied to the synthetic indicators.

1.3.3. Synthetic indicators based on statistical analysis

The previous sectoral (and geographic) indicators do not fully exploit all the individual information. Specifically, in the sectoral indicators proposed, a local reference has not been used, unlike some pre-existing indicators, in a strict sense, meaning that they are not normalised by the productive density of the surrounding area, and therefore, the distinction between join location (first-order concentration) and colocation (second-order concentration) is not adequately made. Similarly, when calculating geographic indicators, the different sectoral composition should be considered, as it could alter the value. To obtain these “refined” indicators, a least-squares regression is used to estimate a measure of production agglomeration for each sector and geographic area simultaneously, thus controlling for the other dimension. That is, for sectoral indicators, this approach controls for the spatial location of production points, correcting for the varying concentration of productive activities across the territory, and for geographic indicators, it controls for sectoral composition, correcting for the differing concentration of productive activities. In a way, the procedure follows the ideas of the Gibbs model[15], but it is aimed at constructing an agglomeration indicator rather than explaining concentration. Therefore, from the individual spatial concentration indicators $SCI(x_i, r)$ and $SISCI(x_i)$, it is possible to obtain sectoral and geographic average indicators by estimating expressions [14.a] and [14.b].

$$SCI(x_i, r) = \alpha + \sum_s \alpha_s^1 d_{si} + \sum_q \beta_q^1 d_{qi} + \delta^1 z_i + \varepsilon_{isr}^1 \qquad [14.a]$$

$$SISCI(x_i) = \alpha + \sum_s \alpha_s^2 d_{si} + \sum_q \beta_q^2 d_{qi} + \delta^2 z_i + \varepsilon_{isr}^2 \qquad [14.b]$$

Where d_{si} and d_{qi} represent sector and geographical (neighbourhood, rasters, etc.) dummies in which each point is located; z_i captures establishment characteristics that may influence and distort the concentration measure, such as belonging to a shopping centre. In expression [14.a], the dependent variable corresponds to the $SCI(x_i, r)$. In this case, for each production point, there are measures for different buffers, but dummies related to these different buffers are not included since, by the construction of the indicators, the average of all of them for each radius considered is equal to one. However, when using indicators for different buffers in the same estimation, expression [14.a] is estimated using weighted least squares, with the weights based on the function $f(r)$. The weighting for each buffer in this indicator is identical to that used in constructing the synthetic point-level indicators.

From the coefficients α_s^1 or α_s^2, which are equivalent in value though they have

[15] A description of the model and its statistical properties can be found in Sweeney & Gómez-Antonio (2016).

different standard deviations, the Statistical Synthetic Sectoral Spatial Concentration Indicators[16](*SSSSCI*) are constructed. Similarly, β_q^1 and β_q^2, which are also equivalent in value though their standard deviations differ, allow for the calculation of the Statistical Synthetic Geographical Spatial Concentration Indicators[17](*SSGSCI*). From these indicators, the attractiveness of a neighbourhood can be determined.

Finally, it is important to assess to what extent the family of indicators proposed here, inspired by point pattern theory and following Ripley's K-density function, meet the desirable conditions for agglomeration indicators as outlined by Duranton & Overman (2002, 2005). First, the designed indicators meet the criterion of comparability across industries, which is achieved by first establishing a reference within the sector in step 2 and a local-global reference in step 3.

The second condition, control for overall agglomeration trends across industries in the territory, is again addressed in steps 3 and 4. In step 3, by establishing a local but globally scoped reference (i.e., normalising based on how companies are globally concentrated in the study area at a certain distance), the average trend of geographic concentration of productive activity in the studied territory is taken into account. Additionally, in step 4, for the statistical econometric indicators used to obtain the *SSSSCI* and *SSGSCI*, the first (sectoral) controls for the concentration between different administrative areas and the distribution of productive activity. In the second (geographic), it controls for the sectoral composition within each administrative area for which the geographic indicator is calculated. In the previously mentioned indicators, the distinction between sectoral and geographic concentration is also clearly separated (third condition).

All the indicators, being continuous, are unbiased with respect to the degree of spatial aggregation (fourth condition). Finally, the fifth condition is addressed and simplified

[16] To obtain the final value of the Statistical Synthetic Sectoral Spatial Concentration Indicator (*SSSSCI*) for each sector, the coefficient obtained from the corresponding dummy, the constant term, and the weighted average of the neighbourhood dummies must be added together.
[17] Similarly, to obtain the final value of the Statistical Synthetic Geographical Spatial Concentration Indicator (*SSGSCI*), the coefficient obtained from the corresponding dummy for each area, the constant term, and the weighted average of the dummies for the various sectors must be added together.

in step 5 by establishing a procedure for creating confidence intervals for the indicators, as well as by testing the coefficients in the statistical-econometric analysis of the aggregated indicators derived from individual ones, although the latter assumes normality.

These indicators also meet the criteria proposed by Kopczewska et al. (2019), at least theoretically, although in some cases this depends on their implementation and the available data. In principle, these indicators are sensitive to different spatial distributions in the sense they propose, depending on the detail or granularity with which the measurement of neighbouring points is conducted. For example, a measurement that does not introduce a sufficient number of buffers, or enough based on the observed distribution of production points, may not detect these differences in the distribution of activity across the territory.

The second criterion also has the same type of constraints as the first, as it depends on the specific application. While geographic location will always be taken into account by design, the size of the production points will only be considered in the case of weighted indicators (see Chapter 2). Where this family of indicators undoubtedly excels is in the understanding and interpretation of each of its tools. Additionally, it offers unique numerical indicators, independent of the measurement buffer, at the production point, industry, and geographic area levels, as these authors advocate.

Finally, they also meet the two additional criteria introduced by Kominers (2008), as they can be measured using available information (though more sophisticated versions require a greater amount of data), and they are justified by an appropriate theoretical model, being part of the *K*-Ripley density function family.

1.4. Data

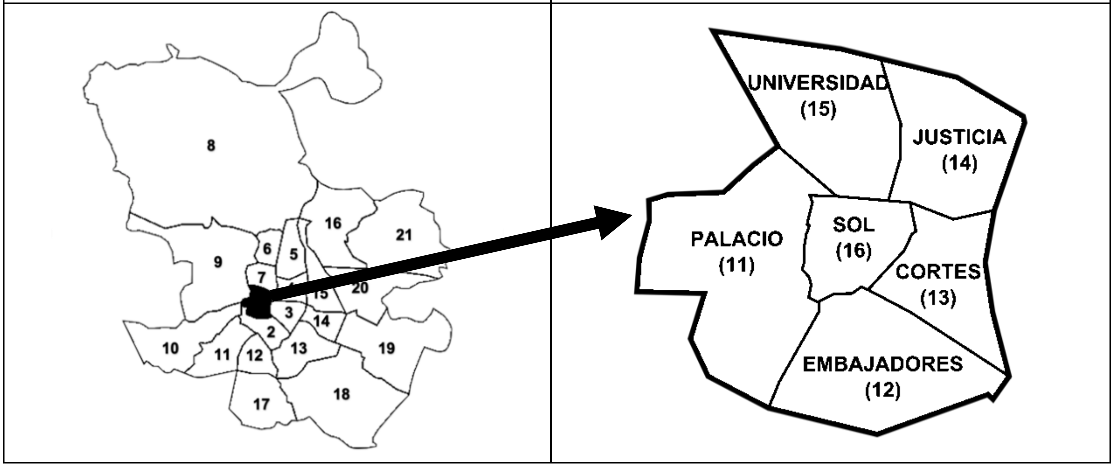
1.4.1. Study area

The municipality of Madrid is located in the centre of the Iberian Peninsula and is the capital of Spain. It is the most populated municipality in Spain, according to INE (2023).

It consists of 131 neighbourhoods (since 2017) and is administratively divided into 21 districts[18] (Figure 1.3). It has an area of 604.3 square kilometres.

Figure 1.3 Administrative units in Madrid

Districts of Madrid	Neighbourhoods in Centro District



1.4.2. Data source

The data used for measuring agglomeration in the city of Madrid come from the Census of Premises and Activities of the Madrid City Council. This administrative database contains records of all premises open at street level in Madrid, with information on the associated activities, including the address, latitude and longitude coordinates based on geocoded street addresses, opening status, activity, etc. This dataset offers current information from March 31, 2014, with monthly updates[19]. It is important to note that this information pertains to establishments, not firms, as a single firm may have multiple establishments. This database relates exclusively to the municipality of Madrid (the city of Madrid). Although the metropolitan area of Madrid, which lacks formal administrative status, might be a more appropriate geographic scope, equivalent data is not freely available from other municipalities within the metropolitan area. Consequently, the database provided by the City Council of Madrid is chosen for its accessibility and unique characteristics. Notably, it offers data at the premise level,

[18] Regulation of the Districts of the City of Madrid, dated December 23, 2004, modified by the Plenary Agreement of October 31, 2017. Throughout this work, this division into 131 neighbourhoods is used.

[19] Two datasets are combined for the analysis. The first dataset contains records of business premises, where each record corresponds to a unique premise. The second dataset consists of activity records, where each entry corresponds to a premise along with an associated activity code. If a premise is linked to multiple activity codes, it appears in the activity dataset as many times as the number of associated codes.

closely aligned with the concept of establishments. The database is comprehensive, encompassing all business types regardless of their legal or fiscal structure, and specifically includes independent entrepreneurs who are not incorporated as commercial companies. This inclusion is particularly important for understanding the small businesses prevalent in urban areas.

To measure and study the change in agglomeration patterns, we have used data from December 2014 and December 2019. We considered 62 activities of certain homogeneity based on the activity classification presented by each establishment, and we removed establishments without activities[20]. Appendix A provides specific details of the sectoral aggregation carried out. The final sample includes 99,939 production points for 2014 and 110,997 production points for 2019.

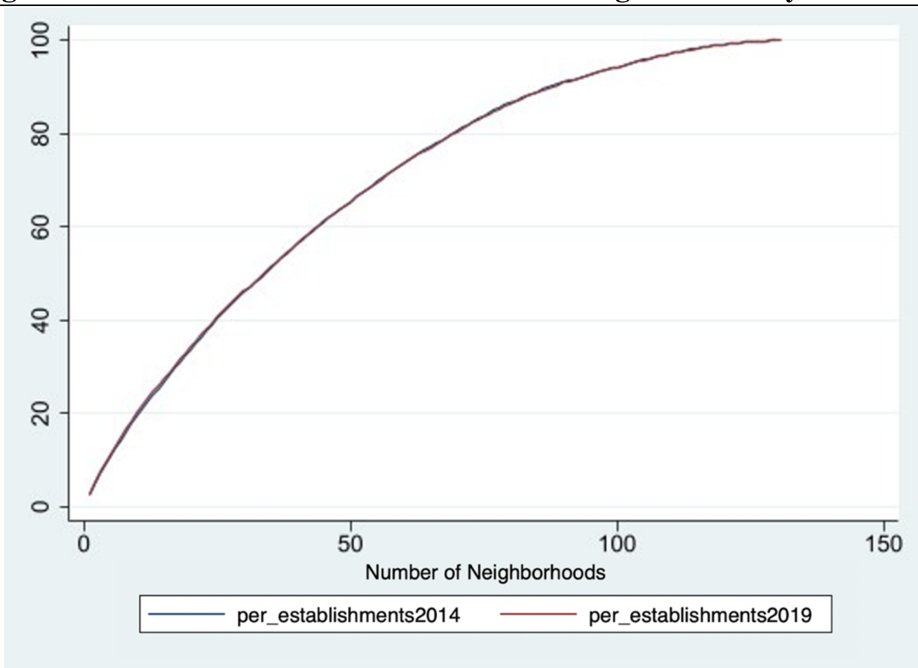
Regarding the geographic units of analysis, the location of each establishment is known through its geographic coordinates (UTM coordinates in meters)[21], which identify the main entrance from the street to the establishment. In addition to the exact location and activities carried out in each establishment, it is known whether the establishment is part of any type of geographical grouping (e.g., a commercial centre). However, there is no economic or accounting information available for each of the establishments. In Figure 1.4, the distribution of production points throughout the city can be observed. In the upper part, the cumulative percentage of establishments in each neighbourhood is represented once they are ordered from highest to lowest in terms of the number of establishments. The concave shape of the curve indicates a certain concentration of

[20] A single establishment with more than one activity becomes two production points.

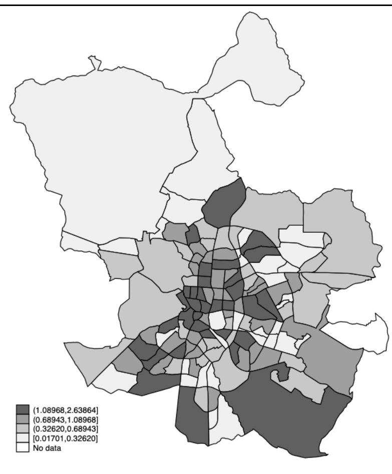
[21] The UTM (Universal Transverse Mercator) geographic coordinate system is used to reference any point on the Earth's surface, utilizing a cylindrical projection to represent the Earth on a plane (UTM coordinates in meters). It is important to note that before September 15, 2017, the reference system employed was ED-50 (European Datum 1950). From that date onward, it transitioned to ETRS89 (European Terrestrial Reference System 1989). For the estimation, the reference system from December 2014 was converted to the European Terrestrial Reference System 1989. The data have been organised in a georeferenced database linked to a geographic information system (ArcGIS), facilitating the management of statistical and cartographic information and its interface with STATA. Regarding the geographic units of analysis, the location of each establishment is determined by its geographic coordinates, which correspond to the main street entrance of the establishment. In addition to precise location and activity details for each establishment, the dataset provides information about the neighbourhood where the establishment is located and whether it is part of a commercial centre. However, no economic or accounting data are available for the establishments.

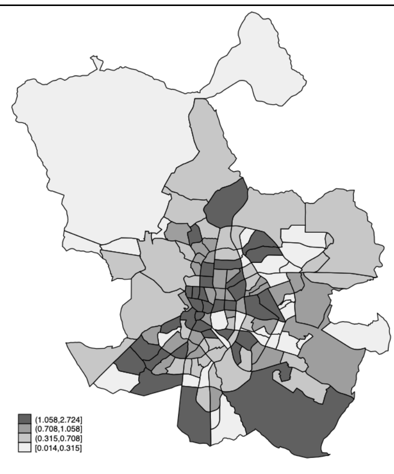
activities in some neighbourhoods, especially in the centre (maps in the lower part). Additionally, there is significant stability in the geographic distribution between 2014 and 2019.

Figure 1.4 Distribution of establishments throughout the city of Madrid



2014	2019





1.5. Results

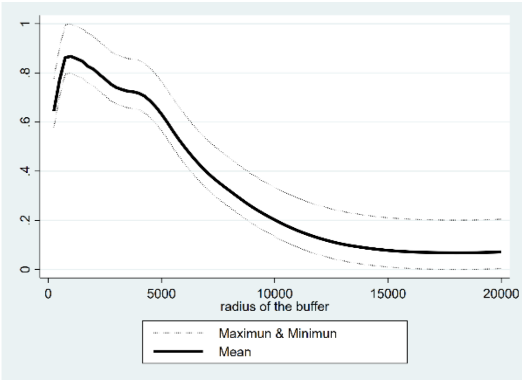
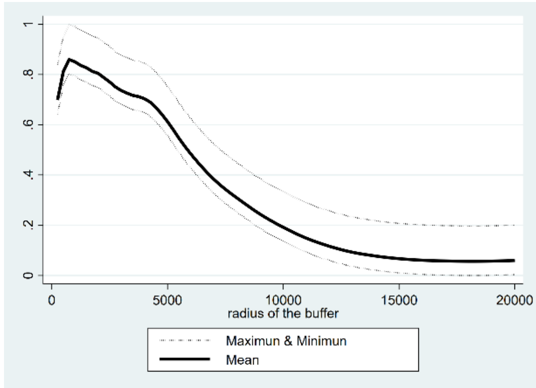
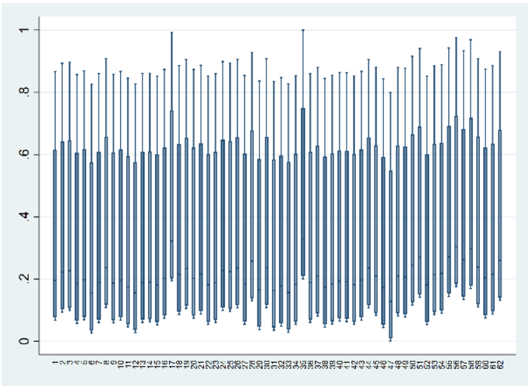
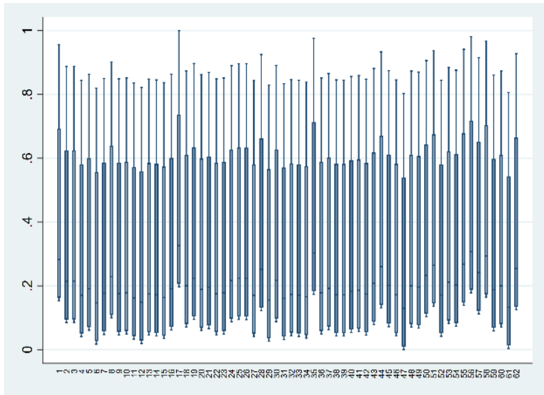
1.5.1. Calculation of the distance-based point-level indicators for sectoral agglomeration

The results will be discussed following the order in which the methodology was developed. The number of establishments around each activity point within a maximum distance is calculated for buffers ranging from 250 meters to 20,000 meters, with increments of 250 meters[22] (these are the 80 different buffers that will be considered).

To estimate the exponent γ_{sr}, which modifies the usual correction by the inverse of the overlap between the considered buffer and the administrative area, expression [7.d] is estimated separately for each year. The only known characteristic of the establishment that is incorporated is whether it is grouped with others (e.g., in a shopping centre). Additionally, sectoral dummies (61 corresponding to all sectors considered, minus one to avoid collinearity) and neighbourhood dummies (130 corresponding to all neighbourhoods, minus one) are introduced. Finally, the value of the obtained coefficients is normalised between 0 and 1. The results for this exponent are presented in Figure 1.5. As shown in Panel A of the figure, as the buffer size increases, the exponent becomes smaller, although it does not reach 0 on average. This indicates that, as expected, the correction by the inverse of the overlap is exaggerated and depends on the buffer size considered. Thus, when the buffers are small, the correction may be acceptable. As the buffer size increases, and with it the probability that the study area extends beyond the administrative boundary, and more importantly, that the overlap between the buffer and the administrative area (λ_{ir}) decreases, the correction becomes less necessary. Conversely, as seen in Panel B of this Figure 1.5, which presents box plots by sector, activities do not seem to affect the average correction. Additionally, the results are quite similar for the two selected years.

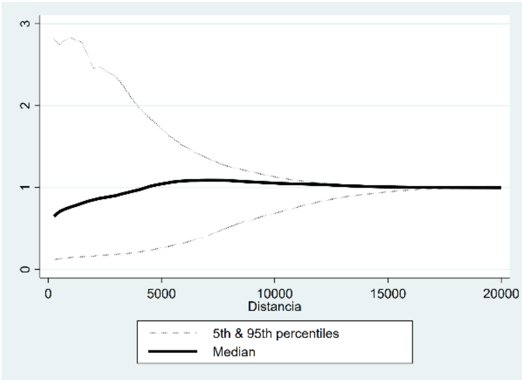
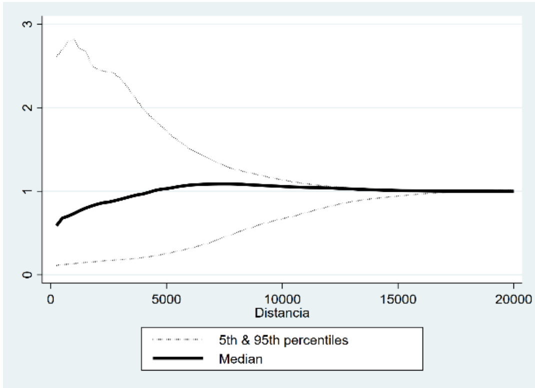
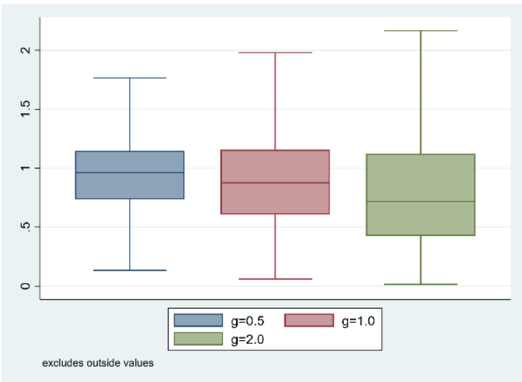
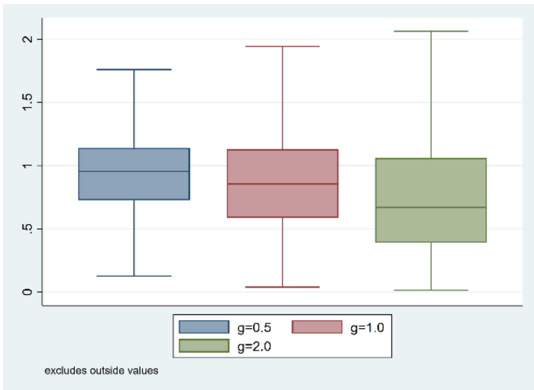
[22] 20,000 meters corresponds to the 99th percentile of the bilateral distances between establishments in the city of Madrid.

Figure 1.5 Overlap correction moderated by buffer width and sector

2014	2019
A. Evolution of overlap correction moderator (γ_{sr}) by buffer width	
	
B. Boxplots of the overlap correction moderator (γ_{sr}) by sector of activity	
	

Once the estimation of these exponents is available, it is possible to calculate sectoral agglomeration indicators at the establishment or production point level, both for each of the buffers (Panel A of Figure 1.6) and the synthetic indicators, considering different weighting functions ($g \in \{0.5, 1.0, 2.0\}$) for the indicators obtained for each buffer in the construction of the synthetic indicator (Panel B of Figure 1.6).

Figure 1.6 Distance-based point-level indicators for sectoral agglomeration

2014	2019
A. Individual spatial concentration indicators, $SCI(x_i, r)$	
	
B. Synthetic individual spatial concentration indicators, $SISCI(x_i)$	
	

Regarding the $SCI(x_i, r)$, it is observed that the values converge to 1 as the buffer size increases. This phenomenon occurs by design, as increasing the buffer size includes a larger number of activity points from the same sector in the numerator of the indicator, thus making the proportion of points around each establishment tend toward one. On the other hand, it is also observed that the median is below one, at least until buffers of 7.5 Kilometres. Additionally, it is noted that the values of the 95th percentile converge toward the reference value more quickly than those of the 5th percentile. Both effects are related to the fact that the values of the indicator for smaller buffers are greater than one and are also further from one than those for larger buffers, which creates this asymmetry.

Regarding the synthetic indicators, $SISCI(x_i)$, the box plots are shown in Panel B of Figure 1.6. It is worth noting that the selected weighting function significantly affects the distribution of the indicators. The higher the exponent of the function (results are shown for 0.5, 1, and 2), the lower the median of the indicator and the greater the dispersion between production points. Henceforth, only the results obtained with the unitary exponent will be used.

1.5.2. Results at sectoral level

1.5.2.1. Sectoral agglomeration curves

The individual spatial concentration indicators, $SCI(x_i, r)$, can be aggregated to obtain Sectoral Spatial Concentration Indicators, $SSCI(r)$. By arranging these according to the buffer size, it is possible to plot sectoral agglomeration curves, which can be found for the 62 activities considered in Table A.1 of Appendix A.

The shapes obtained are summarised in Table 1.4. It shows that most establishments are in sectors that exhibit stable agglomeration behaviour, i.e., whose curves do not cross the unit, regardless of the buffer considered (35 sectors in both 2014 and 2019). In general, sectors showing dispersion dominate (35 sectors in 2014 and 33 in 2019), compared to those showing agglomeration behaviour (22 sectors in 2014 and 21 in 2019). Since the behaviour of establishments in each sector is compared to the average of all existing ones, this result indicates that either the agglomerated sectors have a greater number of establishments, or that agglomeration behaviour (the value of the indicator) is more intense on average than dispersion. Once again, the results are similar in both years.

Table 1.4 Typology of sectoral agglomeration curves obtained with establishments in Madrid

		Decreasing or L	U or V	Inverse-U or V	Double-U and V	Increasing or inverse L
Do not cross the unit		**Type 1(A)**	**Type 2(D)**	**Type 3(A)**	**Type 4(D)**	**Type 5(D)**
	2014	11 sectors	2 sectors	4 sectors	1 sector	17 sectors
	2019	12 sectors	4 sectors	2 sectors	1 sector	16 sectors
Do cross the unit		**Type 6(A)**	**Type 7(D)**	**Type 8**	**Type 9**	**Type 10(D)**
	2014	7 sectors	12 sectors	3 sectors	2 sectors	3 sectors
	2019	7 sectors	8 sectors	5 sectors	3 sectors	4 sectors

Note: (A) represents agglomeration behaviour; (D) refers to dispersion behaviour

1.5.2.2. Aggregated indicators

The availability of sectoral agglomeration curves may not answer the question of whether a specific activity is agglomerated or not, as in many cases, i.e., sectors where the curve crosses the unit, it depends on the distance at which the diagnosis is made. Therefore, aggregated indicators are proposed, which, on the one hand, complement the curves (as the curves likely contain richer information) and, on the other hand, simplify the data, as they summarise, albeit in a limited way, the information contained in them into a single value. Two alternative methods allow for the calculation of these aggregated sectoral agglomeration indicators. The first is by aggregating the synthetic concentration indicators at the point level, following expression [12], and obtaining the Synthetic Sectoral Spatial Concentration Indicators (*SSSCI*). These indicators are obtained through a double average. First, a weighted arithmetic mean of the individual spatial concentration indicator values for different buffers is calculated, weighting by an inverse function of distance, to obtain an indicator for each production point ($SISCI(x_i)$). A second simple arithmetic mean is then calculated for all the previous indicators for all the production points in the sector. The results for each sector, for the specific case of Madrid in the two selected years, can be found in the first two columns of Table A.2 in Appendix A.

The second method consists of estimating expressions [14.a] or [14.b]. In the first expression, the individual spatial concentration indicators, obtained by considering different buffers, are regressed against the characteristics of the establishment, which in this case consists of an indicator variable for whether the establishment belongs to a shopping centre, along with sector and neighbourhood dummies. This estimation is done in a weighted manner, using the function $f(r)$ as the weight. In this way, to get the indicator for each sector, the coefficient obtained from the corresponding dummy is added to the constant term and the weighted average of the neighbourhood dummies.

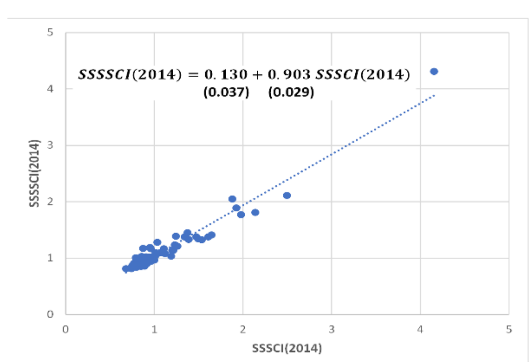
The results for the Statistical Synthetic Sectoral Spatial Concentration Indicators (*SSSSCI*) for the two years under study are presented in the third and fourth columns of Table A.2 in Appendix A[23]. There is an important theoretical difference between the first indicators, obtained by aggregation, and the second, which are obtained by estimation. While the former reflect a simple average across all data, the latter are sectoral indicators that eliminate the location bias. In other words, if a sector has a greater presence in a neighbourhood due to either its relative location in the city (e.g., the city centre) or some idiosyncratic characteristic of the geographic area that favours economic activity (e.g., the development of industrial zones), these effects will be removed in the calculation of the estimated indicator.

For this reason, the equality between both indicators for the same year is rejected at the usual significance levels (see Figures 1.7a and 1.7b). Specifically, the regression line between the two indicators in both years shows that the constant is statically different from zero, with a slope less than one. This means that the *SGSSCI* are higher for less agglomerated sectors and lower for more agglomerated ones, possibly as a result of the geographic effect correction. However, in both cases, there is significant temporal stability (Figures 1.7c and 1.7d). In fact, the statistical similarity between the *SSSCI* and the *SSSSCI* for the two years analysed is accepted, indicating temporal stability during this period in Madrid.

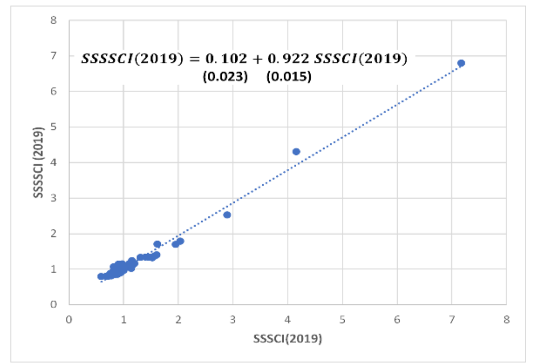
[23] The results obtained from the estimation of equation [14b] are, by construction, the same as the previous ones; only the statistical significance of the obtained indicators changes, as fewer data points are available, i.e. just one per production point, compared to 80 per point in the first case. Table A.1 includes a single value for the coefficient and its two significance levels based on the estimation.

Figure 1.7 Relationship between synthetic sectoral agglomeration indicators

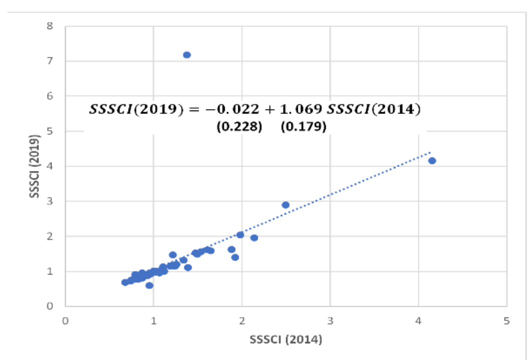
a. *SSSCIs* and *SSSSCIs* in 2014



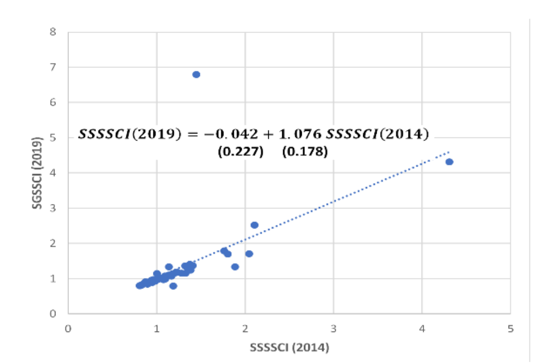
b. *SSSCIs* and *SSSSCIs* in 2019



c. *SSSCIs* in 2014 and 2019



d. *SSSSCIs* in 2014 and 2019



Note: SSSCI refers to Synthetic Sectoral Spatial Concentration Indicators, SSSSCI refers to Statistical Synthetic Sectoral Spatial Concentration Indicators.

In Table 1.5, the sectors in 2014 and 2019 are classified based on their level of aggregated agglomeration using the *SSSSCI*[24]. When analysing this classification, it is observed that sectors with restricted locations in the city, especially manufacturing sectors, those that use land intensively, those with high added value, and those related to tourism and leisure are the ones that exhibit a high level of urban geographic concentration. This concentration is not necessarily in the same areas of the city, as the first two types of activities are concentrated in specific areas on the periphery, while the latter two are concentrated in the city centre. On the opposite end, with greater dispersion, are those sectors related to everyday consumption, which require proximity to consumers.

[24] The values selected to classify the sectors are taken ad-hoc

Table 1.5 Classification of activity sectors in 2014 and 2019 according to the intensity of establishment concentration

2014 \ 2019	High agglomeration	Medium-high agglomeration	No agglomeration No dispersion	Medium-high dispersion	High dispersion
High agglomeration	Food (M) Wood & paper (M) Chemicals (M) Transport materials (M) Wholesale trade Accommodation Entertainment Retail clothing Jewelry stores Sex-shops Street markets Restaurants Bars with shows	Furniture (M) Supplies Fast food			
Medium-high agglomeration	Retail beverages	Textile (M) Graphic arts Metal products (M) Software Media Travel Agencies Libraries & Museums Petrol stations	Other manufacturing Automobile (S) Insurance Building services Bookstores	Public Administration	Banquet & dinning rooms
No agglomeration No dispersion		Residences	Mail offices Architecture	Research & Marketing	
Medium-high dispersion				Consulting Veterinarians Social services Gaming Fresh food retail trade Retail electronic commerce Perfumeries Garden & Pets	Construction Transportation Act. Auxiliary Comp. Hardware & DIY
High dispersion				Bakeries	Electrical & electronic prod. Telecommunications Real state Education Health care Sports Associations Personal services Non-specialized retail trade Stores Ice-cream Pharmacies Bars

Note: High-agglomerated industries: SSSSCI ≥ 1.25; Medium-high agglomerated industries: 1.00 < SSSSCI < 1.25; No agglomeration, no dispersion: SSSSCI = 1; Medium-high dispersed industries: 0.90 ≤ SSSSCI < 1; High-dispersed industries: SSSSCI < 0.90.

1.5.3. Results at neighbourhood level

1.5.3.1. Agglomeration curves at the neighbourhood level

One contribution of this chapter is the construction of agglomeration indicators with a geographical perspective based on the calculation of indicators at the production point level. Similar to how sectoral agglomeration indicators are calculated, it is possible to build aggregated agglomeration indicators for firms/establishments located in an administrative context, in this case, the neighbourhoods that make up a city. Therefore, the Geographical Spatial Concentration Indicators ($GSCI(r)$) are first obtained for each neighborhood in the city of Madrid and for each buffer size, following expression [11.b]. By arranging the indicators obtained for each neighbourhood by buffer size, Geographical Concentration Curves are obtained. In this case, the agglomeration curves would be interpreted as the ability of production points located in each neighbourhood to attract production points from the same productive activities at a certain distance. The peculiarity of this tool is that it combines production points from different activities that share the same location within a defined area (in this case, neighbourhoods, but it could be rasters, districts, or any other geographical area), and these curves can be interpreted as a characteristic of the geographical area for which they are calculated. The curves obtained for each neighbourhood in Madrid can be found in Table A.3 of Appendix A.

The typology of curves as the buffer radius increases is similar to that found in the sectoral case, although the distribution among types is different. Specifically, in Table 1.6, it is shown that curves indicating some form of dispersion dominate (72 neighbourhoods in 2014 and 71 in 2019). This suggests that the agglomeration process is concentrated in a very small number of neighbourhoods within the city.

Table 1.6 Typology of agglomeration curves for Madrid's neighbourhoods obtained from establishments

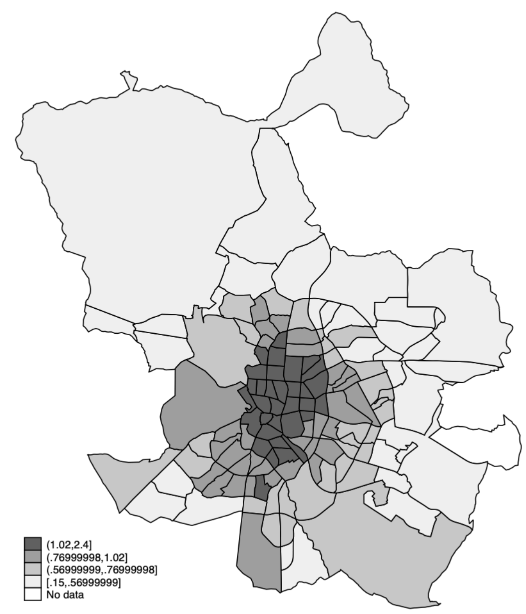
		Decreasing or L	**U or V**	**Inverse U or V**	**Double U and V**	**Increasing or inverse L**
Do not cross the unit		**Type 1 (A)**	**Type 2 (D)**	**Type 3 (A)**	**Type 4 (D)**	**Type 5 (D)**
	2014	2 neighb.	23 neighb.	13 neighb.	-	29 neighb.
	2019	2 neighb.	24 neighb.	10 neighb.	-	29 neighb.
Do cross the unit		**Type 6 (A)**	**Type 7 (D)**	**Type 8**	**Type 9**	**Type 10 (D)**
	2014	-	6 neighb.	29 neighb.	14 neighb.	14 neighb.
	2019	1 neighb.	4 neighb.	32 neighb.	15 neighb.	14 neighb.

Note: (A) represents agglomeration behaviour; (D) refers to dispersion behaviour

1.5.3.2. Aggregated indicators the neighbourhood level

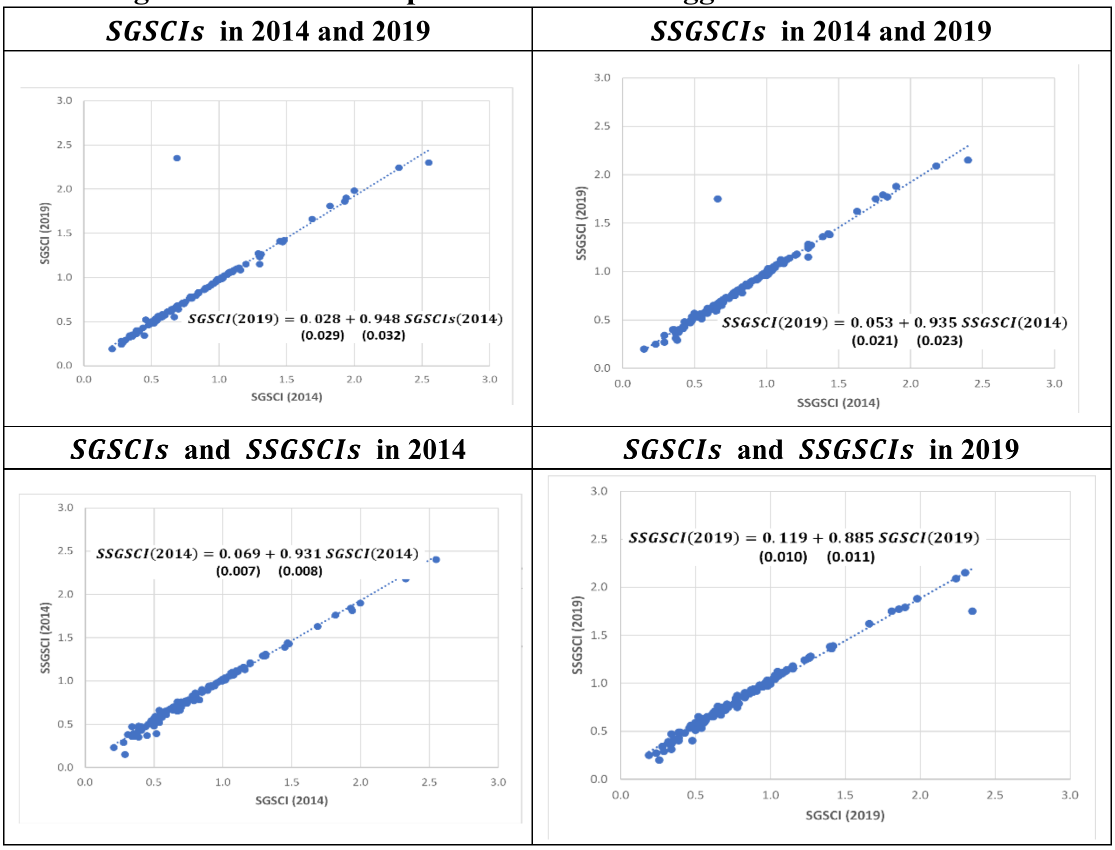
Similar to the sectoral indicators, synthetic indicators can also be obtained for geographic units, in this case, neighbourhoods. The Synthetic Geographical Spatial Concentration Indicators (*SGSCI*), constructed by aggregating the $GSCI(r)$ or $SISCI(x_i)$, and the Statistical Synthethic Geographical Spatial Concentration Indicators (*SSGSCI*) for each neighbourhood, estimated from the individual and synthetic spatial concentration indicators, are presented in Table A.4 of Appendix A and shown in Figure 1.8. The figure highlights two key observations: first, the strong stability in the ability of central neighbourhoods to attract productive activity between 2014 and 2019, and second, the growth of new production centres in the south, mainly manufacturing, and some in the east of the city, which have increased their capacity to attract productive activities that are restricted from the city centre.

Figure 1.8 Statistical synthetic geographical spatial indicators for the neighbourhoods of Madrid

2014	2019
	

On the other hand, the comparison between the *SGSCI* and the *SSGSCI*, both with each other and in their temporal evolution (Figure 1.9), yields a noteworthy result. While equality is accepted for the *SGSCI* between the two years analysed, this is not the case for the *SSGSCI*. This is because the two indicators do not measure exactly the same thing, as the latter removes the sectoral effect. Therefore, it seems that between 2014 and 2019, there is some degree of spatial concentration of activity, but with a certain sectoral behaviour. When both indicators are compared across the two years, the pattern found in the sectoral indicators becomes more pronounced. Once again, this difference highlights the ability of the "statistical" indicators to remove the sectoral composition bias in diagnosing the attraction capacity of production points located in each neighbourhood.

Figure 1.9 Relationship between sectoral agglomeration indicators



Note: SGSCI refers to Synthetic Geographical Spatial Concentration Indicators, SSGSCI refers to Statistical Synthetic Geographical Spatial Concentration Indicators.

1.6. Conclusions

The main objective of this chapter is to propose a new distance-based indicator for sectoral agglomeration at the point level, which, following the family of indicators based on the K-Ripley density function, allows for both the traditional sectoral interpretation and a geographical one once aggregated. Additionally, a unique quantitative assessment is developed for each sector or geographic area, moving away from current indicators that provide this information based on the buffer used in the measurement. Furthermore, the goal of simplifying the process of testing the null hypothesis of no concentration is also addressed. To achieve this, the methodology used in previous studies is streamlined, following the proposal of Marcon & Puech (2017). The construction of the indicator at the production point level requires some changes to the traditional method of calculating these measures, which use microdata from production points.

Once the process of developing the indicator is described, a simple statistical analysis is used based on the individual point-level indicators. By using sector and geographic area dummies, this allows for the calculation of more precise indicators for each of these two dimensions, controlling for geographic concentration in the first case and sectoral concentration in the second.

In response to the final objective, an alternative procedure to the current Monte Carlo simulations for calculating confidence intervals of the null hypothesis, which are based on the randomness of the location of productive activities, is developed. Instead, confidence intervals are created around the mean values of each sector based on the individual indicators, introducing the opposite assumption: that activities restrict their location to their current ones. This leads to significant savings in computation time. Non-randomness in location is particularly relevant in the urban context, where location possibilities are restricted due to competition between productive, residential, and service activities, as well as for safety, health, and social considerations (e.g., noise). The reduction in computation time is especially important when analysing large amounts of production data, as the procedure developed here reduces computation times to a hundredth or even a thousandth of the time, depending on the number of simulations performed.

To illustrate the capabilities of this new set of indicators, data from the Census of Premises and Activities of the Madrid City Council for December 2014 and 2019 are used. The application of this set of indicators allows for the measurement of productive agglomeration in the context of a city. In this specific case, a significant concentration is found in activities with important location restrictions, such as manufacturing, that intensively use land, are of high added value, and are related to tourism. The first two are located in specific areas on the periphery, while the latter two are concentrated in the city centre. On the other hand, there is significant dispersion in activities related to everyday consumption, which require proximity to consumers.

From a geographical perspective, there is significant persistence in the high concentration of activity in the central neighbourhoods of the city and in some peripheral areas where industrial parks have been developed.

Appendix A.

Table A. Sectoral classification

Sector	Name	Description
1	Food (Manufacturing)	Manufacture of food, beverages and tobacco
2	Textile (Manufacturing)	Textile, clothing, leather and footwear industry
3	Wood and paper (Manufacturing)	Wood, cork, paper and basketry industry
4	Graphic arts	Graphic arts and reproduction of recorded media
5	Chemicals (Manufacturing)	Chemical, pharmaceutical, rubber and plastic industry
6	Metal Products (Manufacturing)	Manufacture of iron, steel, metal products and other non-metallic mineral products,
7	Electrical and electronic products (Manufacturing)	Manufacture of computer, electronic and optical products, electrical material and equipment and repair and installation of machinery, equipment, computers and personal and household goods
8	Transport Material (Manufacturing)	Manufacture of motor vehicles, trailers, semi-trailers and other transport material
9	Furniture (Manufacturing)	Manufacture of furniture
10	Other Manufacturing	Other manufacturing industries
11	Supplies	Electricity, gas, steam and air conditioning supply; water collection, purification and distribution; wastewater collection and treatment; waste collection, treatment and disposal; decontamination activities and other waste management services
12	Construction	Building development and construction; Civil engineering; Specialized construction and building completion activities
13	Automobile (Trade)	Sale and repair of motor vehicles and motorcycles
14	Wholesale Trade	Wholesale trade and commission trade activities, except of motor vehicles and motorcycles
15	Transportation	Land, pipeline and air transportation; warehousing and support activities for transportation
16	Postal and delivery sector	Postal and courier activities
17	Accommodation	Accommodation services
18	Software	Publishing (including computer software); programming, consulting and other computer-related activities
19	Media	Motion picture, video and television programme activities, sound recording and music publishing activities; radio and television programming and broadcasting activities; information service activities
20	Telecommunications	Telecommunications
21	Insurance	Financial, insurance, reinsurance and pension fund activities, except compulsory social security; activities auxiliary to insurance and financial services
22	Real estate	Real estate activities; rental and leasing activities
23	Consulting	Legal and accounting activities

Table A. Sectoral classification(continued)

24	Auxiliary services	Activities of head offices; business management consultancy activities; employment related activities; office administrative and other business support activities
25	Architecture	Architectural and engineering activities; technical testing and analysis
26	Research and marketing	Research and development; advertising and market research; other professional, scientific and technical activities; security and investigation activities
27	Veterinarians	Veterinary activities
28	Travel Agencies	Travel agency, tour operator, tour reservation service and related activities
29	Building Services	Building services and gardening activities
30	Public Administration	Public administration, defense, compulsory social security; activities of extra-territorial organizations and bodies
31	Education	Education
32	Health care activities	Health care activities
33	Residences	Assistance in residential establishments
34	Social Services	Social work activities without accommodation
35	Entertainment	Creative, artistic and entertainment activities
36	Libraries and museums	Activities of libraries, archives, museums and other cultural activities
37	Gaming	Gambling and betting activities
38	Sports	Sporting, recreational and entertainment activities
39	Associations	Associative activities
40	Personal Services	Other personal service activities
41	Non-specialized retail trade	Retail trade in non-specialized establishments, self-service, department stores, convenience stores (24h), bazaars and similar establishments
42	Fresh food retail trade	Retail sale of fruits and vegetables, butcheries, delicatessen, poultry, eggs, offal, fish, seafood and frozen food.
43	Bakeries and bakeries	Retail sale of bread, bakery products, buns, pastry, confectionery and prepared dishes
44	Retail beverages	Retail sale of milk, dairy products, soft drinks, wines and spirits without consumption
45	Stores	Retail sale of tobacco and smokers' articles
46	Ice cream and dried fruit shops	Retail sale of ice cream, candy, dried fruits, variants, potato chips, churro shop, coffee, herbal infusions and chocolate
47	Petrol stations	Retail sale of automotive fuel in specialized stores
48	Retail electronic commerce	Retail trade of office machinery and equipment, computer products (computers, software, peripheral equipment and consumables), telephone and telecommunications products, audio-visual equipment and devices, electronic equipment, photographic and photographic material and optical material

Table A. Sectoral classification(continued)

49	Hardware and DIY stores	Retail trade of household textiles, hardware products, construction materials, doors, windows and shutters, DIY stores, sanitary equipment, electrical equipment, carpets, rugs, wall and floor coverings, household appliances, furniture, household equipment, kitchen furniture, mattresses and lighting equipment.
50	Bookstores and leisure	Retail sale of books, newspapers, magazines and stationery, audiovisual products in specialized stores (CDs, DVDs,...), bicycles, musical instruments recreational boats and yachts, sporting goods, games and toys, stamps, coins and medals, weapons, and exhibition halls and art galleries
51	Retail clothing stores	Retail sale of clothing, footwear and leather goods
52	Pharmacies	Pharmacy, parapharmaceutical products, orthopedic articles, medical and orthopedic instruments and herbal products
53	Perfumeries and drugstores	Retail trade of perfumery, cosmetic goods and drugstore products
54	Garden and pets	Retail trade of seeds, fertilizers, plants and cut flowers, pet supplies, pet food, pet animals
55	Jewelry stores	Retail trade of jewelry, watches and costume jewellery
56	Sex shops and others	Sex-Shop and second-hand goods
57	Street markets	Retail trade in stalls and at flea markets, by correspondence, internet, home delivery, vending machines, street food sales
58	Restaurants	Restaurant
59	Fast food and self-service	Fast food and self-service restaurants
60	Bars	Bar restaurant, cafeteria, chocolate shop/tea room, ice cream parlor, wine cellar with consumption, tavern, bar without kitchen and cyber-café
61	Banquet and colective caterings	Banqueting halls and catering for events and collective servicies
62	Bars with shows	Special bar and show cafes

Source: Own elaboration based on data from the Census of Premises and Activities of the Madrid City Council.

Table A.1. Sectoral agglomeration curves

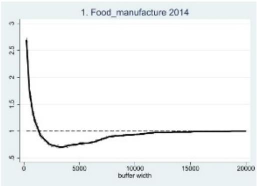
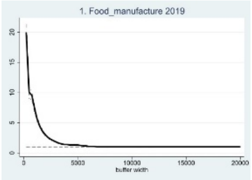
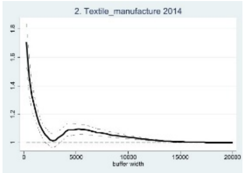
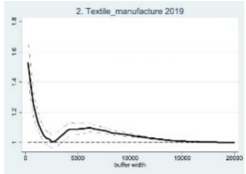
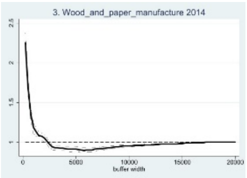
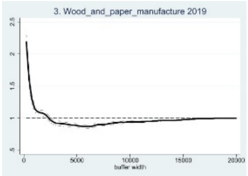
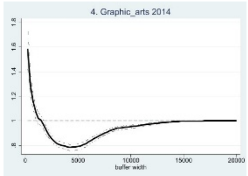
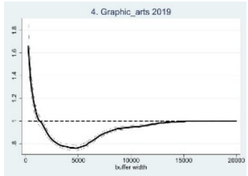
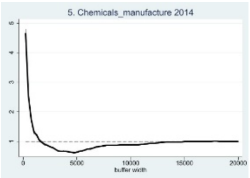
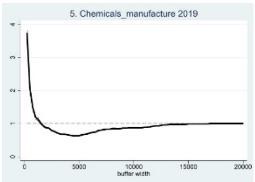
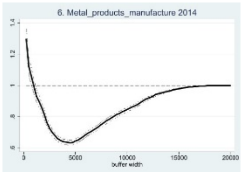
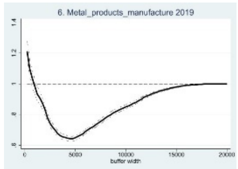
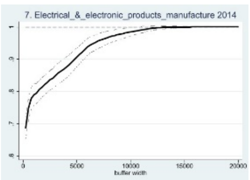
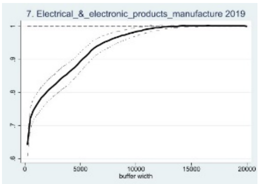
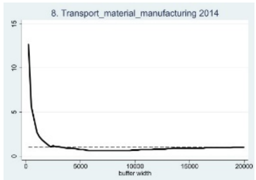
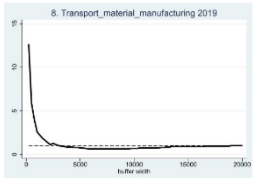
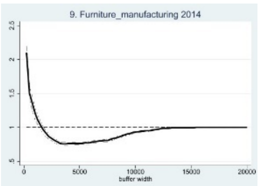
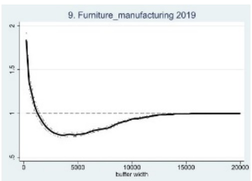
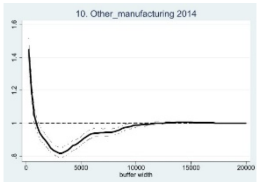
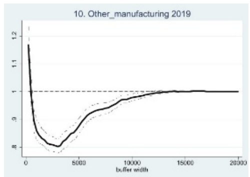
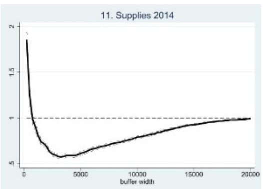
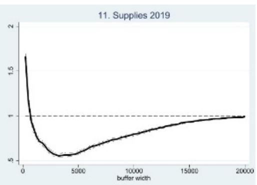
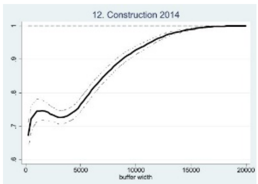
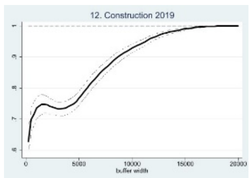
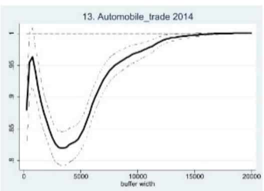
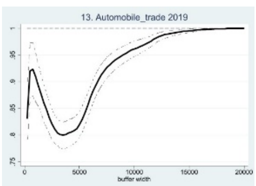
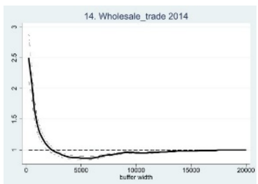
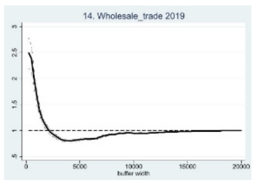
2014	2019		2014	2019
 Type 6	 Type 1		 Type 1	 Type 1
 Type 6	 Type 6		 Type 6	 Type 6
 Type 6	 Type 6		 Type 7	 Type 7
 Type 5	 Type 5		 Type 1	 Type 1
 Type 6	 Type 6		 Type 7	 Type 7
 Type 7	 Type 7		 Type 5	 Type 5
 Type 4	 Type 4		 Type 6	 Type 6

Table A.1. Sectoral agglomeration curves (continued)

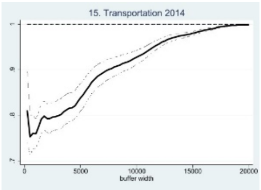
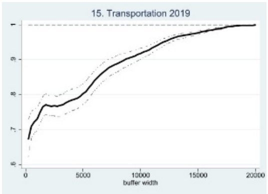
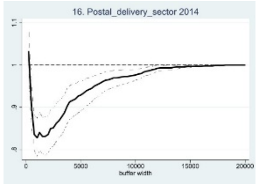
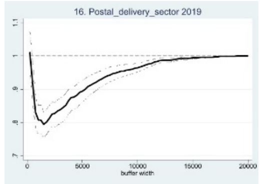
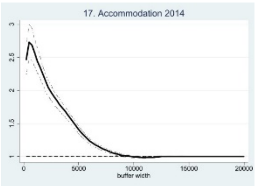
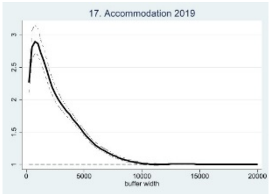
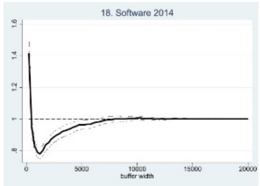
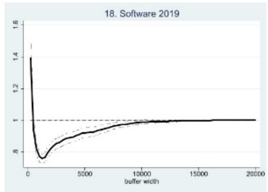
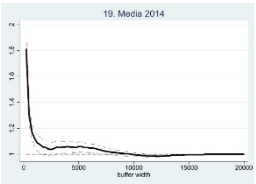
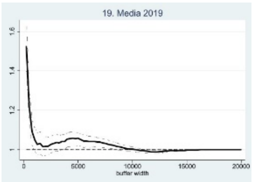
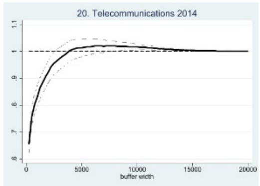
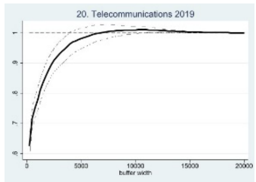
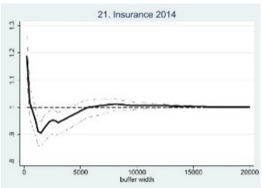
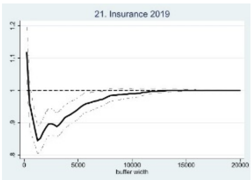
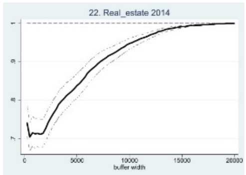
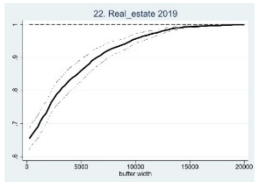
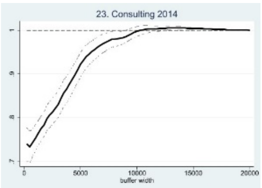
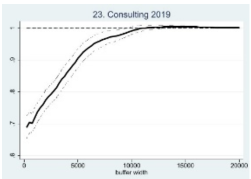
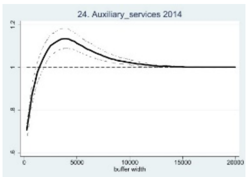
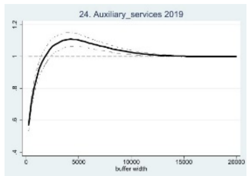
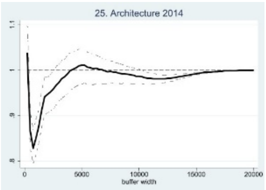
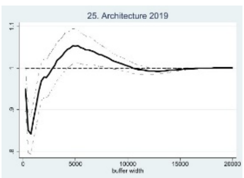
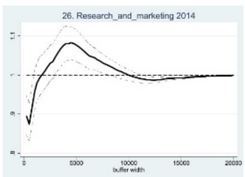
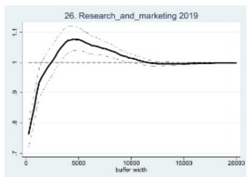
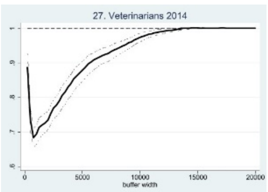
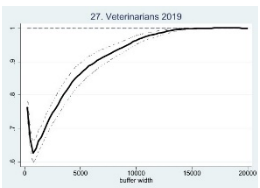
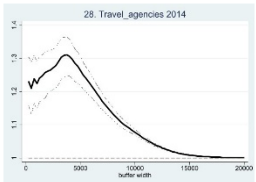
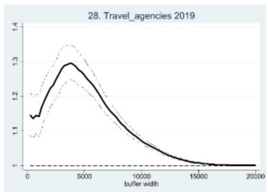
2014	2019		2014	2019
 Type 5	 Type 5		 Type 7	 Type 2
 Type 1	 Type 1		 Type 7	 Type 7
 Type 1	 Type 1		 Type 5	 Type 5
 Type 7	 Type 7		 Type 5	 Type 5
 Type 5	 Type 5		 Type 8	 Type 8
 Type 7	 Type 9		 Type 8	 Type 8
 Type 2	 Type 2		 Type 3	 Type 3

Table A.1. Sectoral agglomeration curves (continued)

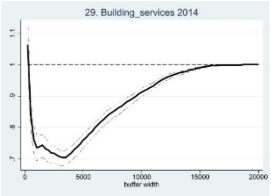
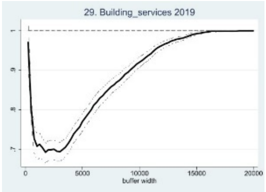
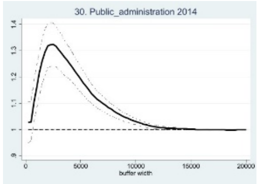
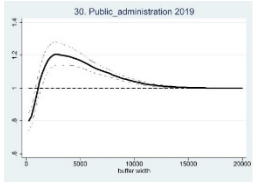
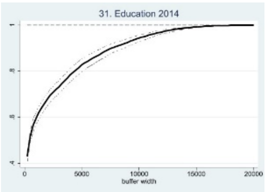
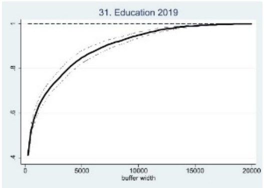
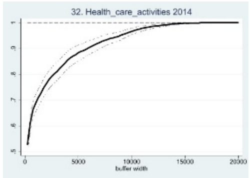
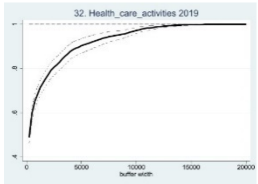
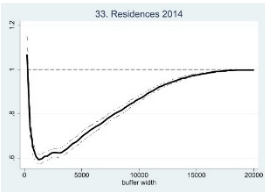
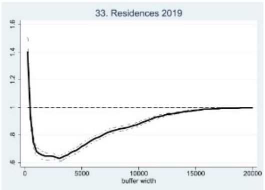
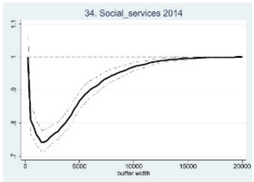
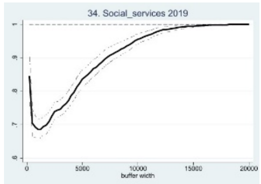
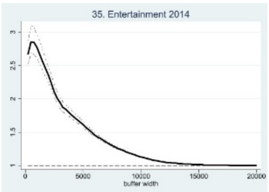
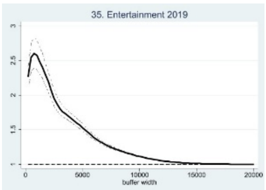
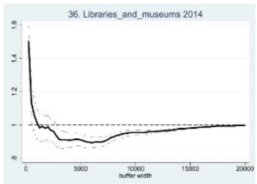
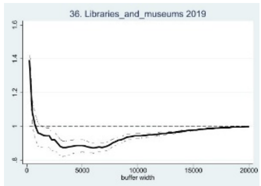
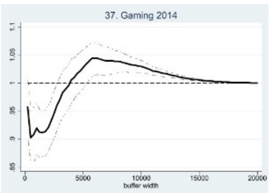
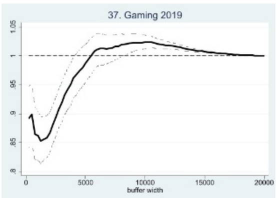
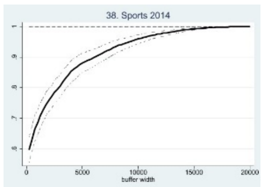
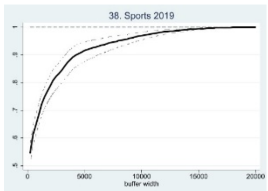
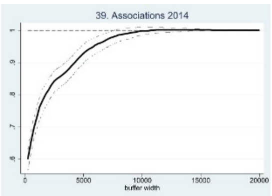
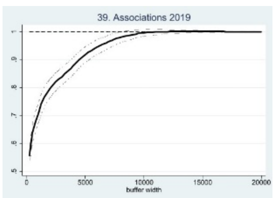
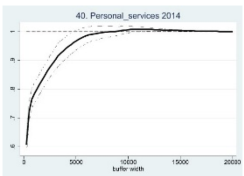
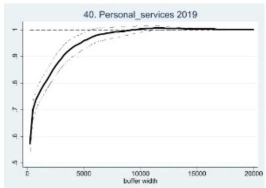
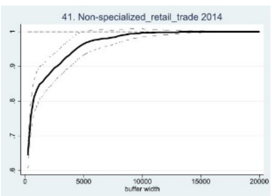
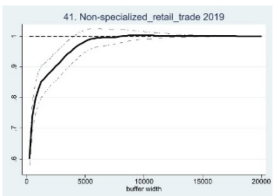
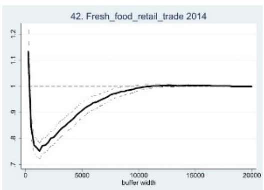
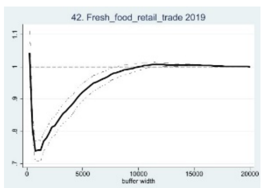
2014	2019		2014	2019
 Type 7	 Type 2		 Type 3	 Type 8
 Type 5	 Type 5		 Type 5	 Type 5
 Type 7	 Type 7		 Type 2	 Type 2
 Type 1	 Type 1		 Type 6	 Type 6
 Type 9	 Type 9		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 7	 Type 7

Table A.1. Sectoral agglomeration curves (continued)

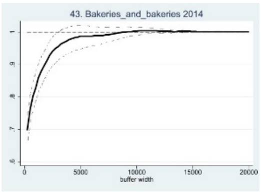
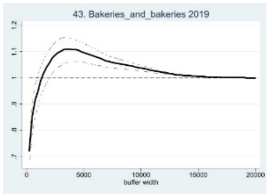
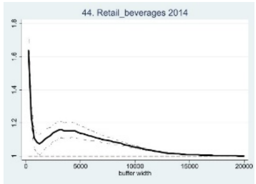
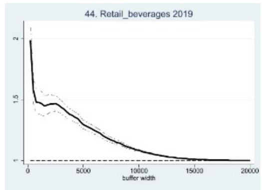
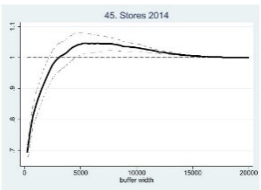
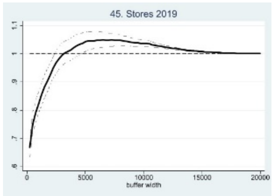
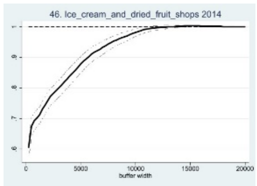
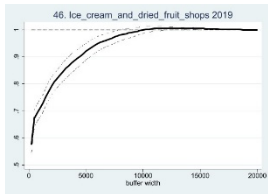
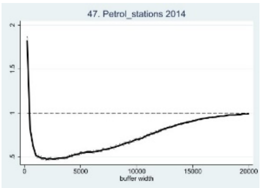
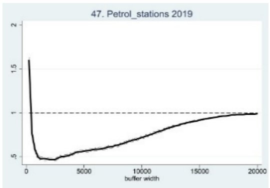
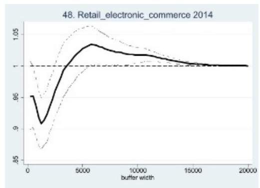
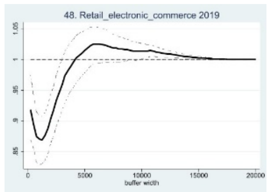
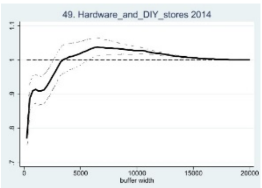
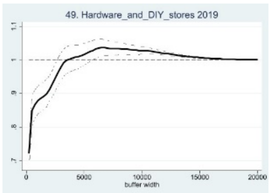
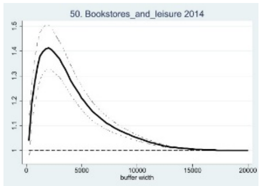
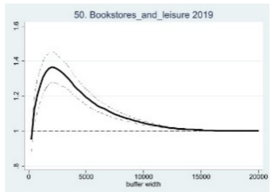
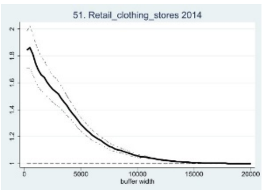
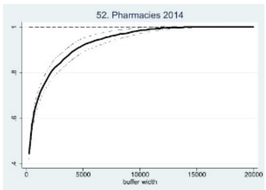
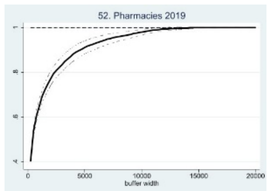
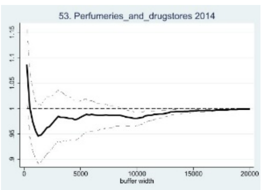
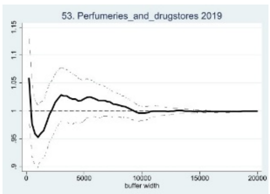
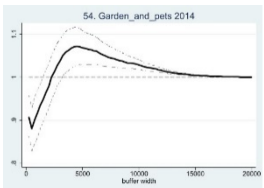
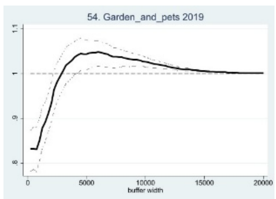
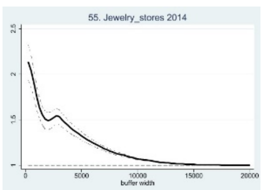
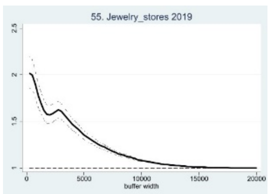
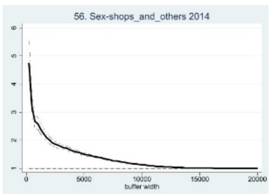
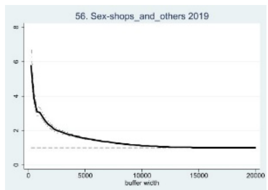
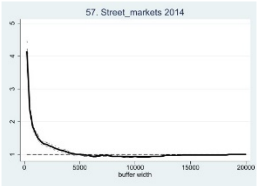
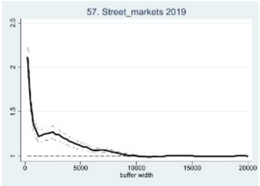
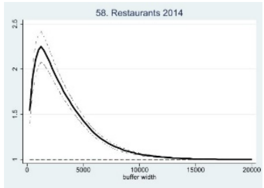
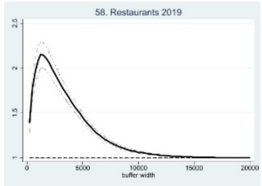
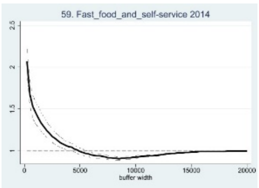
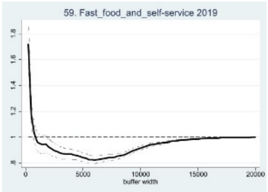
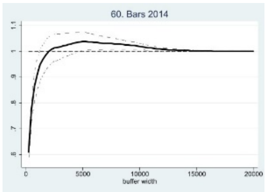
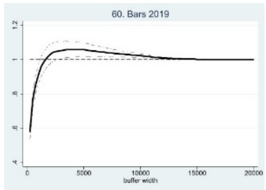
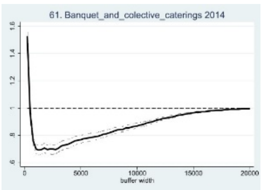
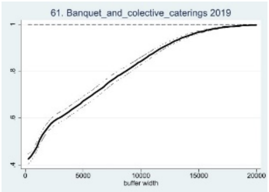
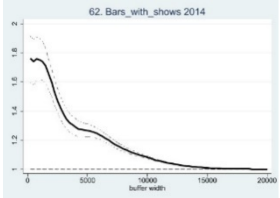
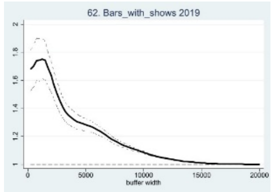
2014	2019		2014	2019
 Type 5	 Type 10		 Type 1	 Type 1
 Type 10	 Type 10		 Type 5	 Type 5
 Type 7	 Type 7		 Type 9	 Type 9
 Type 10	 Type 10		 Type 3	 Type 8
 Type 1	 Type 1		 Type 5	 Type 5
 Type 5	 Type 6		 Type 8	 Type 8
 Type 1	 Type 1		 Type 1	 Type 1

Table A.1. Sectoral agglomeration curves (continued)

2014	2019		2014	2019
				
Type 1	Type 1		Type 3	Type 3
				
Type 5	Type 5		Type 10	Type 10
				
Type 7	Type 5		Type 1	Type 1

Note: X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Table A.2. Sectoral agglomeration measures

	Synthetic sectoral spatial concentration indicators $SSSCI$			Statistical synthetic sectoral spatial concentration indicators $SSSSCI$	
Name	**2014**	**2019**		**2014**	**2019**
1. Food (M)	1.38***	7.18***		1.45***	6.80***
2. Textile (M)	1.26***	1.20***		1.21***	1.16***
3. Wood and paper (M)	1.34***	1.31***		1.37***	1.33***
4. Graphic arts	1.11***	1.12***		1.16***	1.10***
5. Chemicals (M)	1.88***	1.62***		2.05***	1.71***
6. Metal Products (M)	0.96***	0.94***		1.16***	1.11***
7. Electrical and electronic (M)	0.83***	0.81***		0.88***	0.86***
8. Transport Material (M)	4.16***	4.16***		4.31***	4.31***
9. Furniture (M)	1.25***	1.15***		1.38***	1.24***
10. Other Manufacturing	1.07***	0.96***		1.10***	0.99
11. Supplies	1.03**	0.98**		1.28***	1.15***
12. Construction	0.77***	0.76***		0.91***	0.88***
13. Automobile (Trade)	0.91***	0.88***		1.02**	0.99
14. Wholesale Trade	1.47***	1.53***		1.37***	1.33***
15. Transportation	0.83***	0.78***		0.90***	0.85***
16. Postal and delivery sector	0.93***	0.91***		1.01	0.99
17. Accommodation	1.98***	2.04***		1.77***	1.79***
18. Software	1.02	0.99		1.10***	1.07***
19. Media	1.23***	1.15***		1.23***	1.18***
20. Telecommunications	0.87***	0.84***		0.88***	0.87***
21. Insurance	1.02	0.97		1.05***	1.00
22. Real estate	0.79***	0.78***		0.89***	0.85***
23. Consulting	0.83***	0.81***		0.92***	0.91***
24. Auxiliary services	0.94***	0.89***		0.95***	0.89***
25. Architecture	0.96***	0.96***		1.01	1.00
26. Research and marketing	0.97	0.93		0.98	0.93***
27. Veterinarians	0.84***	0.79***		0.95***	0.90***
28. Travel Agencies	1.20***	1.16***		1.14***	1.10***
29. Building Services	0.86***	0.83***		1.03**	0.99
30. Public Administration	1.12***	1.00***		1.08***	0.97***
31. Education	0.68***	0.69***		0.81***	0.80***
32. Health care activities	0.75***	0.75***		0.86***	0.85***
33. Residences	0.79***	0.91***		1.01	1.14***
34. Social Services	0.88***	0.81***		0.95***	0.93***
35. Entertainment	2.14***	1.95***		1.81***	1.69***
36. Libraries and museums	1.09***	1.04***		1.10***	1.07***
37. Gaming	0.96**	0.93**		0.94***	0.95***

Table A.2. Sectoral agglomeration measures(continued)

38. Sports	0.76***	0.76***		0.84***	0.83***
39. Associations	0.80***	0.78***		0.84***	0.82***
40. Personal Services	0.83***	0.81***		0.87***	0.86***
41. Non-special retail trade	0.85***	0.84***		0.85***	0.84***
42. Fresh food retail trade	0.92***	0.90***		0.98*	0.96***
43. Bakeries and bakeries	0.87***	0.95***		0.87***	0.91***
44. Retail beverages	1.22***	1.47***		1.14***	1.34***
45. Stores	0.89***	0.88***		0.90***	0.90***
46. Ice cream and dried fruit	0.78***	0.79***		0.85***	0.84***
47. Petrol stations	0.87***	0.82***		1.17***	1.07***
48. Electronics stores	0.97**	0.94**		0.96***	0.96***
49. Hardware and DIY stores	0.92***	0.90***		0.91***	0.89***
50. Bookstores and leisure	1.19***	1.14***		1.03***	1.02
51. Retail clothing stores	1.54***	1.55***		1.32***	1.36***
52. Pharmacies	0.74***	0.72***		0.81***	0.80***
53. Perfumeries and drugstores	1.00	1.01		0.97***	0.97***
54. Garden and pets	0.97*	0.93*		0.96***	0.93***
55. Jewelry stores	1.61***	1.61***		1.38***	1.40***
56. Sex shops and others	2.50***	2.90***		2.11***	2.52***
57. Street markets	1.92***	1.40***		1.89***	1.34***
58. Restaurants	1.65***	1.58***		1.41***	1.37***
59. Fast food and self-service	1.39***	1.11***		1.33***	1.16***
60. Bars	0.89***	0.89***		0.85***	0.86***
61. Colective caterings	0.95***	0.60***		1.19***	0.79***
62. Bars with shows	1.50***	1.48***		1.34***	1.33***

*Note: *Indicates that the estimation is statistically significant at the 90% level, ** at the 95% level, and *** at the 99% level.*

Table A.3. Agglomeration curves for each neighbourhood

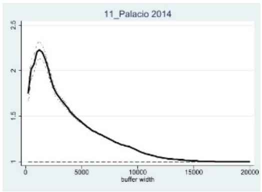
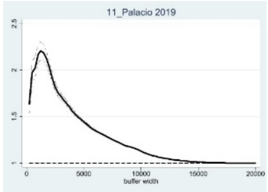
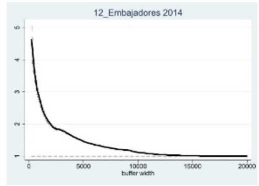
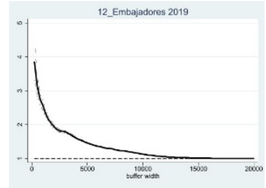
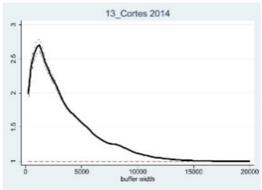
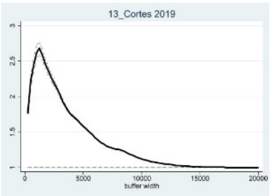
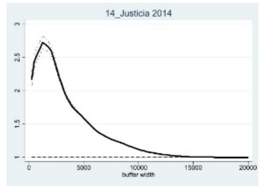
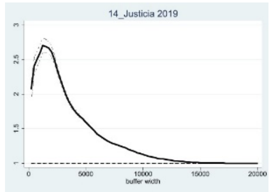
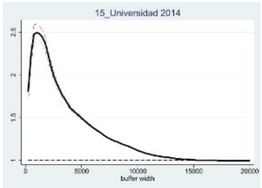
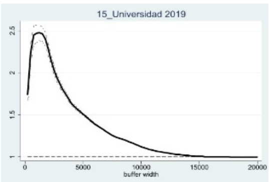
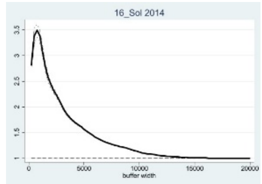
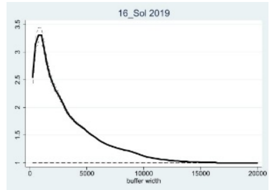
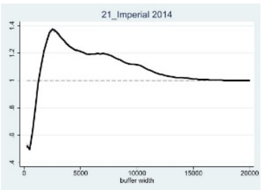
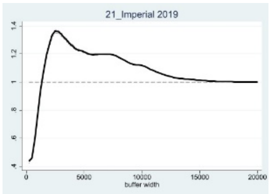
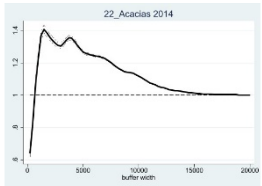
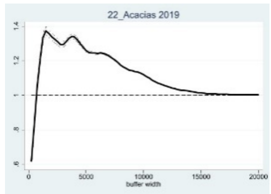
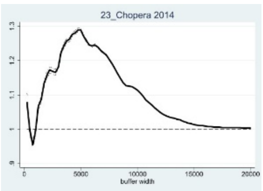
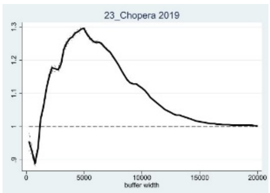
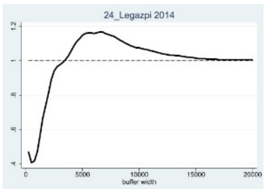
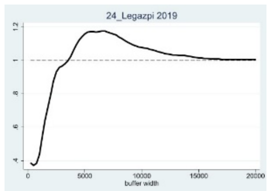
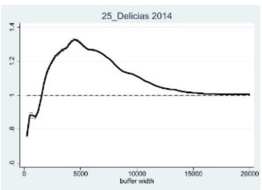
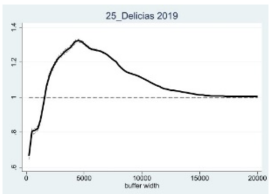
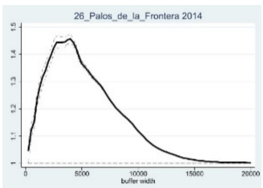
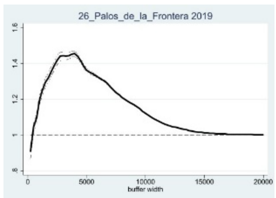
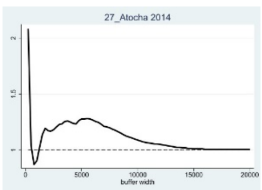
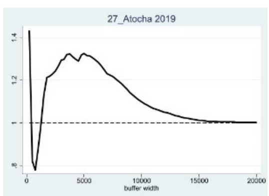
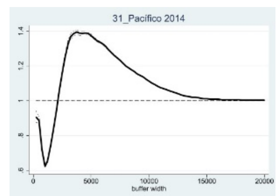
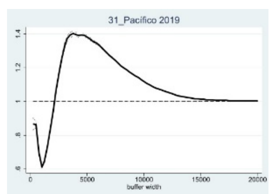
2014	2019		2014	2019
 Type 3	 Type 3		 Type 1	 Type 1
 Type 3	 Type 3		 Type 3	 Type 3
 Type 3	 Type 3		 Type 1	 Type 1
 Type 8	 Type 8		 Type 8	 Type 8
 Type 8	 Type 8		 Type 8	 Type 8
 Type 8	 Type 8		 Type 3	 Type 8
 Type 9	 Type 9		 Type 9	 Type 9

Table A.3. Agglomeration curves for each neighbourhood (continued)

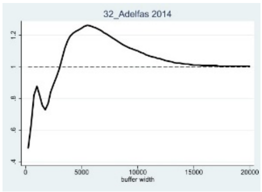
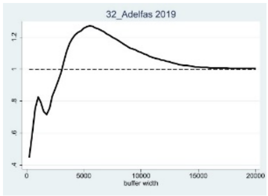
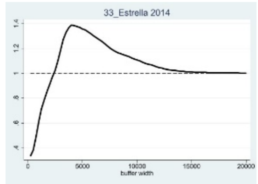
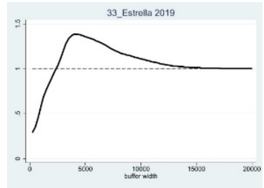
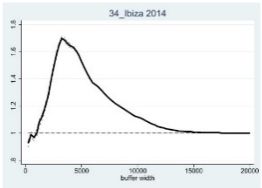
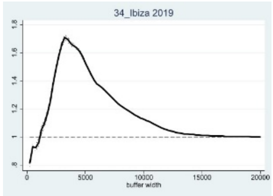
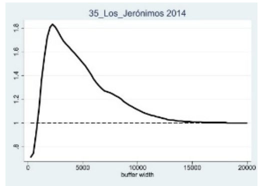
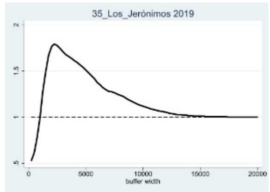
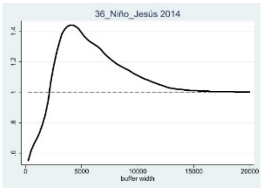
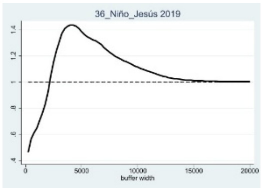
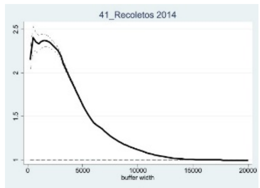
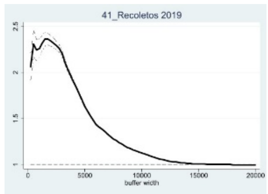
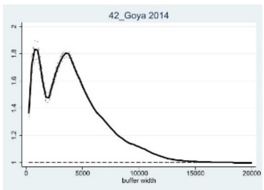
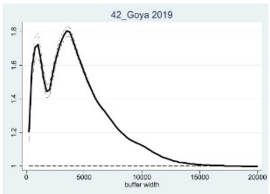
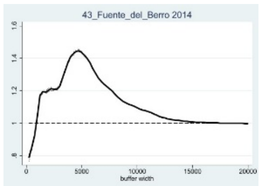
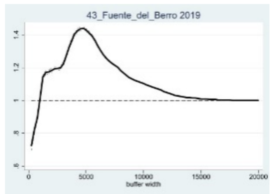
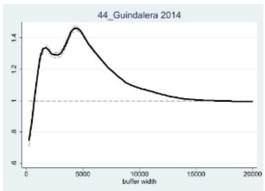
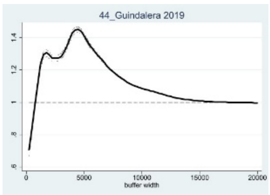
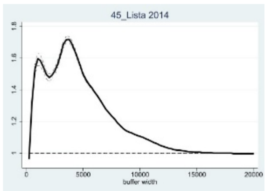
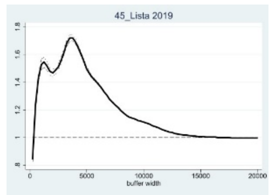
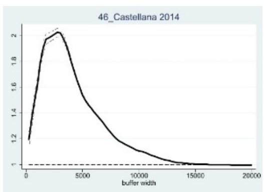
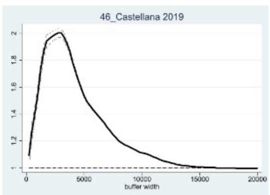
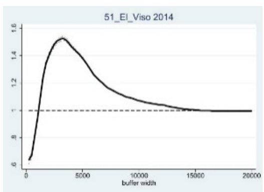
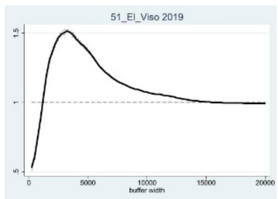
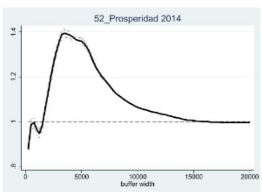
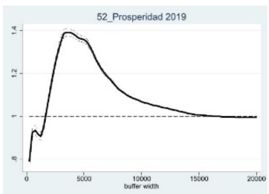
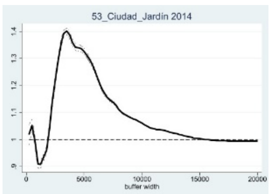
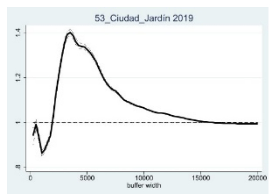
2014	2019		2014	2019
 Type 8	 Type 8		 Type 8	 Type 8
 Type 8	 Type 8		 Type 8	 Type 8
 Type 8	 Type 8		 Type 3	 Type 3
 Type 3	 Type 3		 Type 8	 Type 8
 Type 8	 Type 8		 Type 8	 Type 8
 Type 3	 Type 3		 Type 8	 Type 8
 Type 8	 Type 8		 Type 9	 Type 9

Table A.3. Agglomeration curves for each neighbourhood (continued)

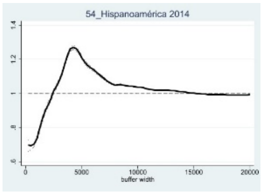
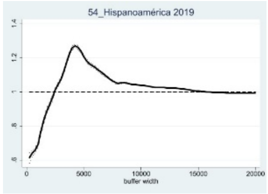
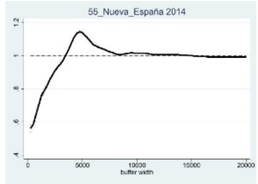
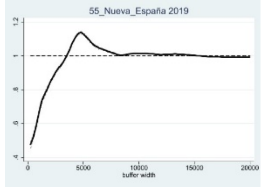
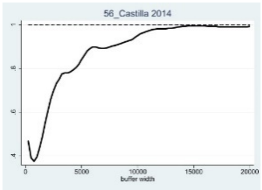
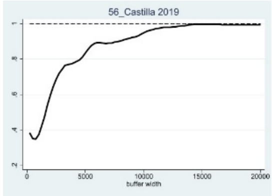
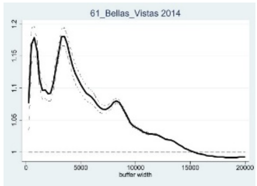
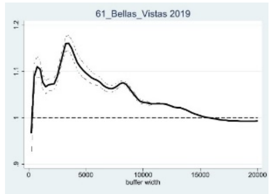
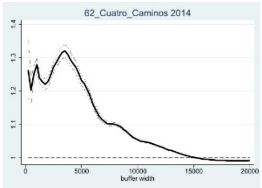
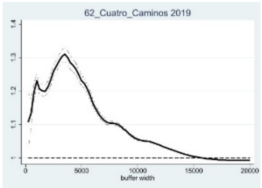
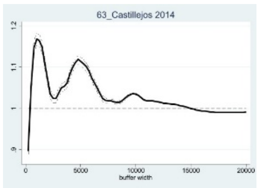
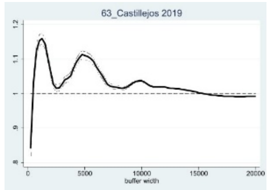
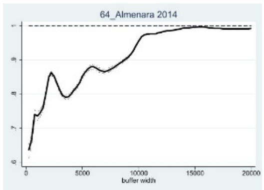
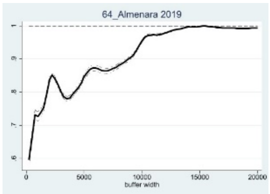
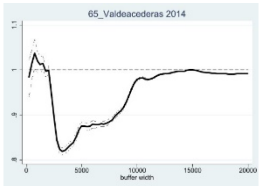
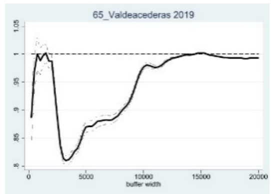
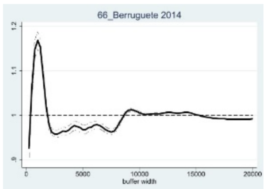
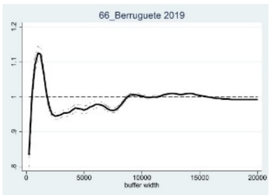
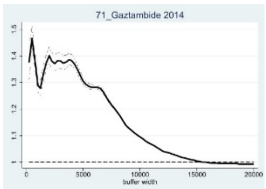
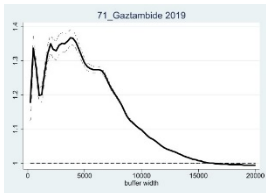
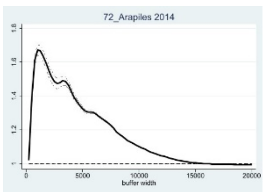
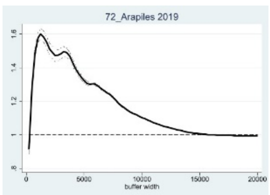
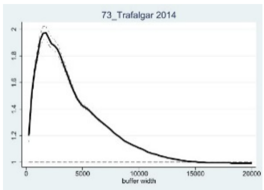
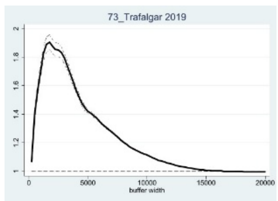
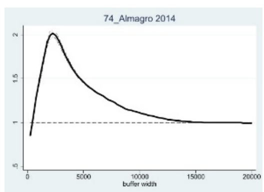
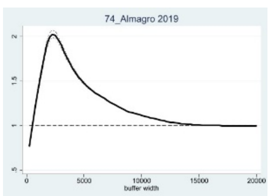
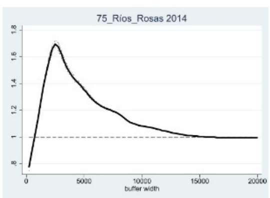
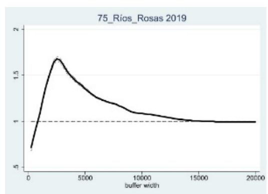
<table>
<tr><th>2014</th><th>2019</th><th></th><th>2014</th><th>2019</th></tr>
<tr>
<td>Type 8</td>
<td>Type 8</td>
<td></td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
<tr>
<td>Type 5</td>
<td>Type 5</td>
<td></td>
<td>Type 3</td>
<td>Type 8</td>
</tr>
<tr>
<td>Type 3</td>
<td>Type 3</td>
<td></td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
<tr>
<td>Type 5</td>
<td>Type 5</td>
<td></td>
<td>Type 7</td>
<td>Type 9</td>
</tr>
<tr>
<td>Type 9</td>
<td>Type 9</td>
<td></td>
<td>Type 3</td>
<td>Type 3</td>
</tr>
<tr>
<td>Type 3</td>
<td>Type 8</td>
<td></td>
<td>Type 3</td>
<td>Type 3</td>
</tr>
<tr>
<td>Type 8</td>
<td>Type 8</td>
<td></td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
</table>

Table A.3. Agglomeration curves for each neighbourhood (continued)

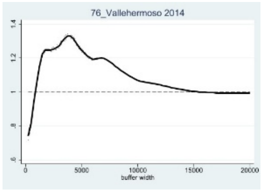
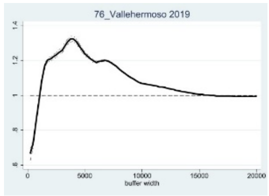
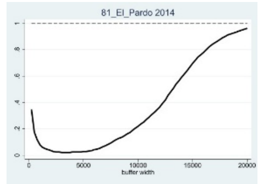
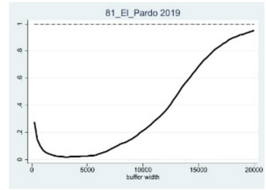
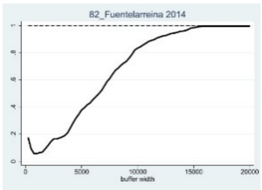
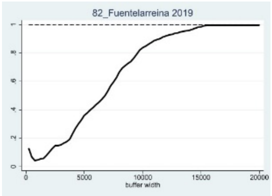
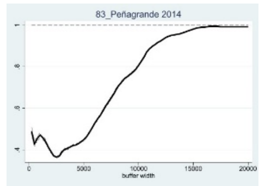
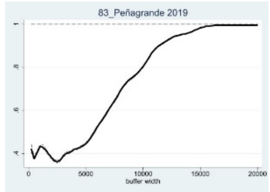
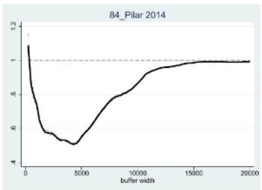
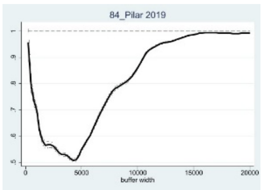
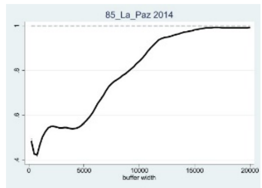
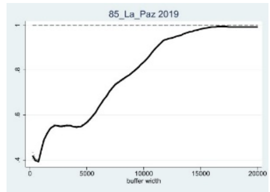
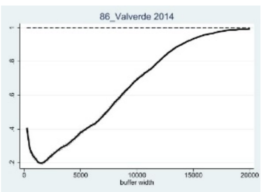
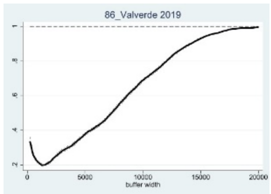
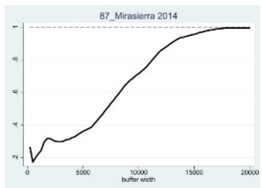
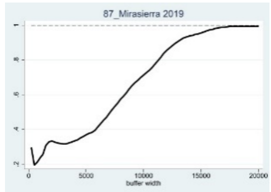
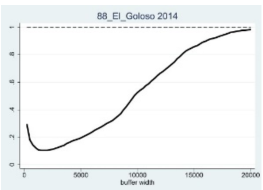
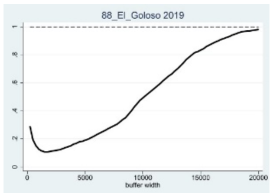
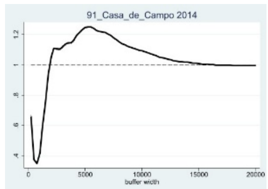
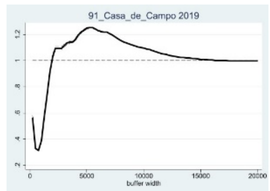
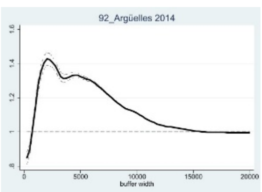
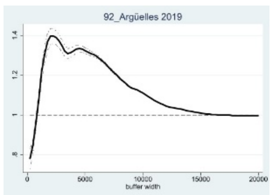
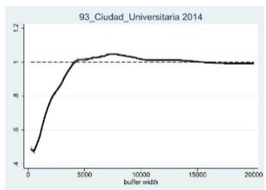
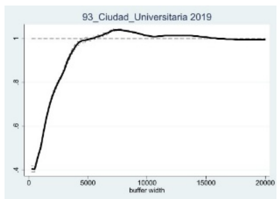
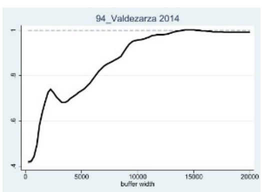
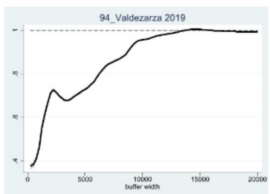
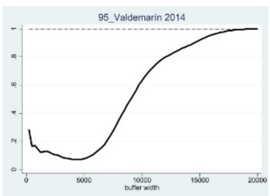
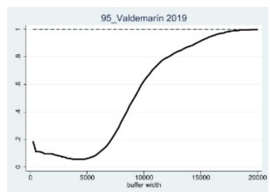
2014	2019		2014	2019
 Type 8	 Type 8		 Type 2	 Type 2
 Type 5	 Type 5		 Type 5	 Type 5
 Type 7	 Type 2		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 9	 Type 9
 Type 8	 Type 8		 Type 10	 Type 10
 Type 5	 Type 5		 Type 5	 Type 5

Table A.3. Agglomeration curves for each neighbourhood (continued)

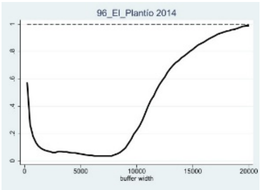
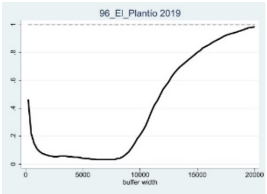
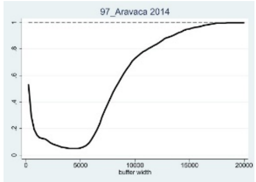
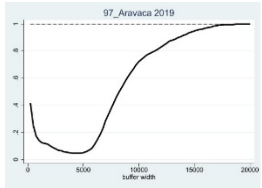
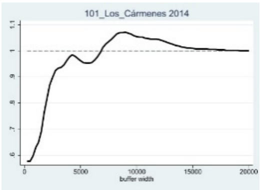
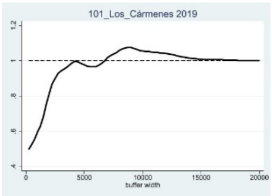
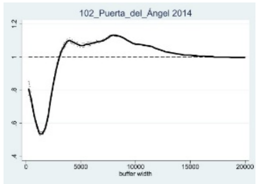
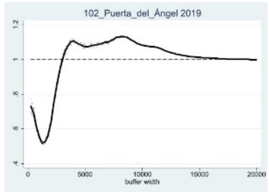
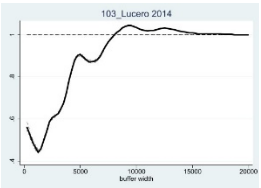
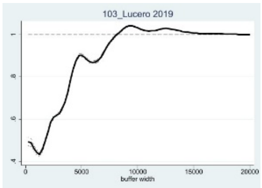
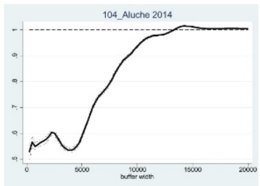
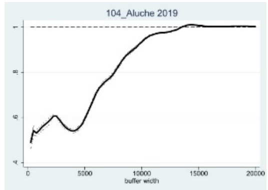
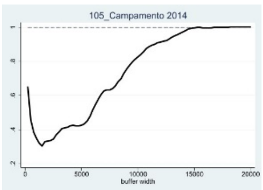
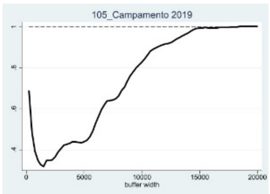
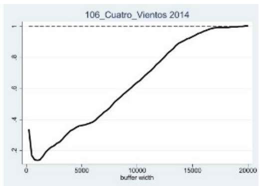
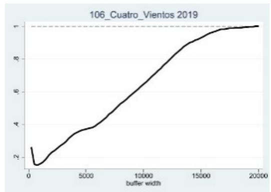
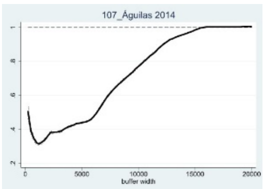
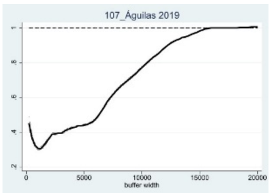
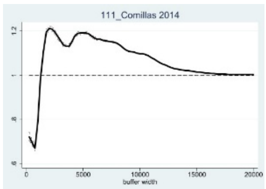
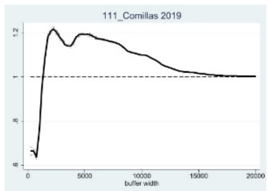
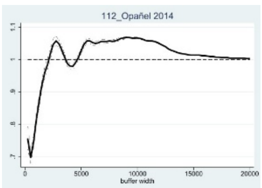
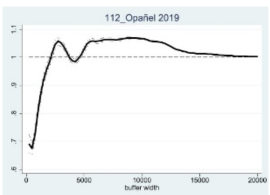
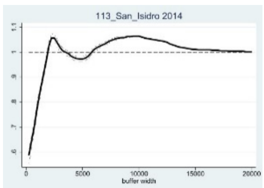
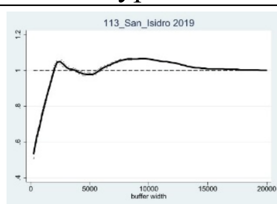
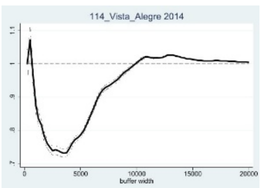
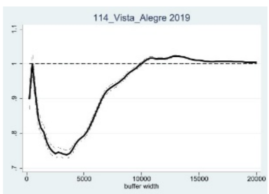
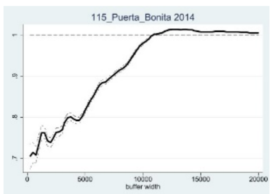
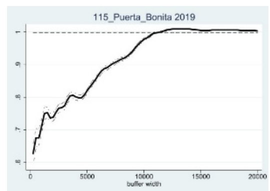
2014	2019		2014	2019
 Type 2	 Type 2		 Type 2	 Type 2
 Type 10	 Type 10		 Type 9	, Type 9
 Type 10	 Type 10		 Type 5	 Type 5
 Type 2	 Type 2		 Type 5	 Type 5
 Type 10	 Type 10		 Type 8	 Type 8
 Type 8	 Type 8		 Type 8	 Type 8
 Type 7	 Type 7		 Type 10	 Type 10

Table A.3. Agglomeration curves for each neighbourhood (continued)

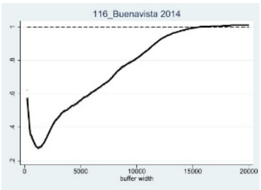
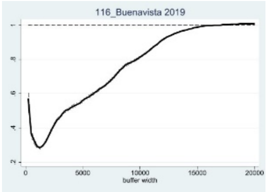
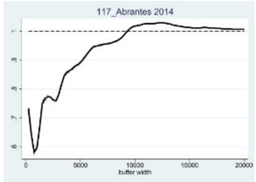
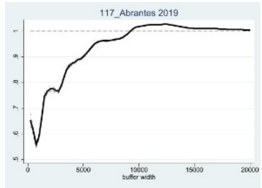
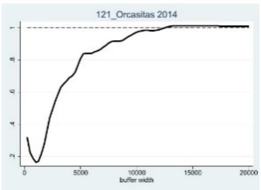
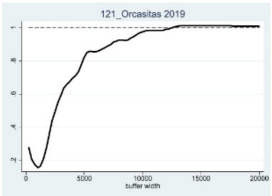
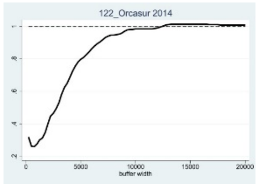
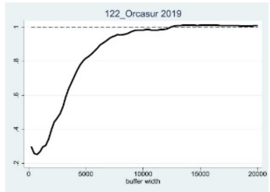
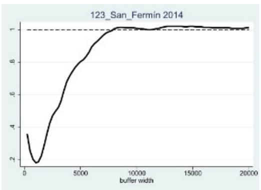
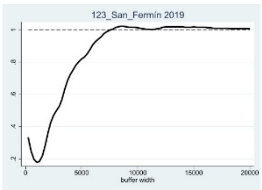
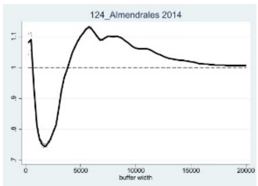
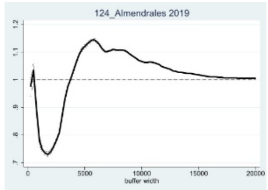
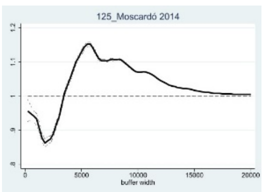
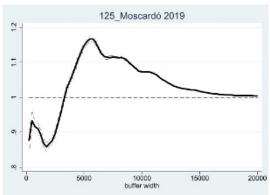
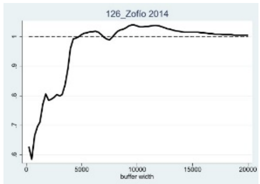
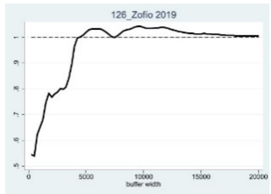
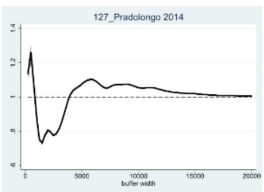
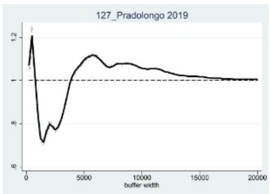
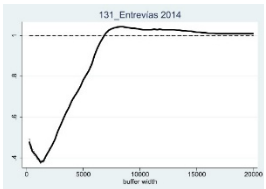
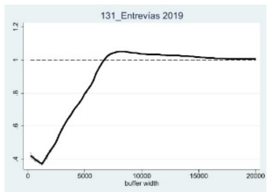
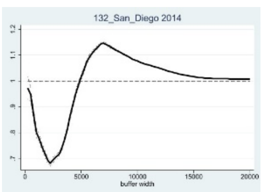
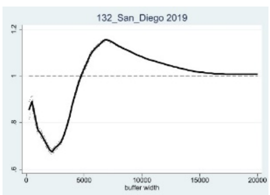
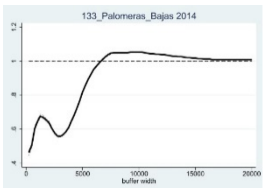
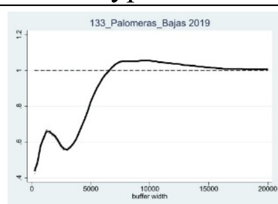
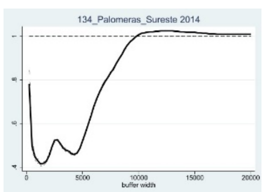
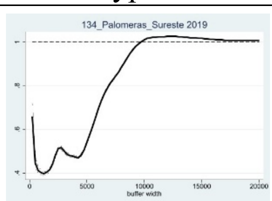
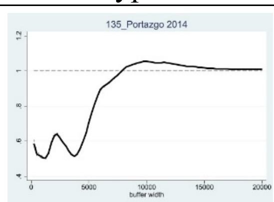
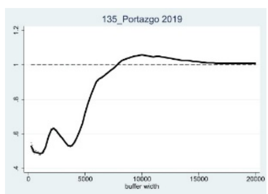
2014	2019		2014	2019
 Type 2	 Type 2		 Type 10	 Type 10
 Type 10	 Type 10		 Type 10	 Type 10
 Type 2	 Type 2		 Type 9	 Type 9
 Type 9	 Type 9		 Type 10	 Type 10
 Type 9	 Type 9		 Type 10	 Type 10
 Type 9	 Type 9		 Type 9	 Type 9
 Type 7	 Type 7		 Type 7	 Type 7

Table A.3. Agglomeration curves for each neighbourhood (continued)

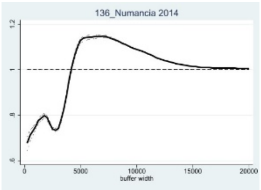
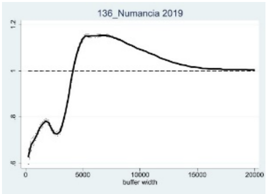
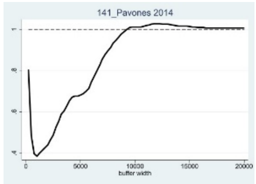
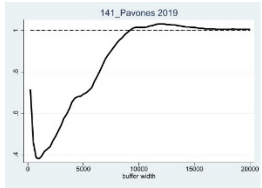
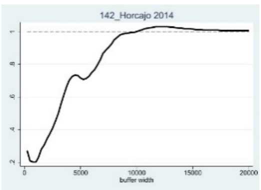
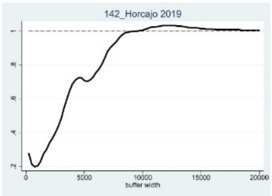
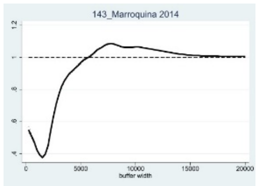
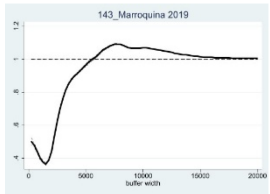
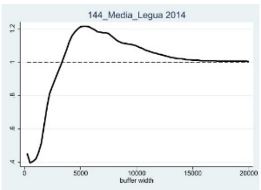
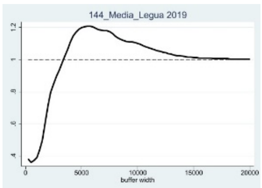
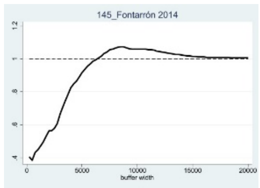
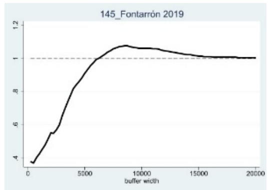
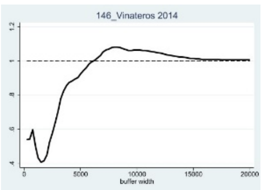
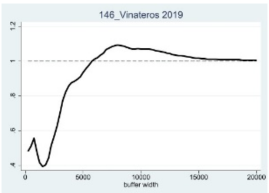
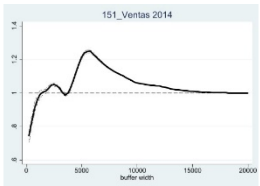
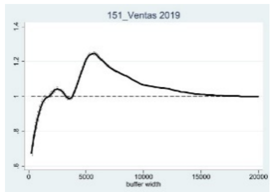
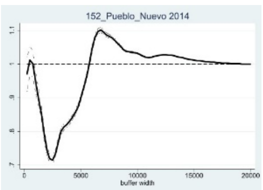
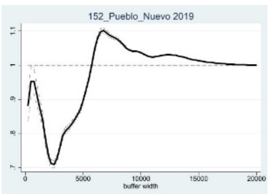
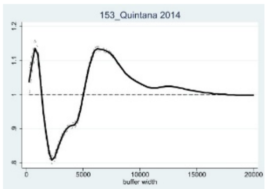
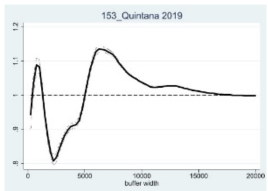
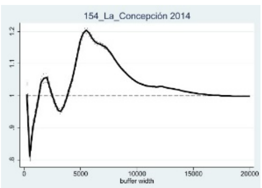
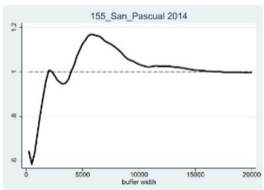
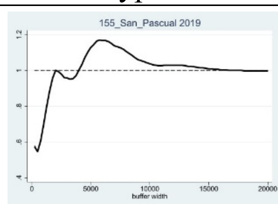
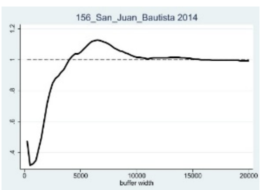
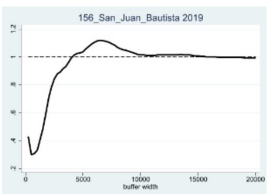
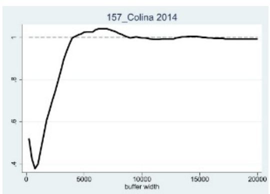
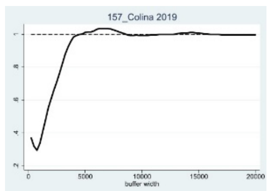
2014	2019		2014	2019
 Type 8	 Type 8		 Type 2	 Type 2
 Type 5	 Type 5		 Type 10	 Type 10
 Type 8	 Type 8		 Type 5	 Type 5
 Type 10	 Type 10		 Type 8	 Type 8
 Type 9	 Type 9		 Type 9	 Type 9
 Type 9	 Type 9		 Type 8	 Type 8
 Type 10	 Type 10		 Type 10	 Type 10

Table A.3. Agglomeration curves for each neighbourhood (continued)

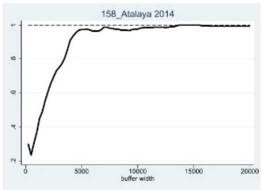
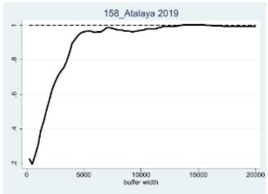
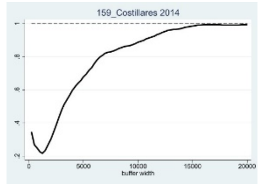
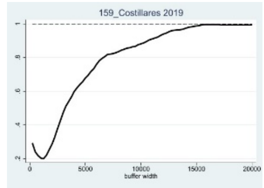
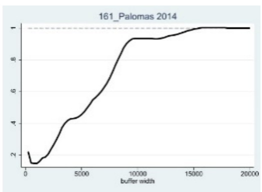
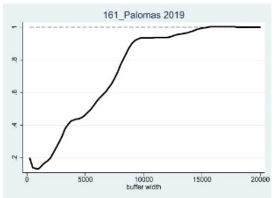
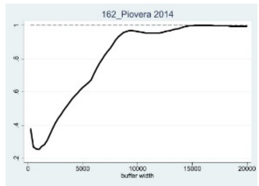
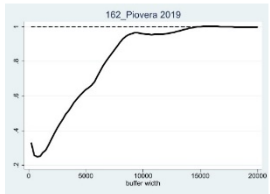
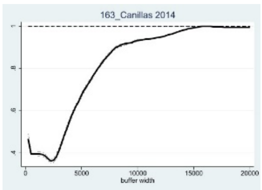
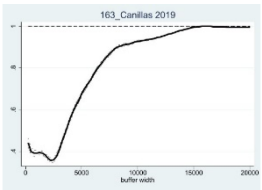
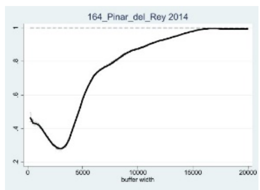
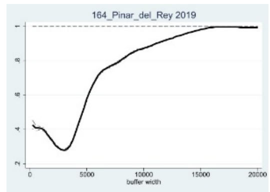
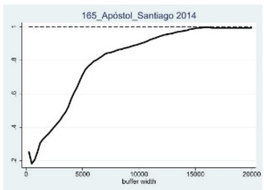
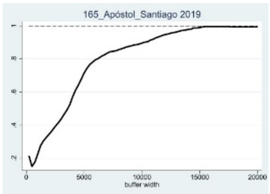
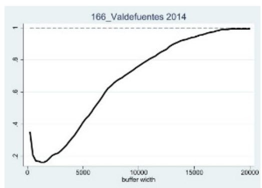
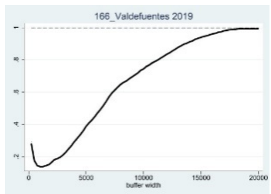
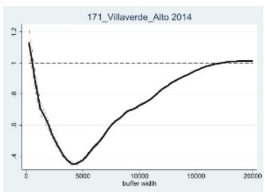
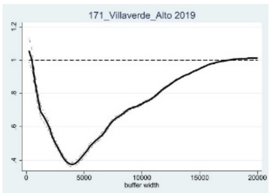
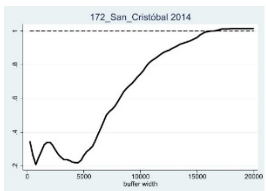
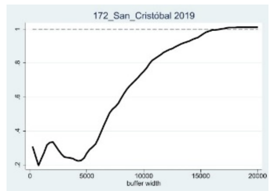
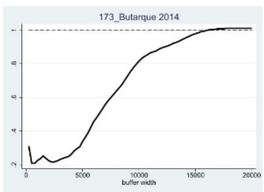
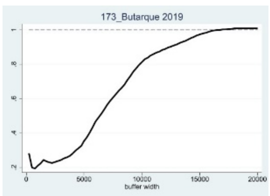
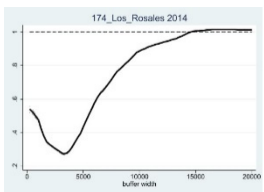
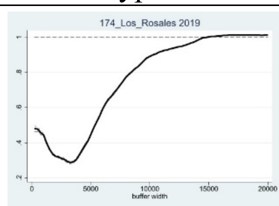
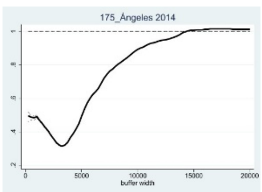
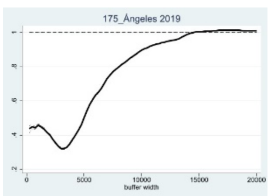
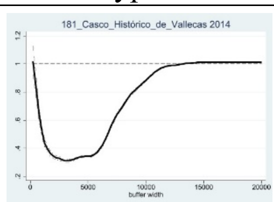
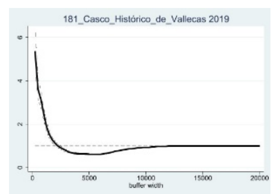
2014	2019		2014	2019
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 2	 Type 2
 Type 5	 Type 5		 Type 2	 Type 2
 Type 7	 Type 7		 Type 5	 Type 5
 Type 5	 Type 5		 Type 2	 Type 2
 Type 2	 Type 2		 Type 2	 Type 6

Table A.3. Agglomeration curves for each neighbourhood (continued)

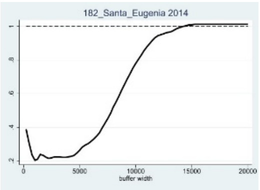
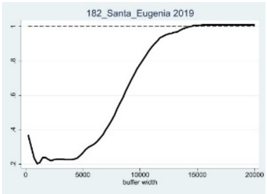
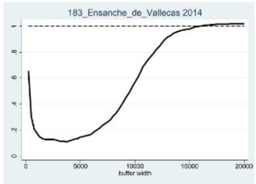
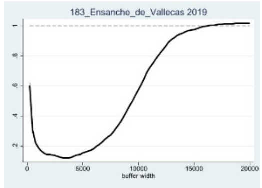
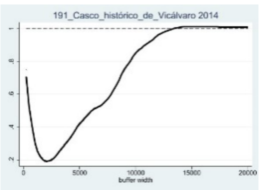
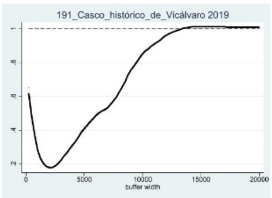
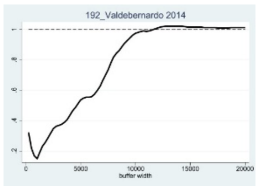
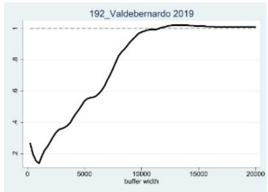
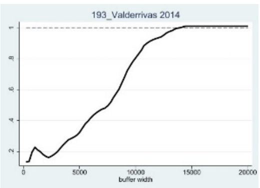
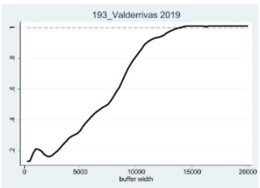
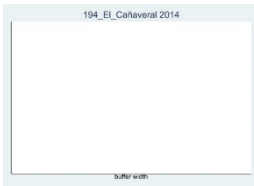
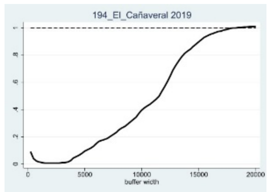
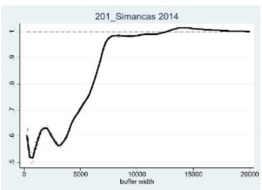
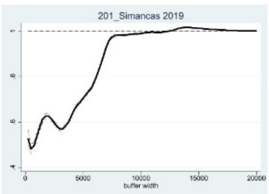
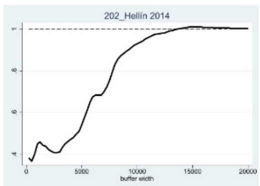
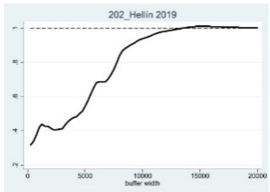
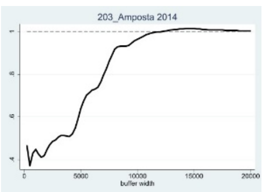
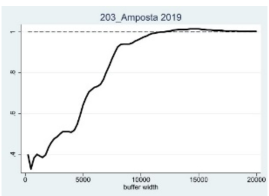
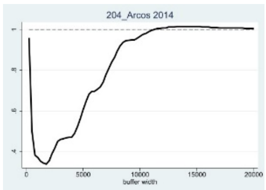
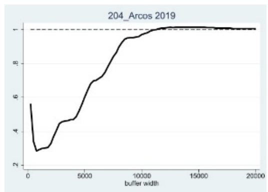
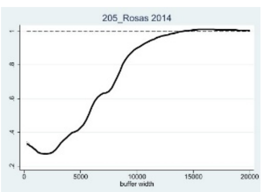
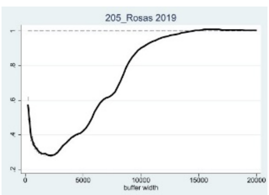
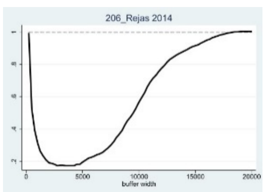
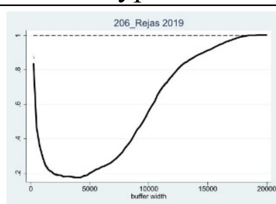
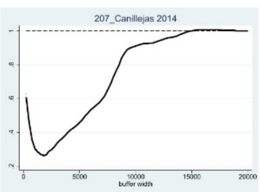
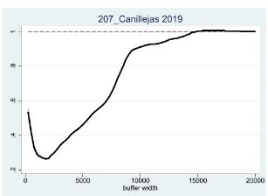
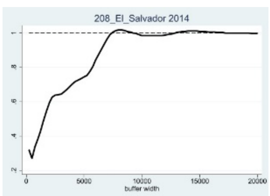
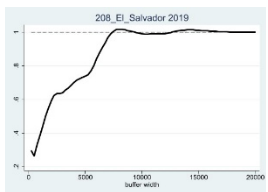
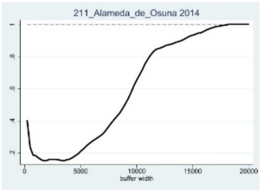
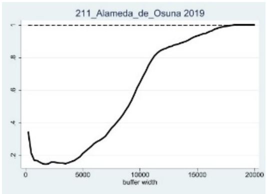
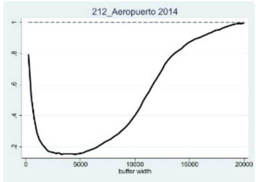
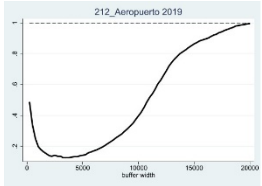
2014	2019		2014	2019
 Type 2	 Type 2		 Type 2	 Type 2
 Type 2	 Type 2		 Type 5	 Type 5
 Type 5	 Type 5		 -	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 2	 Type 2
 Type 5	 Type 2		 Type 2	 Type 2
 Type 2	 Type 2		 Type 5	 Type 5

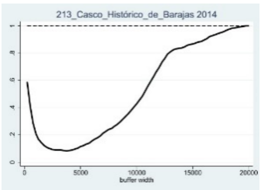
Table A.3. Agglomeration curves for each neighbourhood (continued)

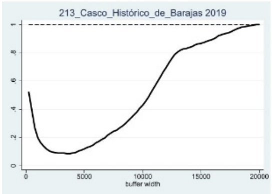
<table>
<tr><th>2014</th><th>2019</th><th></th><th>2014</th><th>2019</th></tr>
<tr>
<td>

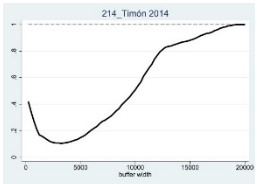
Type 2</td>
<td>

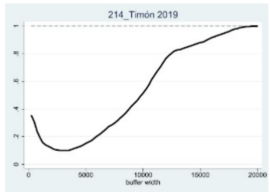
Type 2</td>
<td></td>
<td>

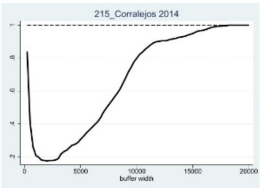
Type 2</td>
<td>

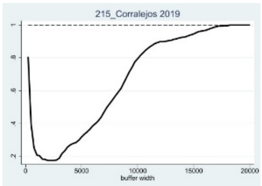
Type 2</td>
</tr>
<tr>
<td>

Type 2</td>
<td>

Type 2</td>
<td></td>
<td>

Type 2</td>
<td>

Type 2</td>
</tr>
<tr>
<td>

Type 2</td>
<td>

Type 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</table>

Note: X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Table A.4. Neighbourhood agglomeration measures

	Synthetic geographical spatial concentration indicators $SSSCI$			Statistical synthetic geographical spatial concentration indicators $SSGSCI$	
	2014	2019		2014	2019
11. Palacio	1.69***	1.66***		1.63***	1.62***
12. Embajadores	2.55***	2.30***		2.40***	2.15***
13. Cortes	1.93***	1.86***		1.84***	1.77***
14. Justicia	2.00***	1.98***		1.90***	1.88***
15. Universidad	1.82***	1.81***		1.76***	1.75***
16. Sol	2.33***	2.24***		2.18***	2.09***
21. Imperial	0.92**	0.89***		0.95***	0.94***
22. Acacias	1.07***	1.05***		1.10***	1.12***
23. Chopera	1.10***	1.06***		1.12***	1.10***
24. Legazpi	0.75***	0.72***		0.78***	0.75***
25. Delicias	1.00	0.97		1.02***	1.01
26. Palos de la Frontera	1.20***	1.15***		1.21***	1.18***
27. Atocha	1.30***	1.15***		1.29***	1.15***
31. Pacífico	0.99	0.98		1.01	1.03***
32. Adelfas	0.85***	0.83***		0.88***	0.87***
33. Estrella	0.80***	0.78***		0.86***	0.87***
34. Ibiza	1.15***	1.11***		1.16***	1.14***
35. Los Jerónimos	1.16***	1.08***		1.13***	1.11***
36. Niño Jesús	0.90***	0.87***		0.94***	0.93***
41. Recoletos	1.94***	1.90***		1.81***	1.79***
42. Goya	1.48***	1.42***		1.43***	1.39***
43. Fuente del Berro	1.06***	1.03***		1.09***	1.08***
44. Guindalera	1.08***	1.06***		1.09***	1.08***
45. Lista	1.31***	1.26***		1.29***	1.26***
46. Castellana	1.45***	1.41***		1.39***	1.36***
51. El Viso	1.02***	0.98		1.04***	1.02***
52. Prosperidad	1.06***	1.03***		1.06***	1.04***
53. Ciudad Jardín	1.08***	1.05***		1.07***	1.07***
54. Hispanoamérica	0.89***	0.86***		0.89***	0.89***
55. Nueva España	0.81***	0.78***		0.80***	0.81***
56. Castilla	0.64***	0.61***		0.66***	0.65***
61. Bellas Vistas	1.10***	1.06***		1.12***	1.08***
62. Cuatro Caminos	1.20***	1.15***		1.20***	1.17***
63. Castillejos	1.04***	1.02***		1.05***	1.05***
64. Almenara	0.79***	0.77***		0.82***	0.80***
65. Valdeacederas	0.96***	0.93***		0.98**	0.96***
66. Berruguete	1.01	0.98		1.01	0.97***

Table A.4. Neighbourhood agglomeration measures(continued)

71. Gaztambide	1.30***	1.23		1.29***	1.24***
72. Arapiles	1.31***	1.26***		1.31***	1.27***
73. Trafalgar	1.47***	1.40***		1.44***	1.38***
74. Almagro	1.29***	1.27***		1.29***	1.28***
75. Ríos Rosas	1.13***	1.10***		1.14***	1.12***
76. Vallehermoso	1.03***	0.99		1.04***	1.01
81. El Pardo	0.21***	0.19***		0.23***	0.25***
82. Fuentelarreina	0.31***	0.29***		0.38***	0.29***
83. Peñagrande	0.54***	0.51***		0.59***	0.57***
84. Pilar	0.80***	0.76***		0.80***	0.79***
85. La Paz	0.59***	0.57***		0.61***	0.61***
86. Valverde	0.41***	0.39***		0.43***	0.43***
87. Mirasierra	0.39***	0.40***		0.43***	0.48***
88. El Goloso	0.28***	0.28***		0.29***	0.34***
91. Casa de Campo	0.83***	0.79***		0.78***	0.79***
92. Argüelles	1.11***	1.08***		1.11***	1.10***
93. Ciudad Universitaria	0.74***	0.70***		0.74***	0.72***
94. Valdezarza	0.64***	0.62***		0.70***	0.69***
95. Valdemarín	0.28***	0.24***		0.29***	0.27***
96. El Plantío	0.29***	0.26***		0.15***	0.20***
97. Aravaca	0.36***	0.33***		0.37***	0.36***
101. Los Cármenes	0.78***	0.76***		0.83***	0.78***
102. Puerta del Ángel	0.85***	0.82***		0.87***	0.86***
103. Lucero	0.68***	0.66***		0.72***	0.73***
104. Aluche	0.65***	0.64***		0.68***	0.69***
105. Campamento	0.55***	0.56***		0.59***	0.62***
106. Cuatro Vientos	0.36***	0.35***		0.36***	0.40***
107. Águilas	0.51***	0.49***		0.54***	0.54***
111. Comillas	0.95*	0.93***		0.97***	0.97***
112. Opañel	0.90***	0.88***		0.93***	0.92***
113. San Isidro	0.85***	0.83***		0.88***	0.86***
114. Vista Alegre	0.92***	0.89***		0.93***	0.91***
115. Puerta Bonita	0.80***	0.77***		0.83***	0.84***
116. Buenavista	0.54***	0.54***		0.52***	0.53***
117. Abrantes	0.79***	0.77***		0.83***	0.83***
121. Orcasitas	0.51***	0.50***		0.59***	0.59***
122. Orcasur	0.54***	0.53***		0.60***	0.60***
123. San Fermín	0.53***	0.53***		0.58***	0.59***
124. Almendrales	0.98	0.95		1.00	0.96***
125. Moscardó	0.98	0.96		1.00	0.99
126. Zofío	0.80***	0.77***		0.83***	0.83***
127. Pradolongo	1.02	1.00		1.02*	0.99

Table A.4. Neighbourhood agglomeration measures(continued)

131. Entrevías	0.62***	0.61***		0.68***	0.67***
132. San Diego	0.92***	0.89***		0.93***	0.91***
133. Palomeras Bajas	0.69***	0.68***		0.72***	0.73***
134. Palomeras Sureste	0.65***	0.62***		0.69***	0.65***
135. Portazgo	0.68***	0.66***		0.70***	0.71***
136. Numancia	0.85***	0.83***		0.87***	0.85***
141. Pavones	0.67	0.65***		0.76***	0.76***
142. Horcajo	0.50***	0.50***		0.57***	0.57***
143. Marroquina	0.69***	0.67***		0.74***	0.73***
144. Media Legua	0.73***	0.71***		0.77***	0.78***
145. Fontarrón	0.65***	0.63***		0.70***	0.71***
146. Vinateros	0.70***	0.68***		0.76***	0.75***
151. Ventas	0.96*	0.93***		0.98**	0.98***
152. Pueblo Nuevo	0.94**	0.91***		0.94***	0.92***
153. Quintana	1.02***	0.99		1.01	1.01
154. La Concepción	1.00	0.97		1.00	1.00
155. San Pascual	0.85***	0.83***		0.90***	0.90***
156. San Juan Bautista	0.69***	0.67***		0.66***	0.67***
157. Colina	0.70***	0.64***		0.70***	0.68***
158. Atalaya	0.60***	0.57***		0.66***	0.63***
159. Costillares	0.49***	0.47***		0.54***	0.56***
161. Palomas	0.39***	0.39***		0.42***	0.43***
162. Piovera	0.51***	0.50***		0.51***	0.55***
163. Canillas	0.56***	0.56***		0.58***	0.59***
164. Pinar del Rey	0.54***	0.52***		0.56***	0.57***
165. Apóstol Santiago	0.48***	0.46***		0.54***	0.53***
166. Valdefuentes	0.39***	0.36***		0.42***	0.41***
171. Villaverde Alto	0.79***	0.78***		0.77***	0.75***
172. San Cristóbal	0.41***	0.40***		0.47***	0.49***
173. Butarque	0.39***	0.39***		0.48***	0.49***
174. Los Rosales	0.55***	0.53***		0.59***	0.59***
175. Ángeles	0.56***	0.54***		0.61***	0.61***
181. Casco Histórico de Vallecas	0.69***	2.35***		0.66***	1.75***
182. Santa Eugenia	0.40***	0.40***		0.46***	0.48***
183. Ensanche de Vallecas	0.39***	0.39***		0.35***	0.40***
191. Casco histórico de Vicálvaro	0.53***	0.50***		0.55***	0.51***
192. Valdebernardo	0.44***	0.43***		0.47***	0.48***
193. Valderrivas	0.34***	0.34***		0.47***	0.47***
194. El Cañaveral	-	0.17***		-	0.29***
201. Simancas	0.69***	0.66***		0.69***	0.68***
202. Hellín	0.54***	0.52***		0.66***	0.65***
203. Amposta	0.58***	0.55***		0.65***	0.63***

Table A.4. Neighbourhood agglomeration measures(continued)

204. Arcos	0.67***	0.55***		0.65***	0.59***
205. Rosas	0.46***	0.52***		0.50***	0.57***
206. Rejas	0.52***	0.48***		0.39***	0.40***
207. Canillejas	0.53***	0.51***		0.56***	0.56***
208. El Salvador	0.58***	0.58***		0.63***	0.65***
211. Alameda de Osuna	0.35***	0.34***		0.39***	0.37***
212. Aeropuerto	0.45***	0.34***		0.37***	0.31***
213. Casco Histórico de Barajas	0.37***	0.35***		0.38***	0.39***
214. Timón	0.34***	0.32***		0.36***	0.39***
215. Corralejos	0.50***	0.49***		0.48***	0.53***

*Note: *Indicates that the estimation is statistically significant at the 90% level, ** at the 95% level, and *** at the 99% level.*

Chapter 2: A weighted distance-based point-level indicator for measuring sectorial agglomeration in the case of Chinese manufacturing industry

2.1. Introduction

The manufacturing industry is a crucial driver of growth in China. The "Fourteenth Five-Year Plan[25]" emphasises the full implementation of the "Manufacturing Strong Nation" initiative (Asian Development Bank, 2021). Furthermore, the 20th National Congress Report highlighted the priority of focusing on the real economy for economic development (Xi, 2022). The report advocated for promoting new forms of industrialisation and channelling various resources toward strengthening the real economy, particularly the manufacturing sector. Additionally, at the third Belt and Road Forum in 2023, China announced plans to lift all restrictions on foreign investment access in the manufacturing sector to support an open world economy. The global pandemic not only underscored the importance of the manufacturing sector worldwide but also affirmed China's role as a pivotal link in the global supply chain (BRF, 2023). These facts highlight the importance of the Chinese manufacturing sector not only for China's economic policy but also for the global economy as a whole.

In this context, opposing forces are reshaping the spatial distribution of productive activities within China. On one hand, the concentration of activity generates necessary productivity gains in a high-growth setting, shaping the contribution of industry to national growth. On the other hand, growing concentration drives up costs that companies must bear (e.g., rent) and wages, which may eventually slow this concentration. Additionally, following the reform and opening-up policies, the implementation of fiscal decentralisation has intensified local protectionism, leading to regional market segmentation (Wei, 1999; Wei, 2001; Zhao & Zhang, 1999; Bai et al., 2004; He & Wang, 2012). As a result, a certain redistribution of activity has taken place,

[25] The Fourteenth Five-Year Plan for National Economic and Social Development (2021–2025).

rebalancing income territorially across the country. This has led to an uneven spatial distribution of economic activities, with varying degrees of spatial agglomeration across different regions and industries. As domestic market reform and globalisation deepen, the spatial distribution of industrial location gradually aligns with the development advantages across administrative boundaries (Mao et al., 2013; Brandt et al., 2017). The objective of this study is framed within this context, addressing the measurement of agglomeration in China's manufacturing sector following more than three decades of substantial economic changes in the country.

This study has two main contributions. Theoretically, it introduces a new weighted distance-based method to measure the industrial agglomeration patterns of all manufacturing sectors, supported by point-level indicators. This indicator, derived from that presented in the previous chapter, enables analysis from both sectoral and geographical perspectives and shares the same advantages. This approach overcomes the Modifiable Areal Unit Problem[26] (Openshaw & Taylor, 1979) associated with traditional agglomeration measurement indicators, widely used in Chinese manufacturing, providing a more accurate representation of industrial spatial agglomeration characteristics. Additionally, compared to the indicator used in the previous chapter, a weighting element such as employment is introduced, commonly used in the calculation of second-generation indicators. Empirically, this study undertakes the measurement of spatial concentration for the Chinese manufacturing sector, utilising data on 1.7 million firms with actual workers in two-digit manufacturing industries across the country, using both provincial (31) and prefecture-level (368) geographic disaggregations. This approach avoids several limitations of previous studies that included only large firms. This computational undertaking is significant, as previous investigations were typically limited to specific cities or regions, focused on a smaller number of industries or on larger firms, and did not account for the edge effect. More importantly, and with practical implications for the results, the

[26] The Modifiable Areal Unit Problem (MAUP) was first introduced by Openshaw & Taylor (1979). For a comprehensive analysis of MAUP, also known as the "dots to boxes" problem, see Briant et al. (2010).

Orbis database indicates that 90% of Chinese manufacturing companies are micro or small enterprises. Although their share of total employment is considerably lower, their exclusion could significantly affect the results, particularly in sectors where such companies hold greater relevance. Additionally, we used more recent data from 2021.

To the best of our knowledge, no existing research comprehensively reflects the overall agglomeration effects of China's manufacturing industry or the agglomeration attractiveness of each region. Consequently, a significant gap remains in research on the agglomeration dynamics within China's manufacturing sector.

In addition, China's spatial agglomeration of industries is influenced not only by natural environmental differences and the economic effects of agglomeration but also by factors related to the national system. We also examine the role of Economic Development Zones[27], a location-based policy tool frequently employed by both central and local governments, in the evolution of China's manufacturing agglomeration. These zones have emerged as highly concentrated areas for the manufacturing sector across the country. Existing studies have verified that development zone policies have local effects on the agglomeration of certain manufacturing industries, but their nationwide impact remains unknown (Wang, 2013; Alder et al., 2016; Zheng et al., 2017; Lu et al., 2019; Liu et al., 2023). Therefore, we combine microdata from China's manufacturing sector with spatial distribution data of development zones, considering whether firms are located within a development zone to assess the impact on the geographical agglomeration of different manufacturing activities. However, the extensive use of this instrument throughout China may generate more localised than regional effects, potentially limiting its impact on country-level agglomeration measures.

[27] The establishment of development zones began in 1984, and after nearly 40 years of expansion, they have rapidly spread from the initial eastern coastal region to the central and western regions, now nearly covering the entire country. According to the latest *Catalogue of China's Development Zones (2018 Edition)*, the State Council has approved a total of 2,543 development zones, including 552 state-level development zones and 1,991 provincial-level zones (autonomous regions and municipalities directly under the central government).

The rest of this chapter is organised as follows. After this brief introduction, an overview of existing research on the geographic concentration of China's manufacturing industry is provided. The third and fourth sections describe the proposed weighted indicator and the data used. In the fifth section, results are discussed. The main conclusions of the chapter are summarised in the final section.

2.2. Literature review

Since the 1990s, an increasing number of methods for measuring industrial agglomeration have been put into practice. Research on the agglomeration of Chinese manufacturing has also evolved in line with the development of these indices (the most representative studies are shown in Table 2.1). As discussed in the previous chapter, Duranton & Overman (2005) have categorised the evolution of agglomeration measures into three generations. The first-generation indices measure the degree of industry concentration using aggregated data and generally by adapting pre-existing indicators. Interestingly, in China's case, the original data were sometimes aggregated from firm-level microdata sources. Without a doubt, the Gini index is probably the most widely used measure, or the Hoover index, which is closely related to it, although some studies also employ the Herfindahl index. The most common disaggregation level has been at the 2-digit level for the industrial (or manufacturing) sector and at the provincial level geographically. Some studies offer further sectoral disaggregation, up to the 3-digit or 4-digit level, to assess the impact that aggregation has on firm-level results. Additionally, as most articles use firm-level data, different administrative disaggregation levels (prefectures and counties) are also used to address the headquarters effect. The time period analysed in this type of research spans from the 1980s to the first half-decade of the 21st century.

Additionally, a wide variety of variables are used in these measurements (number of establishments or firms, production, added value, and employment). Notably, most studies consider only large and medium-sized firms. It is also worth mentioning that only a few studies undertake a dynamic analysis of spatial concentration, for which it is always necessary to have comparable databases that are sufficiently spaced out over time.

Five key findings emerge from these studies: (i) a growing trend of geographic agglomeration and industrial concentration, especially following economic reforms (Wen, 2004; He & Zhu, 2009; Long & Zhang, 2012). Some studies analysing periods before the reforms identify a U-shaped trend (Bai et al., 2004); (ii) a high concentration in several eastern coastal regions (Wen, 2004); (iii) significant heterogeneity among sectors, with this increase being more pronounced in the more liberalised and globalised sectors (He & Zhu, 2009). This agglomeration also occurs in labour-intensive industries (Fan & Scott, 2003); (iv) a positive relationship between agglomeration and productivity (Fan & Scott, 2003); and (v) considerable sensitivity of agglomeration indicators to the level of disaggregation used: greater agglomeration at smaller geographical units and in more disaggregated industries (He et al., 2007).

The second generation of indicators is mainly limited to the use of the index proposed by Ellison & Glaeser (1997). Interestingly, firm-level data are also used and then aggregated, except in the case of Lu (2010), who uses establishment-level data. The availability of these individual data allows for greater geographical granularity, down to counties and ZIP codes, and sectoral disaggregation. A common feature among all these studies is the use of employment as the reference variable. The time period considered is restricted to the decade from 1996 to 2001.

The main findings can be summarised in four points: (i) a trend toward increased industrial agglomeration, although still below the levels of developed countries (Lu & Tao, 2009); (ii) greater protection in inland China poses an obstacle to agglomeration (Lu & Tao, 2009); in fact, state-owned enterprises are observed to be less concentrated than private ones, likely because they are used as a tool for regional policy (Lu, 2010); (iii) a higher concentration in technology industries (Hong & Fu, 2011); and (iv) certain sector-specific findings, such as that of Lin et al. (2011) for the textile industry, which shows an inverted U-shaped relationship between agglomeration and productivity.

The third-generation index, or distance-based methods, has relatively strict data requirements, needing information such as the location or size of firms, and involves a complex measurement process. Consequently, applying the distance-based method to study industrial agglomeration in China presents considerable challenges. For this reason, the application of the third-generation index in China has been relatively limited. In recent years, however, a few scholars have made progress in expanding industrial agglomeration measurement methods.

All studies conducted with this methodology use the indicator proposed by Duranton & Overman (2005), which, based on the K-Ripley function, estimates density indicators by using kernel functions to weight the distance of other firms from the point of analysis. Generally, constructing these measures poses a computational challenge due to the amount of bilateral information (between production points) that must be calculated, introducing at least four types of restrictions or limitations to reduce the points considered: (i) focusing the analysis on specific regions; (ii) also limiting these calculations to certain sectors rather than to manufacturing or the economy as a whole; (iii) restricting the analysis to large and medium-sized firms, which can reduce the sample of analysis points by up to 90%; and (iv) implementing bootstrapping strategies to create smaller, operationally manageable, representative subsamples. Finally, it should be noted that all these studies use firm-level data. The period analysed covers the first two decades of the 21st century.

The findings in this case are specific to the analysed regions or sectors, but overall indicate: (i) a recent downward trend in agglomeration across the manufacturing sector (Huang et al., 2022), with a reversal particularly among larger and new entrant firms; (ii) significant sectoral differences, as services, high-tech industries, and labour-intensive manufacturing activities tend to cluster, while personal services and capital-intensive industries are more dispersed (Li et al., 2015a; Huang et al., 2022); and (iii) a substantial difference between analyses focused on urban agglomerations versus the national economy. Consequently, excluding smaller firms may reverse these results

(Long & Zhang, 2012), and focusing on urban areas likely overestimates industrial agglomeration in China (Li et al., 2019; Liu et al., 2020; Jiao, 2021; Wu et al., 2022).

Table 2.1 Main studies evaluating the agglomeration of the manufacturing industry in China

	Authors	Measure	Zone	Source	Original data	Sector	Units	Period	Variable
First generation	(Fan & Scott, 2003)	Herfindahl index	China	Statistical yearbooks of the Chinese provinces	Sector/province	2-digit industrial sectors	Provinces	2000	Establishments Employment
	(Bai et al., 2004)	Hoover index on location quotient	China	CSY, CSYIE, CIC, CSB	Firm level data Various types of firms	2-digit Industrial sectors	Provinces	1985-1997	Output
	(Wen, 2004)	Gini coefficient	China	CIC	Firm level data Large and medium size firms	3-digit Industrial sector (large and medium firms)	Provinces	1995	Value added
	(He et al., 2007)	Gini coefficient	China	CEC	Firm level data Large and medium size firms	2-digit and 167 Manufacturing industries	Provinces and counties	2004	Employment
	(He & Zhu, 2009)	Gini coefficient	China	CSYIE, ASIF	Firm level data Large and medium size firms	2-digit Manufacturing industries	Provinces	1980-2004	Employment, gross output, value added
	(Long & Zhang, 2012)	Gini coefficient	China	CIC, CEC	Firm level data All sizes	2-digit, 3-digit, 4-digit industrial sectors	Provinces, prefectures and counties	1995, 2004	Output, number of firms

Table 2.1 Main studies evaluating the agglomeration of the manufacturing industry in China (continued)

Second generation	(Lu & Tao, 2009)	EG index	China	ASIF	Firm level data Large and medium size firms	2-digit, 3-digit, 4-digit manufacturing industries	Provinces, cities and counties	1998-2005	Employment
	(Lu, 2010)	EG index	China	CESC	Establishment level data All establishment	Broad, 2-digit, 3-digit, 4-digit primary, secondary and tertiary industries	Provinces, cities and counties	1996, 2001	Employment
	(Hong & Fu, 2011)	EG index	China	CEC	Firm level data All sizes	4-digit manufacturing industries	Provinces, cities, counties and ZIP codes	2004	Employment
	(Lin et al., 2011)	EG index	China	ASIF	Firm level data Large and medium size firms	3-digit textiles industries	Cities	2000-2005	Employment
Third generation	(Li et al., 2015)	Kd	Beijing	AIC	Firm level data All sizes Bootstrapping samples	4 industrial industries and 4 service industries	Firms	2010	Points
	(Brakman et al., 2017)	Kd	China	ASIF	Firm level data Large and medium size	2-digit manufacturing industry	Firms	2002,2008	Points
	(Tian et al., 2017)	Kd	Shanghai	AIC	Firm level data All sizes	2-digit manufacturing	Firms	1990-2010	Points
	(Huang et al., 2022)	Kd	Beijing-Tianjin-Hebei region	AIC	Firm level data All sizes	3-digit manufacturing industries	Firms	2004-2013	Points
	Huang et al., 2024)	Kd	Guangdong	Qichacha	Firm level data All sizes	1 manufacturing industry	Firms	2000-2022	Points

CSY: China Statistical Yearbook; CSYIE: China Statistical Yearbook on Industrial Economy; CIC: China Industrial Census; CSB: China Statistics Bureau; CEC: China Economic Census; ASIF: Annual Survey of Industrial Firms; CESC: Chinese Establishment Census; AIC: Administration for Industry and Commerce.

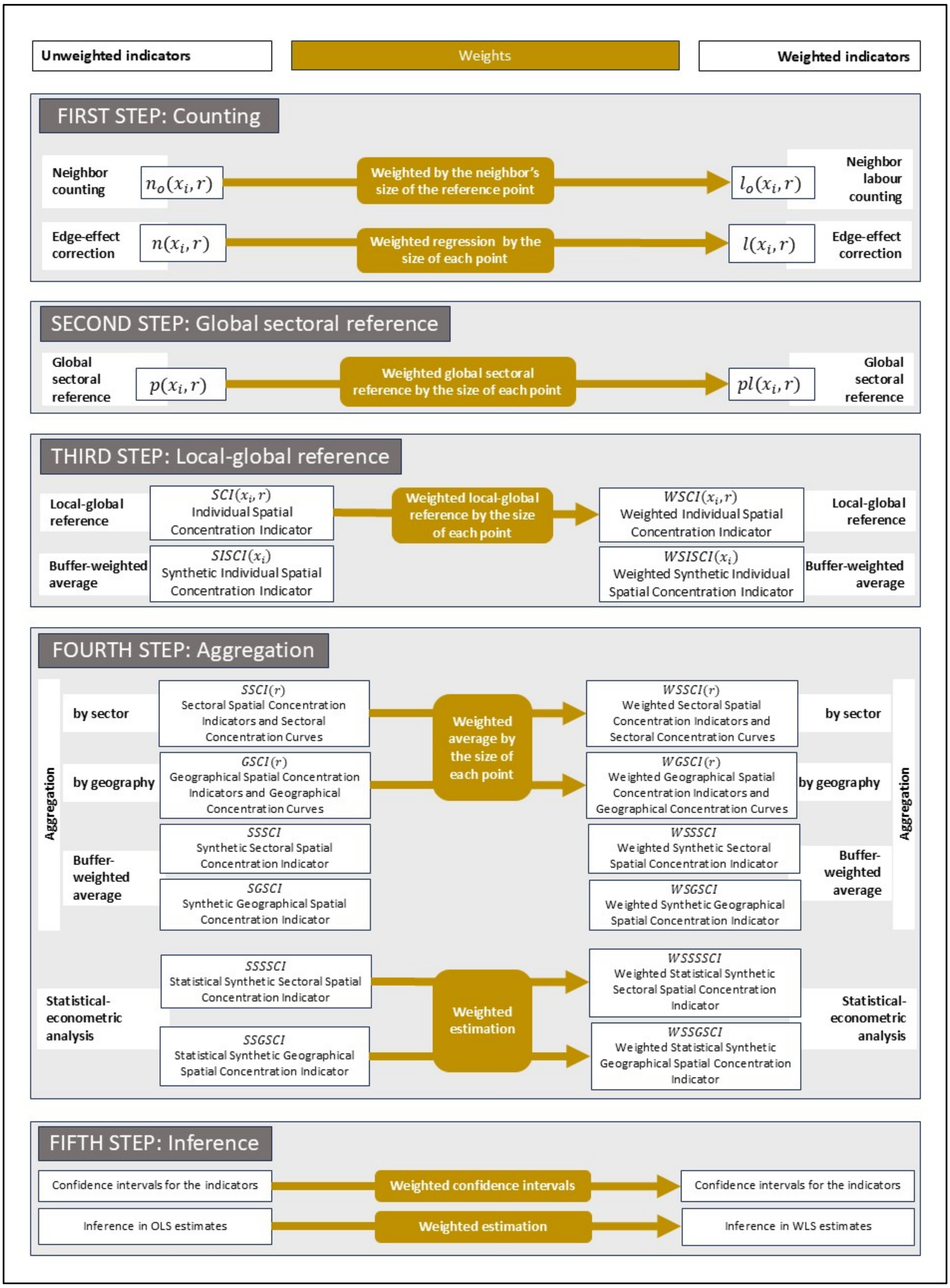
2.3. Weighted distance-based point-level and derived aggregate indicators for sectoral agglomeration

The stylised procedure outlined in the previous chapter serves to explain the methodology for obtaining sectoral and geographical agglomeration indicators, grounded in the typology proposed by Marcon & Puech (2017).

The main difference and contribution of this research is that previous proposals measured an unweighted indicator, whereas here, all agglomeration indicators are weighted by activity size, specifically, by the number of employees, which is the most commonly used weighting variable.

Moreover, the availability of information on the activity size of firms allows for a more accurate measurement of productive agglomeration, rather than relying solely on the number of production points as done in the previous chapter. This availability enables each production point to be evaluated based on its significance, making the indicators more aligned with the concept of productive agglomeration. Additionally, by considering only active firms (in this case, those with employment), it avoids counting "zombie" activity, i.e., firms that remain registered but are inactive or maintain "fictitious" activity for tax reasons. Figure 2.1 shows the differences between the unweighted indicator (proposed in the first chapter) and the weighted indicator throughout the calculation process.

Figure 2.1 Relationship between the proposed Unweighted and Weighted distance-based indicators



2.3.1. Weighted distance-based point-level indicators for sectoral agglomeration

Let us assume there is a certain sector of activity with N_s geographically referenced production points that, in total, employs L_s workers, all located within an administrative area with a total of N production points and L employment. For each of the administrative divisions within the studied area, N_g and L_g refer to the total production points and employment across all sectors within that division.

The first step in defining a weighted agglomeration indicator is to count the employees of neighbouring points from the same sector around each reference point within a maximum distance of r, following expression [1].

$$l_0(x_i, r) = \sum_{\substack{j \neq i \\ i,j \in s}} l_j\left(\left\|x_j - x_i\right\| \leq r\right) \qquad [1]$$

The reference point is denoted as x_i, representing establishment i of sector s, while its neighbours in the same sector are x_j; l_j is the employment in the neighbouring establishment x_j. Therefore, $l_0(x_i, r)$ measures the employment around establishment x_i in the same sector of activity s within a maximum distance of r. Expression [1] is a simplification of the standard stylised expression for measuring neighbours developed in the previous chapter (expression [1.a]). However, for simplicity, the measurement is developed in an unweighted manner by density ($w(x_j) = 1$), weighted by the activity size of the neighbour ($z(x_j) = l_j$), and unweighted by the distance between x_j and x_i ($k = 1$), and in this step, without edge-effect correction. Consequently, the agglomeration measurements will depend on the distance considered for the buffer.

However, when the buffers generated extend beyond the administrative geographic boundaries of the database coverage, it becomes unknown how many points would be outside the study area. Therefore, a correction for this edge effect is introduced based on the inverse of the buffer overlap with the studied administrative area (λ_{ir}) but raising the overlap to an exponent that will range between 0 and 1. Thus, the number of employees around each of the points of production within the same sector, after correction, will now be determined by this adjustment.

However, when the buffers generated extend beyond the administrative geographic boundaries of the database coverage, it becomes unknown how many points would be outside the study area. Therefore, we introduce a correction for this edge effect based on the inverse of the buffer overlap with the studied administrative area (λ_{ir}) but raising the overlap to an exponent that will move between 0 and 1 Thus, the number of employees around each of the points of production within the same sector after correction will now be determined by expression [2].

$$l(x_i, r) = \frac{l_0(x_i,r)}{\lambda_{ir}^{\gamma_{sr}}} \quad [2]$$

Where γ_{sr} is the exponent that can vary between sectors and the amplitude of the buffer, as described in the previous chapter. If the exponent γ_{sr} takes on the lower limit of the interval, such correction would not be necessary, while a value equal to 1 would apply the standard correction by the inverse of the buffer overlap. Intermediate values would essentially result in levels of partial correction.

In order to approximate the value of γ_{sr} for each sector and buffer, a relationship is established between the number of corrected employees at each activity point and their qualitative characteristics z_i, which could affect inaccurate measurement of concentration in their surroundings, with fixed effects of sector (d_s) and buffer (d_r) (*cf.* expression [3a]).

$$\ln l(x_i, r) = \alpha + \beta_1 z_i + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \varepsilon_{ir} \quad [3a]$$

Since $l(x_i, r)$ is not observed, it is replaced by expression [2], yielding expression [3b].

$$\ln\ l_0(x_i, r) = \alpha + \beta_1 z_i + \gamma_{sr} \ln(\lambda_{ir}) + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \varepsilon_{ir} \quad [3b]$$

Where the coefficient γ_{sr} is estimated based on the interaction between sector and buffer fixed effects with the overlapped area (*cf.* expression [3c]).

$$\gamma_{sr} \ln(\lambda_{ir}) = \gamma \ln(\lambda_{ir}) + \sum_s \vartheta_s \ln(\lambda_{ir})\, d_s + \sum_r \nu_r \ln(\lambda_{ir})\, d_r \quad [3c]$$

Substituting [3c] into [3b] gives the final expression [3d] to be estimated by weighted least squares, using the employment in x_i as the weight.

$$\ln l_0(x_i,r) = \alpha + \beta_1 z_i + \gamma \ln(\lambda_{ir}) + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \sum_s \vartheta_s \ln(\lambda_{ir})\, d_s + $$
$$+ \sum_r \nu_r \ln(\lambda_{ir})\, d_r + \varepsilon_{ir} \quad [3d]$$

So that $\gamma_{sr} = \gamma + \vartheta_s + \nu_r$.

The second step consists of establishing a sectoral reference, so that firms are comparable across different sectors. To do this, $l(x_i, r)$ is divided by the number of employees within the sector, which is further corrected by the weighted average of the individual correction, similar to expression [2]. In this way, the proportion of jobs in the same sector located around the point x_j within a radius r is obtained according to expression [4].

$$pl(x_i, r) = \frac{l(x_i, r)}{L_s \Big/ \left(\sum_{j \in s} \frac{l_j}{L_s} \lambda_{jr}^{\gamma_{sr}}\right)} \quad [4]$$

In the third step, a Weighted Spatial Concentration Indicator weighted by employment (*WSCI*) at the point level for a buffer r is obtained by normalising expression [4] by the mean for all production points across all sectors of the previous proportion, weighted by their relative importance, considering that the buffer is fixed (a local-global reference), resulting in expression [5].

$$WSCI(x_i, r) = \frac{pl(x_i, r)}{\overline{pl(r)}} \quad [5]$$

Where $\overline{pl(r)} = \sum_i \frac{l_i}{L} pl(x_i, r)$ represents the weighted average, for the entire administrative study area and for all production points, of the relative frequency with which employees of the same sector are found, at a maximum distance of r around each activity point. This produces a relative indicator that is fully comparable across sectors. This correction controls for the first aspect of first-order concentration, specifically the global join-location.

However, the value of the above-mentioned indicator depends on the size of the buffer used and therefore cannot provide a single measure of the concentration level of activities around reference establishments. Therefore, a Weighted Synthetic Individual Spatial Concentration Indicator (*WSISCI*) at the point level is also proposed, which does not rely on distance, by aggregating weighted indicators for different buffers, as in expression [6].

$$WSISCI(x_i) = \frac{\sum_r WSCI(x_i, r) f(r)}{\sum_r f(r)} \qquad [6]$$

Where $f(r) = \left(\frac{r_0}{r}\right)^g$ is a function of distance, with *g* being greater than zero and able to adopt different values such as 1/2, 1, 2, etc. Here, r_0 refers to the radius of the smallest buffer considered, and *r* relates to the entire range of buffers for which the previous indicators are calculated. This way, the results obtained with buffers closer to the analysed point are considered more relevant. Consequently, both the collection of indicators $WSCI(x_i, r)$, and $WSISCI(x_i)$ are measures of agglomeration at the point level.

2.3.2. Weighted sectoral Agglomeration Indicators and Agglomeration Curves

Once the weighted point-level indicators are obtained, the fourth step is to calculate weighted sectoral concentration indicators (or by geographical unit) as a weighted average of the individual indicators for all points within the same sector of activity. Therefore, the Weighted Sectoral Spatial Concentration Indicator *(WSSCI(r))* for each buffer will be constructed as shown in expression [7a][28].

$$WSSCI(r) = \sum_{i \in s} \frac{l_i}{L_s} WSCI(x_i, r) \qquad [7a]$$

Sorting the indicators by buffer size (from smallest to largest) produces sectoral agglomeration curves (or by geographical area). These curves can be classified, as in the previous chapter, according to two criteria (as shown in Table 2.2): a) whether the curve crosses the reference value of one, and b) the shape of the curve, distinguishing between decreasing, U or V shape, inverse U or V, double U or V in opposite directions,

[28] The construction of the Weighted Geographical Spatial Concentration Indicators (*WGSCI(r)*) follows expression [7b], calculated as the weighted average of all points within the same geographical unit (such as a province, city, or more detailed administrative divisions):

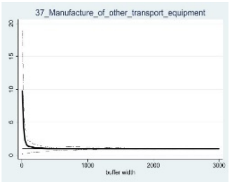
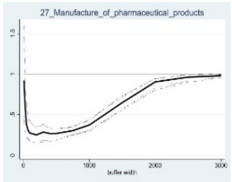
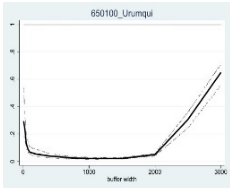
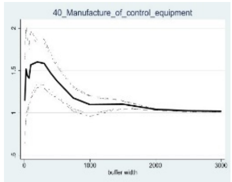
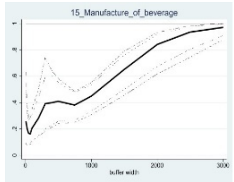
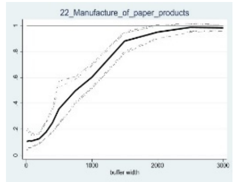
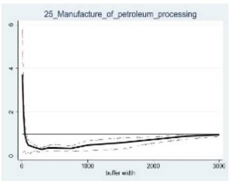
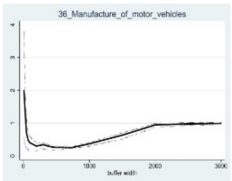
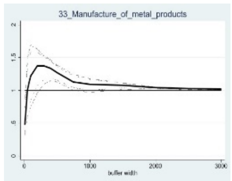
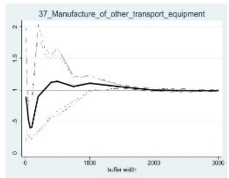
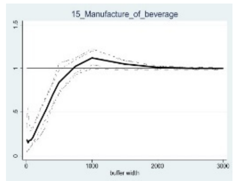
$$WGSCI(r) = \sum_{i \in g} \frac{l_i}{L_g} WSCI(x_i, r) \qquad [7b]$$

Where *g* represents the geographical area for which the indicator is calculated, and L_g is the total number of employees across all activities located within *g*.

and finally, increasing.

Curves consistently above (or below) the unit value indicate that, irrespective of the curve's shape, the activity (or geographical area) displays a stronger attraction (or repulsion) relative to the average. In contrast, curves that cross the unit value suggest that the analysis of activity (or geographic area) concentration depends on the radius considered. Therefore, assessing the concentration level for firms in a sector (or geographic area) needs to account for different distances. Generally, if values within smaller buffer zones are above (or below) the unit, the agglomeration curves for the sector (or geographical area) indicate a relatively agglomerated (or dispersed) behaviour. Conversely, if the unit crossing occurs from initially higher (or lower) values to significantly lower (or higher) ones, the sign may be reversed. Therefore, sectors (or geographical areas) showing type 6 agglomeration curves are expected to demonstrate typical sectoral agglomeration patterns, while types 7 and 10 may indicate dispersion. Type 8 and 9 curves tend to be harder to classify, and their behaviours may vary.

Table 2.2 Typology of spatial concentration curves (*)

	Decreasing or L	U or V	Inverse U or V	Double U and V	Increasing or inverse L
Do not cross the unit	**Type 1(A)**	**Type 2(D)**	**Type 3(A)**	**Type 4(D)**	**Type 5(D)**
		 			
Do cross the unit	**Type 6(A)**	**Type 7(D)**	**Type 8**	**Type 9**	**Type 10(D)**
					

(*) Real examples obtained for the case of China; (A) indicates Agglomeration behaviour, (D) indicates Dispersion behaviour.

The weighted average of the Synthetic Individual Spatial Concentration Indicators yields the Weighted Synthetic Sectoral Spatial Concentration Indicators[29] (*WSSSCI*), expression [8a].

$$WSSSCI = \sum_{i\in s} \frac{l_i}{L_s} WSISCI(x_i) \qquad [8a]$$

In the final step, the statistical significance of both sectoral (or geographic) agglomeration indicators and the global indicators will be evaluated. To perform this test, weighted confidence intervals are constructed around the null hypothesis (in this case, around the value of 1), following the same procedure as in the previous chapter.

As in the previous chapter, a significant number of replicates of the mean indicator for each sector is generated. These replicates are derived from representative subsamples of the indicators at the establishment level.

To determine the number of replicas k_s and the number of point-level indicators (*j*) considered in each of these "alternative" indicators, the minimum required sample size is calculated using expression [9c] for random sampling in finite populations.

$$k_s = \frac{L_s z_\alpha^2 pq}{e^2(L_s - 1) + z_\alpha^2 pq} \qquad [9c]$$

Where Z_a represents the confidence level according to the normal distribution, p is the expected proportion, and $q = 1 - p$. Since an expected proportion is not specified in this case, the sample size is maximised when $p = q = 0.5$. The variable e refers to the allowable error or precision. Using this approach, a number of indicators are calculated, with extreme values removed that represent the top and bottom 2.5 percent of employment. This provides the confidence interval for each sectoral or geographical clustering indicator. Note that the removal of extremes refers to the percentage of employment, not the number of points, meaning that the same number of points may not necessarily be eliminated from both ends, unlike in the unweighted indicator. Additionally, similar to the construction of the agglomeration curve, the confidence

---

[29] The Weighted Synthetic Geographical Spatial Concentration Indicator (*WSGSCI*) is calculated as shown in expression [8b].

$$WSGSCI = \sum_{i\in g} \frac{l_i}{l_g} WSISCI(x_i) \qquad [8b]$$

interval for this curve is obtained from the various considered buffers. Thus, if the value of one fall within this interval, the null hypothesis of no clustering for that sector (or geographic area) within that buffer cannot be rejected. The same process is applied to the weighted global indicators.

2.3.3. Synthetic indicators from statistical-econometric analysis

The previous sectoral and geographic indicators did not fully exploit all individual (point-level) information. Therefore, using least squares regression, the agglomeration intensity for each sector and geographic area is estimated simultaneously from the point-level indicator values $WSCI(x_i, r)$ and $WSISCI(x_i)$, by estimating expressions [10a] and [10b].

$$WSCI(x_i, r) = \alpha + \sum_s \alpha_s^1 d_{si} + \sum_g \beta_g^1 d_{gi} + \delta^1 z_i + \varepsilon_{isr}^1 \qquad [10a]$$

$$WSISCI(x_i) = \alpha + \sum_s \alpha_s^2 d_{si} + \sum_g \beta_g^2 d_{gi} + \delta^2 z_i + \varepsilon_{isr}^2 \qquad [10b]$$

Where d_{si} and d_{qi} represent sector and geographical dummies (prefecture-level divisions) in which each point is located; z_i captures establishment characteristics, which may in some way alter the measure of concentration. Expression [10a] uses the $WSCI(x_i, r)$ for all points and for different buffers; for this reason, we estimate by weighted least squares, weighted by the product of the function $f(r)$ and the relative weight of the employment of each production point with respect to the total employment of the studied area. The second factor constitutes the weights in the case of expression [10b].

From the coefficients α_s^1 or α_s^2, which are equivalent in value but have different standard deviations, the Weighted Statistical Synthetic Sectoral Spatial Concentration Indicator[30] (*WSSSSCI*) is obtained. In turn, β_q^1 and β_q^2, which are also equivalent in value though with differing standard deviations, assess for agglomeration effects that are location-specific rather than sector-specific, thus constituting the Weighted

[30] To obtain the final value of the Statistical Weighted Synthetic Sectoral Spatial Concentration Indicator (*SWSSSCI*) for each sector, the coefficient obtained in the corresponding dummy is summed with the constant term and the weighted average of the prefecture-level division dummies.

Statistical Synthetic Geographical Spatial Concentration Indicators (*WSSGSCI*). From these indicators, it is possible to determine the attractiveness of each prefecture-level division[31]. This procedure accounts for the second aspect of first-order concentration: the local join-location. Consequently, the sectoral indicator controls for local concentration within each administrative division, meaning that the resulting measure will capture only the second-order concentration, or colocation. Similarly, for the geographical indicator, the measure controls for the sectoral composition and spatial concentration derived from it.

2.4. Data

2.4.1. Study area

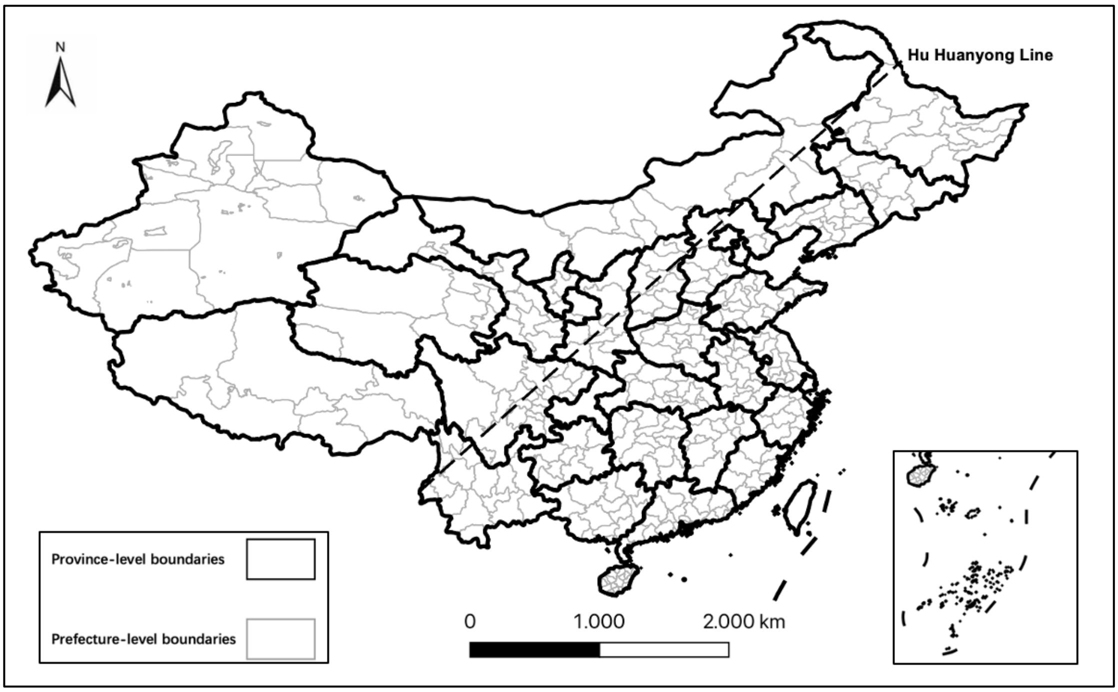
This chapter examines all manufacturing sectors within mainland China, excluding Hong Kong, Macao, and Taiwan. China spans approximately 5,200 Kilometres from east to west and 5,500 Kilometres from north to south, covering an estimated area of around 9.6 million square Kilometres. Administratively, China is divided across multiple levels[32] (see Figure 2.2). The first-level administrative divisions include 31 provincial-level units, consisting of 4 municipalities directly under the central government, 22 provinces, and 5 autonomous regions. The provincial level is considered too aggregated to capture clustering effects within provinces, particularly in larger provinces where substantial differences between cities exist. Prefecture-level divisions serve as the second-level administrative divisions and comprise 368 prefecture-level units (293 prefecture-level cities, 7 prefectures, 30 autonomous prefectures, 3 leagues, and 35 special administrative units not included in typical

[31] Similarly, to obtain the final Statistical Weighted Synthetic Geographical Spatial Concentration Indicator (*SWSGSCI*), the coefficient obtained from the dummy variable for each area is added to the constant term and the weighted average of the sector dummy variables.

[32] China's administrative divisions are organised hierarchically by the state, but high-quality, public data on these regions is limited. To work with this data, we use the 1:1 million public version of basic geographic data from the National Geomatics Centre of China as a base, which is segmented. Python is used to pre-process and merge the vectors. By adding the Ministry of Civil Affairs' "2021 Administrative Division Summary Table" as an attribute, the data is consolidated, vector errors are corrected, and the coordinate system is converted from CGCS2000 to WGS84. This process provides a detailed dataset with national boundaries and attributes for prefecture-level administrative divisions.

statistics, such as municipalities directly under the central government, provincial-level municipalities, and provincial-level counties). These units belong to the county level in the administrative hierarchy but lack separate prefecture-level structures. Applying the methodology from the previous section at finer administrative levels could introduce challenges in accurately evaluating geographic indicators with sufficient data.

Figure 2.2 Chinese administrative divisions: Provinces and prefectures



2.4.2. Data source and pre-processing

The data used in this paper were sourced from the Orbis database (Bureau van Dijk), which contains comprehensive records on companies, including identifiers, location, activity, industry classification, and detailed financial information. Orbis is widely used in economic research and is considered one of the most robust resources for firm-level data. To ensure data quality and study reliability, necessary preprocessing steps were conducted. Specifically, the focus was on manufacturing data from 2021, selecting firms with active status and reclassifying the information based on the Chinese Industrial Classification for National Economic Activities (GB/T4754-2017)[33]. This analysis considers 31 manufacturing industries at the 2-digit level.

[33] The GB/T 4754-2017 code closely aligns with the ISIC Rev. 4 code. For further reference, see Table B.1.1 of Appendix B.1.

The distance-based method requires precise geographical information (i.e., the geographical coordinates) of firms' locations. Notably, each firm's location is determined by the main address of its headquarters, with latitude and longitude coordinates obtained via Google's Geocoding API[34]. After verifying the accuracy of each firm's location, a sample of 2,796,388 manufacturing firms was obtained. However, firms without employee data were excluded, resulting in a final sample of 1,696,392 manufacturing firms with employment information. The available data refer to firms rather than establishments. Nevertheless, in the Chinese manufacturing sector, less than 2% of firms are multi-establishment, so the assumption that all activity occurs at the headquarters location is not especially restrictive (Hong & Fu, 2011).

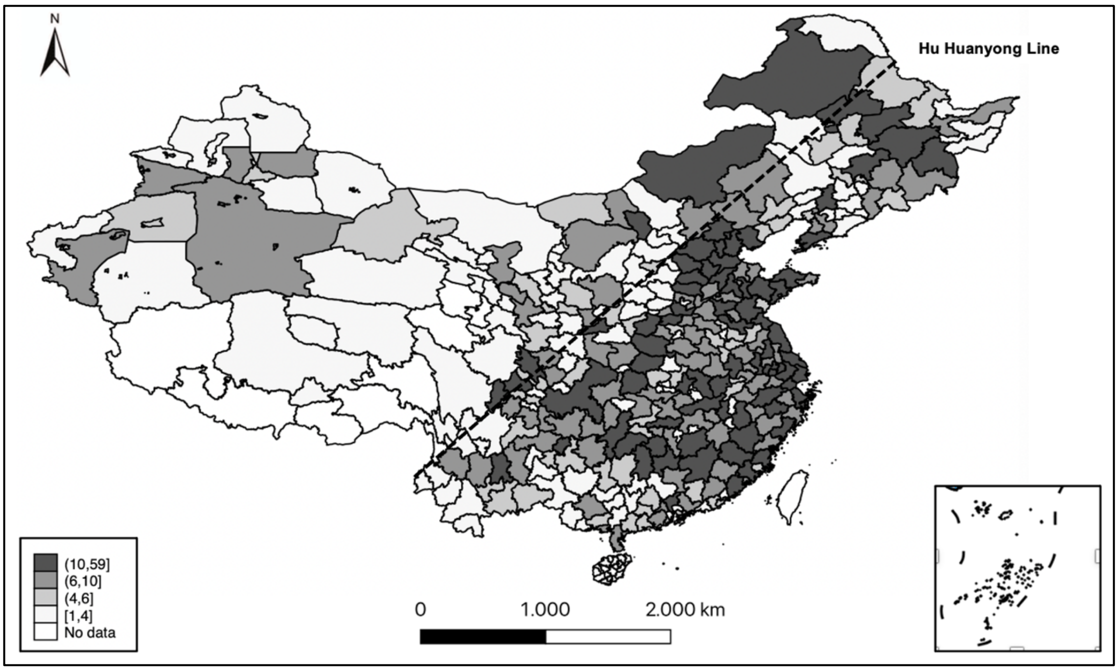
To control for potential factors favouring the geographic concentration of manufacturing activity, specific information on Chinese Economic Development Zones (EDZs) is required. These zones implement targeted economic policies, such as investment incentives and foreign direct investment (FDI) promotion policies. Data on EDZs are sourced from the 2018 edition of the Catalogue of China's Development Zones[35], published by various national agencies, including the National Development and Reform Commission. This catalogue provides attributes such as the names of development zones, approval dates, and areas. Baidu Map's Geocoding API[36] was used to determine the geographic coordinates for each development zone, enabling the creation of a polygon map and identifying firms from our database located within each EDZ (see Figure B.2.1 of Appendix B.2 for details). As shown in Figure 2.3, the spatial distribution of development zones is unevenly divided by Hu Huanyong's Line[37]. The

---

[34] The Geocoding API is a service that converts addresses into latitude and longitude coordinates. For this study, Python was used to handle batch geocoding requests. However, geocoding addresses in China presents challenges because the Orbis-provided addresses are based on alphabetic spellings. Therefore, it is essential to supply accurate parameters for address or component lookups, requiring some minor adjustments to the original data.
[35] https://www.gov.cn/zhengce/zhengceku/2018-12/31/5434045/files/6eea5e4b78a645c1a27c231b152792ef.pdf
[36] Baidu Maps, a widely used map service in China, is particularly effective at recognizing addresses in Chinese, which is essential since the attribute information for the development zones is provided in Chinese. Processing these development zones is relatively complex; each development zone's polygon map and boundary must be individually searched on Baidu Maps. Python was used to automate batch geocoding requests and download the polygon maps for each development zone after conversion.
[37] The "Hu Huanyong Line" acts as a significant demographic and economic demarcation in China,

eastern and southern regions host a higher concentration of development zones compared to the western and northern regions.

Figure 2.3 Spatial distribution of development zones above the prefecture-level administrative divisions in China



2.5. Results

2.5.1. Calculation of distance-based point-level indicators for sectoral agglomeration in the Chinese manufacturing sector

To measure the geographic concentration of manufacturing activity in China, 15 buffers were selected, ranging from 10 to 3,000 Kilometres[38]. Given the computational challenges associated with the sample size, a finer measurement was avoided as it would require substantially more processing time without significantly altering the

distinguishing regions with contrasting levels of population density, economic development, and social structure.

[38] The selected distances are 10, 25, 50, 75, 100, 200, 300, 400, 500, 750, 1,000, 1,500, 2,000, 2,500, and 3,000 Kilometres. A 3,000-kilometer radius nearly covers the entire geographic area of China. This selection was made ad-hoc, with finer intervals at shorter distances and broader intervals at longer distances, given the high concentration in the East and the lower concentration in much of central and western China. When constructing these buffers for features spread over a large area (such as China), no projection can provide fully accurate results. Therefore, to account for the curvature of the Earth, critical for a vast country like China, Geodesic Buffers are used, where distances are measured on an ellipsoid or spherical globe.

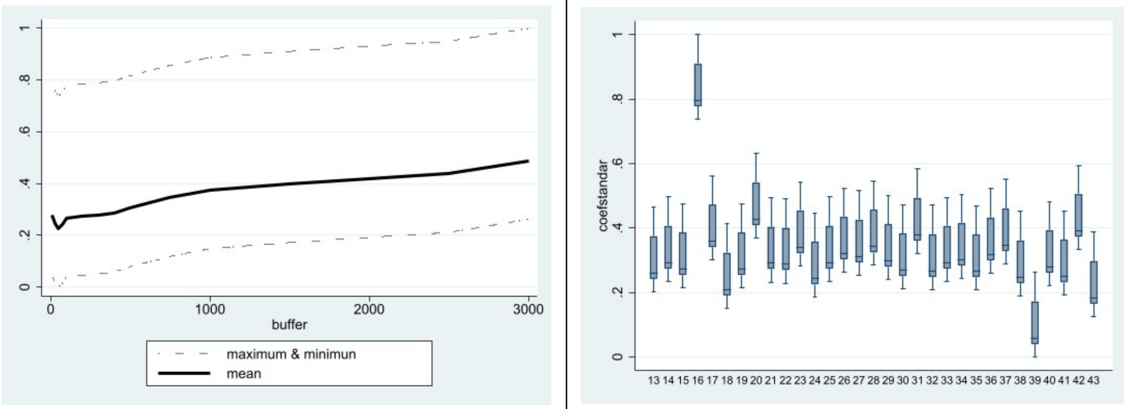
results.

The exponent γ_{sr}, which adjusts the usual edge-effect correction for the inverse overlap between the considered buffer and the administrative area, is estimated by using expression [3.d]. Only sectoral and buffer dummies are introduced. This estimate is made in a weighted manner using the product of two elements: the weight for the size of the firm in terms of employment and the $f(r)$ function in relation to the buffer width. The first is a logical consequence of constructing a weighted-activity measure. The second is specifically suitable for the use of different buffers and addresses some weighting issues in the case of large countries, such as China, with strong agglomeration in only part of the country. In this case, the increase in buffer size leads to a significant drop in company density, especially with very large radius. If this weighting were not applied, larger buffers would dominate the estimation.

The results for the estimated exponents are presented in Figure 2.4. From panel A of the figure, it is observed that the exponents γ_{sr} increase slightly with buffer size. Although this trend is not very pronounced, considering that the increase is measured across buffer extensions up to 3,000 Kilometres, it contrasts with the trend observed in the case of Madrid in the previous chapter. This difference is likely linked to the uneven relative distribution of productive activity in China, where most companies are highly concentrated on one end, the East Coast, and are significantly dispersed in the North and West, unlike Madrid, where activity is centred in the urban core. As seen in panel B of Figure 2.4, which presents the box plots by sector, the results are relatively similar across sectors, though with greater variation than in the case of Madrid. Notably, two sectors, Tobacco (16) and Computer (39), show significant differences compared to the others.

Figure 2.4 Overlap correction moderator by buffer width and sector

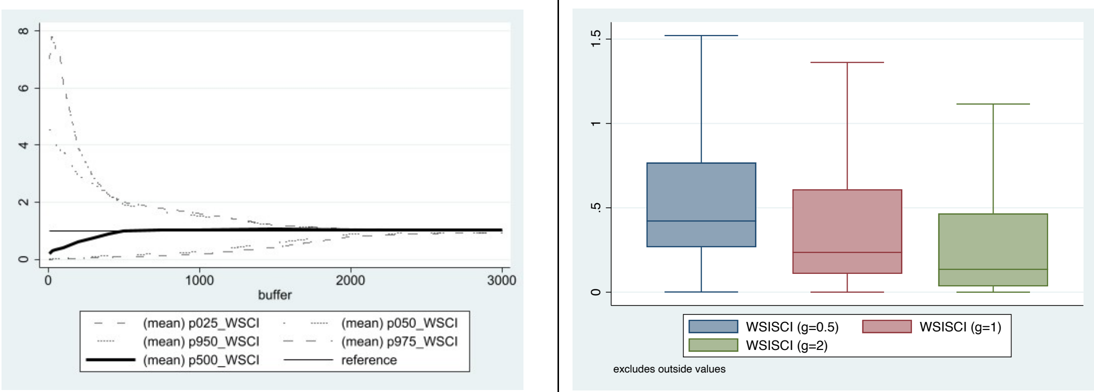
A. Evolution of overlap correction moderator (γ_{sr}) by buffer width	B. Boxplots of the overlap correction moderator (γ_{sr}) by sector of activity



Once these exponents are estimated, it is possible to calculate the Weighted Sectoral Concentration Indicator at the firm level, $WSCI(x_i, r)$, for each buffer (as shown in Panel A of Figure 2.5), as well as the Weighted Synthetic Individual Spatial Concentration Indicator, $WSISCI(x_i)$, (Panel B of Figure 2.5). This calculation considers different weighting functions, specifically $g \in \{0.5, 1.0, 2.0\}$, applied to the indicators for each buffer in constructing the synthetic indicator.

Figure 2.5 Distribution of distance-based point-level indicators for sectoral agglomeration

A. Individual spatial concentration indicators, $WSCI(x_i, r)$	B. Synthetic individual spatial concentration indicators, $WSISCI(x_i)$



The $WSCI(x_i, r)$ indicator values tend to converge to 1 as the buffer size increases. By construction, a larger buffer incorporates a greater number of employees from activity points within the same sector in the indicator's numerator, while the denominator includes all employees in the sector (expression [4]). Consequently, the $WSCI(x_i, r)$ indicator approaches this unit value. Additionally, the median remains below unity, at least up to buffer sizes of 1,000 kilometers. Moreover, the 95th percentile values converge toward the reference value faster than the 5th percentile. These effects are linked to the greater asymmetry in the indicator between positive values (indicating concentration) and negative values (indicating dispersion), particularly for industries and firms with strong concentration (attraction potential). A common characteristic of this type of indicator is also observed here: by construction, dispersion ranges between 0 and 1, while agglomeration can yield values above unity with no upper limit. This effect can be mitigated by using a logarithmic scale for the indicators.

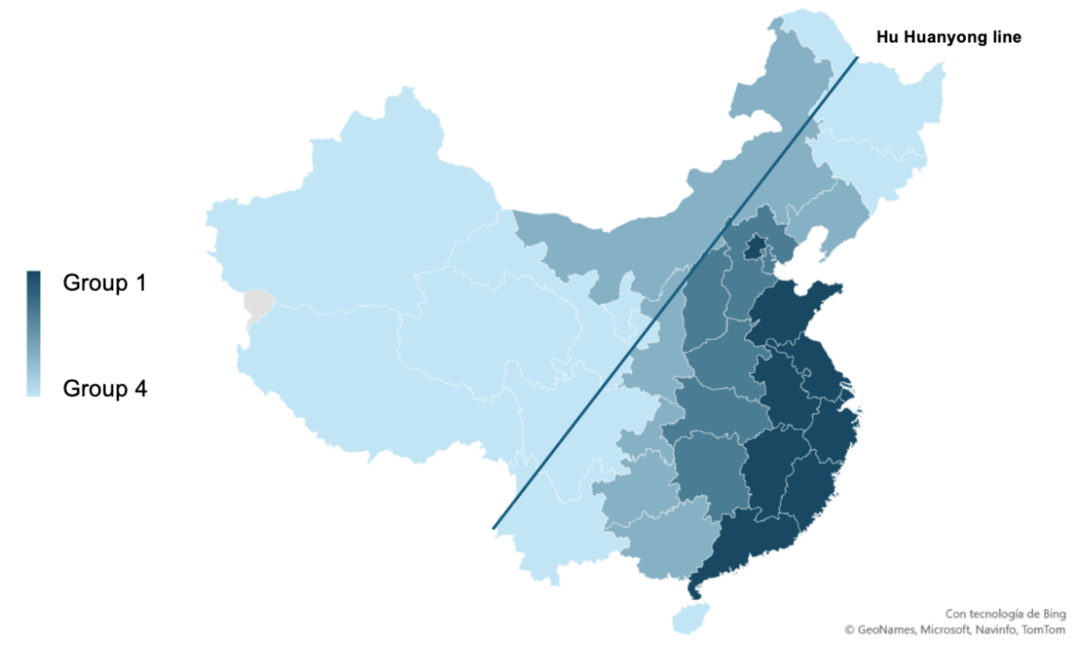
In panel B of Figure 2.5, it is noteworthy that the chosen weighting function significantly affects the distribution of the $WSISCI(x_i)$. As the exponent g increases, both the median of the indicators and their dispersion decrease.

However, the behaviour of the indicators varies across the territory due to the high concentration of activity in China. Therefore, for each Chinese province, the distribution of individual indicators from Figure 2.6. A has been replicated (see Table B.3.1 in Appendix B.3). Based on the shape of these curves, four groups have been defined: (i) provinces where, despite starting with a median below unity, they surpass this threshold within 500 Km (Beijing, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, and Guangdong); (ii) those that follow the general pattern, with the median reaching the unit value between 500 and 1,500 Km but not exceeding this reference (Tianjin, Hebei, Shanxi, Henan, Hubei, and Hunan); (iii) a transitional group between the second and fourth categories (Inner Mongolia, Liaoning, Guangxi, Chongqing, Guizhou, and Shaanxi); and (iv) provinces where the median converges to unity only after 2,000 Km and generally remain well below unity and the general pattern (Jilin, Heilongjiang, Hainan, Sichuan, Yunnan, Tibet, Gansu, Qinghai, Xinjiang, and Ningxia).

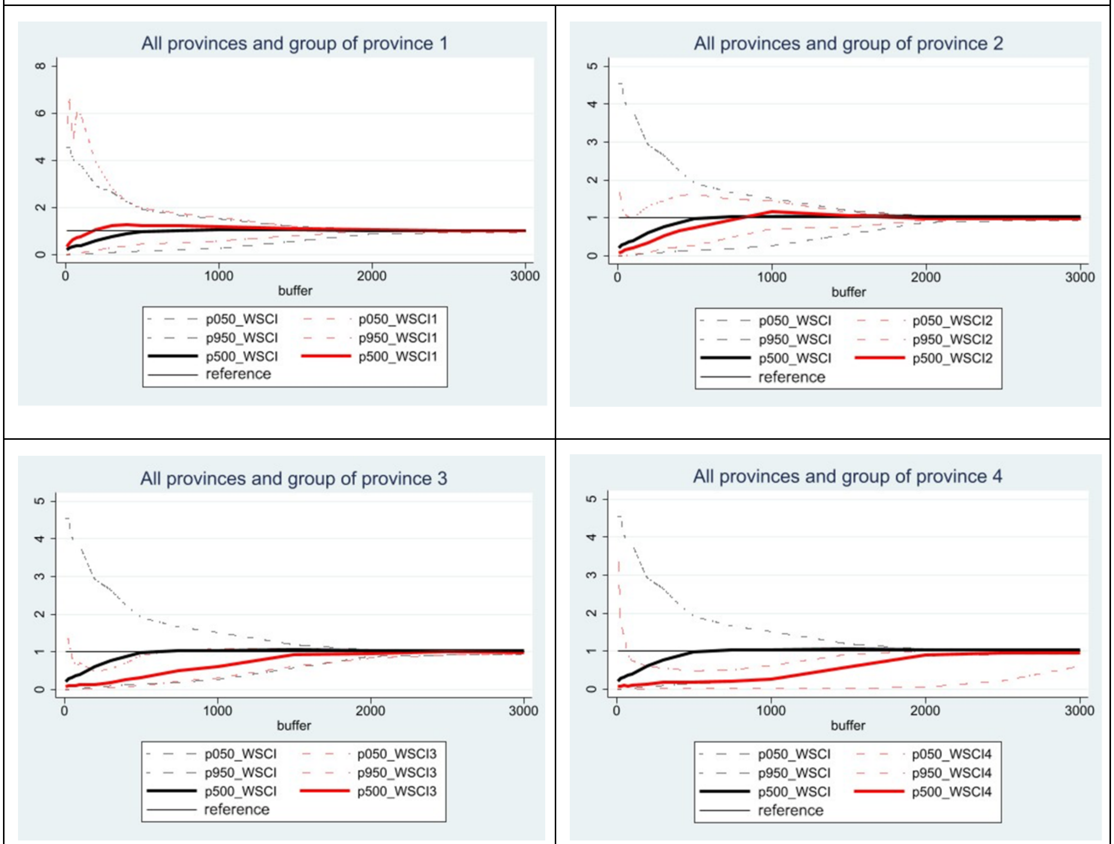
Panel A in Figure 2.6 presents a map of the four provincial groups, while Panel B of Figures 2.6 display the distribution of the W$SCI(x_i, r)$ for each group. As can be observed, the provinces located along the East Coast, except for those in the Northeast, show the highest levels of concentration. Beyond this area, the groups form concentric zones around the high-concentration region. Consequently, the first three groups of provinces are almost entirely located east of the Hu-Huanyong Line. West of this line, all provinces exhibit lower concentration behaviours.

Figure 2.6 The distribution of individual indicators

A. Aggregation of Chinese provinces based on the distribution of individual spatial concentration indicators



B. Distribution of the distance-based point-level indicators for sectoral agglomeration across different aggregations of Chinese provinces



2.5.2. Results at sectoral level

2.5.2.1. Sectoral agglomeration curves

The individual Weighted Spatial Concentration Indicators, $WSCI(x_i, r)$, can be aggregated to derive the Weighted Sectoral Spatial Concentration Indicators, $WSSCI(r)$, as shown in expression [7a]. By arranging these indicators according to buffer size, sectoral agglomeration curves can be plotted. These curves for the 31 considered activities are presented in Table B.3.2 of Appendix B.3. Table 2.3 displays the distribution of these curves following the typology described in the previous chapter, both for the entire country and for the four provincial groups.

Table 2.3 shows that most firms are in sectors exhibiting stable agglomeration behaviours, i.e., these curves do not cross the unit line, regardless of the buffer considered (19 sectors). Additionally, 7 sectors display agglomerated behaviours, while 15 show dispersed patterns. This same analysis, when applied to the provincial groups defined in the previous section, reveals that dispersed behaviour increases as we move westward. This trend is logical since most industrial activity is concentrated on the East Coast, and the indicators are constructed based on neighbour information from the entire economy. Only the Petroleum Manufacturing sector (sector 25) maintains an agglomerated pattern, likely due to its dependence on specific production locations where this natural resource is extracted.

Table 2.3 Typology of sectoral agglomeration curves for each industry in all China and by groups of Chinese provinces

Sector	All provinces		Group 1	Group 2	Group 3	Group 4
13	5 (D)		5 (D)	8 (-)	5 (D)	5 (D)
14	5 (D)		2 (D)	5 (D)	5 (D)	7 (D)
15	5 (D)		5 (D)	10 (D)	2 (D)	5 (D)
16	6 (A)		9 (-)	6 (A)	5 (D)	9 (-)
17	8 (-)		8 (-)	8 (-)	5 (D)	2 (D)
18	8 (-)		8 (-)	8 (-)	5 (D)	5 (D)
19	4a (A)		4a (A)	5 (D)	2 (D)	2 (D)
20	9 (-)		9 (-)	8 (-)	5 (D)	5 (D)
21	9 (-)		8 (-)	5 (D)	5 (D)	2 (D)
22	5 (D)		10 (D)	5 (D)	5 (D)	5 (D)
23	4b (D)		8 (-)	5 (D)	2 (D)	2 (D)
24	8 (-)		8 (-)	5 (D)	5 (D)	2 (D)
25	9 (-)		9 (-)	9 (-)	9 (-)	6 (A)
26	5 (D)		8 (-)	8 (-)	5 (D)	5 (D)
27	7 (D)		7 (D)	9 (-)	2 (D)	2 (D)
28	4a (A)		4a (A)	9 (-)	5 (D)	5 (D)
29	8 (-)		8 (-)	5 (D)	5 (D)	2 (D)
30	2 (D)		7 (D)	8 (-)	5 (D)	5 (D)
31	8 (-)		8 (-)	8 (-)	2 (D)	5 (D)
32	5 (D)		5 (D)	5 (D)	5 (D)	2 (D)
33	5 (D)		8 (-)	5 (D)	5 (D)	2 (D)
34	5 (D)		8 (-)	5 (D)	2 (D)	2 (D)
35	5 (D)		8 (-)	9 (-)	2 (D)	2 (D)
36	7 (D)		3a (A)	7 (D)	2 (D)	7 (D)
37	1 (A)		1 (A)	9 (-)	2 (D)	2 (D)
38	8 (-)		3a (A)	5 (D)	2 (D)	2 (D)
39	3a (A)		3a (A)	9 (-)	2 (D)	2 (D)
40	3a (A)		3a (A)	9 (-)	2 (D)	2 (D)
41	5 (D)		10 (D)	5 (D)	5 (D)	5 (D)
42	5 (D)		5 (D)	4b (D)	5 (D)	2 (D)
43	1 (A)		8 (-)	1 (A)	5 (D)	2 (D)

Table 2.3. Typology of sectoral agglomeration curves for each industry in all China and by groups of Chinese provinces (cont.)

Number of sectors in each type of agglomeration curve						
Type	**Number**		**Number**	**Number**	**Number**	**Number**
1 (A)	2		1	1	-	-
2 (D)	1		1	-	12	7
3 (A)	2		3	-	-	-
4a (A)	2		2	-	-	-
4b (D)	1		-	1	-	-
5 (D)	11		4	12	18	10
6 (A)	1		-	1	-	1
7 (D)	2		2	1	-	2
8 (-)	6		12	7	-	-
9 (-)	3		4	7	1	1
10 (D)	-		2	1	-	-

Types of sectoral agglomeration curves:

	Do not cross the unit	**Do cross the unit**
Decreasing or L	1	6
U or V	2	7
Inverse U or V	3	8
Double U and V	4	9
Increasing or inverse L	5	10

(A): Agglomeration behaviour; (D) Dispersion behaviour.

2.5.2.2. Aggregated indicators

Sectoral agglomeration curves may not clearly indicate whether a specific activity is agglomerated or not, especially in cases where the curves cross the unit value, as this assessment can depend on the chosen analysis distance. Even for curves that do not cross this threshold, no single indicator fully captures the degree of concentration or dispersion in production. Therefore, aggregated indicators are suggested as a simplified summary, condensing the curve information for practical use in empirical and econometric analyses. To derive these aggregated sectoral agglomeration indicators, two alternative methods are available. One approach is to obtain the Weighted Synthetic Sectoral Spatial Concentration Indicators (*WSSSCI*) by aggregating the Weighted Global Agglomeration Indicators at the point level, following expression [8a]. These indicators result from a double weighted average and do not distinguish between first-order concentration (joint location) and second-order concentration (colocation). The first column of Table B.3.3 in Appendix B.3 provides the *WSSSCI* results for each manufacturing sector of China at the 2-digit disaggregation level.

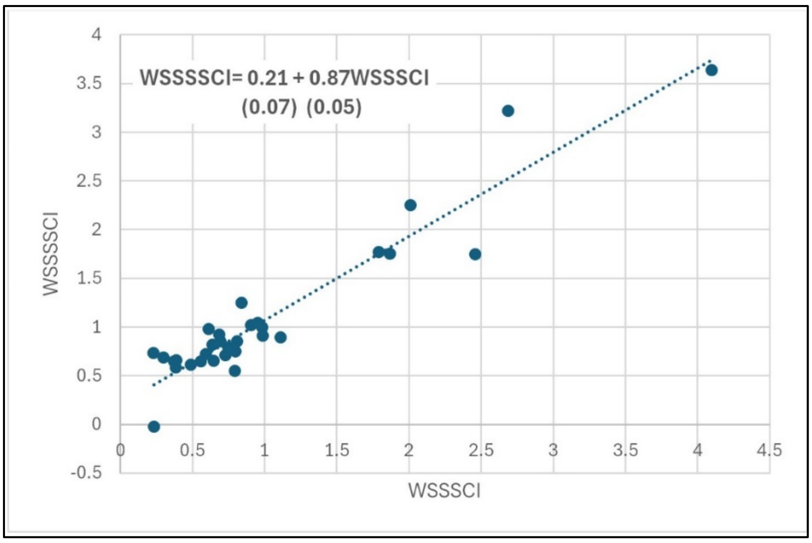
The other method involves estimating expression [10a], where individual spatial concentration indicators are regressed against sector and prefecture-level division dummies. This estimation is weighted by the product of the function $f(r)$ and the size of each firm in terms of employment. To obtain the Weighted Statistical Synthetic Sectoral Spatial Concentration Indicators (*WSSSSCI*), the coefficient obtained in the corresponding industry dummy, the constant term, and the weighted average of prefecture-level division dummies are summed. The results are shown in the second column of Table B.3.3 in Appendix B.3[39]. There is a significant theoretical difference between *WSSSCI* and *WSSSSCI* indicators. While *WSSSCI* represents a simple average of all data, *WSSSSCI* provides sectoral indicators that remove location bias. Consequently, it solely captures second-order concentration by fully excluding first-

[39] It is also possible to obtain these indicators by estimating expression [10b], starting with the Weighted Synthetic Individual Spatial Concentration Indicators, in this case weighting only by each firm's size. The results are equivalent, with only changes in significance levels. For simplicity, only the results from the weighted estimation of equation [10a] are presented.

order concentration. This is achieved by eliminating the "generic joint location" aspect from expression [5] and controlling for "local joint location" through the addition of prefecture-level dummies. In essence, if a sector has a greater presence in a prefecture-level division due to its relative location within the city or due to unique characteristics that enhance economic activity and generate an attraction effect, this effect is neutralised in the calculation of the estimated indicator.

Despite their theoretical differences, these two types of indicators show certain similarities, except in specific industries (see Figure 2.7). For example, sectors where the *WSSSSCI* values are lower (higher) than the *WSSSCI* values are those that cluster a large number of firms in areas of high (low) manufacturing concentration. Therefore, by eliminating or controlling for first-order concentration, the agglomeration indicator decreases (increases). Among the sectors where the statistical indicator shows a substantial increase compared to the average-based indicator are Manufacture of Agricultural Products (13), Raw Chemical Materials (26), and Repair of Machinery and Equipment (43). Conversely, sectors such as Manufacture of Stationery and Educational Products (24) and Computer and Communications (39) show reduced values. The former sectors are more geographically concentrated in areas with lower manufacturing concentration, while the latter display the opposite pattern.

Figure 2.7 Relationship between global sectoral agglomeration indicators (*WSSSCI* vs *WSSSSCI*)



Note: SSSCI refers to Synthetic Sectoral Spatial Concentration Indicators; SSSSCI refers to Statistical Synthetic Sectoral Spatial Concentration Indicators. Standard errors are provided in parentheses.

Table 2.4 categorises the manufacturing sector into four groups based on the value of *WSSSSCI*: High-agglomerated (*WSSSSCI* ≥ 2), Agglomerated ($1 \leq$ *WSSSSCI* < 2), Dispersed ($0.75 \leq$ *WSSSSCI* < 1), and Highly-dispersed (*WSSSSCI* < 0.75)[40]. As the developed indicators show the spatial concentration relative to the entire Chinese manufacturing activity, both the results obtained and this sectoral classification should be interpreted accordingly. The sectors classified as agglomerated and highly-agglomerated generally fall into three categories: those dependent on natural resources, those with high capital intensity, and some high-technology sectors. The concentration of natural resource-dependent and capital-intensive sectors is justified by their proximity to abundant resource-rich areas. High capital intensity is also typical of industries with significant entry barriers, resulting in a relatively low number of firms. The geographical concentration of technology-intensive sectors stems from the nature of agglomeration economies, such as knowledge spillovers, specialised labour markets, and intermediate input markets, as well as from Chinese policies supporting these industries and encouraging firms to establish in Development Zones.

Two of the findings are particularly surprising and novel in the Chinese context. The first is that certain technology sectors display a relative dispersion, which is challenging to explain except by considering the vast size of this sector in China and the specific policies developed for some of these industries. These policies may have attracted firms to development zones spread across the country, resulting in this relative dispersion. The second finding, while more aligned with theory, contradicts previous evidence. A relative dispersion is observed in labour-intensive sectors, whereas earlier studies found a trend of relative concentration in labour-intensive industries (Fan & Scott, 2003; He & Zhu, 2009; Li et al., 2015; He et al., 2022, among others).

[40] The values chosen for sector classification are determined on an ad-hoc basis.

Table 2.4 Classification of China's manufacturing industries according to their intensity of spatial concentration

High-agglomerated industries	Agglomerated industries	Dispersed industries	High-dispersed industries
16. Tobacco Products	19. Leather, Furs, Down and Related Products	14. Food Production	13. Agricultural and Sideline Food Processing
37. Railway, Shipbuilding, Air and Spacecraft, and Other Transport Equipment	20. Wood Processing, Bamboo, Rattan, Palm Fibre and Straw Products, except Furniture	17. Textiles Industry	15. Beverage Production
43. Repair of Fabricated Metal Products, Machinery and Equipment	25. Petroleum processing, coking Products, and Gas Production	18. Wearing Apparel and Other Related Products, except Fur Apparel	22. Papermaking and Paper Products
	28. Chemical Fibres	21. Furniture Manufacturing	23. Printing and Reproduction of Recorded Media
	36. Motor Vehicles	27. Medical and Pharmaceutical Products	24. Stationery, Educational and Sports Goods
	39. Computer, Communications and Other Electric Equipment	30. Non-metallic Mineral Products	26. Raw Chemical Materials and Chemical Products
		31. Smelting and Pressing of Ferrous Metals	29. Rubber and Plastic Products
		38. Electrical Equipment and Machinery	32. Smelting and Pressing of Non-ferrous Metals
		40. Measuring, Testing, Navigating and Control Equipment; Watches and Clocks	33. Metal Products, except Machinery and Equipment
			34. General Machinery and Equipment
			35. Specialized Machinery and Equipment
			41. Other Manufacturing n.e.c.
			42. Treatment and Disposal of Metal and Non-Metal Waste Materials

Note: High-agglomerated industries: WSSSSCI ≥ 2; Agglomerated industries: 1 ≤ WSSSSCI < 2; Dispersed sectors: 0.75 ≤ WSSSSCI < 1; High-dispersed sectors: WSSSSCI < 0.75.

2.5.3. Results at prefecture-level divisions

2.5.3.1. Agglomeration curves at prefecture-level divisions

Similar to the construction of sectoral agglomeration curves, agglomeration indicators from a geographical perspective can be computed by aggregating point-level indicators. These aggregated indicators can be calculated for firms within an administrative division (in this case, prefecture-level divisions). According to expression [7b], Weighted Geographical Spatial Concentration Indicators ($WGSCI(r)$) are defined for each administrative division in China and for each buffer radius. Sorting these indicators by buffer radius allows for the creation of geographic concentration curves. In this context, agglomeration curves indicate the extent to which production points within an administrative division attract other production points of the same activity within a given radius. In this indicator, the firms within the geographic area (whose individual indicators are averaged) belong to various sectors, but the counted neighbouring firms are of the same sector as the firm located in the administrative division. The curves for each prefecture-level division in China are presented in Table B.3.4 of Appendix B.3.

The types of curves resemble those observed in the sectoral case, though their distribution varies. As shown in Table 2.5, the majority of curves display a clear dispersion pattern (256 administrative divisions). This suggests that the agglomeration process is concentrated in a limited number of prefecture-level divisions within China.

Table 2.5 Classification of sectoral agglomeration curves for prefecture-level divisions obtained with Manufacture firms in China

	Decreasing or L	U or V	Inverse U or V	Double U and V	Increasing or inverse L
Do not cross the unit	Type 1(A)	Type 2(D)	Type 3(A)	Type 4(D)	Type 5(D)
	1 PLADs	51 PLADs	6 PLADs	20 PLADs	113 PLADs
Do cross the unit	Type 6(A)	Type 7(D)	Type 8	Type 9	Type 10(D)
	2 PLADs	6 PLADs	85 PLADs	18 PLADs	66 PLADs

Note: PLADs refers to Prefecture-level administrative divisions.

2.5.3.2. Aggregated indicators at prefecture-level administrative divisions

In a manner symmetrical to the sectoral perspective, the Weighted Synthetic Geographical Spatial Concentration Indicators (*WSGSCI*) are calculated following expression [8b], and the Weighted Statistical Synthetic Sectoral Spatial Concentration Indicators (*WSSSSCI*) are obtained from estimating expressions [10a] and [10b]. Both indicators possess properties similar to those of sectoral indicators but are interpreted from a geographical perspective. The results of both indicators can be found in Table B.3.5 of Appendix B.3.

The comparison of both indicators (Figure 2.8) reveals a noteworthy finding. Although the second indicator (the statistical one) removes the sectoral composition bias within each prefecture, both indicators show certain similarities, though these are less pronounced for prefectures with values below one. The results of the Weighted Statistical Synthetic Geographical Spatial Indicators align closely with findings from previous research (Wen, 2004; He et al., 2007; He and Wang, 2010, among others). The spatial distribution map (Figure 2.9) provides a more intuitive view, showing that, overall, the eastern and southern administrative divisions exhibit a stronger capacity for agglomeration compared to the south-western regions. This geographical pattern is particularly prominent in the Beijing-Tianjin-Hebei region, the Yangtze River Delta, and the Pearl River Delta[41]. Indeed, urban agglomeration areas generally demonstrate greater clustering potential, and these regions also contain a higher density of development zones. It is worth noting that Jiuquan City demonstrates significant geographical clustering attractiveness. As a Satellite Launch Center, Jiuquan is an important node city within the Silk Road Economic Belt. With favorable logistics and strong external connectivity, combined with substantial policy support, Jiuquan provides notable investment incentives that drive the growth of the manufacturing industry.

[41] The Beijing-Tianjin-Hebei region includes Beijing, Tianjin, and Hebei Province. The Yangtze River Delta encompasses Shanghai, Jiangsu Province, Zhejiang Province, and Anhui Province, while the Pearl River Delta refers specifically to Guangdong Province. Hong Kong and Macao are excluded from the Pearl River Delta in this context, as manufacturing is not a primary industry in these regions.

Figure 2.8 Relationship between geographical agglomeration indicators at prefecture level (*WSSSSCI* vs *WSGSCI*)

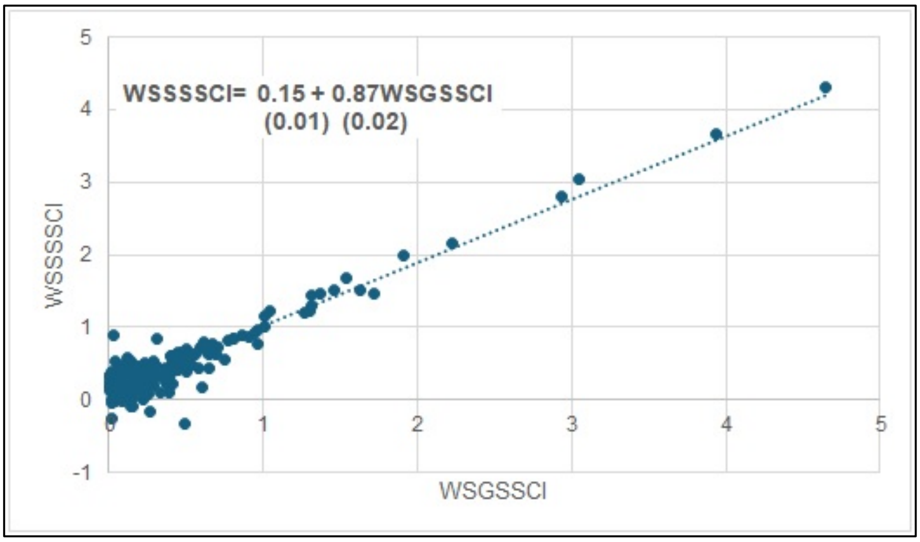
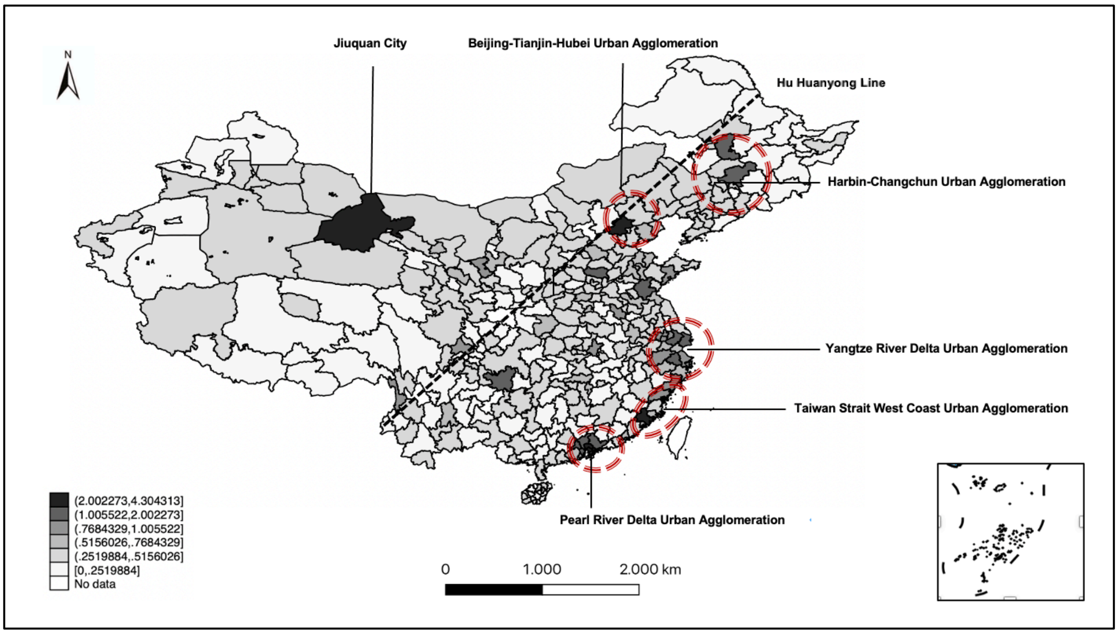


Figure 2.9 Statistical synthetic geographical spatial indicators for the prefecture-level administrative divisions of China



In the central and south-western regions, it is notable that Chongqing shows slightly lower geographical clustering attractiveness compared to two nearby cities, Zunyi in Guizhou Province and Shiyan in Hubei Province. However, as a crucial component of China’s historic industrial base and the only municipality directly under central government administration in southwest China, Chongqing holds a unique and significant place in the national manufacturing landscape. Its internal industrial structure and distribution warrant further in-depth examination. Additionally, some regions with a high density of development zones do not exhibit strong clustering attractiveness. These characteristics, as detailed in Table C.5 in Appendix C, also deserve further study as they may indicate a need for enhanced city planning to optimise resource utilisation.

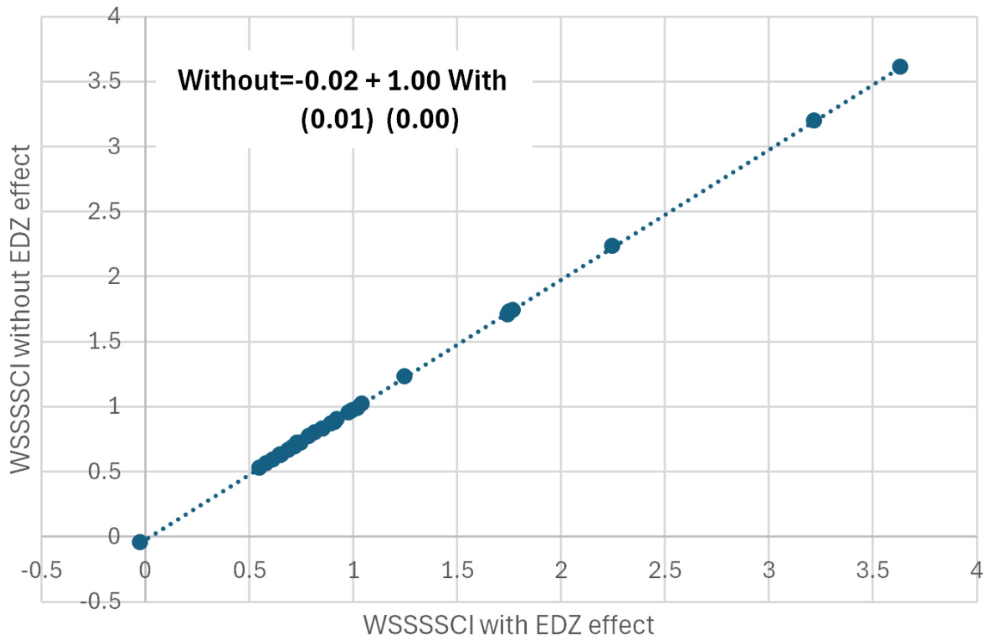
2.5.4. The effect of EDZs on the agglomeration of Chinese manufacturing

EDZs are areas where companies receive certain benefits if they choose to establish themselves, stemming from public support policies such as subsidies, tax breaks, and installation facilities. These zones primarily aim to attract businesses, particularly foreign firms, and are spread throughout the country. This widespread distribution could theoretically have a neutral effect on the overall concentration of productive activities. Specifically, as of 2018, there were over 2,500 EDZs across China. Approximately 16.7% of the valid Orbis sample of firms in China are located within an EDZ. However, this distribution varies across sectors (see Table B.3.6 in Appendix B.3), ranging from 12.4% in Wood Processing (sector 20) to nearly double, at 24.2%, for Medical and Pharmaceutical Products (sector 27). Geographically, there are even more significant variations between provinces, with the percentage of firms in EDZs spanning from 2.5% in Tibet (province 54) to 29.1% in Shanghai (province 31).

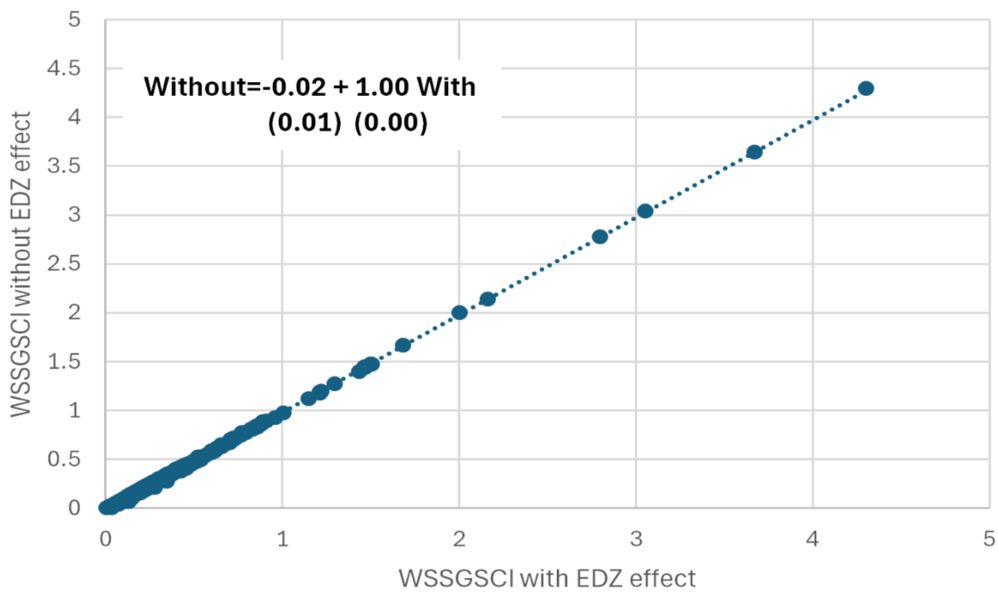
To tentatively evaluate the extent to which EDZs affect the process of manufacturing agglomeration, expression [10a] is re-estimated by incorporating a dummy variable, which takes a value of one if a company's headquarters is located within an EDZ. The coefficient associated with this dummy is 0.07106 (with a standard deviation of 0.00076), indicating a statistically positive and significant effect at the firm level on the agglomeration process, as expected. However, taking the average presence of firms in EDZs within the sample, the average effect is 0.01188, this would be the incremental increase in sectoral indicators due to EDZs. It is likely that the large number of EDZs can attract foreign firms and may even have a localised concentration effect. However, when broader buffers are considered, these effects may partially offset each other. Specifically, when recalculating the statistical indicators (*WSSSSCI and WSSGSCI*) without the EDZ effect and comparing them to those that include it, the results at the sectoral and prefecture levels remain relatively unchanged (see Figure 2.10).

Figure 2.10 Relationship between Statistical Global Indicators with and without EDZ effects

A. Weighted Statistical Synthetic Sectoral Spatial Concentration Indicators (*WSSSSCI*). 31 industries at 2-digits



B. Weighted Statistical Synthetic Geographical Spatial Concentration Indicators (*WSSGSCI*). 268 prefectures



2.6. Conlusions

In this chapter, a weighted version of the distance-based point-level sectoral agglomeration indicators is introduced to measure industrial agglomeration. The construction of the unweighted indicator, as presented in the previous chapter, allows for transformation into its weighted form. This transformation is not trivial and requires reformulation of the entire calculation process at each step. The resulting weighted indicator retains all the desirable properties of the unweighted indicator from which it is derived.

From an empirical perspective, this research is the first to use an indicator of these characteristics to measure the geographic concentration of Chinese manufacturing without any limitations, beyond those imposed by the availability of required data. Previous studies on this issue in China were typically restricted to specific sectors, regions, or types of firms to significantly reduce the number of production points, making this type of analysis feasible. The computational ease of this indicator and its associated statistical measures supports expanding the sample of production points.

The empirical analysis conducted for Chinese manufacturing reveals at least five interesting findings:

- Geographic concentration is primarily observed along the country's East Coast, gradually dissipating toward the West and North.
- Productive agglomeration is concentrated in a few sectors and geographic areas (prefectures). This result is partly driven by the significant variation in concentration processes across activities and, more importantly, across prefectures. Additionally, the indicators are measured relative to the average concentration level.
- Manufacturing activities related to natural resources, capital-intensive industries, and certain technology sectors exhibit higher degrees of spatial concentration, while labour-intensive sectors display relative dispersion.
- Prefectures with significant urban and metropolitan profiles have a greater

ability to attract manufacturing activity overall.

- Finally, it appears confirmed that EDZs do not significantly distort the overall concentration process of Chinese manufacturing. This outcome is explained by the extensive geographic distribution of EDZs across China, resulting in localised effects rather than global ones. Their limited impact is more noticeable in geographic terms, i.e., among prefectures, than among sectors, at least at the level of disaggregation used in this chapter.

While this chapter demonstrates that the proposed family of distance-based indicators for sectoral agglomeration can be adapted to a measure weighted by the activity size at each point, there are many improvements that should continue to be incorporated for better measurement, which are beyond the scope of this study.

From a theoretical perspective, although the developed indicators isolate (eliminate) the two aspects of first-order concentration and enable the measurement of second-order concentration or colocation, they may still exaggerate the concentration. This measurement error may stem from calculating simple or weighted averages based on individual indicators that might be affected by neighbouring points responsible for the concentration. Thus, the individual indicators could incorporate the concentration from the causative point into the neighbouring points, potentially exaggerating it. However, this effect might be offset in the calculation of aggregate indicators through normalisation against the mean. Therefore, a method based on identifying the chain of points responsible for productive agglomeration could provide a more accurate measurement.

From an empirical perspective, though with significant theoretical implications, the analysis presented here does not cover other weighting components within the neighbour count discussed in the first chapter. Some of these were omitted due to the lack of additional information, while others were excluded because their implementation would be too complex given available resources. Including these components would yield more accurate measures.

In any case, improving the initial dataset, such as having information at the establishment level rather than at the firm level or incorporating terrestrial transportation networks to define realistic buffers, could significantly enhance the results and provide more precise answers to the empirical questions posed in this chapter.

Appendix B.1. Sectoral classification

Industrial Classification for National Economic Activities (GB/T 4754-2017) structure and correspondences with International Standard Industrial Classification (ISIC Rev.4)

The latest edition of China's Industrial Classification for National Economic Activities (GB/T 4754-2017) continues to align closely with international standards in classification principles and methods. It is essentially identical to the International Standard Industrial Classification (ISIC Rev.4) published by the United Nations, allowing for mutual conversion. Since data collection, statistics, and descriptions in this chapter are all based on the GB/T 4754-2017 codes, a correspondence conversion has been conducted between the two classifications. This approach ensures that this research results align with international standards and remain comparable with previous studies in China.

Table B.1.1: GB/T 4754-2017 structure and correspondences with ISIC Rev.4

GB/T 4754-2017	ISIC Rev.4
13 Manufacture of agricultural and sideline food processing	10 Manufacture of food products
14 Manufacture of food production	
15 Manufacture of beverage production	11 Manufacture of beverages
16 Manufacture of tobacco products	12 Manufacture of tobacco products
17 Manufacture of textiles	13 Manufacture of textiles
18 Manufacture of wearing apparel and other related products, except fur apparel	14 Manufacture of wearing apparel
19 Manufacture of leather, furs, down and related products, including footwear production	15 Manufacture of leather and related products
20 Manufacture of wood processing, bamboo, rattan, palm fiber and straw products, except furniture	16 Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
21 Manufacture of furniture	31 Manufacture of furniture
22 Manufacture of papermaking and paper products	17 Manufacture of paper and paper products
23 Printing and reproduction of recorded media	18 Printing and reproduction of recorded media
24 Manufacture of stationery, educational and sports goods	32 Other manufacturing
25 Manufacture of petroleum processing, coking products, and gas production	19 Manufacture of coke and refined petroleum products
26 Manufacture of raw chemical materials and chemical products	20 Manufacture of chemicals and chemical products
27 Manufacture of medical and pharmaceutical products	21 Manufacture of basic pharmaceutical products and pharmaceutical preparations
28 Manufacture of chemical fibers	20 Manufacture of chemicals and chemical products
29 Manufacture of rubber and plastic products	22 Manufacture of rubber and plastics products
30 Manufacture of non-metallic mineral products	23 Manufacture of other non-metallic mineral products
31 Manufacture of smelting and pressing of ferrous metals	24 Manufacture of basic metals
32 Manufacture of smelting and pressing of non-ferrous metals	
33 Manufacture of metal products, except machinery and equipment	25 Manufacture of fabricated metal products, except machinery and equipment
34 Manufacture of general machinery and equipment	28 Manufacture of machinery and equipment n.e.c.
35 Manufacture of specialized machinery and equipment	

Table B.1.1: GB/T 4754-2017 structure and correspondences with ISIC Rev.4(continued)

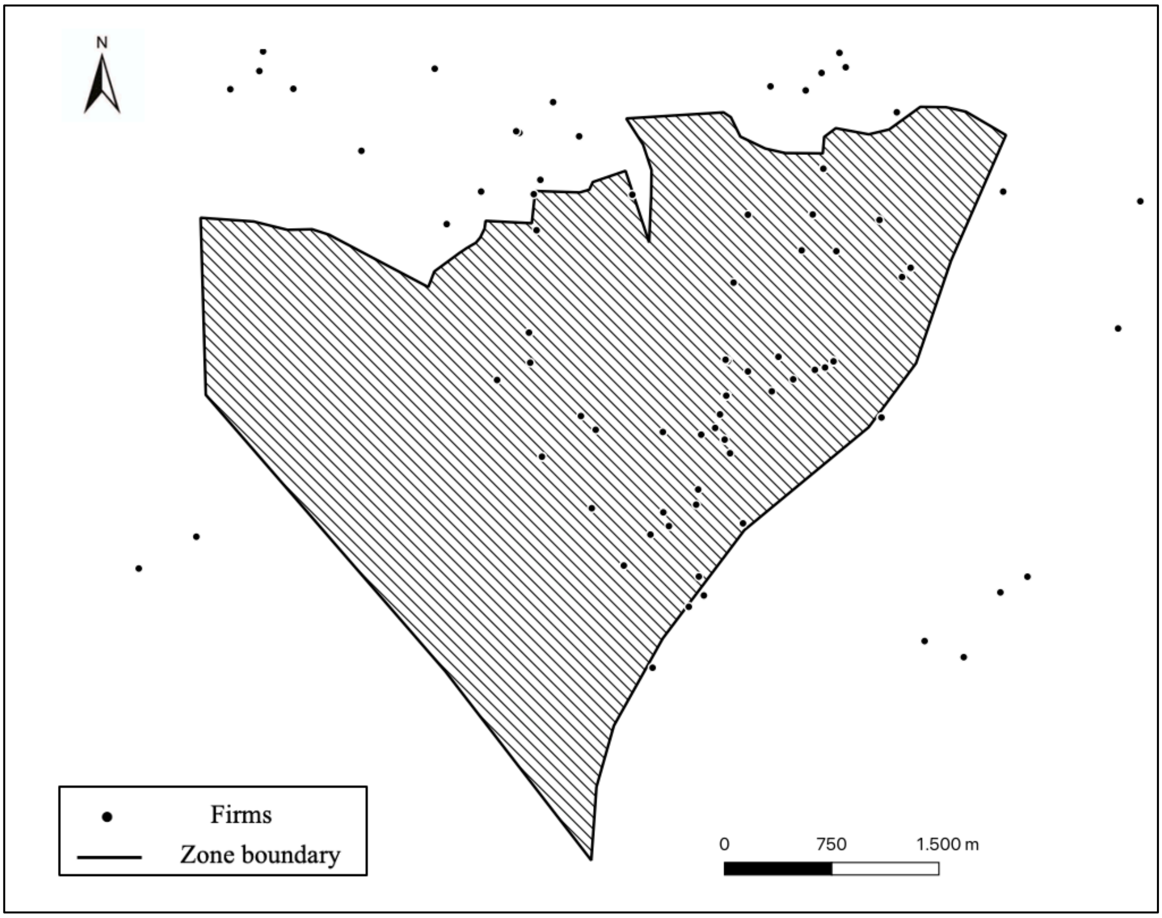
36 Manufacture of motor vehicles	29 Manufacture of motor vehicles, trailers, and semi-trailers
37 Manufacture of railway, shipbuilding, air and spacecraft, and other transport equipment	30 Manufacture of other transport equipment
38 Manufacture of electrical equipment and machinery	27 Manufacture of electrical equipment
39 Manufacture of computer, communications, and other electronic equipment	26 Manufacture of computer, electronic and optical products
40 Manufacture of measuring, testing, navigating and control equipment; watches and clocks	
41 Other manufacturing	32 Other manufacturing
42 Manufacture of treatment and disposal of metal and non-metal waste materials	38 Waste collection, treatment, and disposal activities; materials recovery
43 Repair of fabricated metal products, machinery, and equipment	33 Repair and installation of machinery and equipment

Source: Own elaboration based on GB/T 4754-2017 and ISIC Rev.4 classification.

Appendix B.2: Identifying firms located in Development Zones

Each development zone is assigned a unique administrative identification code. For instance, the code for the Guangzhou Economic and Technological Development Zone is G441034, which is similar to but distinct from the standard administrative division codes. However, it is not possible to determine if a firm is located within a development zone solely based on this code. To accurately identify firms situated in development zones, ArcGIS is used to analyse spatial relationships, marking firms located within each development zone. This process is applied across all 2,543 development zones.

Figure B.2.1. Guangzhou Economic and Technological Development Zone



Source: Own elaboration based on the actual distribution of firms within Guangzhou Economic and Technological Development Zone.

Appendix B.3: Results of agglomeration measurement

Table B.3.1. Distribution of the distance-based point-level indicators for sectoral agglomeration across Chinese provinces

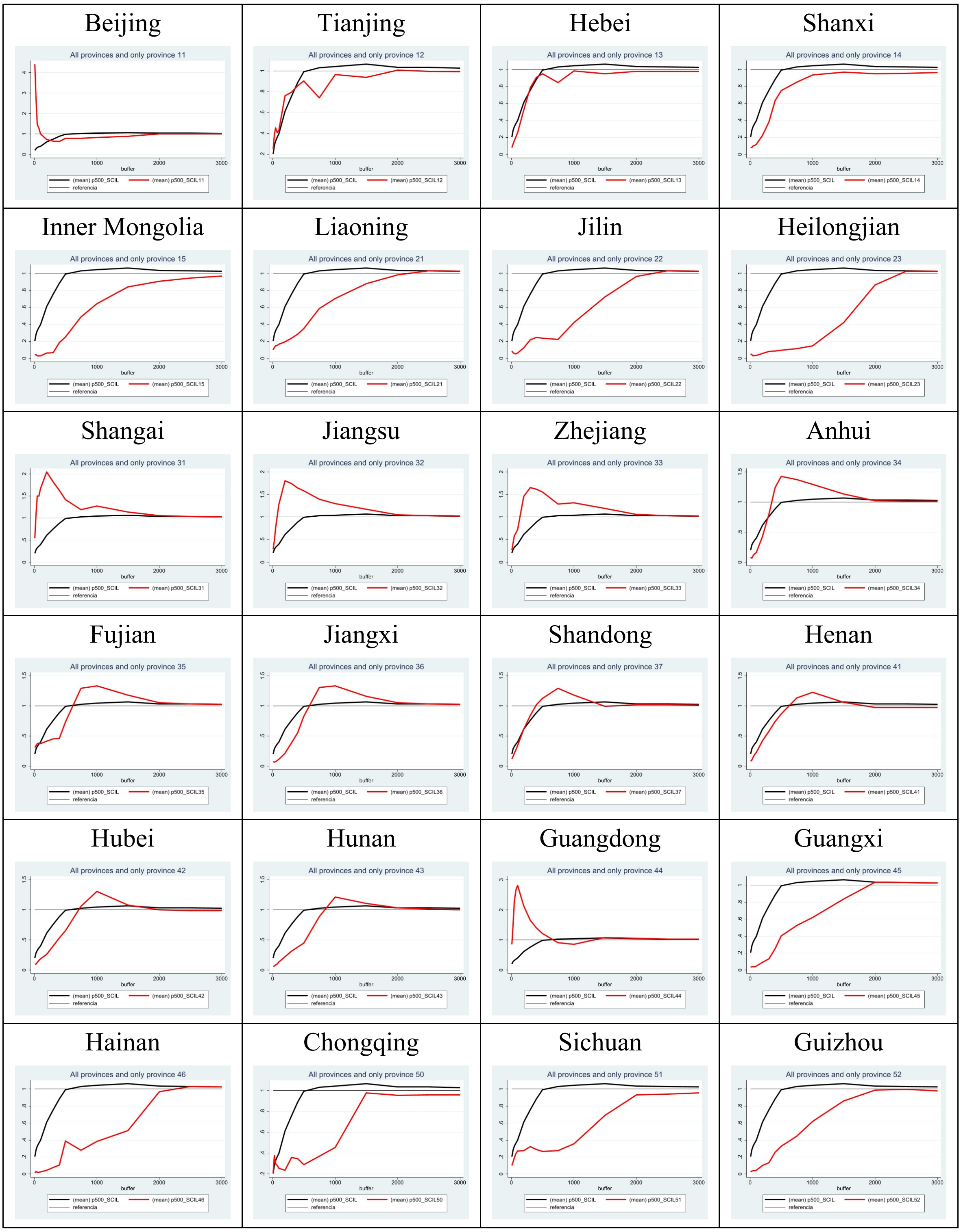
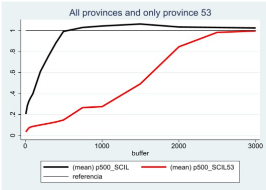
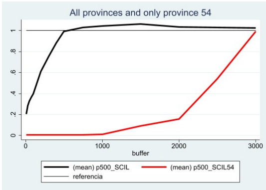
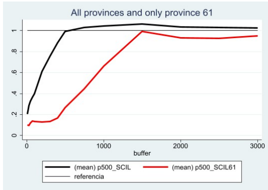
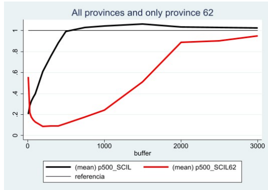
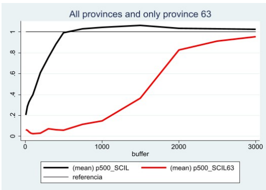
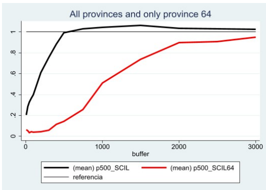
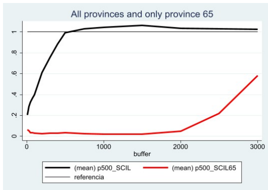


Table B.3.1. Distribution of the distance-based point-level indicators for sectoral agglomeration across Chinese provinces (continued)

<table>
<tr>
<td>Yunnan

</td>
<td>Tibet

</td>
<td>Shaanxi

</td>
<td>Gansu

</td>
</tr>
<tr>
<td>Qinghai

</td>
<td>Ningxia

</td>
<td>Xinjiang

</td>
<td></td>
</tr>
</table>

Note:X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Values for all provinces

Values only for the corresponding province

The reference value

Table B.3.2. Sectoral agglomeration curves for All-China and provincial groups

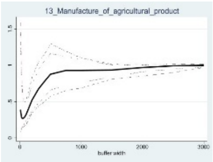
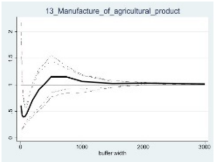
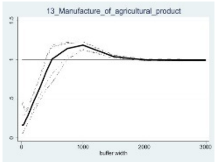
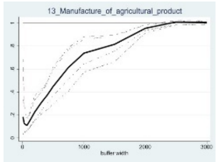
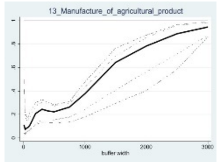
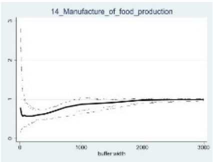
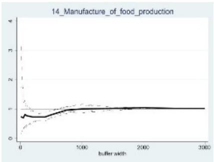
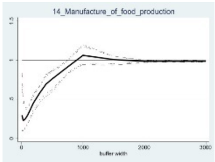
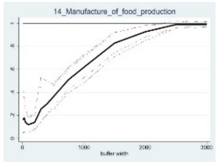
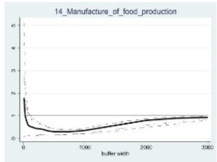
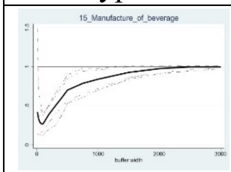
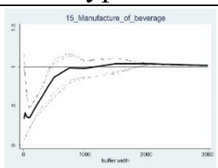
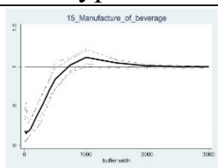
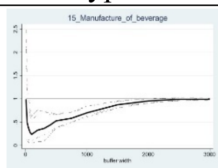
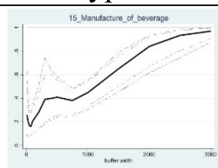
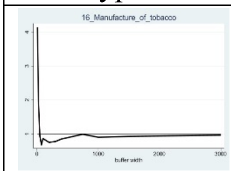
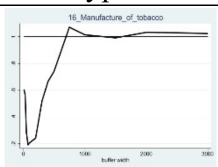
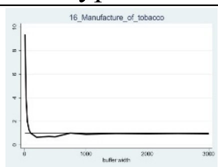
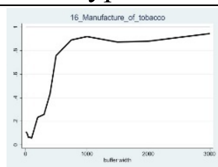
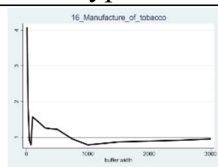
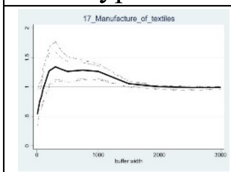
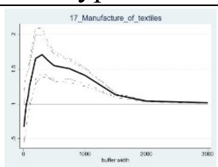
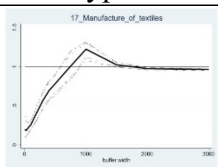
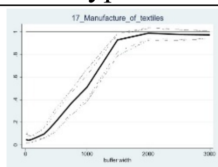
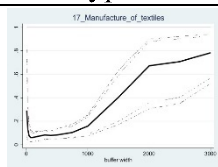
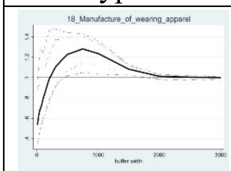
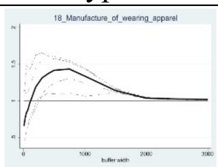
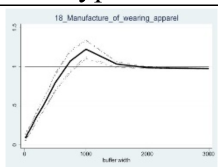
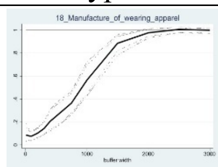
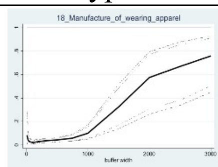
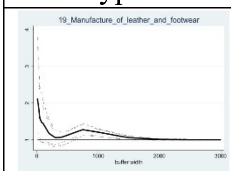
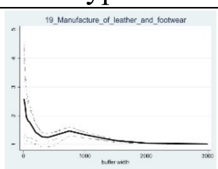
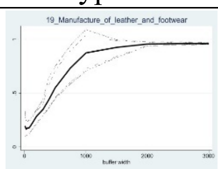
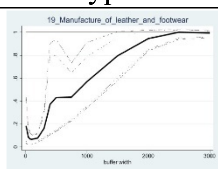
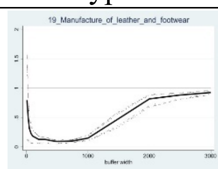
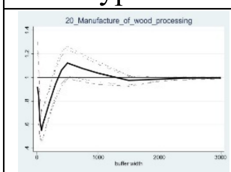
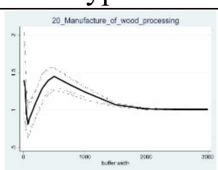
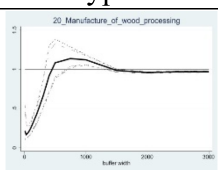
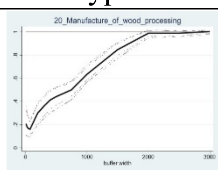
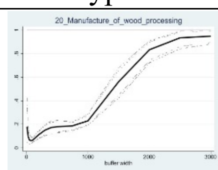
All China	Group 1	Group 2	Group 3	Group 4
 Type 5	 Type 10	 Type 8	 Type 5	 Type 5
 Type 2	 Type 2	 Type 10	 Type 5	 Type 7
 Type 5	 Type 5	 Type 10	 Type 2	 Type 4
 Type 6	 Type 10	 Type 6	 Type 5	 Type 6
 Type 8	 Type 8	 Type 10	 Type 5	 Type 2
 Type 8	 Type 8	 Type 10	 Type 5	 Type 2
 Type 1	 Type 1	 Type 5	 Type 2	 Type 2
 Type 9	 Type 9	 Type 8	 Type 5	 Type 2

Table B.3.2. Sectoral agglomeration curves for All-China and provincial groups (continued)

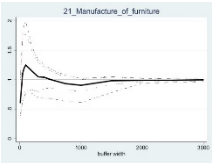
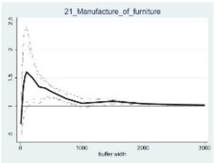
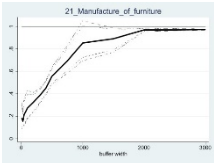
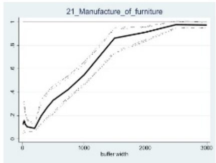
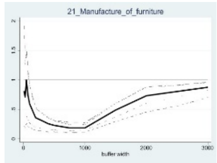
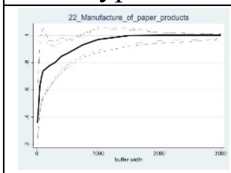
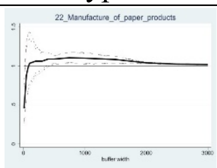
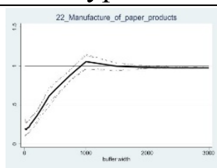
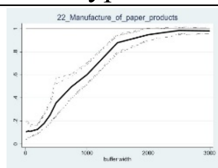
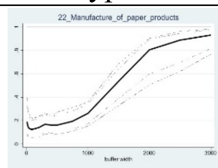
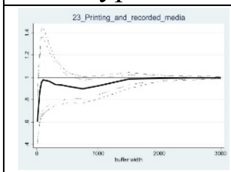
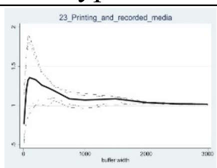
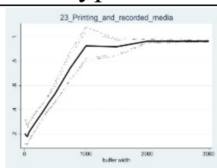
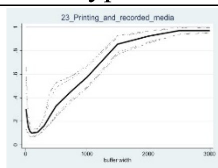
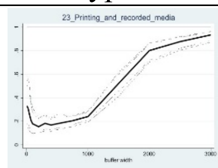
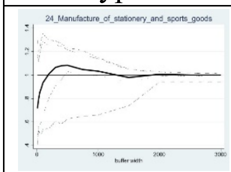
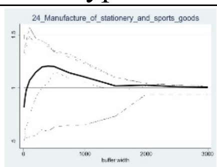
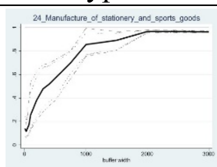
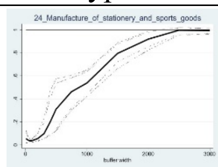
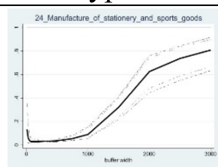
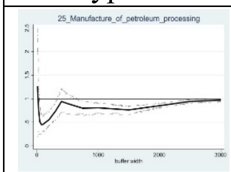
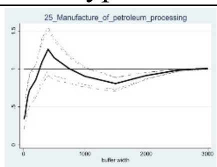
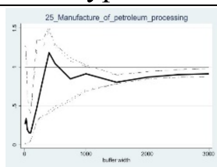
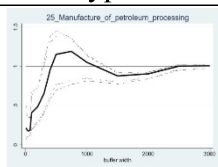
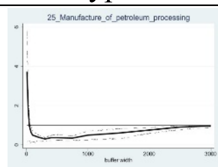
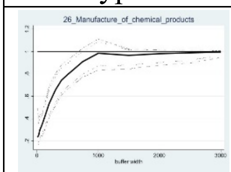
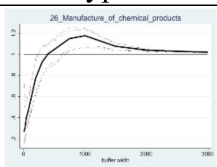
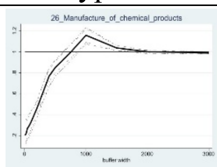
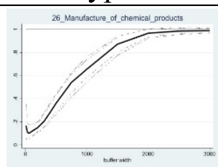
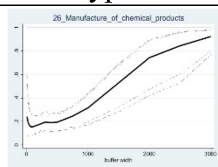
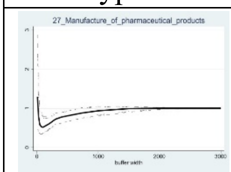
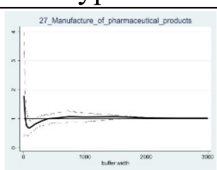
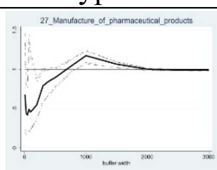
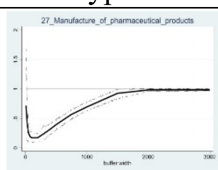
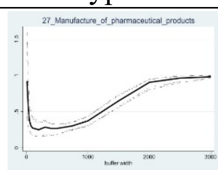
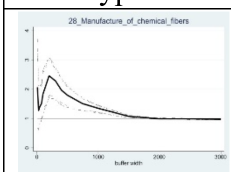
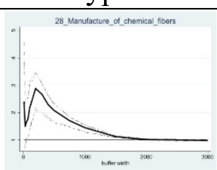
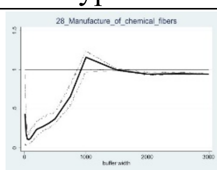
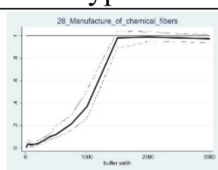
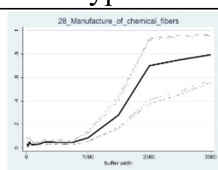
 Type 8	 Type 8	 Type 5	 Type 5	 Type 2
 Type 5	 Type 10	 Type 10	 Type 5	 Type 5
 Type 5	 Type 8	 Type 5	 Type 2	 Type 2
 Type 10	 Type 8	 Type 5	 Type 5	 Type 2
 Type 9	 Type 8	 Type 10	 Type 8	 Type 6
 Type 5	 Type 10	 Type 10	 Type 5	 Type 2
 Type 6	 Type 6	 Type 10	 Type 2	 Type 2
 Type 3	 Type 3	 Type 10	 Type 5	 Type 5

Table B.3.2. Sectoral agglomeration curves for All-China and provincial groups (continued)

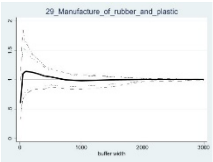
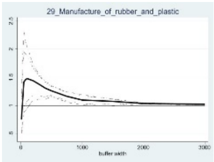
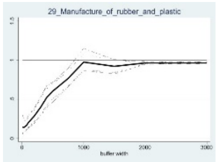
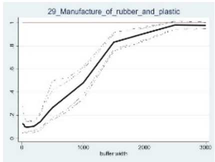
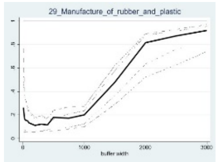
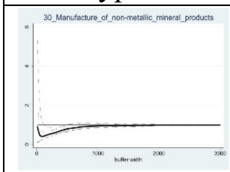
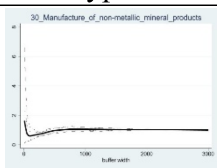
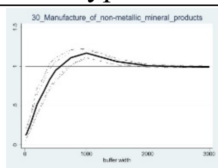
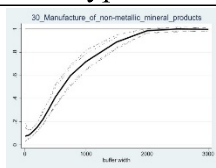
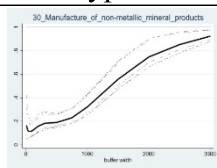
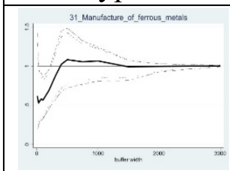
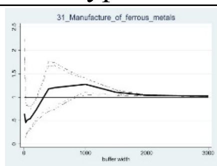
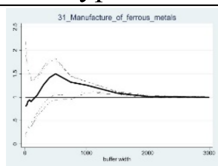
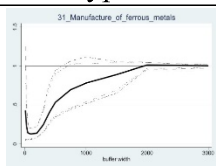
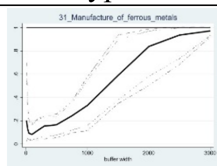
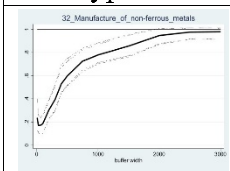
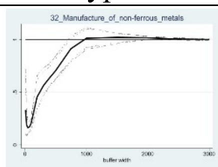
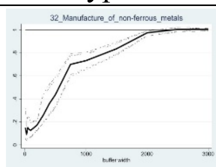
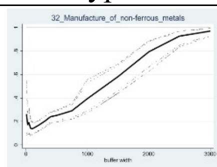
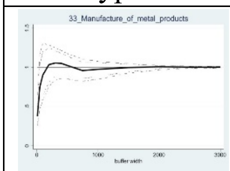
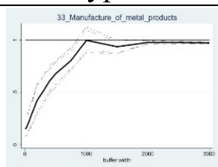
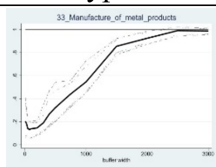
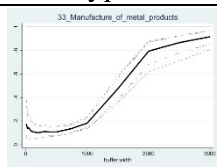
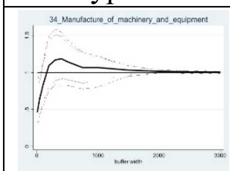
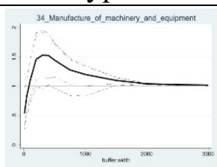
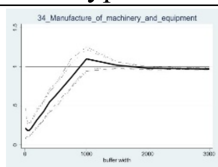
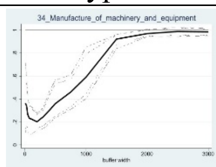
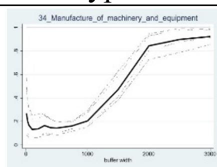
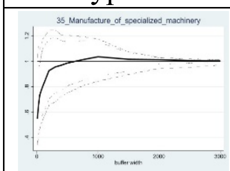
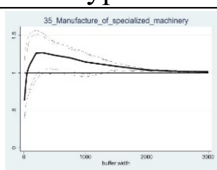
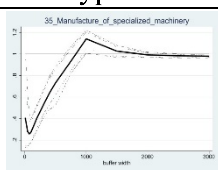
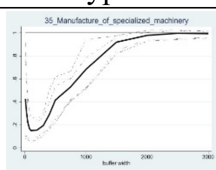
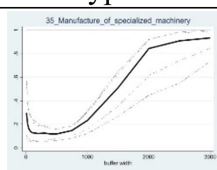
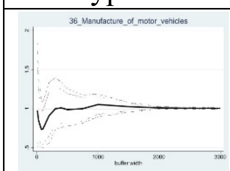
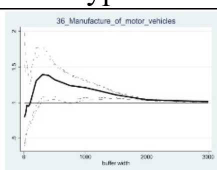
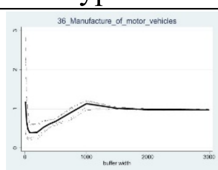
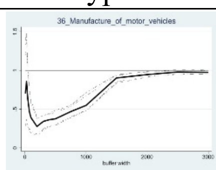
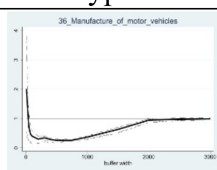
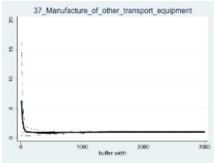
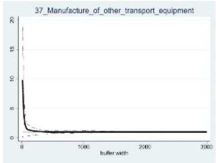
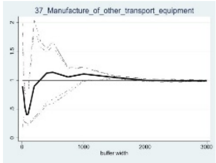
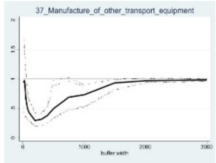
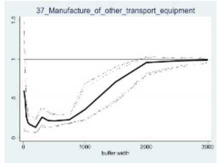
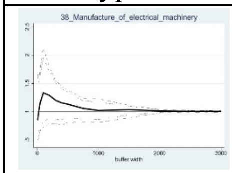
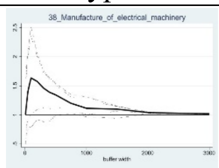
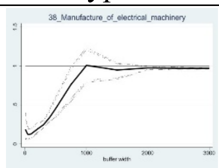
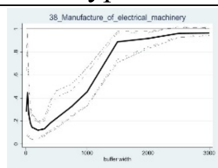
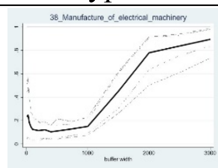
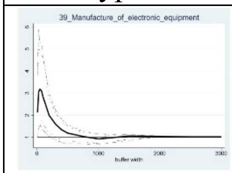
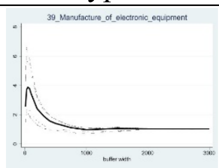
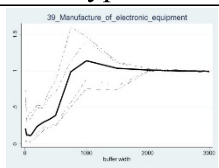
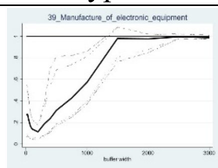
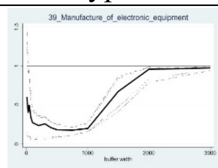
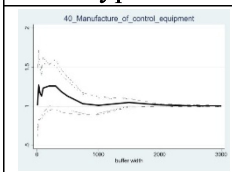
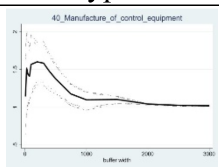
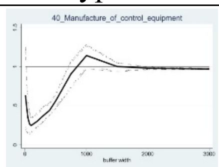
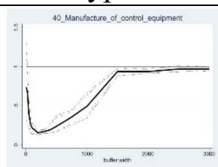
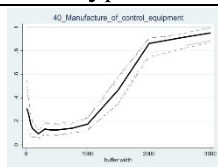
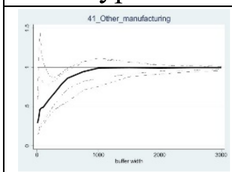
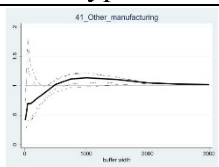
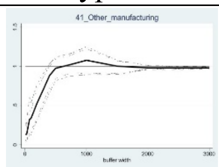
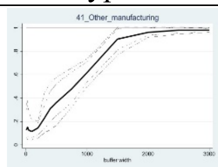
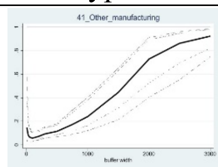
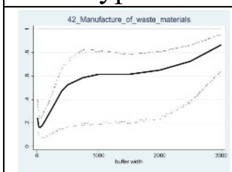
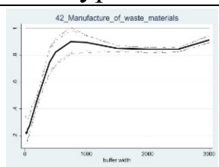
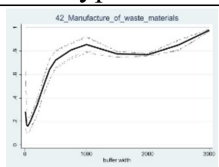
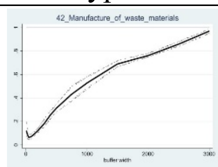
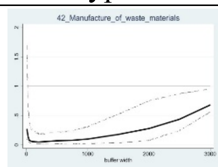
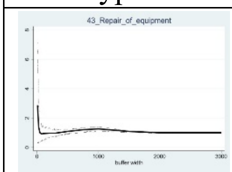
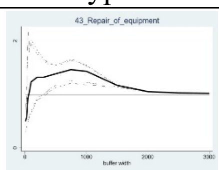
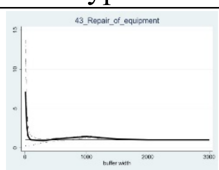
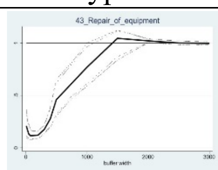
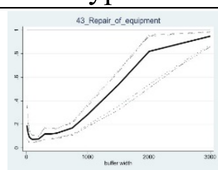
 Type 8	 Type 8	 Type 5	 Type 5	 Type 2
 Type 2	 Type 7	 Type 10	 Type 5	 Type 5
 Type 10	 Type 10	 Type 8	 Type 2	 Type 2
 Type 5	 Type 5	 Type 5	 Type 5	 Type 2
 Type 10	 Type 8	 Type 5	 Type 5	 Type 2
 Type 10	 Type 8	 Type 10	 Type 2	 Type 2
 Type 10	 Type 8	 Type 10	 Type 2	 Type 2
 Type 9	 Type 8	 Type 7	 Type 2	 Type 7

Table B.3.2. Sectoral agglomeration curves for All-China and provincial groups (continued)

				
Type 1	Type 1	Type 9	Type 2	Type 2
				
Type 8	Type 3	Type 5	Type 2	Type 2
				
Type 3	Type 3	Type 10	Type 5	Type 2
				
Type 3	Type 3	Type 10	Type 2	Type 2
				
Type 5	Type 10	Type 10	Type 5	Type 2
				
Type 5	Type 5	Type 5	Type 5	Type 2
				
Type1	Type 8	Type 1	Type 10	Type 2

Note:X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Table B.3.3. Sectoral agglomeration measures

Sector	Weighted Synthetic Sectoral Spatial Concentration Indicators *WSSSCI*	Weighted Statistical Synthetic Sectoral Spatial Concentration Indicators *WSSSSCI*	
		With EDZ	Without EDZ
13. Agricultural and Sideline Food Processing	0.37	0.65***	0.63***
14. Food Production	0.70	0.85***	0.83***
15. Beverage Production	0.39*	0.66***	0.63***
16. Tobacco Products	2.69***	3.22***	3.20***
17. Textiles Industry	0.69*	0.92***	0.90***
18. Wearing Apparel and Other Related Products, except Fur Apparel	0.64*	0.82***	0.80***
19. Leather, Furs, Down and Related Products	1.87***	1.75***	1.73***
20. Wood Processing, Bamboo, Rattan, Palm Fiber and Straw Products	0.84	1.25***	1.23***
21. Furniture Manufacturing	0.81	0.85***	0.83***
22. Papermaking and Paper Products	0.49***	0.61***	0.59***
23. Printing and Reproduction of Recorded Media	0.73***	0.71***	0.69***
24. Stationery, Educational and Sports Goods	0.80	0.55***	0.53***
25. Petroleum processing, coking Products, and Gas Production	0.95	1.04**	1.02***
26. Raw Chemical Materials and Chemical Products	0.30***	0.69***	0.66***
27. Medical and Pharmaceutical Products	0.98	1.00	0.97
28. Chemical Fibers	1.79	1.77***	1.74***
29. Rubber and Plastic Products	0.80	0.75***	0.72***
30. Non-metalic Mineral Products	0.75	0.79***	0.77***
31. Smelting and Pressing of Ferrous Metals	0.61	0.98**	0.96***
32. Smelting and Pressing of Non ferrous Metals	0.23***	-0.02***	-0.04***
33. Metal Products, except Machinery and Equipment	0.56***	0.64***	0.62***
34. General Machinery and Equipment	0.59**	0.72***	0.70***
35. Specialized Machinery and Equipment	0.65**	0.65***	0.63***
36. Motor Vehicles	0.91	1.02*	0.99
37. Railway, Shipbuilding, Air and Spacecraft, and Other Transport Eq.	4.10	3.63***	3.61***
38. Electrical Equipment and Machinery	0.99	0.91***	0.88***
39. Computer, Communications and Other Electric Equipment	2.46	1.75***	1.71***
40. Measuring, Testing, Navigating and Control Equipment	1.11	0.89***	0.87***
41. Other Manufacturing n.e.c.	0.39***	0.58***	0.56***
42. Treatment and Disposal of Metal and Non-Metal Waste Materials	0.23***	0.73***	0.72***
43. Repair of Fabricated Metal Products, Machinery and Equipment	2.01	2.25***	2.23***

*Note: *Indicates that the estimation is statistically significant at the 90% level, ** at the 95% level, and *** at the 99% level.*

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture

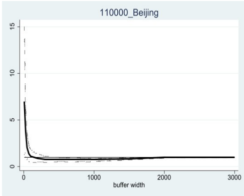
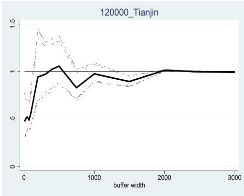
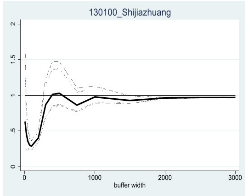
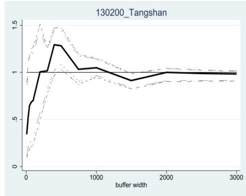
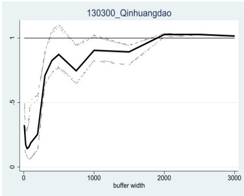
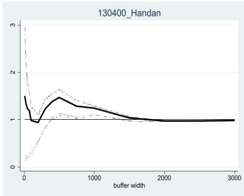
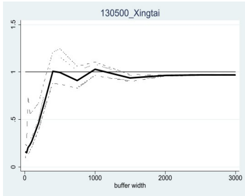
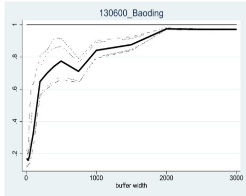
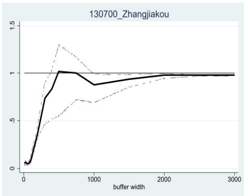
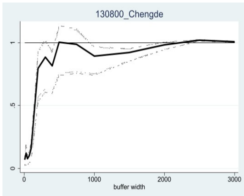
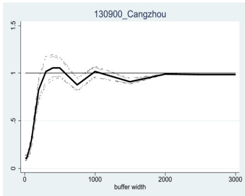
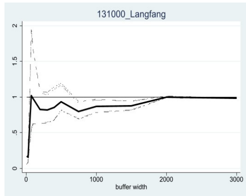
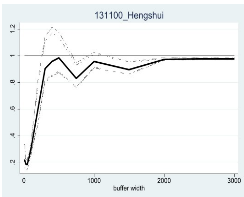
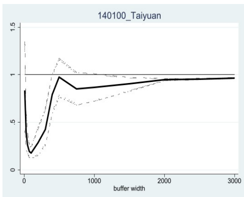
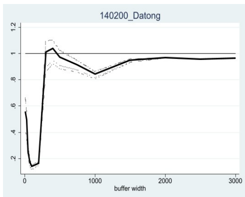
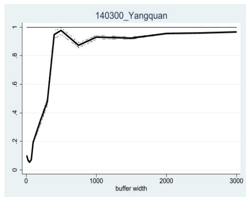
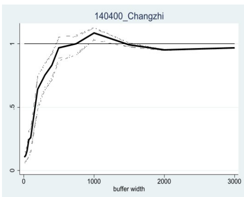
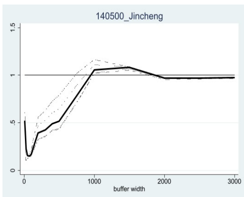
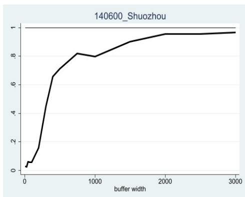
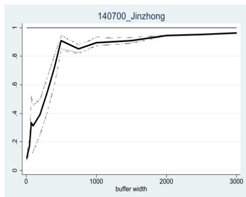
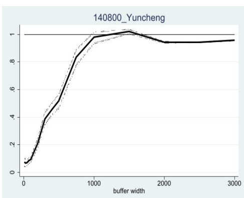
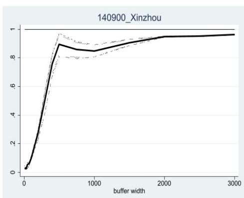
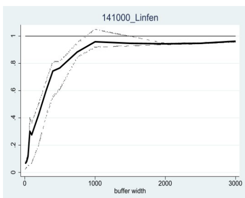
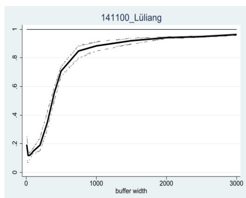
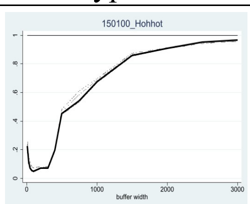
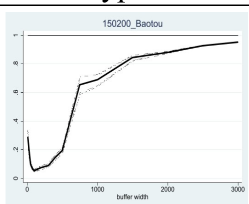
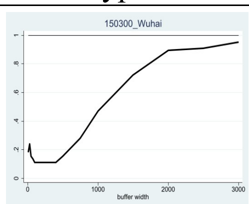
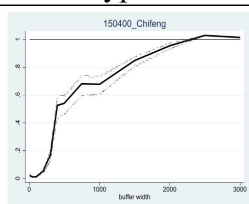
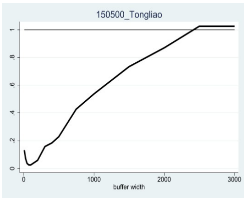
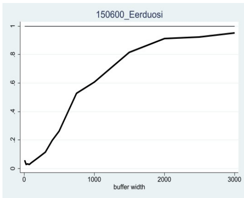
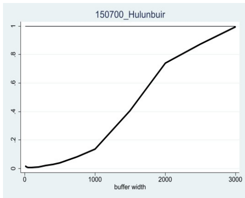
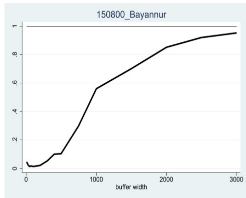
 Type 1	 Type 10	 Type 4	 Type 8
 Type 4	 Type 3	 Type 5	 Type 5
 Type 5	 Type 5	 Type 10	 Type 5
 Type 5	 Type 2	 Type 10	 Type 5
 Type 10	 Type 10	 Type 5	 Type 5
 Type 10	 Type 5	 Type 5	 Type 5
 Type 2	 Type 2	 Type 2	 Type 5

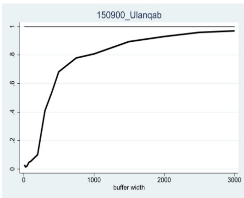
Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

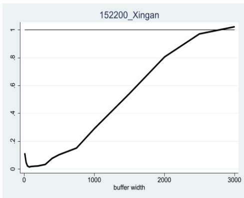
<table>
<tr>
<td>

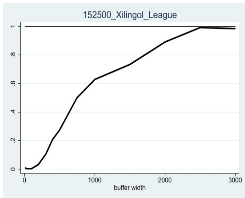
Type 5</td>
<td>

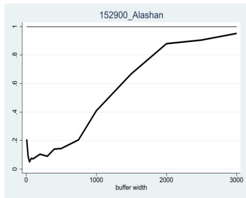
Type 5</td>
<td>

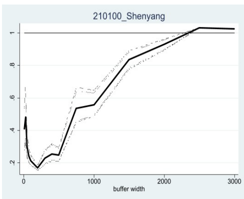
Type 5</td>
<td>

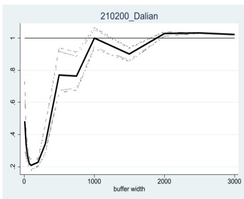
Type 5</td>
</tr>
<tr>
<td>

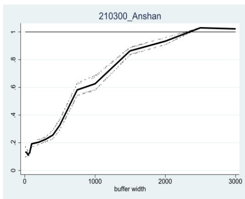
Type 5</td>
<td>

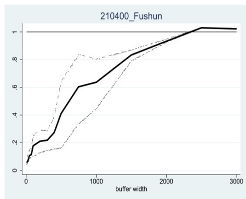
Type 5</td>
<td>

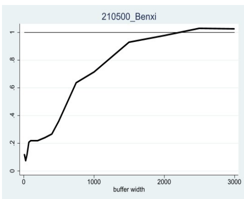
Type 5</td>
<td>

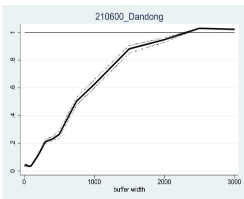
Type 2</td>
</tr>
<tr>
<td>

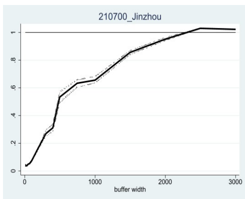
Type 2</td>
<td>

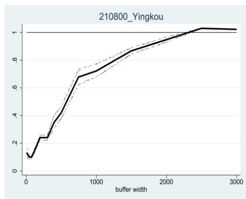
Type 4</td>
<td>

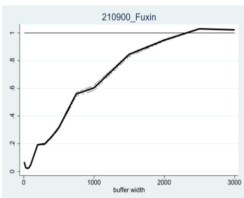
Type 5</td>
<td>

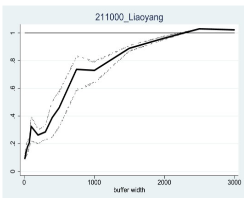
Type 5</td>
</tr>
<tr>
<td>

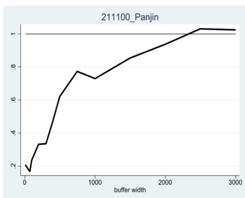
Type 5</td>
<td>

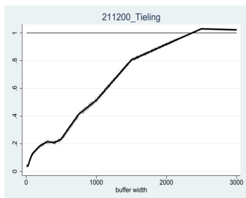
Type 5</td>
<td>

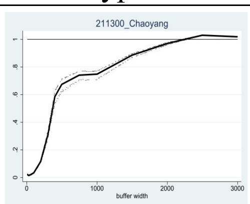
Type 5</td>
<td>

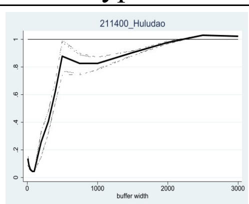
Type 5</td>
</tr>
<tr>
<td>

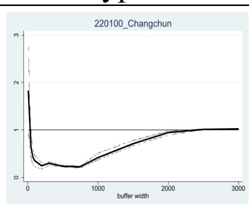
Type 5</td>
<td>

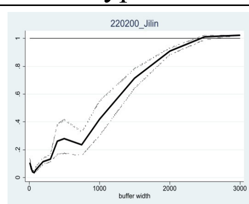
Type 5</td>
<td>

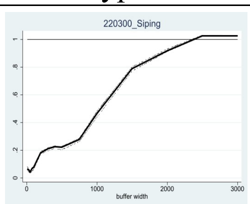
Type 5</td>
<td>

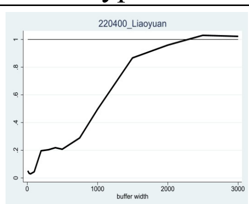
Type 5</td>
</tr>
<tr>
<td>

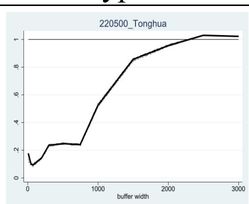
Type 5</td>
<td>

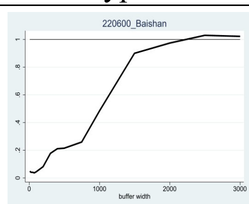
Type 5</td>
<td>

Type 7</td>
<td>

Type 5</td>
</tr>
<tr>
<td>

Type 5</td>
<td>

Type 5</td>
<td>

Type 5</td>
<td>

Type 5</td>
</tr>
</table>

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

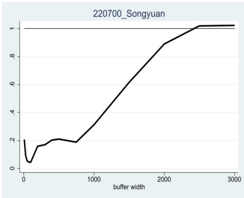
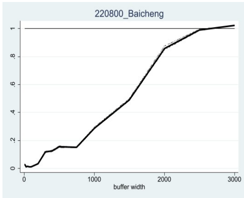
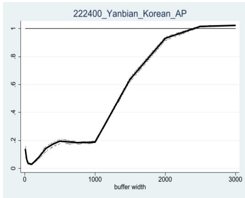
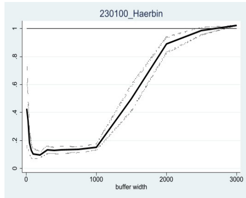
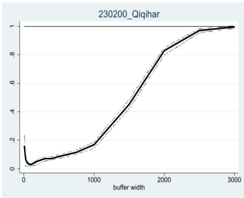
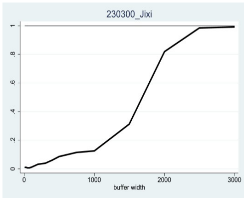
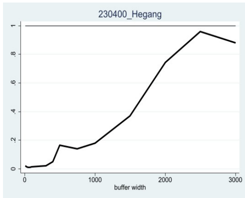
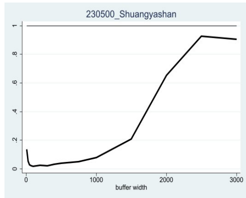
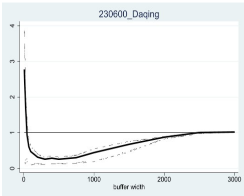
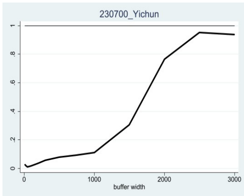
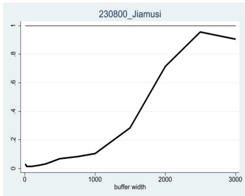
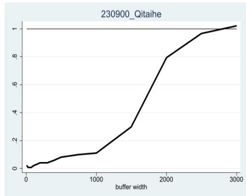
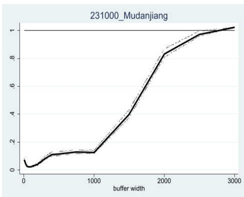
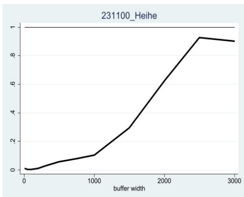
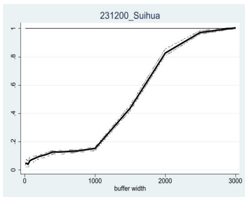
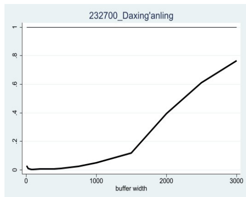
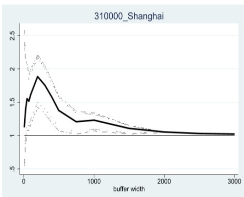
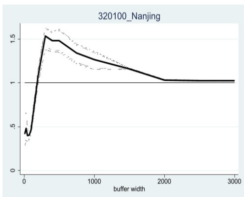
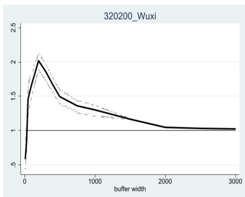
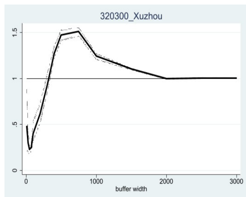
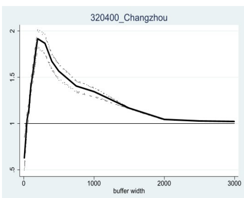
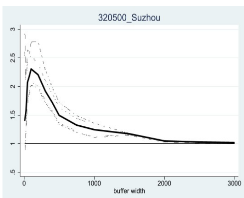
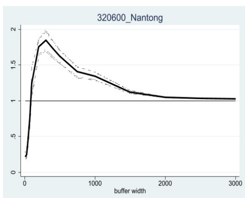
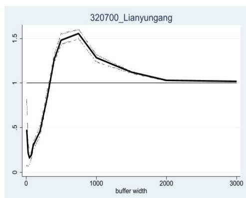
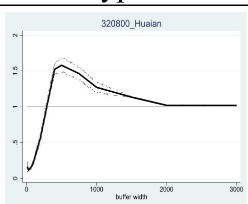
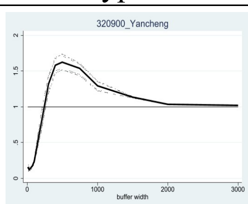
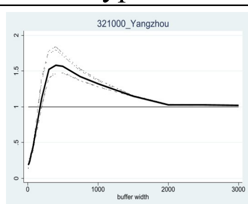
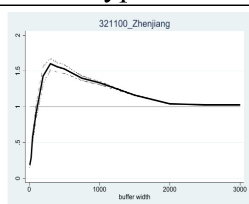
<table>
<tr>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 2</td>
</tr>
<tr>
<td>Type 2</td>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 2</td>
</tr>
<tr>
<td>Type 7</td>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 5</td>
</tr>
<tr>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 5</td>
<td>Type 5</td>
</tr>
<tr>
<td>Type 3</td>
<td>Type 8</td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
<tr>
<td>Type 8</td>
<td>Type 3</td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
<tr>
<td>Type 8</td>
<td>Type 8</td>
<td>Type 8</td>
<td>Type 8</td>
</tr>
</table>

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

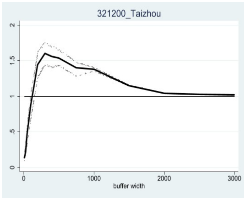
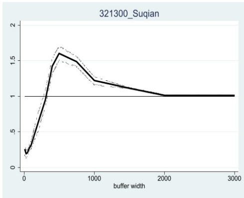
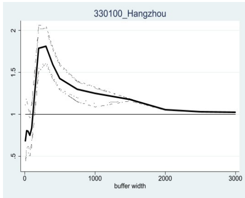
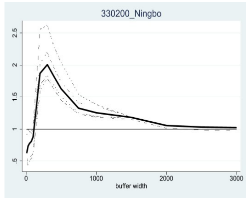
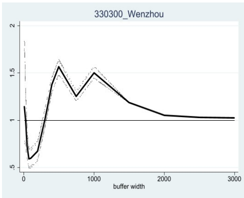
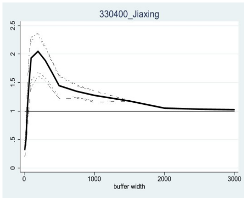
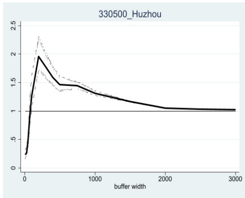
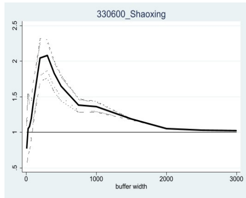
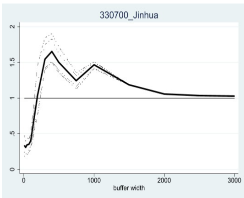
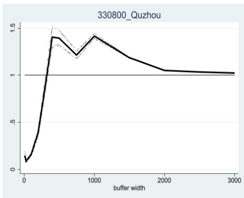
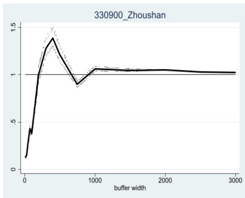
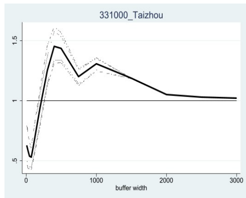
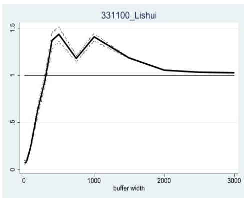
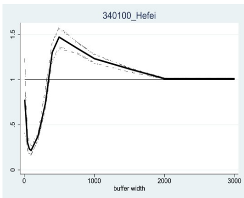
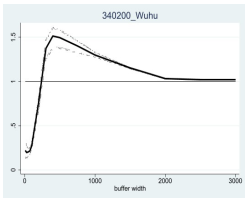
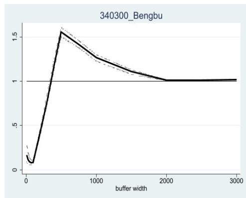
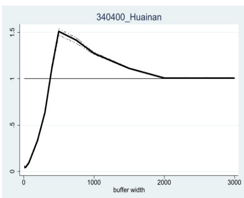
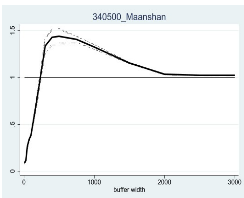
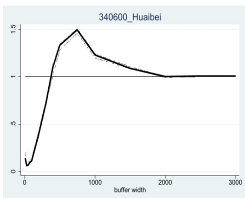
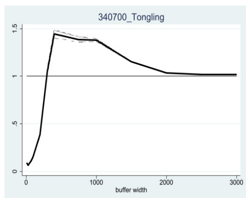
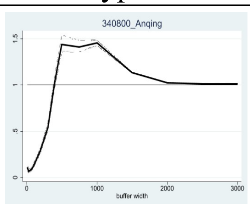
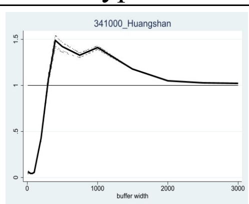
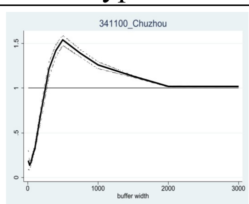
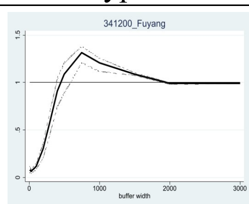
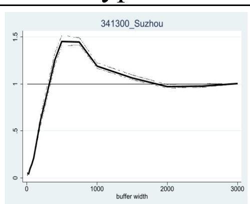
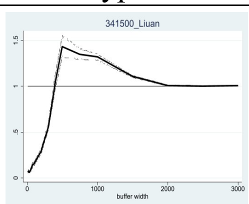
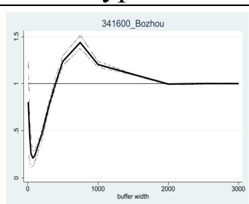
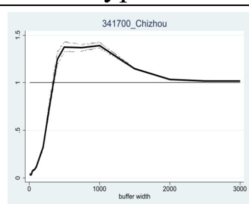
 Type 8	 Type 8	 Type 8	 Type 8
 Type 9	 Type 8	 Type 8	 Type 8
 Type 8	 Type 8	 Type 8	 Type 8
 Type 8	 Type 9	 Type 8	 Type 8
 Type 8	 Type 8	 Type 8	 Type 8
 Type 8	 Type 8	 Type 8	 Type 8
 Type 8	 Type 8	 Type 9	 Type 8

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

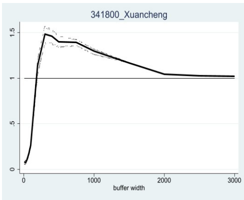
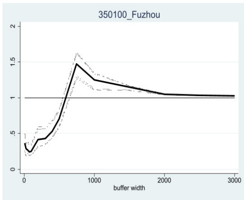
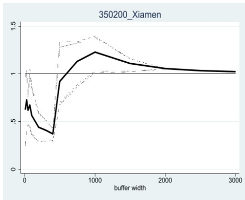
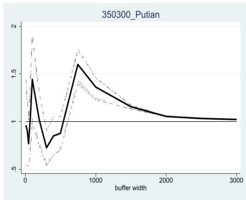
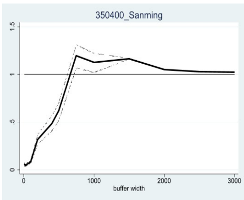
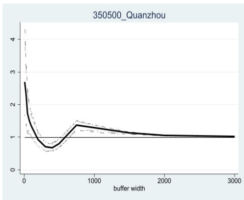
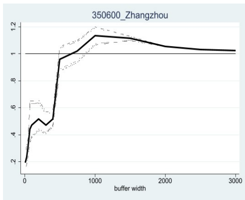
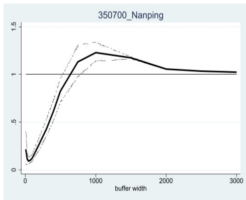
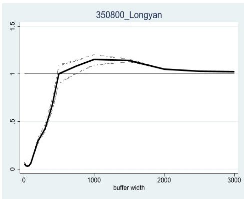
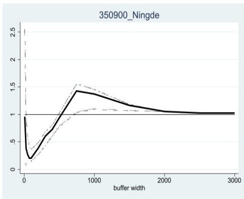
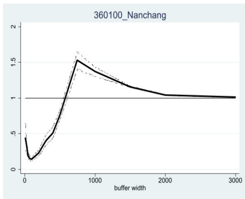
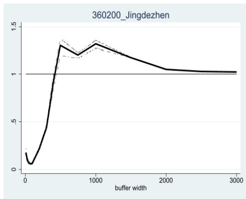
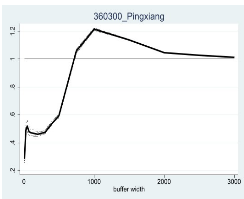
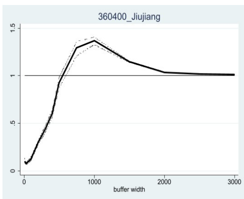
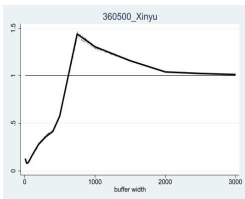
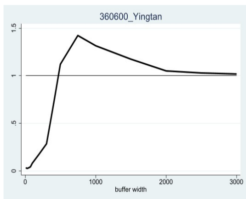
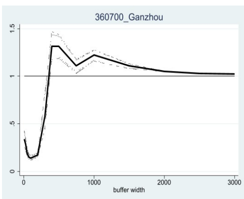
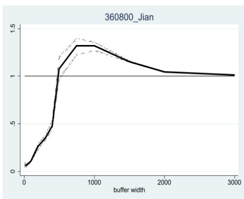
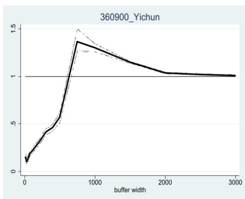
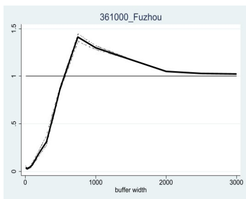
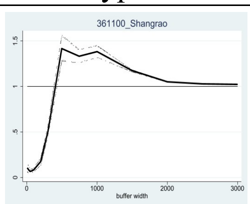
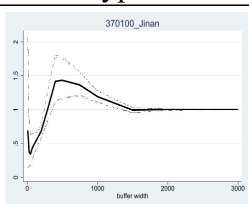
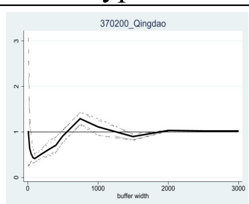
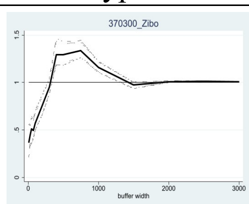
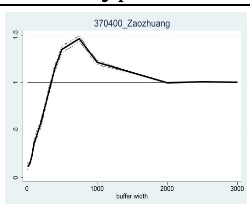
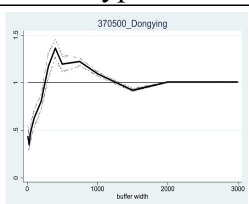
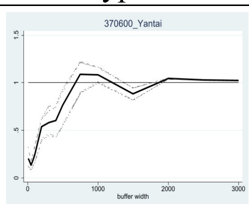
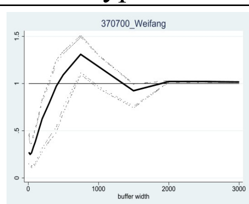
<table>
<tr>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 9</td>
<td>
Type 9</td>
</tr>
<tr>
<td>
Type 10</td>
<td>
Type 6</td>
<td>
Type 10</td>
<td>
Type 10</td>
</tr>
<tr>
<td>
Type 10</td>
<td>
Type 9</td>
<td>
Type 8</td>
<td>
Type 8</td>
</tr>
<tr>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 8</td>
</tr>
<tr>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 8</td>
</tr>
<tr>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 9</td>
<td>
Type 8</td>
</tr>
<tr>
<td>
Type 8</td>
<td>
Type 8</td>
<td>
Type 10</td>
<td>
Type 8</td>
</tr>
</table>

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

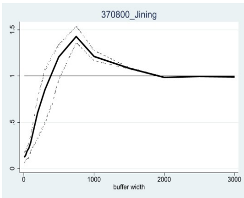
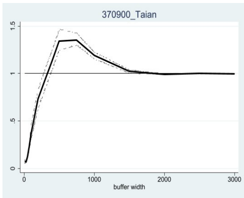
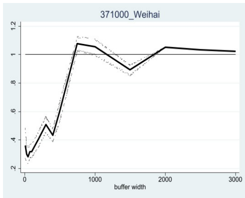
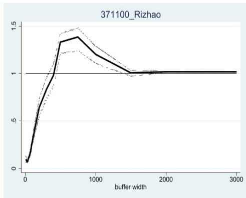
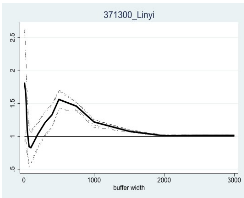
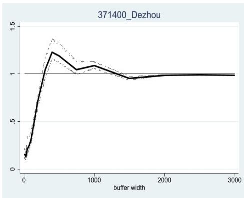
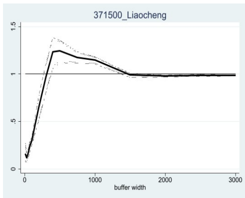
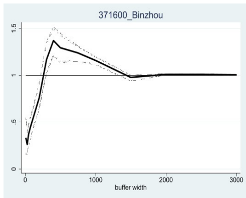
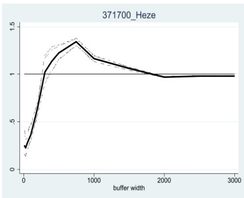
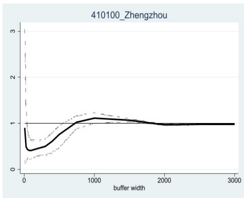
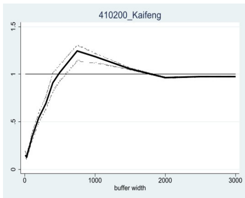
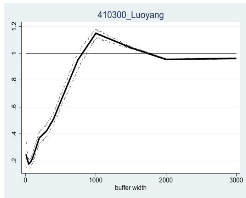
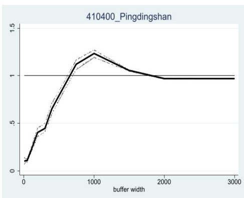
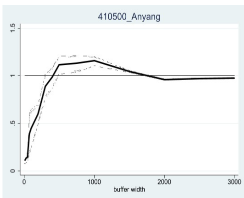
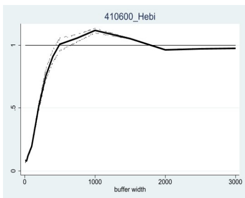
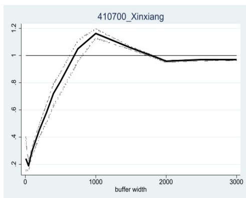
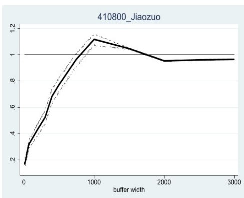
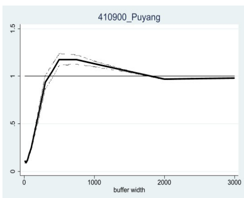
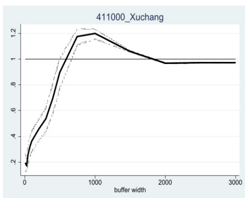
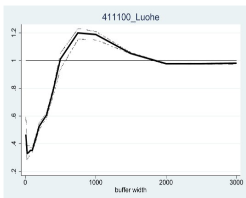
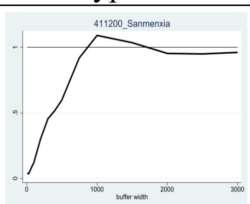
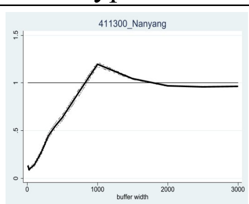
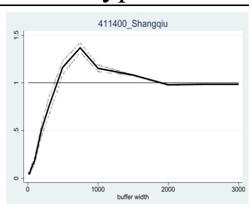
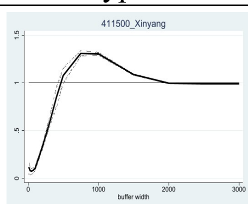
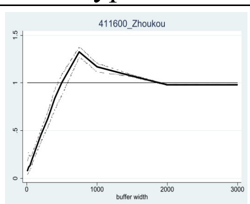
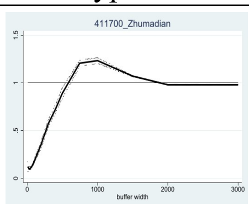
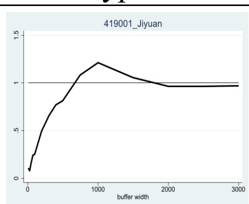
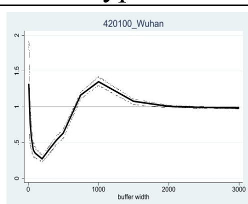
 Type 8	 Type 8	 Type 10	 Type 8
 Type 9	 Type 10	 Type 8	 Type 8
 Type 8	 Type 2	 Type 10	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10
 Type 10	 Type 10	 Type 8	 Type 8
 Type 10	 Type 10	 Type 8	 Type 8
 Type 10	 Type 10	 Type 10	 Type 9

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

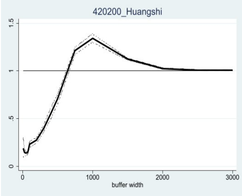
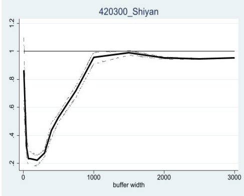
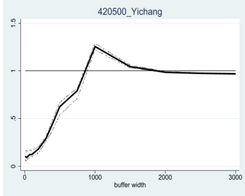
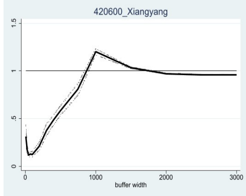
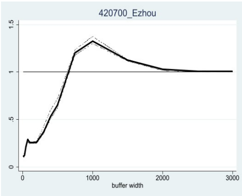
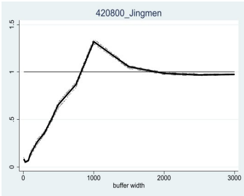
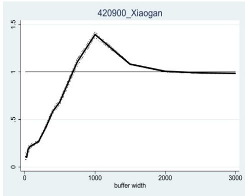
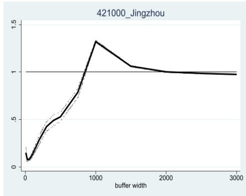
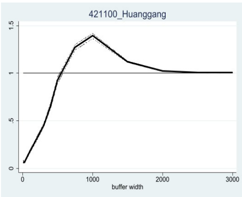
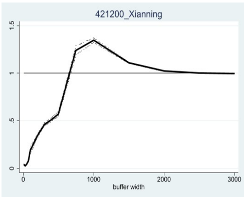
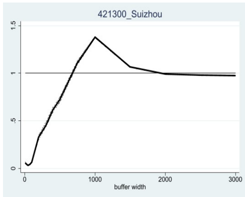
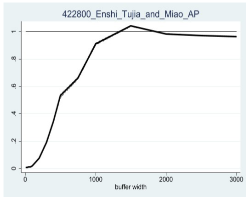
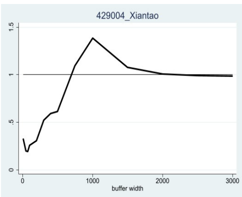
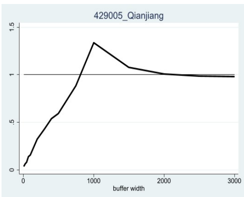
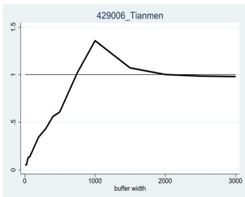
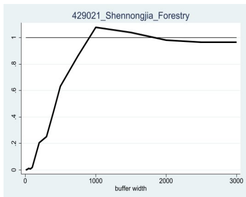
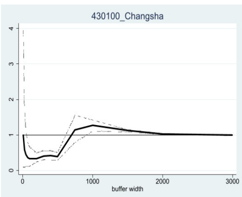
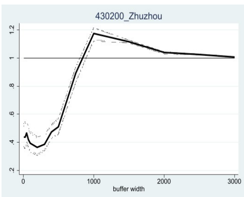
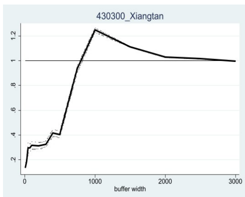
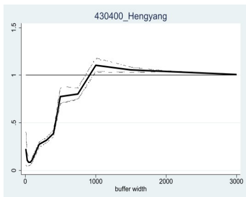
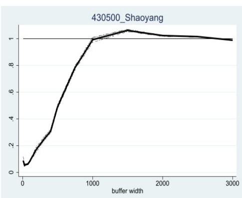
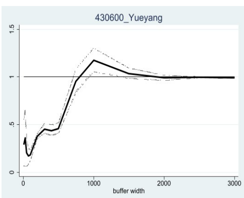
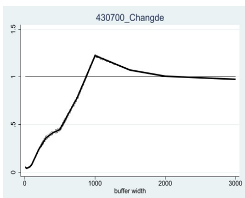
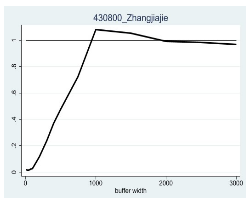
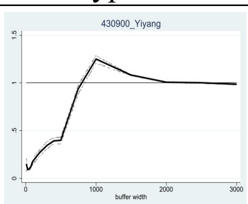
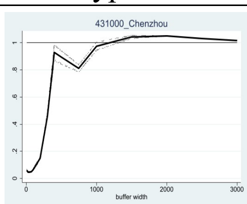
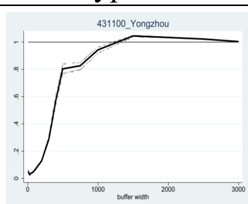
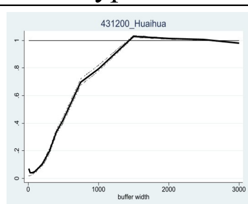
 Type 8	 Type 2	 Type 10	 Type 10
 Type 8	 Type 8	 Type 8	 Type 8
 Type 8	 Type 8	 Type 8	 Type 10
 Type 8	 Type 8	 Type 8	 Type 10
 Type 7	 Type 8	 Type 8	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10
 Type 8	 Type 10	 Type 10	 Type 10

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

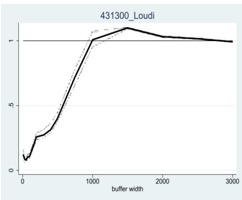
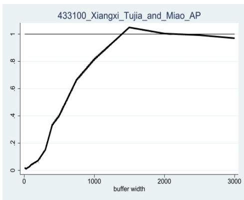
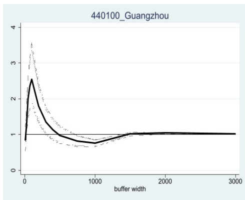
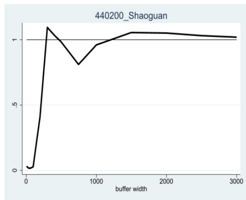
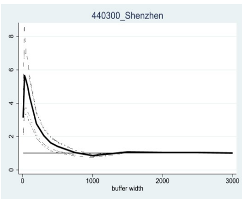
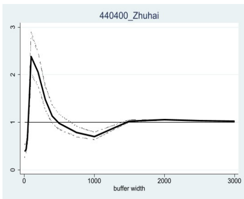
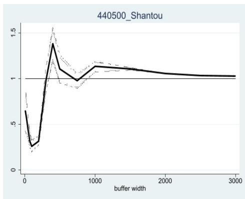
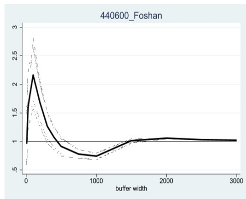
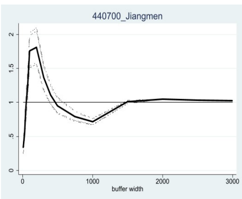
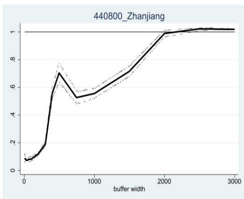
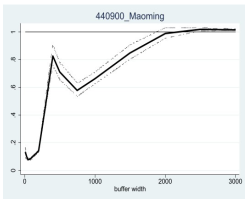
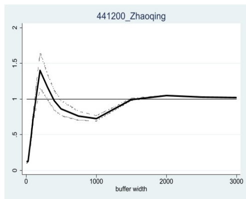
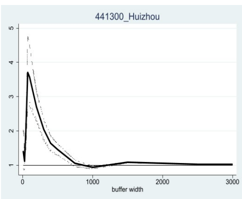
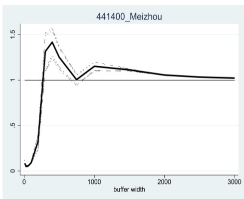
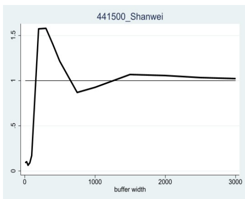
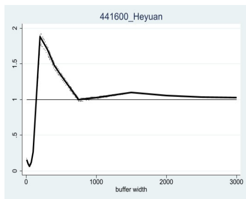
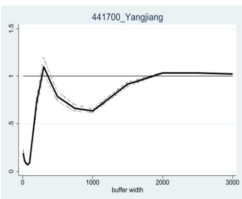
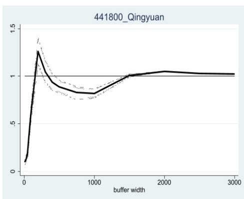
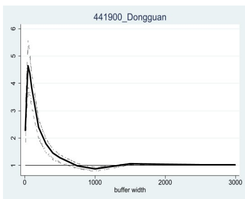
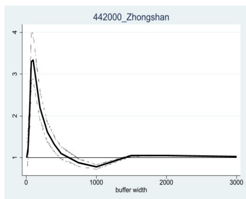
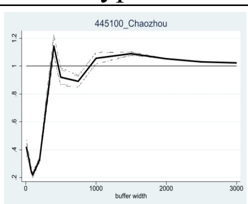
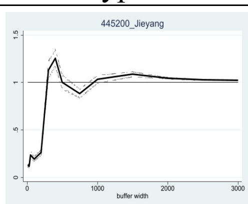
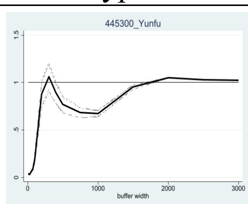
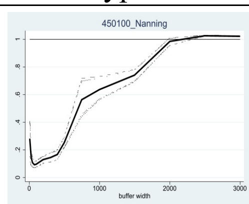
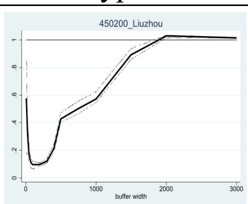
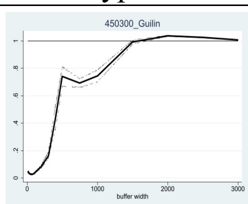
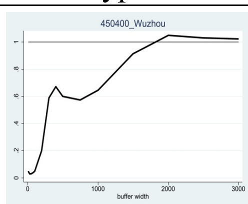
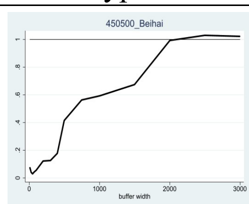
 Type 10	 Type 10	 Type 9	 Type 9
 Type 3	 Type 9	 Type 9	 Type 9
 Type 9	 Type 4	 Type 4	 Type 9
 Type 3	 Type 8	 Type 8	 Type 8
 Type 9	 Type 9	 Type 3	 Type 8
 Type 8	 Type 8	 Type 8	 Type 2
 Type 2	 Type 10	 Type 10	 Type 10

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

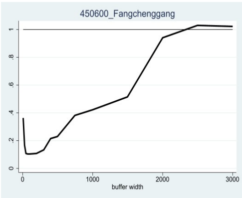
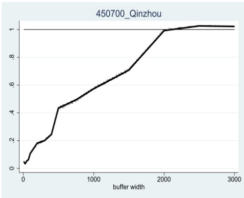
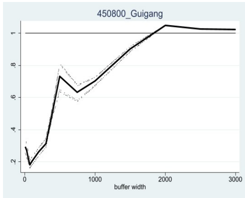
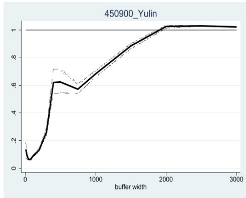
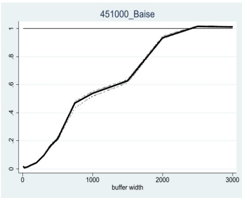
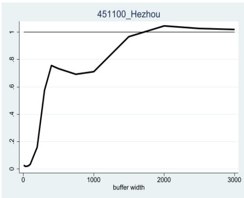
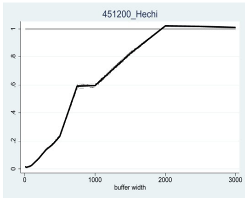
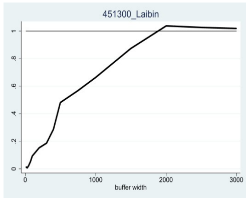
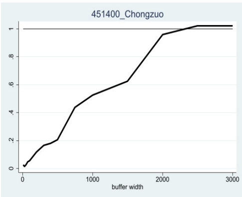
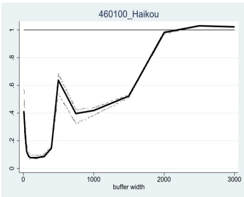
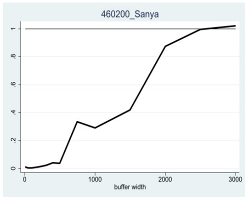
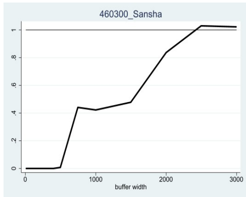
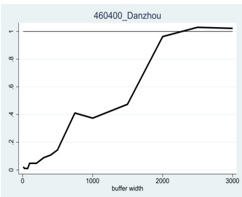
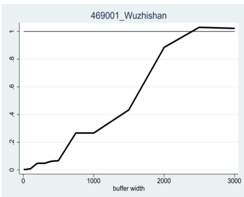
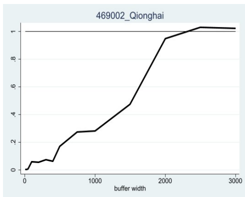
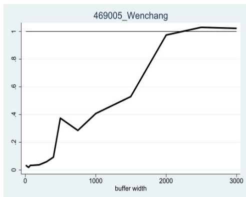
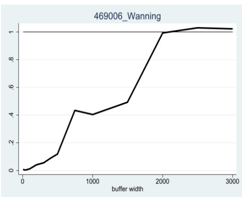
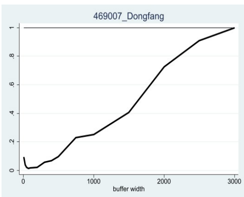
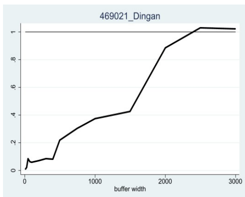
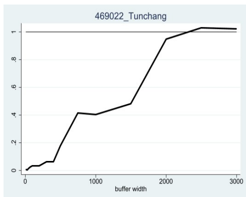
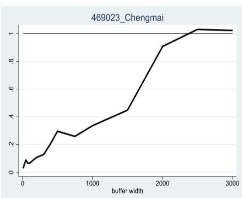
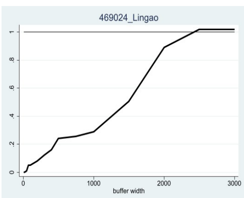
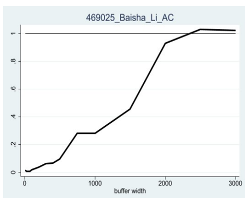
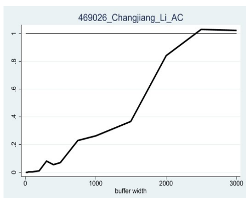
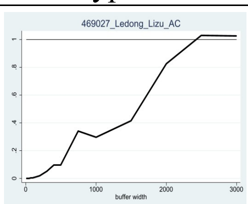
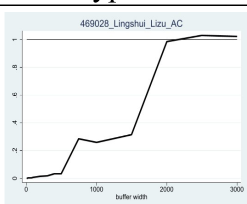
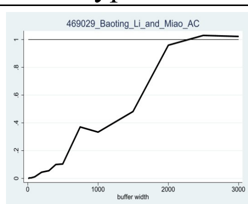
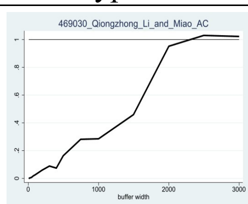
 Type 2	 Type 10	 Type 10	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10
 Type 5	 Type 4	 Type 5	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10
 Type 10	 Type 2	 Type 10	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10
 Type 10	 Type 10	 Type 10	 Type 10

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

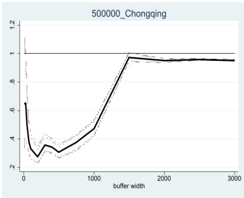
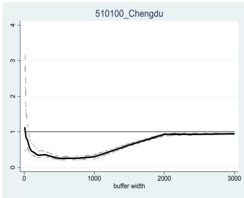
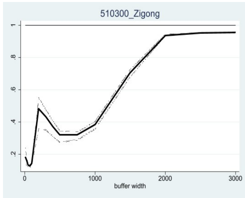
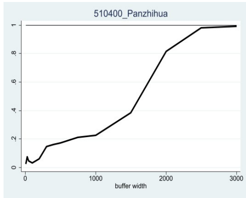
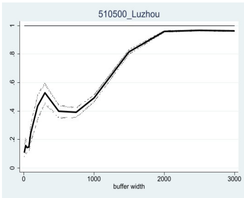
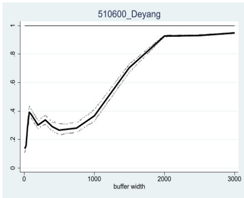
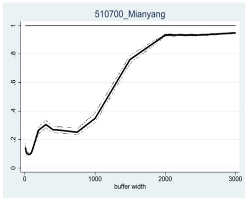
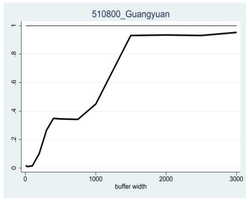
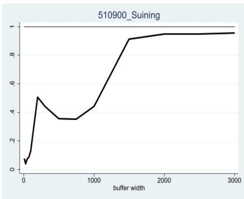
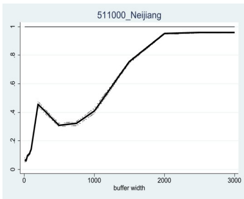
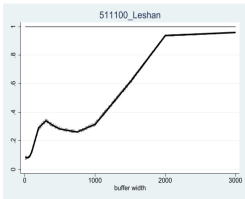
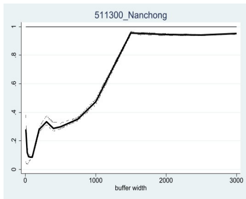
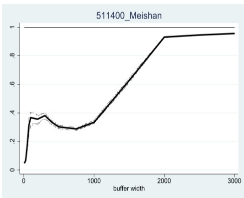
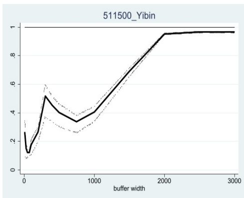
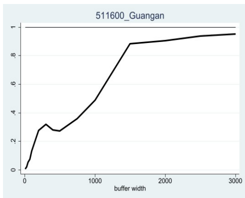
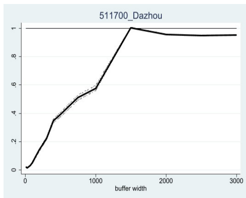
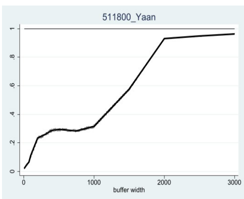
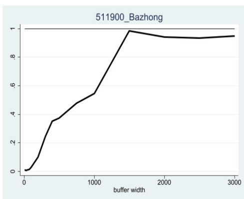
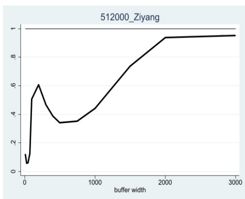
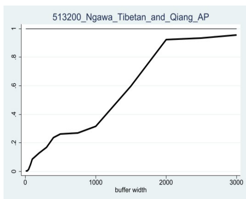
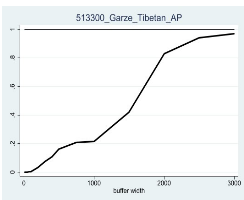
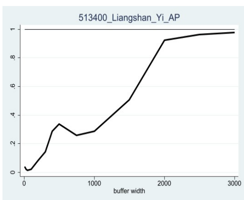
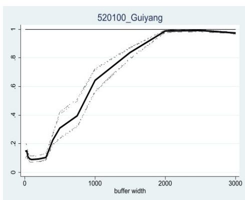
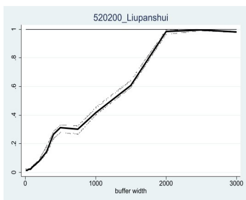
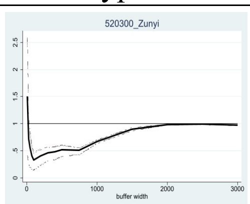
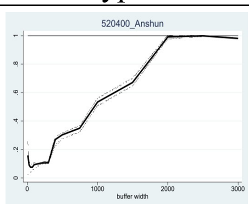
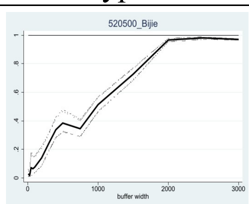
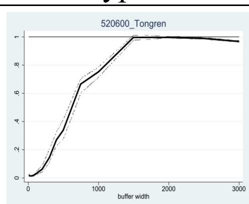
 Type 2	 Type 7	 Type 4	 Type 5
 Type 4	 Type 4	 Type 4	 Type 5
 Type 4	 Type 4	 Type 4	 Type 4
 Type 5	 Type 4	 Type 5	 Type 5
 Type 5	 Type 5	 Type 4	 Type 5
 Type 5	 Type 5	 Type 5	 Type 5
 Type 7	 Type 5	 Type 5	 Type 5

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

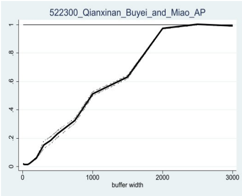
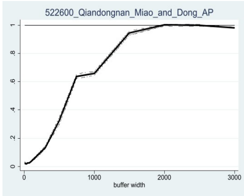
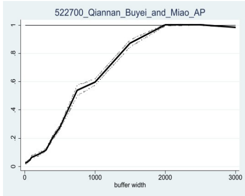
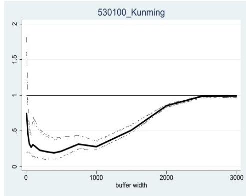
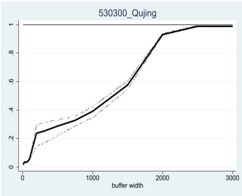
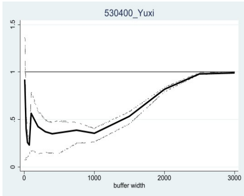
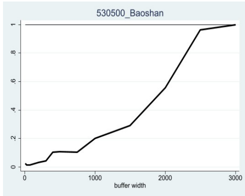
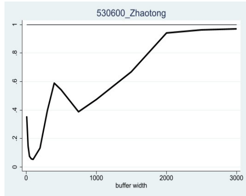
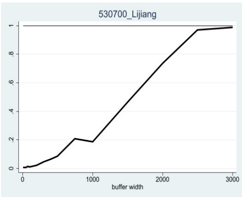
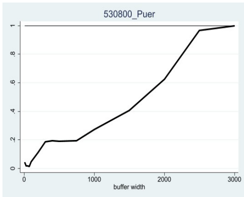
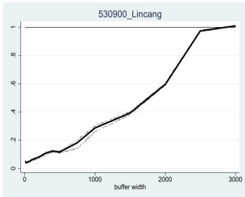
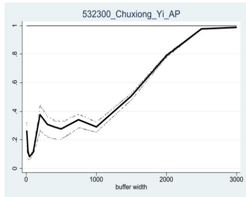
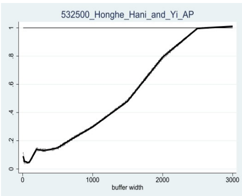
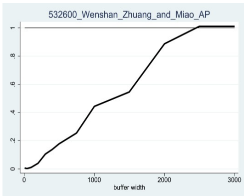
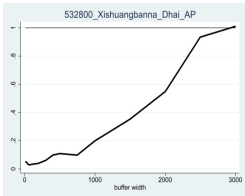
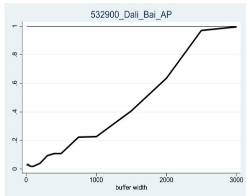
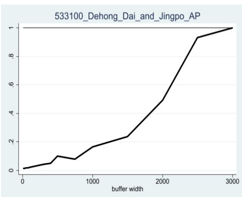
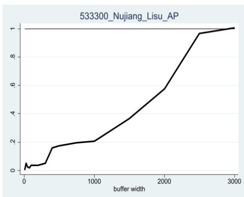
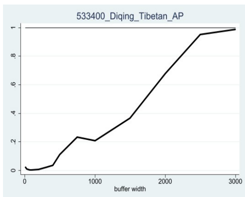
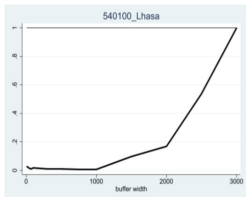
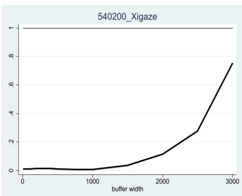
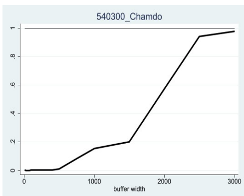
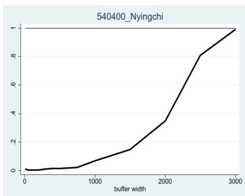
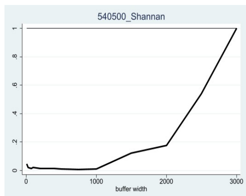
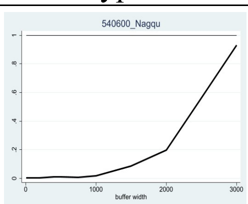
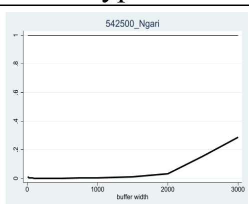
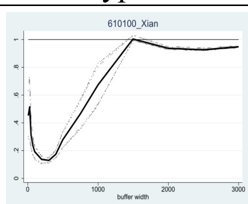
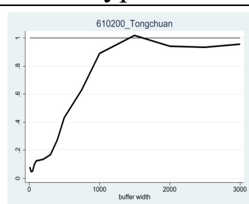
 Type 5	 Type 5	 Type 5	 Type 2
 Type 5	 Type 4	 Type 5	 Type 4
 Type 5	 Type 5	 Type 5	 Type 4
 Type 5	 Type 5	 Type 5	 Type 5
 Type 5	 Type 5	 Type 5	 Type 2
 Type 2	 Type 5	 Type 5	 Type 2
 Type 2	 Type 2	 Type 2	 Type 5

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

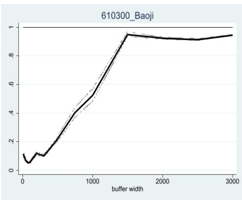
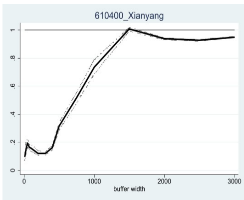
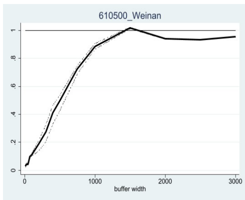
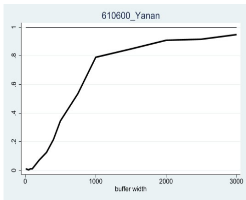
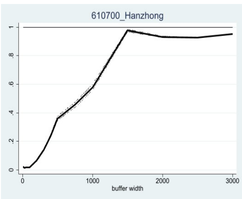
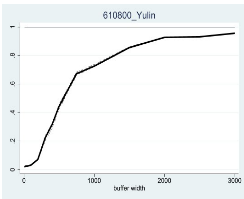
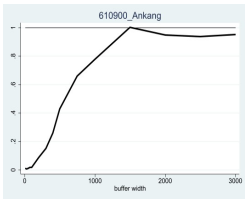
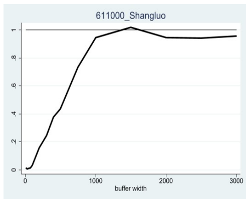
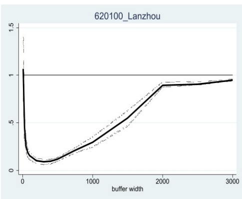
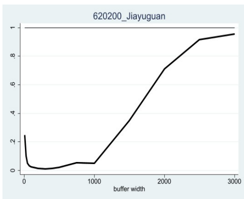
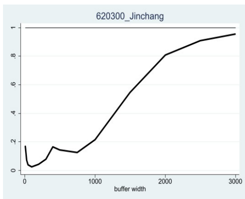
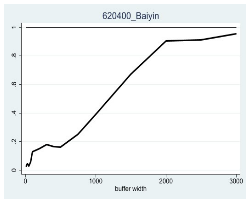
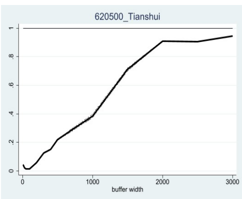
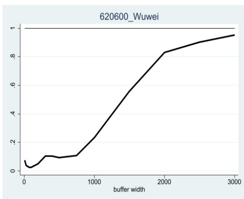
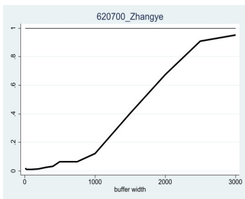
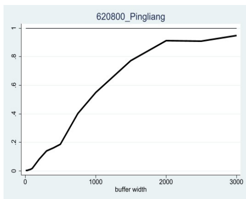
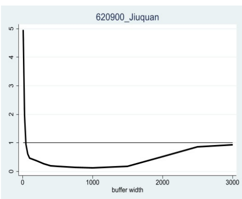
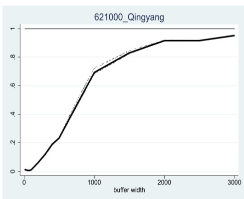
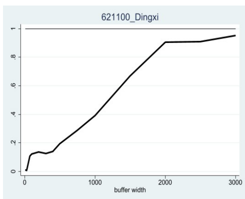
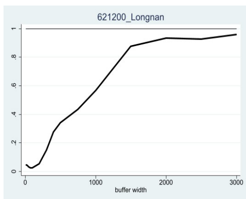
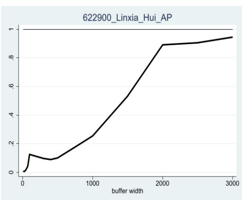
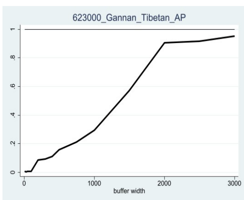
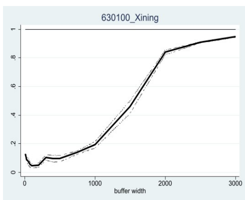
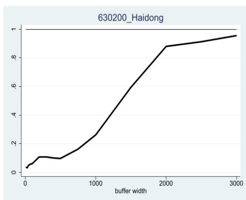
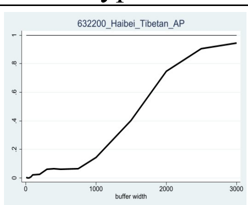
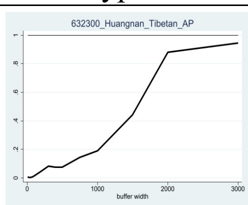
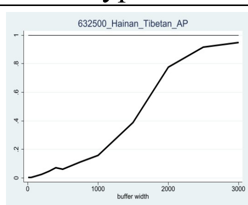
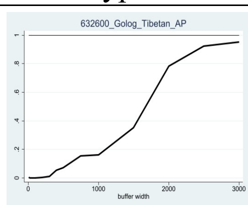
 Type 5	 Type 5	 Type 5	 Type 5
 Type 5	 Type 5	 Type 5	 Type 5
 Type 7	 Type 2	 Type 4	 Type 5
 Type 5	 Type 5	 Type 5	 Type 5
 Type 6	 Type 5	 Type 5	 Type 5
 Type 5	 Type 5	 Type 5	 Type 5
 Type 5	 Type 5	 Type 5	 Type 5

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

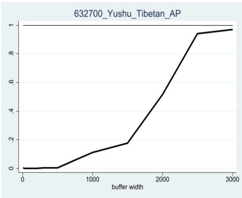
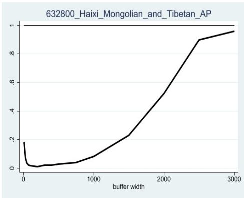
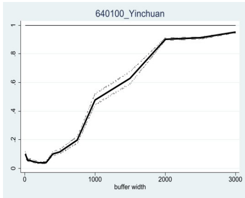
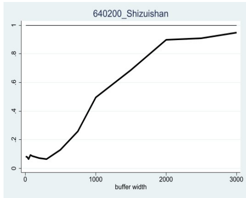
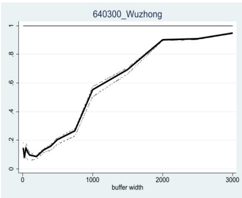
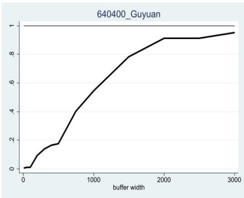
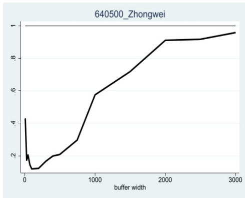
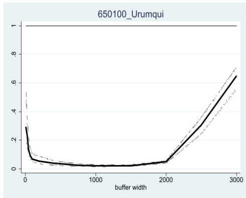
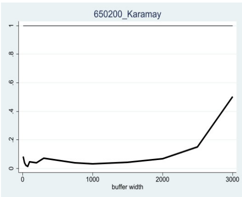
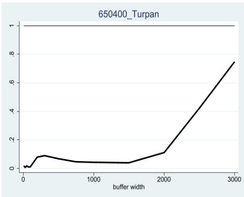
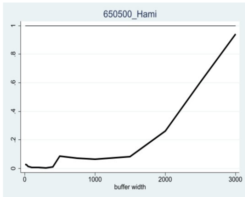
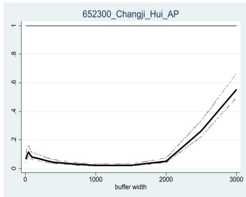
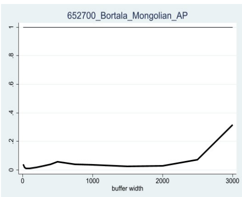
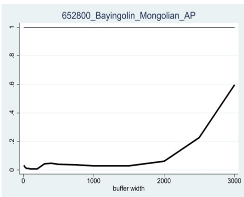
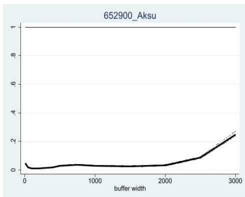
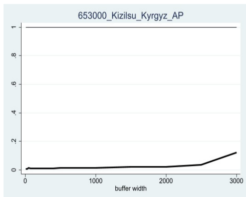
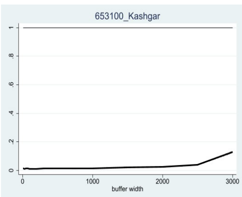
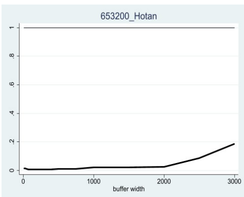
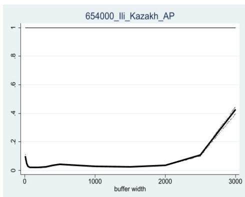
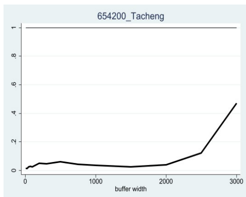
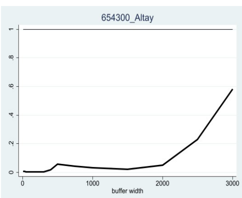
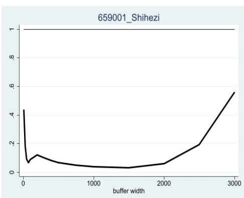
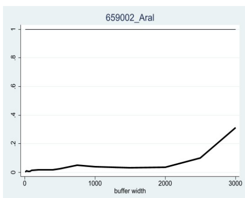
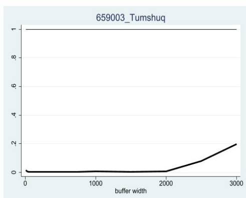
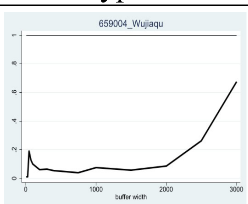
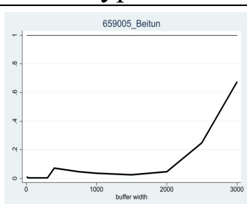
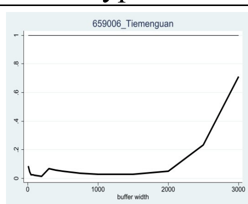
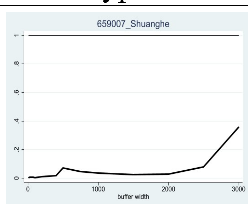
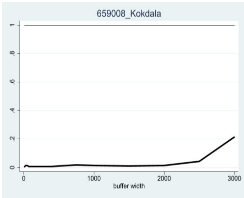
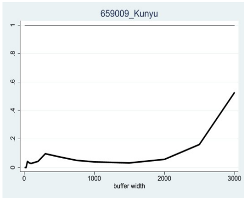
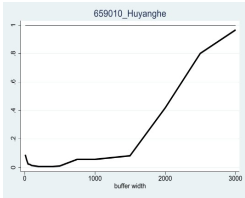
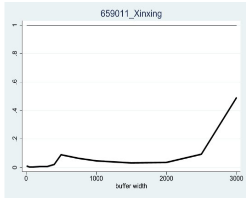
 Type 5	 Type 2	 Type 5	 Type 5
 Type 5	 Type 5	 Type 2	 Type 2
 Type 2	 Type 2	 Type 2	 Type 2
 Type 2	 Type 2	 Type 2	 Type 2
 Type 2	 Type 2	 Type 2	 Type 2
 Type 2	 Type 2	 Type 2	 Type 2
 Type 2	 Type 2	 Type 2	 Type 2

Table B.3.4. Sectoral agglomeration curves for each Chinese prefecture(continued)

 Type 2	 Type 2	 Type 2	 Type 2

Note:X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
110000	Beijing	Beijing	4.65	4.30	4.29	27
120000	Tianjin	Tianjin	0.53	0.49	0.45	33
130100	Shijiazhuang	Hebei	0.54	0.64	0.61	20
130200	Tangshan	Hebei	0.51	0.54	0.52	16
130300	Qinhuangdao	Hebei	0.29	0.26	0.21	6
130400	Handan	Hebei	1.37	1.47	1.45	18
130500	Xingtai	Hebei	0.23	0.33	0.32	17
130600	Baoding	Hebei	0.23	0.35	0.34	17
130700	Zhangjiakou	Hebei	0.11	0.24	0.24	11
130800	Chengde	Hebei	0.16	0.30	0.29	10
130900	Cangzhou	Hebei	0.2	0.36	0.35	15
131000	Langfang	Hebei	0.36	0.44	0.41	12
131100	Hengshui	Hebei	0.27	0.41	0.39	11
140100	Taiyuan	Shanxi	0.6	0.17	0.15	6
140200	Datong	Shanxi	0.49	-0.34	-0.34	1
140300	Yangquan	Shanxi	0.14	0.35	0.35	1
140400	Changzhi	Shanxi	0.19	0.38	0.37	2

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
140500	Jincheng	Shanxi	0.39	0.09	0.05	1
140600	Shuozhou	Shanxi	0.08	0.35	0.34	1
140700	Jinzhong	Shanxi	0.18	0.44	0.43	3
140800	Yuncheng	Shanxi	0.11	0.34	0.33	4
140900	Xinzhou	Shanxi	0.09	0.33	0.31	2
141000	Linfen	Shanxi	0.15	0.42	0.41	2
141100	Lüliang	Shanxi	0.19	0.44	0.42	4
150100	Hohhot	Inner Mongolia	0.17	0.43	0.37	9
150200	Baotou	Inner Mongolia	0.22	0.03	-0.01	9
150300	Wuhai	Inner Mongolia	0.19	0.46	0.41	1
150400	Chifeng	Inner Mongolia	0.05	0.28	0.25	10
150500	Tongliao	Inner Mongolia	0.11	0.46	0.44	5
150600	Eerduosi	Inner Mongolia	0.06	0.27	0.24	10
150700	Hulunbuir	Inner Mongolia	0.02	0.22	0.2	15
150800	Bayannur	Inner Mongolia	0.05	0.33	0.29	6
150900	Ulanqab	Inner Mongolia	0.07	0.31	0.29	3
152200	Xingan	Inner Mongolia	0.08	-0.02	-0.04	3
152500	Xilingol League	Inner Mongolia	0.03	0.31	0.28	8
152900	Alashan	Inner Mongolia	0.15	0.46	0.41	2

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
210100	Shenyang	Liaoning	0.38	0.28	0.26	17
210200	Dalian	Liaoning	0.41	0.35	0.33	14
210300	Anshan	Liaoning	0.15	0.32	0.3	7
210400	Fushun	Liaoning	0.1	0.28	0.26	3
210500	Benxi	Liaoning	0.15	0.34	0.32	4
210600	Dandong	Liaoning	0.06	0.16	0.15	3
210700	Jinzhou	Liaoning	0.07	0.26	0.24	5
210800	Yingkou	Liaoning	0.15	0.31	0.29	7
210900	Fuxin	Liaoning	0.07	0.30	0.28	4
211000	Liaoyang	Liaoning	0.16	0.43	0.43	2
211100	Panjin	Liaoning	0.23	0.41	0.4	5
211200	Tieling	Liaoning	0.08	0.30	0.28	4
211300	Chaoyang	Liaoning	0.06	0.27	0.26	7
211400	Huludao	Liaoning	0.14	0.00	-0.02	6
220100	Changchun	Jilin	1.3	1.21	1.18	11
220200	Jilin	Jilin	0.1	0.21	0.19	12
220300	Siping	Jilin	0.08	0.30	0.28	6
220400	Liaoyuan	Jilin	0.07	0.21	0.18	4
220500	Tonghua	Jilin	0.16	0.29	0.27	7

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
220600	Baishan	Jilin	0.06	0.12	0.11	6
220700	Songyuan	Jilin	0.15	0.26	0.25	2
220800	Baicheng	Jilin	0.04	0.20	0.16	5
222400	Yanbian Korean Autonomous Prefecture	Jilin	0.11	0.16	0.13	9
230100	Haerbin	Heilongjiang	0.33	0.11	0.09	16
230200	Qiqihar	Heilongjiang	0.11	0.36	0.35	12
230300	Jixi	Heilongjiang	0.02	0.22	0.21	5
230400	Hegang	Heilongjiang	0.03	0.21	0.2	3
230500	Shuangyashan	Heilongjiang	0.09	0.43	0.42	3
230600	Daqing	Heilongjiang	1.91	2.00	2	5
230700	Yichun	Heilongjiang	0.03	0.07	0.05	5
230800	Jiamusi	Heilongjiang	0.03	0.22	0.22	8
230900	Qitaihe	Heilongjiang	0.03	0.23	0.22	2
231000	Mudanjiang	Heilongjiang	0.06	0.11	0.11	11
231100	Heihe	Heilongjiang	0.02	0.24	0.24	6
231200	Suihua	Heilongjiang	0.06	0.24	0.21	11
232700	Daxing'anling	Heilongjiang	0.02	0.19	0.18	4
310000	Shanghai	Shanghai	1.31	1.30	1.27	59
320100	Nanjing	Jiangsu	0.51	0.39	0.36	17

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
320200	Wuxi	Jiangsu	0.94	0.91	0.89	13
320300	Xuzhou	Jiangsu	0.45	0.54	0.52	14
320400	Changzhou	Jiangsu	0.87	0.88	0.86	12
320500	Suzhou	Jiangsu	1.63	1.51	1.47	25
320600	Nantong	Jiangsu	0.46	0.47	0.44	17
320700	Lianyungang	Jiangsu	0.41	0.52	0.47	11
320800	Huaian	Jiangsu	0.23	0.27	0.25	10
320900	Yancheng	Jiangsu	0.25	0.34	0.31	14
321000	Yangzhou	Jiangsu	0.34	0.34	0.32	11
321100	Zhenjiang	Jiangsu	0.42	0.49	0.47	9
321200	Taizhou	Jiangsu	0.37	0.24	0.21	10
321300	Suqian	Jiangsu	0.31	0.38	0.35	6
330100	Hangzhou	Zhejiang	0.8	0.84	0.82	14
330200	Ningbo	Zhejiang	0.78	0.82	0.8	18
330300	Wenzhou	Zhejiang	1.01	1.01	0.97	11
330400	Jiaxing	Zhejiang	0.71	0.73	0.71	12
330500	Huzhou	Zhejiang	0.48	0.56	0.53	7
330600	Shaoxing	Zhejiang	1	1.15	1.12	11
330700	Jinhua	Zhejiang	0.42	0.52	0.49	14

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
330800	Quzhou	Zhejiang	0.2	0.37	0.34	7
330900	Zhoushan	Zhejiang	0.27	-0.17	-0.19	5
331000	Taizhou	Zhejiang	0.65	0.69	0.66	11
331100	Lishui	Zhejiang	0.18	0.25	0.23	10
340100	Hefei	Anhui	0.64	0.66	0.63	14
340200	Wuhu	Anhui	0.3	0.43	0.39	10
340300	Bengbu	Anhui	0.2	0.20	0.18	6
340400	Huainan	Anhui	0.13	0.29	0.28	6
340500	Maanshan	Anhui	0.24	0.39	0.36	8
340600	Huaibei	Anhui	0.18	0.41	0.4	4
340700	Tongling	Anhui	0.17	0.24	0.2	5
340800	Anqing	Anhui	0.16	0.29	0.28	11
341000	Huangshan	Anhui	0.14	0.29	0.27	7
341100	Chuzhou	Anhui	0.27	0.32	0.3	8
341200	Fuyang	Anhui	0.14	0.17	0.14	8
341300	Suzhou	Anhui	0.15	0.29	0.28	5
341500	Liuan	Anhui	0.14	0.17	0.14	7
341600	Bozhou	Anhui	0.62	0.80	0.77	5
341700	Chizhou	Anhui	0.12	0.28	0.25	5

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
341800	Xuancheng	Anhui	0.2	0.36	0.33	8
350100	Fuzhou	Fujian	0.35	0.36	0.34	17
350200	Xiamen	Fujian	0.65	0.67	0.64	9
350300	Putian	Fujian	0.97	0.77	0.74	6
350400	Sanming	Fujian	0.11	0.24	0.21	11
350500	Quanzhou	Fujian	2.22	2.16	2.14	13
350600	Zhangzhou	Fujian	0.29	0.44	0.41	13
350700	Nanping	Fujian	0.21	0.35	0.33	10
350800	Longyan	Fujian	0.1	0.18	0.15	8
350900	Ningde	Fujian	0.67	0.77	0.77	10
360100	Nanchang	Jiangxi	0.38	0.22	0.19	9
360200	Jingdezhen	Jiangxi	0.19	0.28	0.25	3
360300	Pingxiang	Jiangxi	0.4	0.60	0.58	4
360400	Jiujiang	Jiangxi	0.15	0.17	0.14	13
360500	Xinyu	Jiangxi	0.15	0.27	0.23	2
360600	Yingtan	Jiangxi	0.09	0.21	0.18	3
360700	Ganzhou	Jiangxi	0.33	0.37	0.34	20
360800	Jian	Jiangxi	0.12	0.03	0	14
360900	Yichun	Jiangxi	0.19	0.20	0.18	10

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
361000	Fuzhou	Jiangxi	0.09	0.20	0.18	10
361100	Shangrao	Jiangxi	0.15	0.22	0.19	11
370100	Jinan	Shandong	0.6	0.72	0.69	14
370200	Qingdao	Shandong	0.81	0.84	0.81	14
370300	Zibo	Shandong	0.47	0.66	0.64	12
370400	Zaozhuang	Shandong	0.22	0.43	0.41	7
370500	Dongying	Shandong	0.5	0.70	0.67	10
370600	Yantai	Shandong	0.23	0.34	0.31	14
370700	Weifang	Shandong	0.33	0.36	0.33	15
370800	Jining	Shandong	0.22	0.00	-0.02	15
370900	Taian	Shandong	0.18	0.37	0.35	7
371000	Weihai	Shandong	0.36	0.35	0.32	9
371100	Rizhao	Shandong	0.18	0.26	0.23	5
371300	Linyi	Shandong	1.54	1.68	1.66	14
371400	Dezhou	Shandong	0.24	0.44	0.41	12
371500	Liaocheng	Shandong	0.23	0.50	0.48	9
371600	Binzhou	Shandong	0.4	0.61	0.57	8
371700	Heze	Shandong	0.32	0.40	0.37	10
410100	Zhengzhou	Henan	0.7	0.62	0.58	11

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
410200	Kaifeng	Henan	0.22	0.26	0.24	8
410300	Luoyang	Henan	0.26	0.30	0.29	13
410400	Pingdingshan	Henan	0.16	0.29	0.28	9
410500	Anyang	Henan	0.22	0.34	0.33	8
410600	Hebi	Henan	0.16	0.34	0.3	4
410700	Xinxiang	Henan	0.27	0.40	0.38	11
410800	Jiaozuo	Henan	0.25	0.41	0.39	8
410900	Puyang	Henan	0.19	0.36	0.35	7
411000	Xuchang	Henan	0.26	0.46	0.44	6
411100	Luohe	Henan	0.44	0.66	0.65	4
411200	Sanmenxia	Henan	0.09	0.40	0.38	6
411300	Nanyang	Henan	0.16	0.33	0.31	14
411400	Shangqiu	Henan	0.14	0.26	0.25	9
411500	Xinyang	Henan	0.16	0.26	0.25	10
411600	Zhoukou	Henan	0.17	0.19	0.18	10
411700	Zhumadian	Henan	0.17	0.33	0.31	10
419001	Jiyuan	Henan	0.18	0.45	0.44	2
420100	Wuhan	Hubei	0.96	0.96	0.92	18
420200	Huangshi	Hubei	0.21	0.34	0.3	4

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
420300	Shiyan	Hubei	0.65	0.73	0.71	7
420500	Yichang	Hubei	0.14	0.35	0.31	13
420600	Xiangyang	Hubei	0.26	0.26	0.24	11
420700	Ezhou	Hubei	0.18	0.28	0.25	2
420800	Jingmen	Hubei	0.12	0.28	0.25	5
420900	Xiaogan	Hubei	0.17	0.36	0.33	5
421000	Jingzhou	Hubei	0.16	0.33	0.32	8
421100	Huanggang	Hubei	0.13	0.25	0.23	11
421200	Xianning	Hubei	0.1	0.21	0.19	6
421300	Suizhou	Hubei	0.1	0.15	0.14	1
422800	Enshi Tujia and Miao Autonomous Prefecture	Hubei	0.04	0.21	0.19	8
429004	Xiantao	Hubei	0.32	0.47	0.46	1
429005	Qianjiang	Hubei	0.11	0.29	0.28	1
429006	Tianmen	Hubei	0.12	0.30	0.29	1
429021	Shennongjia Forestry	Hubei	0.05	0.30	0.3	1
430100	Changsha	Hunan	0.75	0.57	0.54	14
430200	Zhuzhou	Hunan	0.45	0.41	0.39	8
430300	Xiangtan	Hunan	0.22	0.36	0.34	7
430400	Hengyang	Hunan	0.2	0.39	0.37	11

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
430500	Shaoyang	Hunan	0.11	0.23	0.21	11
430600	Yueyang	Hunan	0.31	0.44	0.42	11
430700	Changde	Hunan	0.09	0.25	0.24	9
430800	Zhangjiajie	Hunan	0.05	0.16	0.15	3
430900	Yiyang	Hunan	0.16	0.21	0.2	6
431000	Chenzhou	Hunan	0.1	0.26	0.25	12
431100	Yongzhou	Hunan	0.08	0.05	0.04	11
431200	Huaihua	Hunan	0.08	0.28	0.27	12
431300	Loudi	Hunan	0.14	0.29	0.27	5
433100	Xiangxi Tujia and Miao Autonomous Prefecture	Hunan	0.05	0.32	0.29	9
440100	Guangzhou	Guangdong	1.27	1.21	1.19	13
440200	Shaoguan	Guangdong	0.09	0.19	0.17	9
440300	Shenzhen	Guangdong	3.93	3.67	3.64	7
440400	Zhuhai	Guangdong	0.68	0.66	0.64	6
440500	Shantou	Guangdong	0.59	0.71	0.69	6
440600	Foshan	Guangdong	1.32	1.43	1.4	7
440700	Jiangmen	Guangdong	0.65	0.63	0.61	7
440800	Zhanjiang	Guangdong	0.11	0.29	0.27	10
440900	Maoming	Guangdong	0.15	0.26	0.25	6

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
441200	Zhaoqing	Guangdong	0.27	0.33	0.3	7
441300	Huizhou	Guangdong	1.71	1.47	1.44	7
441400	Meizhou	Guangdong	0.15	0.04	0	8
441500	Shanwei	Guangdong	0.2	0.20	0.18	5
441600	Heyuan	Guangdong	0.26	0.15	0.1	6
441700	Yangjiang	Guangdong	0.21	0.43	0.39	5
441800	Qingyuan	Guangdong	0.24	0.25	0.21	4
441900	Dongguan	Guangdong	2.93	2.80	2.78	4
442000	Zhongshan	Guangdong	1.46	1.50	1.47	2
445100	Chaozhou	Guangdong	0.41	0.60	0.58	3
445200	Jieyang	Guangdong	0.21	0.33	0.32	6
445300	Yunfu	Guangdong	0.12	0.24	0.21	5
450100	Nanning	Guangxi	0.22	0.27	0.25	9
450200	Liuzhou	Guangxi	0.41	0.52	0.5	6
450300	Guilin	Guangxi	0.07	0.11	0.09	6
450400	Wuzhou	Guangxi	0.08	0.04	0.02	4
450500	Beihai	Guangxi	0.08	0.07	0.04	6
450600	Fangchenggang	Guangxi	0.26	0.32	0.31	3
450700	Qinzhou	Guangxi	0.08	0.12	0.09	6

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
450800	Guigang	Guangxi	0.29	0.20	0.18	4
450900	Yulin	Guangxi	0.14	0.25	0.23	4
451000	Baise	Guangxi	0.03	0.26	0.24	4
451100	Hezhou	Guangxi	0.07	0.17	0.16	4
451200	Hechi	Guangxi	0.04	0.25	0.23	3
451300	Laibin	Guangxi	0.05	0.19	0.17	2
451400	Chongzuo	Guangxi	0.05	0.21	0.2	4
460100	Haikou	Hainan	0.3	0.47	0.45	2
460200	Sanya	Hainan	0.02	-0.25	-0.27	1
460300	Sansha	Hainan	0.01	0.40	0.4	0
460400	Danzhou	Hainan	0.04	0.20	0.19	2
469001	Wuzhishan	Hainan	0.02	0.11	0.1	0
469002	Qionghai	Hainan	0.03	0.20	0.2	0
469005	Wenchang	Hainan	0.05	0.10	0.09	0
469006	Wanning	Hainan	0.02	0.16	0.15	0
469007	Dongfang	Hainan	0.07	0.28	0.21	1
469021	Dingan	Hainan	0.04	0.52	0.52	0
469022	Tunchang	Hainan	0.03	0.23	0.23	0
469023	Chengmai	Hainan	0.07	0.35	0.27	1

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
469024	Lingao	Hainan	0.03	0.06	0.05	0
469025	Baisha Li Autonomous County	Hainan	0.03	0.24	0.24	0
469026	Changjiang Li Autonomous County	Hainan	0.01	0.17	0.16	0
469027	Ledong Lizu autonomous county	Hainan	0.02	0.27	0.26	0
469028	Lingshui Lizu Autonomous County	Hainan	0.01	0.12	0.12	0
469029	Baoting Li and Miao autonomous county	Hainan	0.02	0.01	0	0
469030	Qiongzhong Li and Miao autonomous county	Hainan	0.02	0.16	0.16	0
500000	Chongqing	Chongqing	0.58	0.43	0.41	49
510100	Chengdu	Sichuan	0.9	0.86	0.83	23
510300	Zigong	Sichuan	0.19	0.43	0.42	4
510400	Panzhihua	Sichuan	0.05	0.12	0.11	3
510500	Luzhou	Sichuan	0.16	0.35	0.34	7
510600	Deyang	Sichuan	0.2	0.37	0.35	7
510700	Mianyang	Sichuan	0.14	0.16	0.12	11
510800	Guangyuan	Sichuan	0.04	0.13	0.12	7
510900	Suining	Sichuan	0.11	0.16	0.14	5
511000	Neijiang	Sichuan	0.11	0.21	0.19	5
511100	Leshan	Sichuan	0.11	0.27	0.26	6
511300	Nanchong	Sichuan	0.21	0.28	0.27	9

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
511400	Meishan	Sichuan	0.13	0.13	0.11	6
511500	Yibin	Sichuan	0.23	0.32	0.31	9
511600	Guangan	Sichuan	0.06	0.33	0.31	6
511700	Dazhou	Sichuan	0.05	0.22	0.21	6
511800	Yaan	Sichuan	0.06	0.39	0.37	8
511900	Bazhong	Sichuan	0.04	0.20	0.18	3
512000	Ziyang	Sichuan	0.16	0.25	0.23	3
513200	Ngawa Tibetan and Qiang Autonomous Prefecture	Sichuan	0.03	0.25	0.25	1
513300	Garze Tibetan Autonomous Prefecture	Sichuan	0.02	0.24	0.23	1
513400	Liangshan Yi Autonomous Prefecture	Sichuan	0.05	0.22	0.2	4
520100	Guiyang	Guizhou	0.15	-0.09	-0.1	9
520200	Liupanshui	Guizhou	0.04	0.13	0.12	3
520300	Zunyi	Guizhou	1.04	1.22	1.19	11
520400	Anshun	Guizhou	0.14	-0.09	-0.12	6
520500	Bijie	Guizhou	0.06	0.16	0.15	7
520600	Tongren	Guizhou	0.04	0.14	0.11	8
522300	Qianxinan Buyei and Miao Autonomous Prefecture	Guizhou	0.04	0.25	0.24	3
522600	Qiandongnan Miao and Dong Autonomous Prefecture	Guizhou	0.05	0.30	0.29	10
522700	Qiannan Buyei and Miao Autonomous Prefecture	Guizhou	0.05	0.27	0.26	7

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
530100	Kunming	Yunnan	0.56	0.63	0.61	14
530300	Qujing	Yunnan	0.05	0.11	0.09	8
530400	Yuxi	Yunnan	0.65	0.43	0.42	7
530500	Baoshan	Yunnan	0.03	0.03	0.02	1
530600	Zhaotong	Yunnan	0.26	0.07	0.06	6
530700	Lijiang	Yunnan	0.02	-0.05	-0.07	3
530800	Puer	Yunnan	0.05	0.11	0.09	4
530900	Lincang	Yunnan	0.06	0.32	0.3	4
532300	Chuxiong Yi Autonomous Prefecture	Yunnan	0.21	0.12	0.09	4
532500	Honghe Hani and Yi Autonomous Prefecture	Yunnan	0.08	0.40	0.39	6
532600	Wenshan Zhuang and Miao Autonomous Prefecture	Yunnan	0.02	0.17	0.16	6
532800	Xishuangbanna Dhai Autonomous Prefecture	Yunnan	0.05	0.30	0.27	2
532900	Dali Bai Autonomous Prefecture	Yunnan	0.04	0.28	0.26	7
533100	Dehong Dai and Jingpo Autonomous Prefecture	Yunnan	0.02	0.21	0.2	4
533300	Nujiang Lisu Autonomous Prefecture	Yunnan	0.03	0.89	0.88	0
533400	Diqing Tibetan Autonomous Prefecture	Yunnan	0.03	0.26	0.24	1
540100	Lhasa	Tibet	0.03	0.22	0.21	3
540200	Xigaze	Tibet	0.01	0.21	0.21	0
540300	Chamdo	Tibet	0.01	0.21	0.2	0

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
540400	Nyingchi	Tibet	0.01	0.21	0.21	0
540500	Shannan	Tibet	0.04	0.14	0.14	0
540600	Nagqu	Tibet	0.01	0.17	0.14	2
542500	Ngari	Tibet	0.01	0.34	0.33	0
610100	Xian	Shaanxi	0.41	0.23	0.2	13
610200	Tongchuan	Shaanxi	0.1	0.48	0.44	1
610300	Baoji	Shaanxi	0.11	0.15	0.13	4
610400	Xianyang	Shaanxi	0.14	0.08	0.07	9
610500	Weinan	Shaanxi	0.08	0.31	0.28	6
610600	Yanan	Shaanxi	0.03	0.00	0	7
610700	Hanzhong	Shaanxi	0.04	-0.02	-0.03	3
610800	Yulin	Shaanxi	0.05	0.28	0.26	2
610900	Ankang	Shaanxi	0.04	0.07	0.06	7
611000	Shangluo	Shaanxi	0.05	0.25	0.23	4
620100	Lanzhou	Gansu	0.68	0.71	0.7	8
620200	Jiayuguan	Gansu	0.16	0.13	0.07	1
620300	Jinchang	Gansu	0.12	0.58	0.55	2
620400	Baiyin	Gansu	0.06	0.44	0.42	6
620500	Tianshui	Gansu	0.05	0.22	0.2	2

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
620600	Wuwei	Gansu	0.07	0.33	0.31	5
620700	Zhangye	Gansu	0.02	0.29	0.25	6
620800	Pingliang	Gansu	0.03	0.28	0.27	7
620900	Jiuquan	Gansu	3.05	3.05	3.04	5
621000	Qingyang	Gansu	0.03	0.19	0.17	6
621100	Dingxi	Gansu	0.05	0.28	0.25	7
621200	Longnan	Gansu	0.06	0.45	0.43	6
622900	Linxia Hui Autonomous Prefecture	Gansu	0.03	0.15	0.14	4
623000	Gannan Tibetan Autonomous Prefecture	Gansu	0.02	0.25	0.25	1
630100	Xining	Qinghai	0.11	0.36	0.34	5
630200	Haidong	Qinghai	0.05	0.32	0.31	4
632200	Haibei Tibetan Autonomous Prefecture	Qinghai	0.01	0.22	0.22	2
632300	Huangnan Tibetan Autonomous Prefecture	Qinghai	0.02	0.30	0.3	0
632500	Hainan Tibetan Autonomous Prefecture	Qinghai	0.01	0.27	0.26	0
632600	Golog Tibetan Autonomous Prefecture	Qinghai	0.01	0.20	0.19	0
632700	Yushu Tibetan Autonomous Prefecture	Qinghai	0.01	0.21	0.21	0
632800	Haixi Mongolian and Tibetan Autonomous Prefecture	Qinghai	0.12	0.42	0.41	4
640100	Yinchuan	Ningxia	0.09	0.16	0.13	4
640200	Shizuishan	Ningxia	0.09	0.38	0.36	3

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
640300	Wuzhong	Ningxia	0.14	0.54	0.52	3
640400	Guyuan	Ningxia	0.03	0.27	0.25	5
640500	Zhongwei	Ningxia	0.31	0.85	0.84	2
650100	Urumqui	Xinjiang	0.22	0.45	0.44	6
650200	Karamay	Xinjiang	0.06	0.25	0.21	4
650400	Turpan	Xinjiang	0.02	0.32	0.31	3
650500	Hami	Xinjiang	0.03	0.35	0.33	1
652300	Changji Hui autonomous prefecture	Xinjiang	0.08	0.32	0.31	8
652700	Bortala Mongolian Autonomous Prefecture	Xinjiang	0.03	0.18	0.16	4
652800	Bayingolin Mongolian Autonomous Prefecture	Xinjiang	0.03	0.26	0.23	7
652900	Aksu	Xinjiang	0.04	0.24	0.21	5
653000	Kizilsu Kyrgyz Autonomous Prefecture	Xinjiang	0.01	0.35	0.34	3
653100	Kashgar	Xinjiang	0.01	0.20	0.19	10
653200	Hotan	Xinjiang	0.01	0.19	0.19	2
654000	Ili Kazakh Autonomous Prefecture	Xinjiang	0.07	0.28	0.27	12
654200	Tacheng	Xinjiang	0.02	0.15	0.13	4
654300	Altay	Xinjiang	0.01	0.24	0.21	3
659001	Shihezi	Xinjiang	0.28	0.54	0.5	3
659002	Aral	Xinjiang	0.01	0.20	0.19	2

Table B.3.5. Neighbourhood agglomeration measures

Prefecture-level division code	Prefectures	Provinces	Weighted Synthetic Geographical Spatial Concentration Indicators $WSGSCI$	Weighted Statistical Synthetic Geographical Spatial Concentration Indicators $WSSGSCI$		Number of EDZs
				With EDZs	Without EDZs	
659003	Tumshuq	Xinjiang	0.01	0.13	0.07	2
659004	Wujiaqu	Xinjiang	0.05	0.18	0.18	2
659005	Beitun	Xinjiang	0.01	0.29	0.28	1
659006	Tiemenguan	Xinjiang	0.06	0.40	0.4	1
659007	Shuanghe	Xinjiang	0.01	0.17	0.17	0
659008	Kokdala	Xinjiang	0.01	0.16	0.15	0
659009	Kunyu	Xinjiang	0.02	0.33	0.33	0
659010	Huyanghe	Xinjiang	0.07	0.35	0.35	0
659011	Xinxing	Xinjiang	0.01	0.21	0.2	0

Table B.3.6. Percentage of firms of the ORBIS sample in EDZs by sectors and provinces in China

Sectors	% firms		Provinces	% firms
13	13.8		11	17.4
14	15.4		12	19.4
15	14.8		13	13.0
16	16.8		14	4.9
17	16.6		15	13.5
18	15.3		21	10.1
19	16.4		22	15.9
20	12.4		23	14.1
21	17.4		31	29.1
22	16.8		32	16.0
23	16.6		33	21.6
24	14.2		34	16.3
25	16.3		35	13.6
26	18.4		36	20.7
27	24.2		37	19.2
28	19.6		41	11.6
29	18.0		42	15.1
30	13.3		43	11.2
31	14.9		44	20.9
32	17.5		45	15.8
33	17.8		46	6.1
34	17.7		50	11.3
35	18.6		51	20.3
36	20.4		52	10.5
37	18.1		53	14.5
38	18.9		54	2.5
39	21.6		61	9.2
40	19.0		62	8.5
41	14.8		63	5.3
42	13.9		64	23.4
43	15.2		65	10.6
Average	**16.7**		**Average**	**16.7**

Chapter 3. Co-Agglomeration distribution of economic activities in the city of Madrid.

3.1. Introduction

The concept of agglomeration economies has been fundamental in understanding the geographic distribution of economic activities, urban development, and regional economics. From Marshall (1890) to the classification by Duranton & Puga (2004), agglomeration economies have evolved to encompass a range of contributing factors. The significance of agglomeration lies in its ability to enhance firm productivity, foster innovation, and attract more businesses and labour to cluster. Specifically, firms within the same industry often locate near each other (Ellison & Glaeser, 1999). Since industry clustering has been validated by various metrics (Ellison & Glaeser, 1997; Duranton & Overman, 2005; Marcon & Puech, 2010; Scholl & Brenner, 2016; Kopczewska et al., 2019), researchers have increasingly turned their attention to the heterogeneity and co-location of diverse activities, rather than solely the agglomeration of firms within the same sector[42].

Various factors contribute to the co-location of different economic activities. One primary reason is the presence of vertical supplier-consumer relationships, where minimising transportation costs is a central consideration (Fujita et al., 1999). Other mechanisms include shared location preferences due to similar natural resource dependencies (Ellison & Glaeser, 1999); proximity to clients, suppliers, or specialised labour markets (Combes & Duranton, 2006; Dahl & Klepper, 2015); enhanced worker-firm matching (Helsley & Strange, 1990); technological spillovers between related sectors (Saxenian, 1996; Glaeser & Kahn, 2001); the co-consumption of technologically distinct products (Tabuchi & Yoshida, 2000); and public policy interventions such as industrial parks and development zones (Nathan & Overman, 2013).

Ellison & Glaeser (1997) introduced the term "co-agglomeration" to describe the spatial

[42] A particularly notable study in this context, conducted by Casanova et al., (2017), examines the co-location of interrelated industries through horizontal and vertical relationships, highlighting the importance of company size as a relevant factor.

concentration of firms across different or associated industries. In much of the literature, intra-industry agglomeration refers to the geographic concentration of firms within the same sector, whereas inter-industry agglomeration involves firms from different sectors clustering geographically. This distinction yields two related but distinct concepts of sectoral co-agglomeration, each with unique implications: global co-agglomeration and bilateral co-agglomeration. In any case, co-agglomeration reflects inter-sectoral synergies, which can be analysed either globally or bilaterally. Bilateral synergies refer to the interactions between two specific sectors, whereas global synergies encompass the relationships between a particular activity and all other sectors collectively.

Global co-agglomeration reflects the ability of firms in one industry to attract productive activity from various sectors without differentiating between them. This approach is particularly useful for industrial policy, serving as a catalyst for business creation, attraction, and urban planning. Bilateral co-agglomeration, on the other hand, restricts this attraction capacity to a single other sector, establishing specific inter-industry relationships. This method is well-suited for targeted industrial policies with specific goals, such as attracting technology firms across various sectors. Its utility lies in offering precise explanations for clustering based on inter-industry connections. Both approaches employ similar measurement methods. In global co-agglomeration, all neighbouring production points are considered, while bilateral co-agglomeration only accounts for those within the targeted sector. These measures are applied to firms within specific industries, adapting second and third-generation agglomeration indicators to assess the phenomenon effectively.

This is exemplified in studies by Ellison & Glaeser (1997) and Devereux et al. (2004), which apply second-generation indicators, as well as in the work of Duranton & Overman (2005, 2008) and Marcon & Puech (2010), who adapt third-generation indicators through the use of microdata. Ellison et al. (2010) compare the results obtained using both second and third-generation indicators, concluding that the differences can be substantial. Their findings underscore the influence of data type on results and the importance of recognising issues such as the modifiable area unit problem in these measurements.

Most studies on this topic have focused on bilateral co-agglomeration analysis, sharing, to some extent, the same limitations and complexities observed in sectoral agglomeration as discussed in the first chapter of this thesis. In response, this chapter adapts the distance-based family of indicators, originally developed at the point level in the first chapter, to address global co-agglomeration specifically. However, these indicators can also be easily applied to measure the bilateral aspect of this phenomenon. Additionally, complementary and equivalent tools to those used for agglomeration measurement are introduced. Empirically, this methodology is applied to data from the Census of Premises provided by the Madrid City Council for the years 2014, the earliest available, and 2019, the last year prior to the COVID-19 pandemic.

The chapter is organised into six sections. Following this introduction, the second section provides an overview of the most widely used methods for analysing co-agglomeration in the literature. The third section details the adaptation of the proposed indicators for measuring global co-agglomeration. The fourth section describes the data used for the empirical application in the case of Madrid. In the fifth section, the results are discussed. The final section presents the main conclusions of the chapter.

3.2. Literature review

The measurement of spatial co-agglomeration of economic activity is a relatively recent focus within regional economics and economic geography. Helsley & Strange (2014) note that the term "co-agglomeration" scarcely appeared in the literature until it was introduced by Ellison & Glaeser (1997) and later used in Dumais et al. (2002). Duranton & Overman (2005) used the term "co-localisation" to describe situations where industries co-agglomerate due to mutual attraction, while reserving "joint-location" for instances where significant co-agglomeration between two industries occurs for other reasons. In this thesis, "co-agglomeration" will be equated with "co-location", identified with second-order concentration, or the joint location of two production points due to productive or demand-based interrelationships. This approach distinguishes it from "joint-location", or first-order concentration, which refers to co-location in a territory driven by opportunity or chance. Duranton & Overman (2002, 2005) underscore the importance of this distinction in agglomeration indicators, recommending that productive agglomeration measures focus exclusively on this second-order concentration.

The measures traditionally used to assess agglomeration, referred to in the literature as intra-industry geographical concentration, can be readily adapted to measure co-agglomeration, or inter-industry geographical agglomeration. This adaptability applies to both discrete indicators (e.g., Ellison & Glaeser, 1997; Helsley & Strange, 2014; Howard et al., 2016; Faggio et al., 2017, 2020; Diodato et al., 2018; Steijn et al., 2022) and distance-based indicators (e.g., Duranton & Overman, 2005; Marcon & Puech, 2010; Lang et al., 2020). Within this framework, Ellison & Glaeser (1997) introduced two distinct concepts: bilateral co-agglomeration and joint agglomeration of related industries. The latter aligns closely with the concept of global co-agglomeration measured here, though with a notable distinction. In Ellison & Glaeser's approach, all establishments from related industries are aggregated into a single sector, and the intra-sector agglomeration is then measured within this unified context. In contrast, the method proposed here evaluates the co-agglomeration capacity of establishments across all activities, centred on the influence exerted by a specific sector.

As noted by Combes et al. (2008), with the extensive micro-economic data available today and the advancements in computational capacity, distance-based approaches should be prioritised when measuring concentration phenomena in productive activity, whether it pertains to intra-sectoral agglomeration, inter-sectoral agglomeration, or overall clustering. Consequently, Table 3.1 highlights the distance-based indicators adapted and employed in the literature specifically for co-agglomeration, though primarily from a bilateral (inter-industry) perspective.

Table 3.1 Main co-agglomeration indicators proposed in the literature and their typology in relation to those previously presented

Measure References	First step					Second step	Third step	Fourth step	Fifth step
	Type counting the neighbours (a)	Weighted by the density (b)	Weighted activity size (c)	Weighted by the distance (d)	Border-effect correction (e)	Average measure	Local reference	Global reference	Null hypothesis (*)
$K_{(A,B)}(d)$ Duranton & Overman (2005)	Density (Bilateral distance)	NO	NO	YES	NO	Non-weighted	Absolute	Number of points	PRAAL
$K_{(n,m)}(d)$ Duranton & Overman (2008)	Density (Bilateral distance)	NO	NO	YES	NO	Non-weighted	Absolute	Number of points	PRAAL
$M_{s_1,s_2}(r)$ $M_{s_2,s_1}(r)$ Marcon & Puech (2010)	Cumulative (Bilateral distance)	NO	YES	NO	NO	Non-weighted	Relative	Global adjustment	PRAAL
$m(r)$ Lang et al. (2020)	Density (Bilateral distance)	NO	YES	YES	NO	Non-weighted	Relative	Global adjustment	PRAAL

(*) PRAAL: Points are redistributed across actual location

3.3. A distance-based point-level and derived aggregate indicators for sectoral co-agglomeration

3.3.1. Distance-based point-level indicators for sectoral co-agglomeration

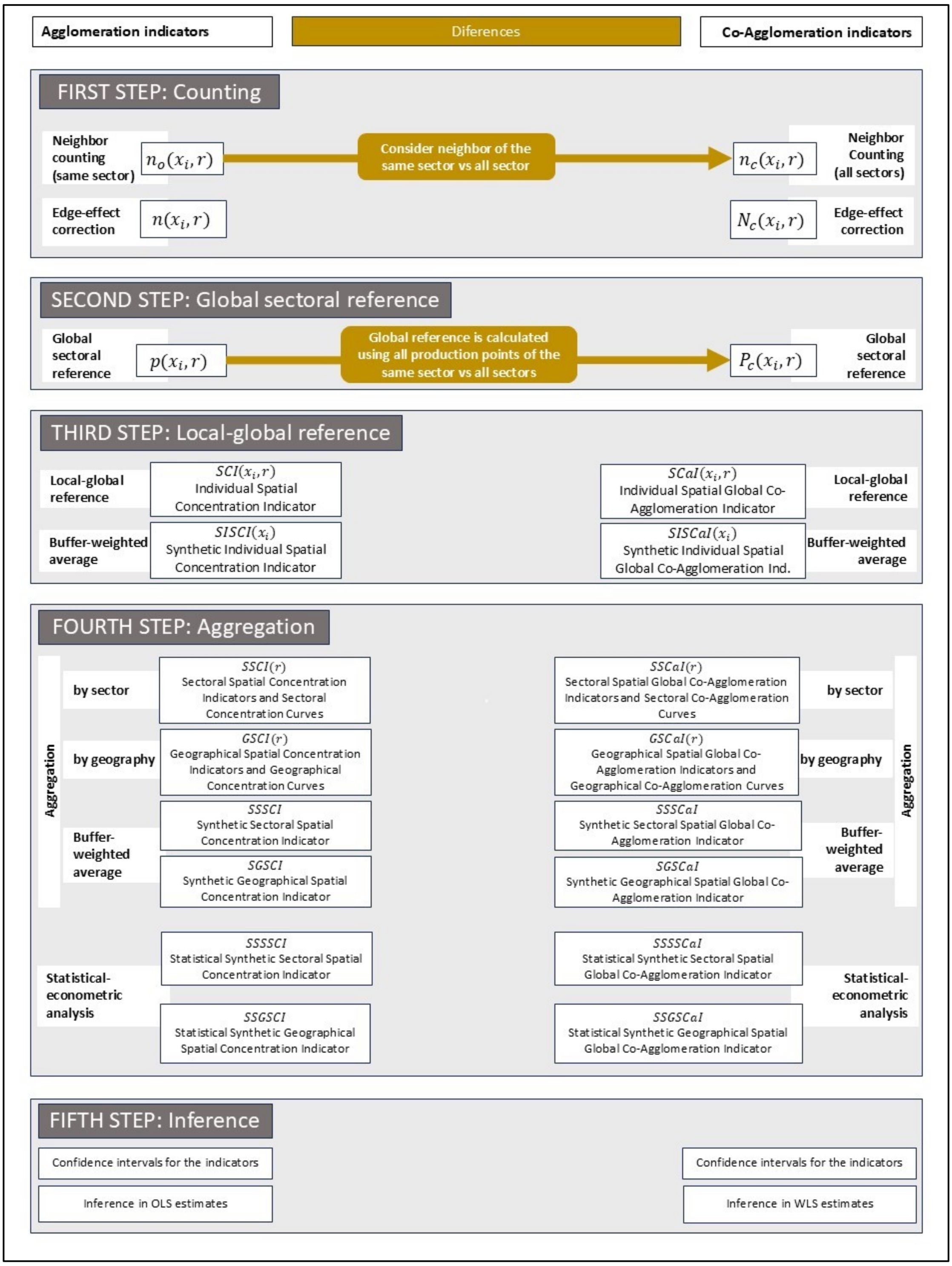
In a manner similar to the agglomeration indicators, global co-agglomeration indicators can be defined at the production point level for each sector and geographical unit (e.g., neighbourhood, district, raster). The procedure for deriving sectoral and geographic global co-agglomeration indicators aligns with the stylised methodology developed in Chapter 1, following the typology outlined by Marcon & Puech (2017). Figure 3.1 illustrates the adaptation of the indicator family from Chapter 1 to the co-agglomeration framework introduced in this chapter.

Consider *A* as an administrative area (city, region, country, etc.) containing *N* geographical points of activity (establishments, premises, local units of activity, etc.) from various sectors, each located within one of the administrative divisions that compose *A*. Let N_s represent the total number of production points for a specific sector within *A*, and N_g denote the number of points within any given geographical division *g* (e.g., neighbourhood) within *A*. To define an indicator of global co-agglomeration, the initial step involves counting the neighbouring points of any sector or activity around each reference point, within a maximum distance *r*, as in expression [1].

$$n_c(x_i, r) = \sum_{\substack{j \neq i \\ i \in s \\ i,j \in A}} 1\left(\left\|x_j - x_i\right\| \leq r\right) \qquad [1]$$

The reference point is denoted as x_i, representing establishment *i* of sector *s*. Its generic neighbours from any sector are denoted as x_j, all located within the administrative area A. Thus, $n_c(x_i, r)$ represents the number of neighbouring points from any sector within area *A* at a maximum distance of *r* around the reference establishment x_i. In relation to the generic expression for counting neighbours (expression [1] in Chapter 1), and for simplicity or due to data limitations, the measure proposed here is unweighted by density ($w(x_j) = 1$), unweighted by the activity size of the neighbour ($z(x_j) = 1$), and unweighted by the distance between x_j and x_i ($k = 1$). These types of weights could be incorporated into the indicator if desired, or it could alternatively be defined as a density measure. Consequently, the global co-agglomeration measurements will depend on the buffer width considered.

Figure 3.1 Main differences in the calculation of agglomeration vs global co-agglomeration indicators



The number of neighbours in expression [1] is counted without edge-effect correction. Similar to the original measures, the inverse of the buffer area overlapping with the administrative study area, denoted as λ_{ir}, is applied to correct for edge effects. This overlap ratio is raised to an exponent ranging between 0 and 1. Consequently, the adjusted number of establishments or production points around the reference point is calculated according to expression [2].

$$N_c(x_i, n) = \frac{n_c(x_i,n)}{\lambda_{ir}^{\gamma_{sr}}} \qquad [2]$$

The exponent γ_{sr} varies among sectors and also depends on the buffer width. If γ_{sr} adopts the lower limit of the interval, such correction would not be required; on the contrary, if the value of γ_{sr} equals to 1, usual correction by the inverse of the area overlapping would be needed. The intermediate values would basically be placed in the mid-correction position. The dependence of γ_{sr} on the buffer width is similar to that in the original indicator developed to measure agglomeration. In other words, it is based on the relationship between the overlapping part of the buffer and the studied administrative area, which changes as the buffer width increases. Consequently, the assumption about the density of establishments outside the administrative area may need adjustment. For global co-agglomeration, however, the relationship with the sector is more debatable. With sufficiently large buffers, all production points in the administrative area will be counted, and a similar proportion of points will fall outside the overlap area for all reference points. For small and medium buffers, which are most relevant for measuring this process, the relative location of reference production points within the administrative area can vary significantly between activities. In any case, as this is a data-driven correction, the data itself will provide the most suitable way to perform it.

The value of γ_{sr} for each sector and buffer is approximated by establishing a relationship between the corrected number of points for each activity and their qualitative characteristics, z_i, which may influence potential measurement errors in concentration levels in their surroundings (e.g., whether the production point is part of a commercial centre), using fixed effects for both sector (d_s) and buffer (d_r). Consistent with the methodology outlined in previous chapters, the final expression to be estimated can be derived, taking the form of expression [3].

$$\ln n_0(x_i, n) = \alpha + \beta_1 z_i + \gamma \ln(\lambda_{ir}) + \sum_s \omega_s d_s + \sum_r \mu_r d_r + \sum_s \vartheta_s \ln(\lambda_{ir})\, d_s +$$

$$+ \sum_r \nu_r \ln(\lambda_{ir})\, d_r + \varepsilon_{ir} \qquad [3]$$

So that, $\gamma_{sr} = \gamma + \vartheta_s + \nu_r$.

The second step involves establishing a global reference, allowing the number of neighbours around the reference point to be compared to the total number of production points within the administrative area. In other words, this step controls for global co-agglomeration across the entire study area and is implemented through expression [4].

$$P_c(x_i, r) = \frac{N_c(x_i, n)}{N \Big/ \left(\frac{1}{N} \sum_{j \in A} \lambda_{ir}^{\gamma_{sr}}\right)} \qquad [4]$$

Therefore, the proportion of production points within the administrative area around the reference point x_i at a distance *r*, $P_c(x_i, r)$, is obtained by dividing the number of neighbours, corrected for edge effects, by the total number of production points in the administrative area, also corrected. This correction is obtained as the average of all corrections applied to each point within the studied area. Note that the reference here is the set of production points in the studied area, as the numerator includes neighbours from all sectors.

In the third step, we obtain a Spatial Global Co-agglomeration Indicator (*SCaI*) at the point level for a buffer *r* by normalising expression [4] by the mean of the same measure for all points within the administrative area, using the same buffer width, i.e., expression [5].

$$SCaI(x_i, r) = \frac{P_c(x_i, r)}{\overline{P_c(r)}} \qquad [5]$$

Where $\overline{P_c(r)} = \frac{1}{N}\sum_{j\in A} P_c\left(x_j, r\right)$ represents the average of the relative frequency with which production points of the administrative area are found, at a maximum distance r around each activity point. With this local-global reference, the first aspect of first-order concentration, i.e., the generic joint-location, is controlled.

The value of the above-mentioned indicator depends on the width of the buffer used. To obtain a unique indicator of co-agglomeration at the point-level, expression [6] provides a Synthetic Individual Spatial Global Co-Agglomeration Indicator ($SISCaI$) at the point-level which does not rely on distance, by aggregating weighted indicators for different buffer widths.

$$SISCaI(x_i) = \frac{\sum_r SCaI(x_i, r) f(r)}{\sum_r f(r)} \quad [6]$$

Where $f(r) = \left(\frac{r_0}{r}\right)^g$ is a function of distance, *g* can take on values greater than zero, such as 0.5, 1, 2, etc. Here, r_0 denotes the radius of the smallest buffer width under consideration, while *r* covers the entire range of buffers used to calculate the preceding indicators. This approach ensures that the results obtained within buffers closer to the analysed point are considered more relevant in determining the degree of co-agglomeration around each point. Consequently, the set of indicators $SCaI(x_i, r)$ and $SISCaI(x_i)$ both serve as point-level measures of co-agglomeration.

3.3.2. Sectoral Global Co-agglomeration Indicators and derived curves

In the fourth step, sectoral global co-agglomeration indicators (or by geographical unit, such as neighbourhood) are obtained by averaging the point-level indicators across all establishments within each sector. This results in the Sectoral Spatial Global Co-Agglomeration Indicator[43] ($SSCaI(r)$) for each buffer width, calculated as in

[43] The construction of Geographical Spatial Global Co-agglomeration Indicators ($GSCaI(r)$) is obtained by aggregating the individual indicators for all production points located within the geographical area for which the indicator is calculated (in this case, by neighbourhood).

$$GSCaI(r) = \frac{1}{N_g}\sum_{i\in g} SCaI(x_i, r) \quad [7b]$$

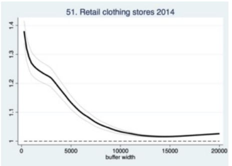
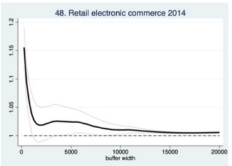
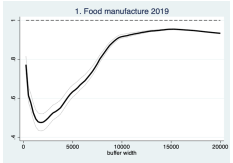
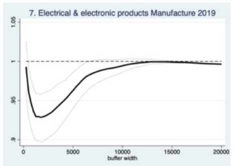
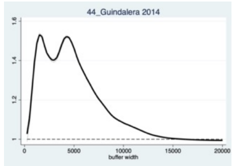
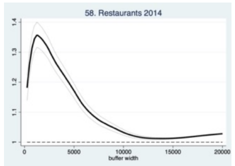
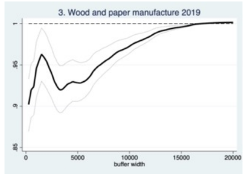
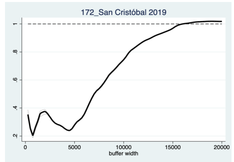
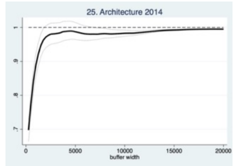
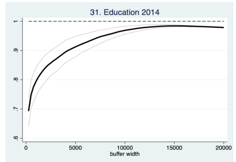
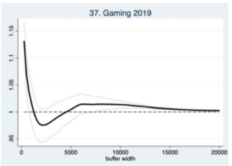
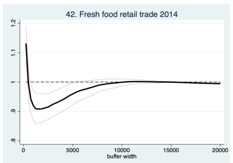
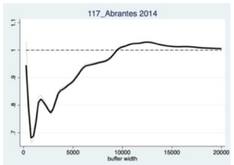
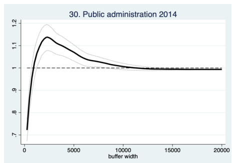
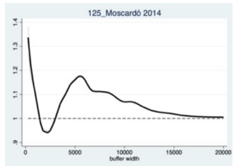
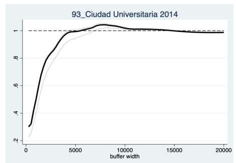
Where g refers to the geographical area (e.g., neighbourhood) for which the indicator is calculated, and N_g represents the total number of firms across all activities within g. It is important to note that each activity may have its own specific geographical indicator.

expression [7].

$$SSCaI(r) = \frac{1}{N_s}\sum_{i \in s} SCaI(x_i, r) \quad [7]$$

For each sector (or geographic area), a relative indicator of spatial global co-agglomeration for productive activity is calculated at each buffer width. By arranging these indicators according to buffer width, global co-agglomeration curves for each activity (or geographic area) can be obtained. Table 3.2 presents the classification of these curves, organised similarly to previous chapters, illustrated with actual co-agglomeration curves for the specific case of Madrid.

Table 3.2 Typology of Spatial Global Co-agglomeration Curves (*)

	Decreasing or L	U or V Double U or V	Inverse U or V Double U or V	Double U and V or Inverse	Increasing or inverse L
Do not cross the unit	**Type 1(A)**	**Type 2(D)**	**Type 3(A)**	**Type 4(D)**	**Type 5(D)**
	 	 	 	 	 
Do cross the unit	**Type 6(A)**	**Type 7(D)**	**Type 8**	**Type 9**	**Type 10(D)**
		 			

Source: Own elaboration.

(*) Real examples obtained for the case of Madrid in 2014 and 2019.

By averaging the Synthetic Individual point-level indicators, it possible to obtain Synthetic Sectoral Spatial Global Co-agglomeration Indicators[44], $SSSCaI$.

$$SSSCaI = \frac{1}{N_s}\sum_{i \in s} SISCaI(x_i) \qquad [8a]$$

The final step involves testing the statistical significance of the previously calculated aggregated indicator. In a similar approach to the indicators introduced in Chapter 1, confidence intervals are constructed for each aggregated co-agglomeration indicator based on production point-level values (rather than around a null hypothesis). This approach checks whether the reference value of no co-agglomeration (in this case, the value one) falls within these confidence intervals, avoiding reliance on Monte Carlo simulations that assume a completely random distribution of production points. Instead, the intervals are derived from representative subsamples of the point-level indicators. Consequently, if the reference value lies within the confidence interval, we cannot reject the hypothesis of no co-agglomeration in the sector. The same procedure applies to synthetic indicators.

3.3.3. Synthetic indicators from statistical-econometric analysis

The aggregated indicators do not fully utilize the individual point-level information. Specifically, in the previous indicator, the second aspect of first-order concentration, i.e., local joint-location, is not controlled. To address this, least squares regressions are estimated using point-level indicator values $SCaI(x_i, r)$ and $SISCaI(x_i)$. By estimating one or both of expressions [9a] and [9b], it is possible to obtain a sectoral and geographic average indicator.

$$SCaI(x_i, r) = \alpha + \sum_s \alpha_s^1 d_{si} + \sum_q \beta_q^1 d_{qi} + \delta^1 z_i + \varepsilon_{isr}^1 \qquad [9a]$$

$$SISCaI(x_i) = \alpha + \sum_s \alpha_s^2 d_{si} + \sum_g \beta_g^2 d_{gi} + \delta^2 z_i + \varepsilon_{isr}^2 \qquad [9b]$$

[44] The Synthetic Geographical Spatial Co-agglomeration Indicators ($SGSCI$) are obtained by averaging the synthetic Individual point-level co-agglomeration indicators across all points within the same geographical unit, as in expression [8b].

$$SGSCaI = \frac{1}{N_g}\sum_{i \in g} SISCaI(x_i) \qquad [8b]$$

The dummy variables d_{si} and d_{qi} represent the sectors and geographical units (such as neighbourhoods) where each point is located. Meanwhile, z_i captures establishment features, such as whether the establishment is part of a shopping centre. The estimation of equation [9a] will be carried out using Weighted Least Squares (WLS), with *f(r)* serving as a weight. The results from the two estimations yield identical coefficients, although the standard errors differ.

From the coefficients α_s^1 or α_s^2, the Statistical Synthetic Sectoral Spatial Global Co-agglomeration Indicator[45] (*SSSSCaI*) is calculated. Similarly, β_q^1 and β_q^2 are used to obtain the Statistical Synthetic Geographical Spatial Global Co-agglomeration Indicators[46] (*SSGSCaI*).

Finally, it is worth noting that the family of indicators developed for measuring global co-agglomeration, similar to those in the first chapter, meets the desirability criteria outlined by Duranton & Overman (2002, 2005), Kominers (2008), and Kopczewska et al. (2019).

3.4. Data

3.4.1. Study area

Madrid, the case study and capital city of Spain, is located in the centre of the Iberian Peninsula and is the country's most populated city (INE, 2023). Since 2017, Madrid has been organized into 21 districts, divided into 131 neighbourhoods, and spans an area of 604.3 square Kilometres.

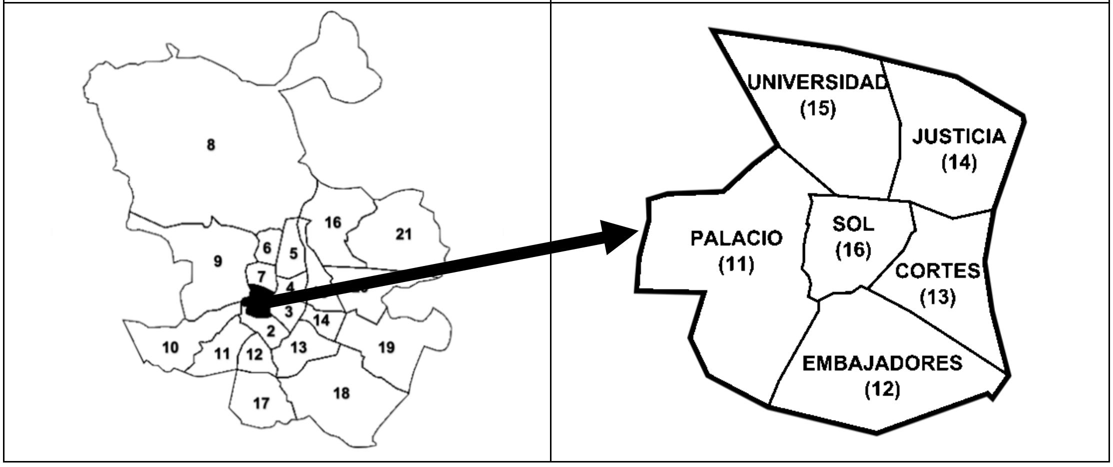
---

[45] To obtain the final value of the Statistical Synthetic Sectoral Spatial Global Co-agglomeration Indicator (*SSSSCaI*) for each sector, the coefficient from the corresponding sectoral dummy is added to the constant term and the weighted average of neighbourhood dummies.

[46] Similarly, to obtain the final Statistical Synthetic Global Geographical Spatial Co-agglomeration Indicator (*SSGSCaI*), the coefficient from each area's corresponding dummy is added together with the constant term and the weighted average of the sector dummies.

Figure 3.2 Administrative units in Madrid

Districts of Madrid	Neighbourhoods in Centre District



3.4.2. Data source

The empirical application uses micro-geographic data from the Census of Premises and Activities of the Madrid City Council. This administrative database includes records of all street-level premises in Madrid, providing information on associated activities, such as address (latitude and longitude coordinates based on geocoded street addresses), opening status and activity category. This dataset offers up-to-date information, with quarterly updates from March 31, 2014, to December 31, 2014, and monthly updates since then. Notably, the database provides detailed information at the establishment level.

The data from December 2014 and December 2019 is used to analyse changes in global co-agglomeration patterns. The dataset has been aggregated into 62 sectors based on their declared activity categories. Establishments with no associated activities were excluded from the sample. Consequently, the final dataset comprises 99,939 establishments for 2014 and 110,997 establishments for 2019. A detailed description of the sectoral aggregation can be found in Appendix A of Chapter 1. In addition to the precise location and activity type of each establishment, the dataset also indicates whether an establishment is part of a geographical grouping (e.g., a commercial centre). However, no accounting or economic data is available for any of the establishments.

3.5. Results

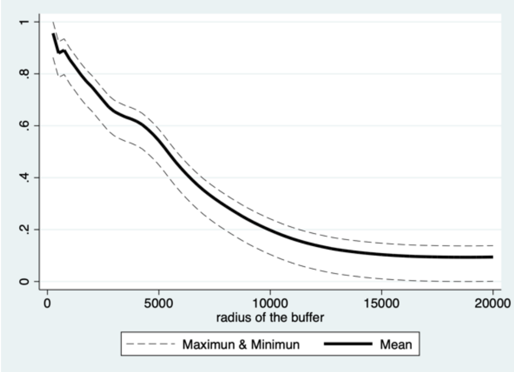
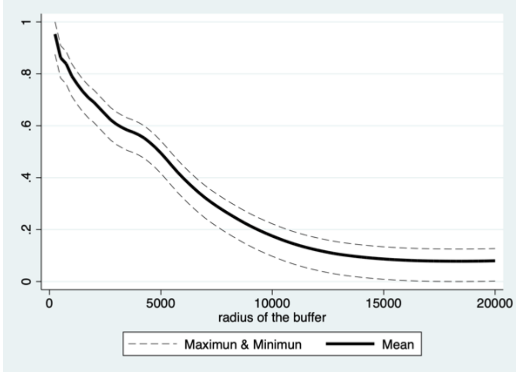
3.5.1. Calculation of the distance-based point-level indicators for sectoral global co-agglomeration

The number of establishments around each activity point within a specified maximum distance is calculated for buffers ranging from 250 meters to 20,000 meters in radius. A total of 80 different buffer widths are considered[47]. To determine the exponent γ_{sr}, which adjusts the standard correction for the inverse of the overlap between the buffer under consideration and the administrative area, expression [3] is estimated independently for each year. The only known characteristic of the establishment included in the model is whether the point is located within a shopping centre. Additionally, sectoral dummies (61 representing all sectors considered, minus one to avoid collinearity) and neighbourhood dummies (130, equivalent to all neighbourhoods minus one) are incorporated. Finally, the resulting coefficient values are normalised to a range of 0 to 1. Figure 3.3 presents the results for this exponent. As shown in Panel A, the exponent decreases as the buffer size increases, although it does not reach zero on average. This indicates that, as expected, the correction for the inverse of the overlap is overstated, and more importantly, it varies with the buffer width. Thus, the standard assumption may be acceptable for smaller buffers. Larger buffer widths reduce the need for correction because a higher proportion of overlapping area (λ_{ir}) increases the probability of encountering lower density outside the buffer. This finding aligns with the context of a monocentric city, where firms tend to cluster in the center. Conversely, the sector of activity appears to have no significant effect on the mean correction, as observed in Panel B, which shows box plots of the exponent by sector. Moreover, the results are consistent across the two years analysed.

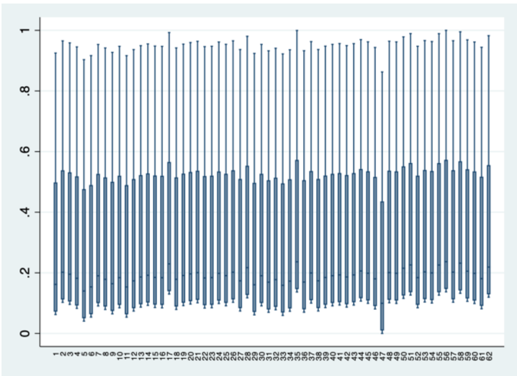
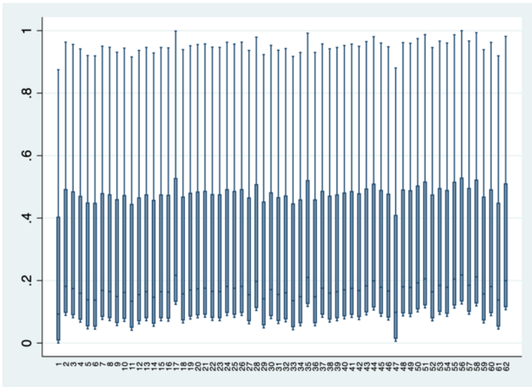
[47] The distance increment is set at 250 meters, starting from 250 meters and then increasing to 500 meters, 750 meters, and so forth. A distance of 20,000 meters corresponds to the 99th percentile of bilateral distances between establishments in the city of Madrid.

Figure 3.3 Overlapping correction moderator by buffer width and sector

A. Evolution of overlapping correction moderator (γ_{sr}) by buffer width

2014	**2019**
	

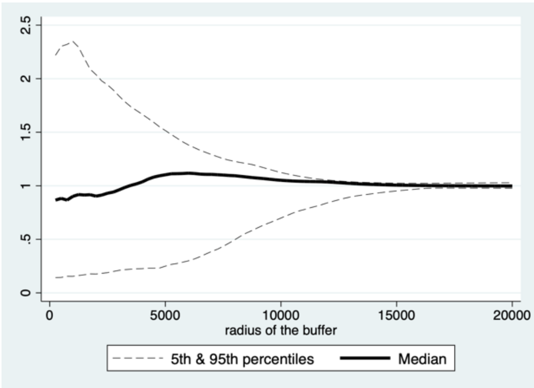
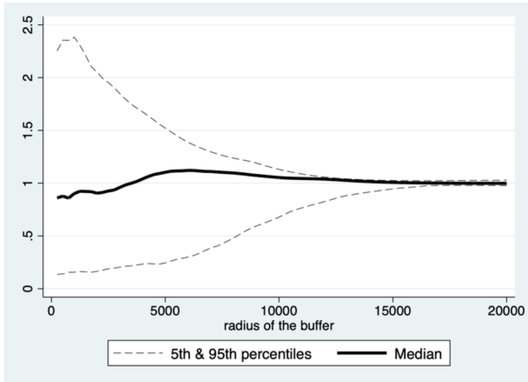
B. Boxplots of the overlapping correction moderator (γ_{sr}) by sector

2014	**2019**
	

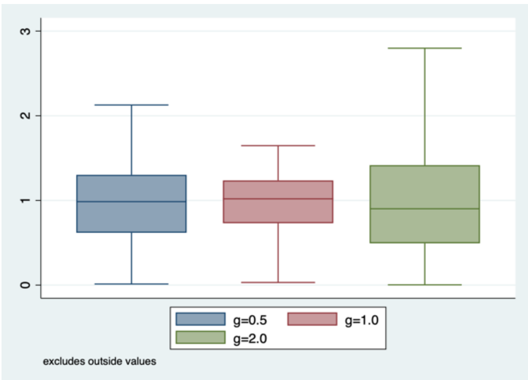
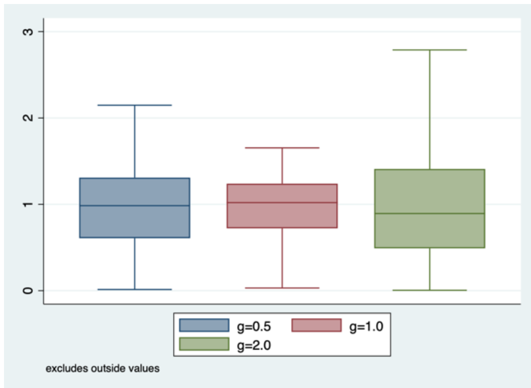
After estimating these exponents, the Individual Spatial Global Co-Agglomeration Indicators (*SCaI*) can be calculated at the establishment or production point level for each buffer (see panel A of Figure 3.4). Additionally, synthetic indicators are constructed by applying different weighting functions ($g \in \{0.5,1.0,2.0\}$) to the indicators obtained for each buffer (see panel B of Figure 4).

Figure 3.4 Distance-based point-level indicators for sectoral co-agglomeration

A. Individual spatial co-agglomeration indicators, $SCaI(x_i, r)$

2014	2019
	

B. Global individual spatial co-agglomeration indicators, $SISCaI(x_i)$

2014	2019
	

As expected, as the buffer size increases, the values of $SCaI(x_i, r)$ converge towards 1. This outcome occurs by design: expanding the buffer width incorporates a larger number of activity points into the numerator of the indicator, while the denominator includes all activity points within the studied area, resulting in a tendency toward unity. However, it is observed that the median value remains below 1, at least up to buffers of 5,000 meters. Additionally, the values at the 95th percentile converge to the reference value faster than those at the 5th percentile. These effects are inherent to the indicator's definition: values below 1 are confined to a range between 0 and 1, while values above 1 have no upper limit. This asymmetry could be mitigated by using logarithmic indicators, which would set zero as the reference point. Moreover, the agglomeration process is concentrated in a few sectors (and production points), with agglomeration effects being more pronounced than dispersion effects.

Panel B of Figure 5 presents the box plots of the Synthetic Individual Spatial Global Co-agglomeration Indicators ($SISCaI(x_i)$). It is important to note that the selected weighting function significantly impacts the distribution of these indicators (results are provided for exponents 0.5, 1, and 2). For both years, when the exponent value is 1, the median of the indicator is higher, with lower dispersion among production points. Conversely, when the exponent value is 2, the median is lower, and the dispersion is greater. Moving forward, only the results obtained with an exponent value of 1 will be used.

3.5.2. Results at sectoral level

3.5.2.1. Sectoral global co-agglomeration curves

The Sectoral Spatial Global Co-agglomeration Indicators ($SSCaI(r)$) are calculated by aggregating the Individual Spatial Global Co-Agglomeration Indicators ($SCaI(x_i, r)$) for all production points within each sector. Using these aggregated indicators, sectoral co-agglomeration curves are constructed for the 62 activities considered, by arranging the indicators based on the buffer width (refer to Table C.1 in the Appendix C).

Table 3.3 presents the results regarding the typology of global co-agglomeration curves. The findings indicate that most establishments are in sectors with stable behaviours, i.e., curves that do not cross the unit threshold, irrespective of the buffer width considered (49 sectors in 2014 and 51 sectors in 2019). Generally, sectors exhibiting global co-dispersive behaviour (37 in 2014 and 38 in 2019) outnumber those displaying global co-agglomerative behaviour (18 in 2014 and 21 in 2019).

This result suggests that globally co-agglomerated sectors have a higher number of establishments, or that the global co-agglomerative behaviour, as captured by the indicator, is, on average, more intense than the dispersive behaviour. This is because the behaviour of establishments in each sector is evaluated relative to the average behaviour of all sectors combined. Notably, the findings are consistent across both years.

Table 3.3 Typology of sectoral co-agglomeration curves obtained for the city of Madrid.

		Decreasing or L	U or V	Inverse U or V	Double U and V	Increasing or inverse L
Do not cross the unit		**Type 1(A)**	**Type 2(D)**	**Type 3(A)**	**Type 4(D)**	**Type 5(D)**
	2014	9 sectors	7 sectors	8 sectors	3 sectors	22 sectors
	2019	9 sectors	9 sectors	9 sectors	2 sectors	22 sectors
Do cross the unit		**Type 6(A)**	**Type 7(D)**	**Type 8**	**Type 9**	**Type 10(D)**
	2014	1 sector	3 sectors	3 sectors	4 sectors	2 sectors
	2019	3 sectors	3 sectors	2 sectors	1 sectors	2 sectors

Note: (A) represents agglomeration behaviour; (D) refers to dispersion behaviour

3.5.2.2. Synthetic indicators

In the same way as with the agglomeration indicators, the behaviour of sectoral global co-agglomeration varies depending on the distance considered in the analysis. To address this, synthetic indicators are proposed. While the curves provide more detailed information, synthetic indicators simplify the analysis by generating a single value that represents the overall behaviour. Two distinct methods are available for calculating these synthetic sectoral global co-agglomeration indicators.

One method involves aggregating the synthetic global co-agglomeration indicators at the point level. This process results in the Synthetic Sectoral Spatial Global Co-Agglomeration Indicators (*SSSCaI*), derived through a two-step averaging process. First, a weighted mean is calculated based on the function $f(r)$ applied to the Individual Spatial Global Co-Agglomeration Indicators across different buffers, as described in expression [6]. Then, a simple arithmetic average of these indicators is computed for all the production points within the sector, as outlined in expression [8a]. The results for each sector, specifically for Madrid in the two selected years, are presented in the first two columns of Table C.2 in the Appendix C.

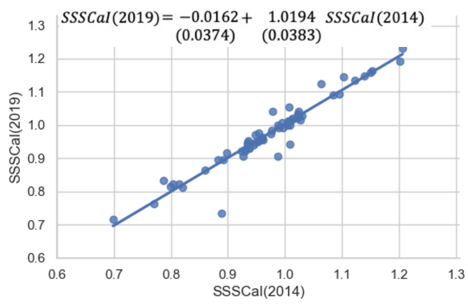
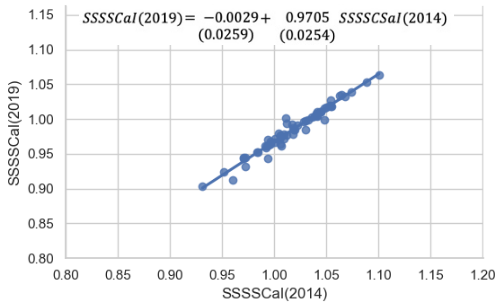
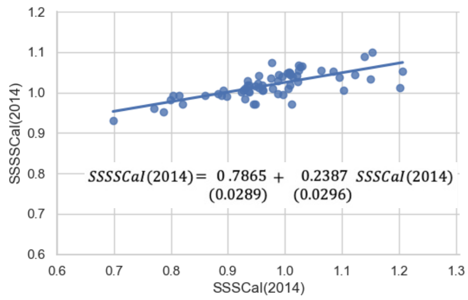
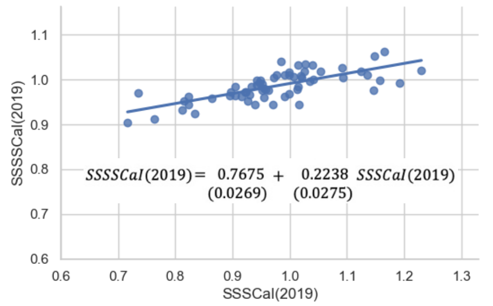
The other method involves estimating expressions [9a] or [9b]. In the first expression, the Individual Spatial Global Co-Agglomeration Indicators for different buffers are regressed against establishment characteristics, such as membership in a commercial centre, and sector and neighbourhood dummies. This estimation is weighted using the function $f(r)$. As a result, the Statistical Synthetic Sectoral Spatial Global Co-Agglomeration Indicator[48] (*SSSSCaI*) is obtained. The results are provided in the third and fourth columns of Table A.2 in the Appendix[49]. The main theoretical difference between the *SSSCaI* and *SSSSCaI* indicators is that the latter eliminates the second aspect of first-order concentration, namely the local joint-location. Consequently, this indicator exclusively captures co-location while entirely removing location bias. For example, if a sector has a greater presence in a neighbourhood due to its proximity to the city centre or some idiosyncratic feature of the geographic area, such as industrial parks, that fosters economic activity and presents a spatial attraction effect (but not an iterative effect), this effect will be excluded in the calculation.

As shown in Figure 3.5, the results of the *SSSCaI* and *SSSSCaI* indicators are statistically similar across the two years, indicating that sectoral co-agglomeration patterns have remained relatively stable between 2014 and 2019. However, a comparison of the two indicators within the same year reveals that they are not equivalent. In particular, the *SSSSCaI* indicators display lower variability than those derived from simple averages. This suggests that part of the global co-agglomeration captured by the *SSSCaI* is due to local joint-location. When this effect is removed, the *SSSSCaI* indicator is significantly adjusted, increasing or decreasing for sectors that co-locate in areas with lower or higher concentration levels, respectively, compared to the previous indicator.

[48] Thus, the Statistical Synthetic Sectoral Spatial Global Co-Agglomeration Indicator (*SSSSCaI*) is obtained by summing the coefficient associated with the corresponding sectoral dummy, the constant term, and the weighted average of neighbourhood dummies.

[49] The results obtained from the estimation of equation [9b] are, by construction, identical to those from the previous approach. However, the statistical significance of the indicators differs due to the reduced data set, i.e., only one value per production point compared to 80 values per production point in the earlier case.

Figure 3.5 Relationship between synthetic sectoral global co-agglomeration indicators

***SSSCaIs* in 2014 and 2019**	***SSSSCaIs* in 2014 and 2019**
	
***SSSCaIs* and *SSSSCaIs* in 2014**	***SSSCaIs* and *SSSSCaIs* in 2019**
	

SSSCaI refers to Synthetic Sectoral Spatial Co-agglomeration Indicators, while SSSSCaI refers to Statistical Synthetic Sectoral Spatial Co-agglomeration Indicators.

Based on the *SSSSCaI* values for 2014 and 2019, Table 3.4 classifies sectors into five groups according to their global co-agglomeration behaviour[50]. Industries related to everyday consumption, proximity to customers, or leisure exhibit higher levels of spatial global co-agglomeration. Conversely, land-intensive industries, particularly manufacturing sectors with high levels of pollution, tend to have lower geographically co-agglomerative attractiveness. Interestingly, creative sectors (such as graphic arts, media, research, and marketing) do not attract other sectors to globally co-agglomerate. This may stem from their lower need for face-to-face interaction with other activities or location-specific image considerations. Supporting this, Coll-Martinez et al. (2019) observed that creative and non-creative industries in Barcelona are not co-located.

[50] The values chosen for sector classification are determined on an ad-hoc basis.

Table 3.4 Classification of Sectors in 2014 and 2019 according to their intensity of global co-agglomeration.

		2019				
		High global co-agglomeration	**Medium-high global co-agglomeration**	**No global co-agglomeration No global co-dispersion**	**Medium-high global dispersion**	**High global dispersion**
2014	**High co-agglomeration**	Retail clothing stores Jewelry stores	Fresh food retail trade Perfumeries & drugstores Electronics stores Gaming Retail beverages Hardware & DIY stores Travel Agencies Sex shops & others Bakeries & bakeries Insurance	Fast food & self-service		
	Medium-high co-agglomeration		Telecommunications Bars with shows Personal Services Ice cream & dried fruit Non-special retail trade Stores	Garden & pets Bookstores & leisure Electrical & electronic (M) Restaurants Real estate Bars Street markets	Wholesale Trade Consulting Textile (M) Automobile (Trade) Other Manufacturing Health care activities Postal & delivery sector Entertainment Pharmacies Transportation Auxiliary services Wood & paper (M) Associations Construction Accommodation Social Services	
	No co-agglomeration, No dispersion				Sports Veterinarians Graphic arts Furniture (M) Research & marketing	Collective caterings Software Transport Material (M) Residences Libraries & museums Architecture Public Administration Supplies
	Medium-high dispersion				Media Food (M) Metal Products (M) Building Services Education	
	High dispersion					Chemicals (M) Petrol stations

Note: High co-agglomeration: $SGSSCI \geq 1.05$*; Medium-high co-agglomeration:* $1 < SGSSCI < 1.05$*; No co-agglomeration, no dispersion* $SGSSCI = 1$*; Medium-high dispersion:* $0.95 < SGSSCI < 1$*; High dispersion:* $SGSSCI \leq 0.95$.

Comparing the results in terms of global co-agglomeration with those obtained in Chapter 1 regarding agglomeration reveals four typologies of sectors and their respective transitions, as reflected in Table 3.5. Two typologies correspond to sectors maintaining the same relative positions in both types of indicators. Sectors that are both agglomerated and globally co-agglomerated typically involve activities related to non-daily shopping and leisure. Conversely, dispersed and globally co-dispersed sectors generally include activities not tied to consumers' everyday lives or, when they are (such as health and education), are characterised by occupying large, land-intensive buildings and facilities.

The other two categories correspond to sectors that exhibit agglomeration (or dispersion) but whose relative positioning is inverted in terms of global co-agglomeration. For instance, agglomerated activities that display global co-dispersion behaviour include sectors with limited establishments across the city, such as manufacturing activities and some land-intensive sectors. Surprisingly, accommodation falls within this group. While accommodation is concentrated in specific areas of the city, it shows notable global co-agglomeration only in the city centre. Other agglomeration points, such as those near the airport, the fairgrounds, and industrial parks, tend to associate more strongly with areas of lower activity density, resulting in an overall pattern of global co-dispersion.

Finally, the last major group comprises activities that are relatively dispersed in terms of agglomeration but demonstrate a global co-agglomeration behaviour. These activities, while spread throughout the city, are predominantly located in areas with a high concentration of other productive activities. This pattern may result from either a tacit or explicit distribution of market areas or from factors related to their location strategies, such as association with shopping centres, co-consumption patterns, the presence of large efficient production sizes, or strategies tied to membership in commercial chains.

Table 3.5 Relationship between agglomeration and global co-agglomeration types in Madrid, 2019

		Agglomeration in 2019		
		Agglomerated	**No agglomerated**	**Dispersed**
Global Co-Agglomeration in 2019	**Global co-agglomerated**	Retail clothing Jewellery stores Sex-shops Bars with shows Retail beverages Travel agencies	Insurance	Gaming Fresh Food retail trade Electronic stores Perfumeries Bakeries Telecommunications Hardware and DIY Non-specialised retail trade Ice-cream
	No global co-agglomerated	Street markets Restaurants Fast-food Graphic Arts	Bookstores	Garden and Pets Bars Electrical and electronics (M) Real estate
	Global dispersed	Food (M) Wood and paper (M) Furniture (M) Textile (M) Metal products (M) Chemicals (M) Transport Materials (M) Supplies Media Wholesale trade Accommodation Entertainment Software Libraries and museums Petrol stations Residences	Mail offices Other manufacturing Automobile (Trade) Building Services Architecture	Research and marketing Consulting Veterinarians Social services Public Administration Construction Transportation Act. Auxiliary Comp. Education Health care Sports Associations Pharmacies Collective caterings

3.5.3. Results at neighbourhood level

3.5.3.1. Global Co-agglomeration curves at neighbourhood level

In the same way that sectoral global co-agglomeration curves are constructed, geographic perspective co-agglomeration indicators can be derived from point-level calculations within neighbourhoods. According to expression [7b], Geographical Spatial Global Co-agglomeration Indicators (*GSCaI(r)*) are defined for each neighbourhood in Madrid and for each buffer width. The geographic co-agglomeration curves are generated by ordering the indicators for each neighbourhood based on buffer width. These curves can be interpreted as the capacity of the production points located in a neighbourhood to attract production points from any productive activities within a given distance. This approach can be regarded as a characteristic of the geographical area being analysed, as it combines the production points of various activities located within the neighbourhood. The co-agglomeration curves for each neighbourhood in Madrid are presented in Table C.3 of the Appendix C.

The types of curves resemble those found in the sectoral case, but their distribution differs. As shown in Table 3.6, the majority of curves (77 neighbourhoods in 2014 and 78 neighbourhoods in 2019) exhibit a clear dispersion pattern. This indicates that the global co-agglomeration process is concentrated in only a small number of neighbourhoods within the City of Madrid.

Table 3.6 Typology of geographical co-agglomeration curves at the neighbourhood level using establishments in Madrid

		Decreasing or L	U or V	Inverse U or V	Double U and V	Increasing or inverse L
Do not cross the unit		**Type 1 (A)**	**Type 2 (D)**	**Type 3 (A)**	**Type 4 (D)**	**Type 5 (D)**
	2014	5 neighb.	24 neighb.	23 neighb.	2 neighb.	27 neighb.
	2019	5 neighb.	24 neighb.	23 neighb.	2 neighb.	28 neighb.
Do cross the unit		**Type 6 (A)**	**Type 7 (D)**	**Type 8**	**Type 9**	**Type 10 (D)**
	2014	-	9 neighb.	14 neighb.	11 neighb.	15 neighb.
	2019	-	9 neighb.	14 neighb.	11 neighb.	14 neighb.

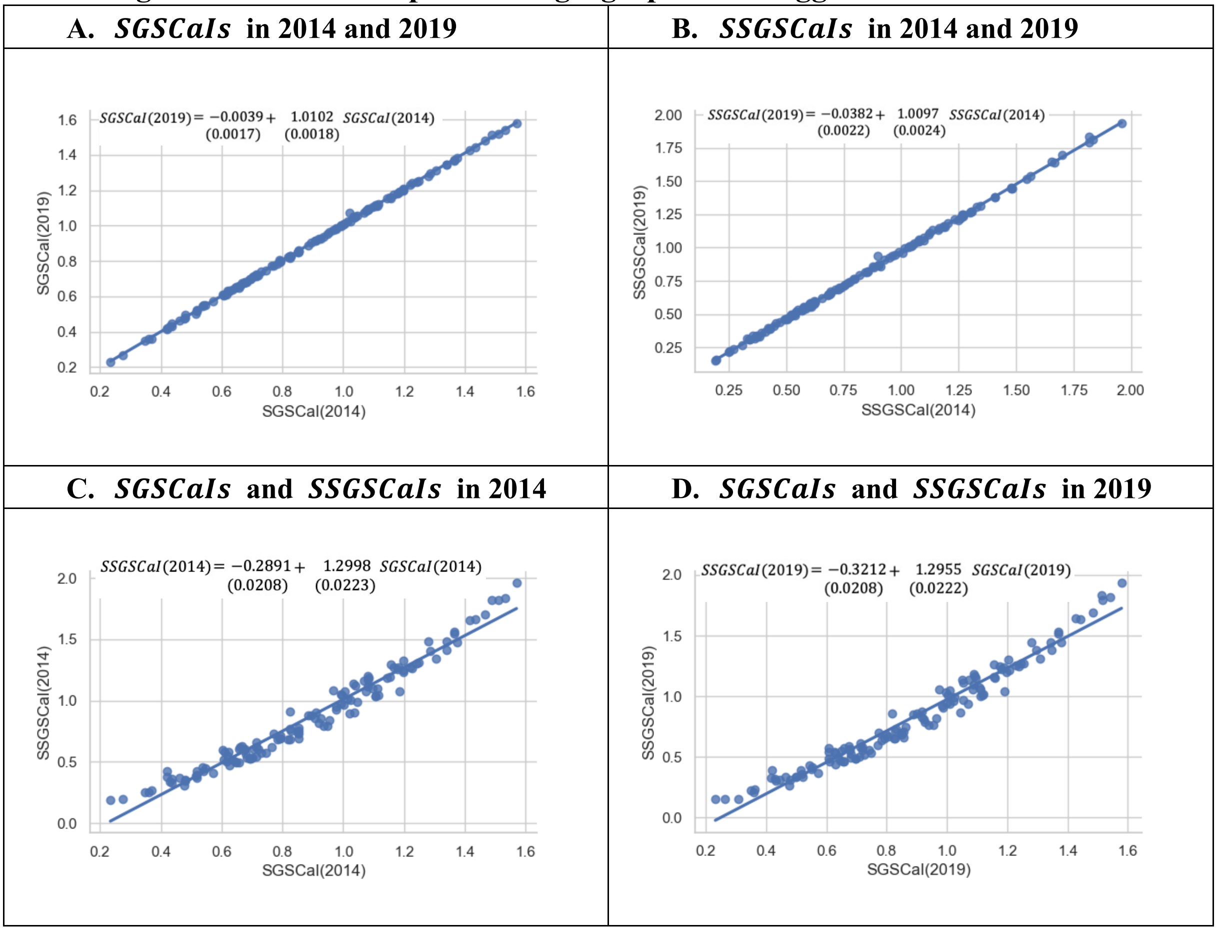
Note: (A) represents agglomeration behaviour; (D) refers to dispersion behaviour

3.5.3.2. Aggregated indicators at neighbourhood level

It is possible to obtain synthetic indicators for geographic units (neighbourhoods), similar to the process used for activities. The Synthetic Geographical Spatial Global Co-agglomeration Indicators (*SGSCaI*) are constructed by aggregating $SISCaI(x_i)$, as shown by expression [8b]. The Statistical Synthetic Geographical Spatial Global Co-agglomeration Indicators (*SSGSCaI*) for each neighbourhood are estimated from the Individual Spatial Global Co-Agglomeration Indicators or the Synthetic Individual Spatial Global Co-Agglomeration Indicators (expressions [9a] and [9b]). Table C.4 in the Appendix C displays the results of these indicators, and Figure 3.6 illustrates the relationship among them.

The equivalence between the indicators for both years is accepted (Figures A and B), indicating that the overall patterns of geographic global co-agglomeration have remained stable over this period. However, when comparing both indicators for the same year (Figures C and D), the regression lines show that the statistical indicators differ. This result suggests that the geographical distribution of production points has sectoral biases. Consequently, to obtain geographically comparable indicators, it is necessary to eliminate this sectoral bias, as achieved in the statistical case.

Figure 3.6 Relationship between geographical co-agglomeration indicators

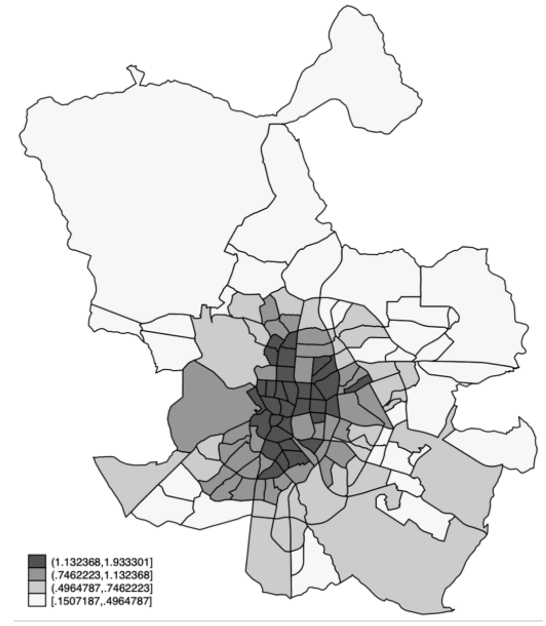


SGSCaI refers to Synthetic Geographical Spatial Co-agglomeration Indicators, and SSGSCI refers to Statistical Synthetic Geographical Spatial Co-agglomeration Indicators.

The results at the neighbourhood level are presented in Figure 7, where darker-coloured areas represent the most globally co-agglomerated neighbourhoods. Comparing the geographic effect over the two years, significant stability is observed in the attractiveness of central neighbourhoods between 2014 and 2019. Unlike agglomeration, which showed shifts towards the south and east of the city during this short analysis period, global co-agglomeration remains entirely stable. This finding suggests that changes in co-agglomeration positions require longer timeframes, while the agglomeration of specific sectors, such as those that are regulated, land-intensive, or related to leisure, exhibits a greater capacity for geographic adjustment.

Figure 3.7 Statistical global geographical spatial co-agglomeration indicators for the neighbourhoods of Madrid

2014	2019



3.6. Conclusions

The aim of this chapter is to adapt, from a theoretical standpoint, the family of indicators proposed in Chapter 1 to the measurement of co-agglomeration, focusing specifically on the relatively unexplored concept of sectoral (or geographical) global co-agglomeration. This concept can be defined as the ability of production points in a sector (or geographical area) to attract other production points, regardless of their activity. The adaptation for global co-agglomeration is relatively straightforward, with guidance on how bilateral co-agglomeration indicators could also be formulated. This highlights the flexibility of the proposed methodology, similar to other distance-based indicators, for measuring phenomena related to agglomeration. Furthermore, this chapter provides tools from this family of indicators to analyse global co-location phenomena among activities, often referred to as inter-industry agglomeration.

From an empirical perspective, this methodology is applied to the analysis of the city of Madrid, using data from the Census of Premises and Activities of the City Council. The results demonstrate significant stability at individual, sectoral, and geographical levels between 2014 and 2019, indicating that global co-agglomeration represents a structural geographical pattern that evolves only over the very long term. Additionally, the findings confirm the importance of the second aspect of first-order concentration in defining aggregated indicators, highlighting the need to correct for local geographical biases in sectoral indicators and for sectoral biases in geographical indicators of global co-agglomeration. Lastly, greater variability is observed among neighbourhoods than among sectors. This reflects both the method of computation and the fact that global co-agglomeration is more pronounced from a geographical perspective than from an activity perspective.

From a sectoral perspective, results reveal significant co-location attractiveness in activities related to daily consumption and leisure, driven by their need for proximity to consumers and co-consumption patterns. In contrast, certain land-intensive industries, such as manufacturing and accommodation, as well as creative industries, are less likely to exhibit co-location attractiveness. These findings contrast with those observed in the case of agglomeration, where some activities display similar relative behaviours while others reverse their positions due to specific restrictions, requirements, or complementarities with other activities tied to co-consumption. From a geographic perspective, these co-located activities are concentrated in the central neighbourhoods of the city, demonstrating notable stability over the two years analysed.

Appendix C. Additional results

Table C.1. Sectoral global co-agglomeration curves

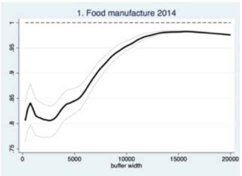
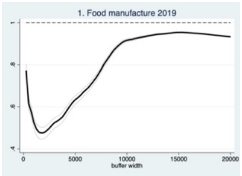
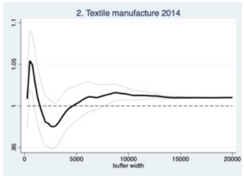
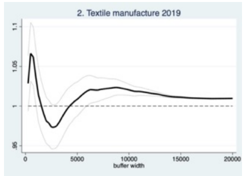
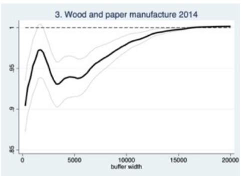
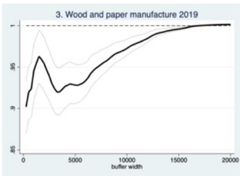
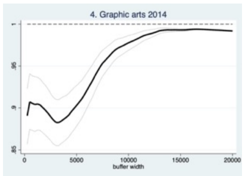
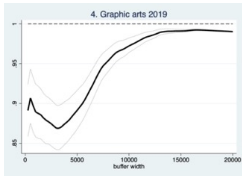
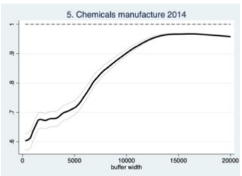
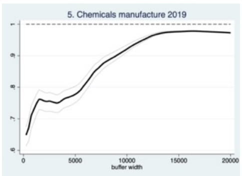
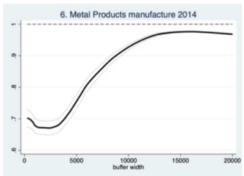
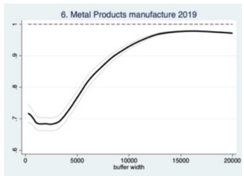
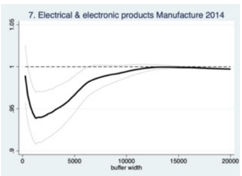
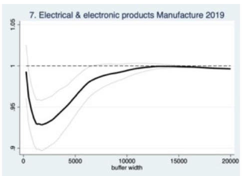
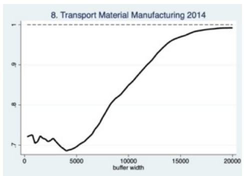
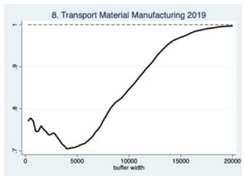
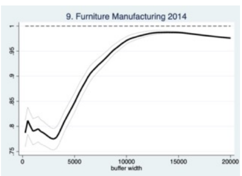
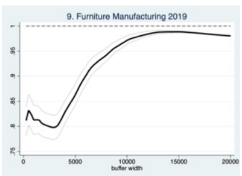
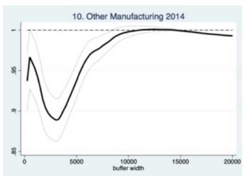
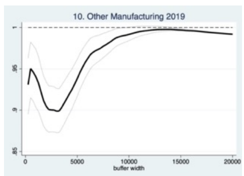
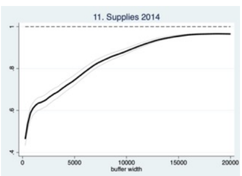
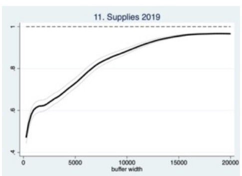
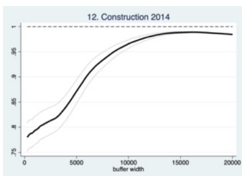
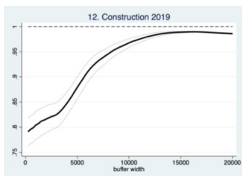
2014	2019		2014	2019
 Type 4	 Type 2		 Type 9	 Type 9
 Type 4	 Type 4		 Type 4	 Type 2
 Type 5	 Type 5		 Type 2	 Type 2
 Type 2	 Type 2		 Type 2	 Type 2
 Type 2	 Type 2		 Type 2	 Type 2
 Type 5	 Type 5		 Type 5	 Type 5

Table C.1. Sectoral global co-agglomeration curves

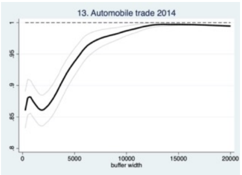
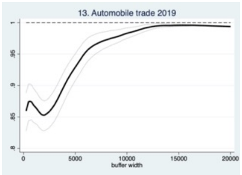
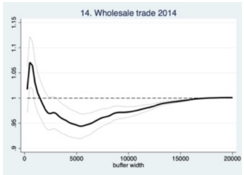
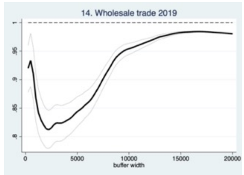
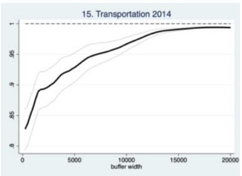
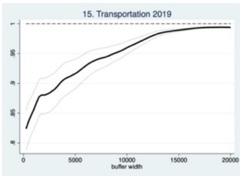
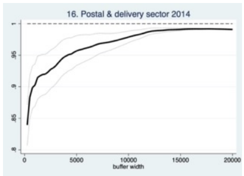
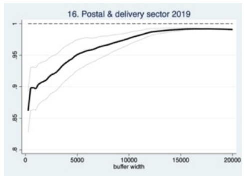
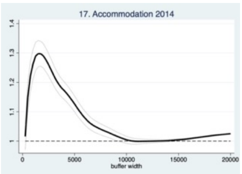
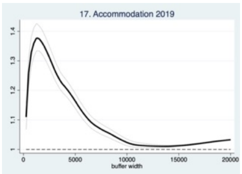
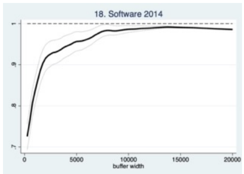
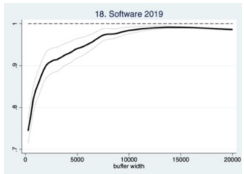
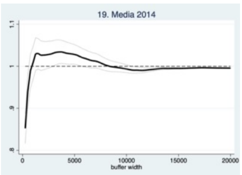
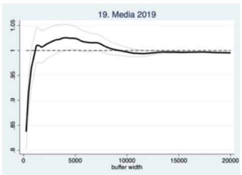
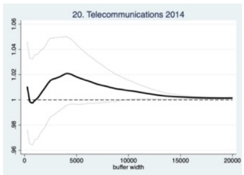
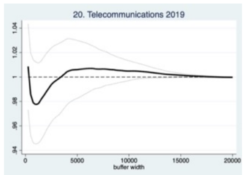
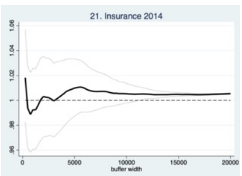
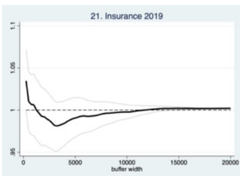
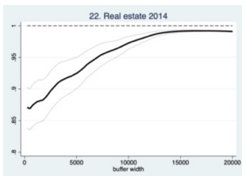
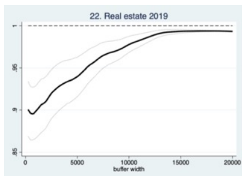
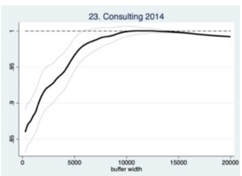
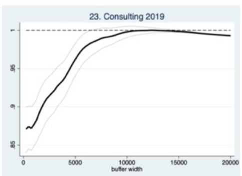
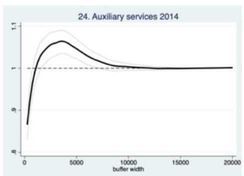
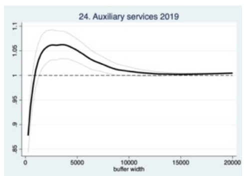
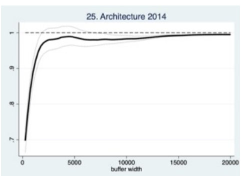
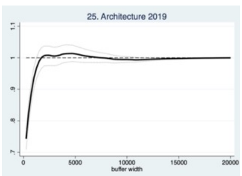
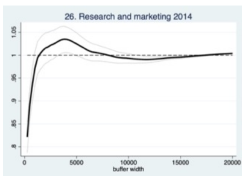
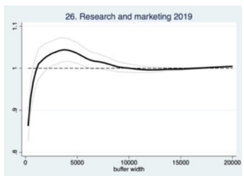
2014	2019		2014	2019
 Type 5	 Type 5		 Type 7	 Type 2
 Type 5	 Type 5		 Type 5	 Type 5
 Type 3	 Type 3		 Type 5	 Type 5
 Type 10	 Type 10		 Type 9	 Type 7
 Type 9	 Type 6		 Type 5	 Type 5
 Type 5	 Type 5		 Type 8	 Type 8
 Type 5	 Type 5		 Type 10	 Type 10

Table C.1. Sectoral global co-agglomeration curves

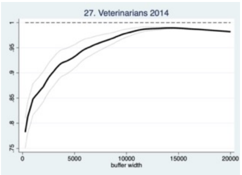
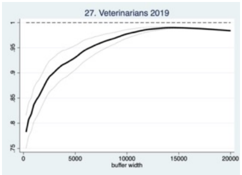
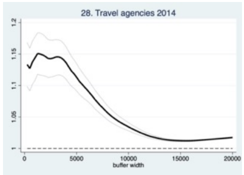
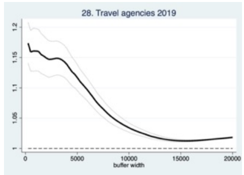
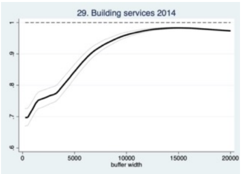
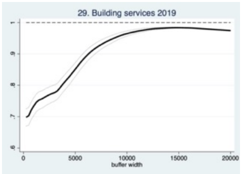
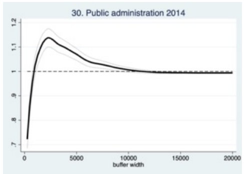
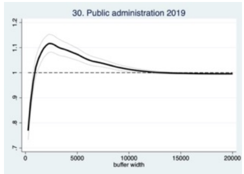
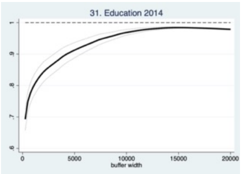
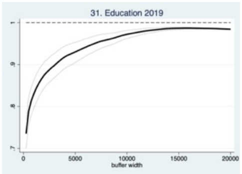
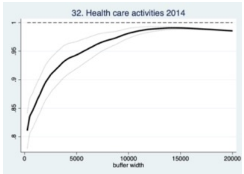
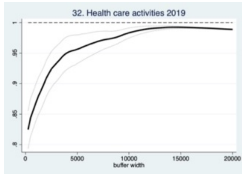
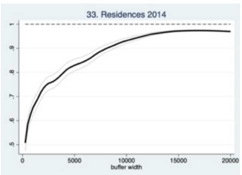
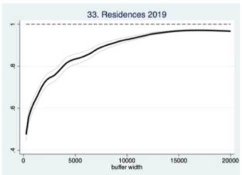
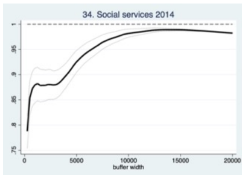
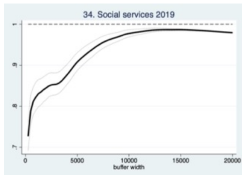
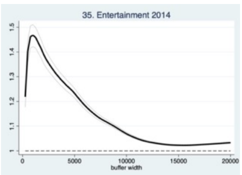
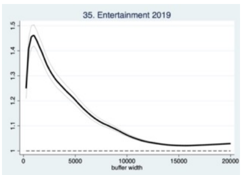
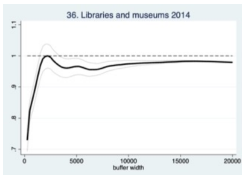
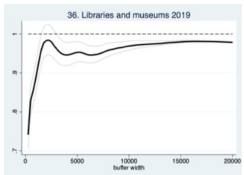
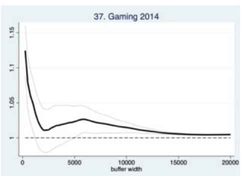
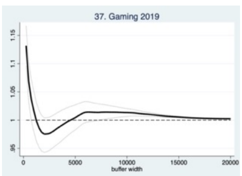
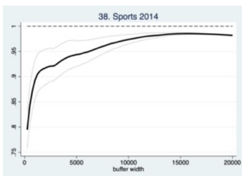
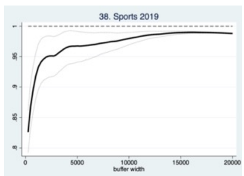
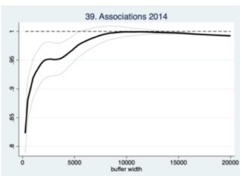
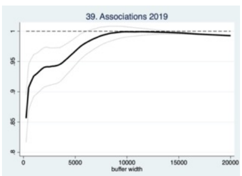
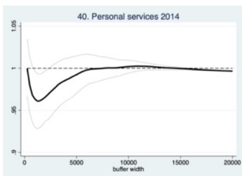
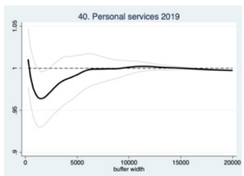
2014	2019		2014	2019
 Type 5	 Type 5		 Type 3	 Type 1
 Type 5	 Type 5		 Type 8	 Type 8
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 3	 Type 3		 Type 5	 Type 5
 Type 1	 Type 6		 Type 5	 Type 5
 Type 5	 Type 5		 Type 2	 Type 7

Table C.1. Sectoral global co-agglomeration curves

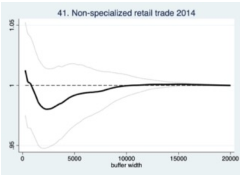
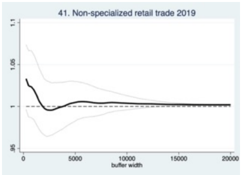
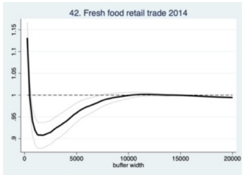
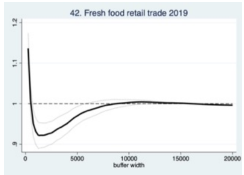
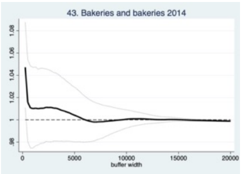
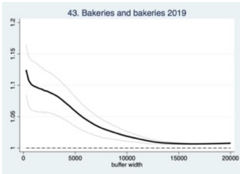
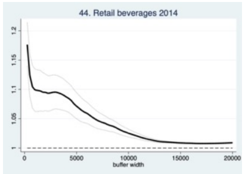
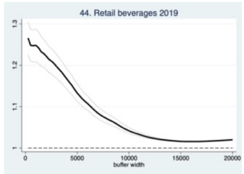
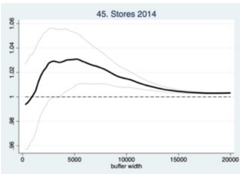
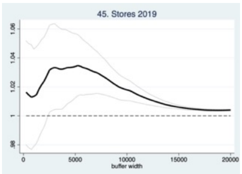
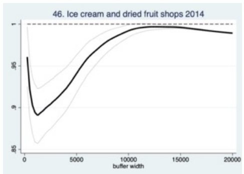
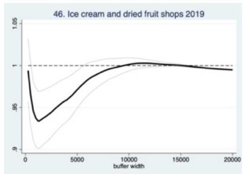
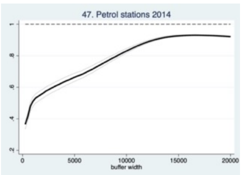
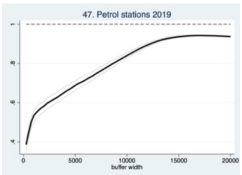
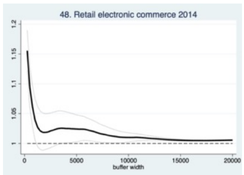
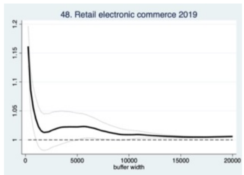
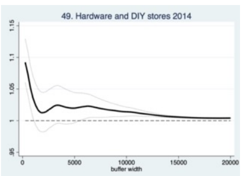
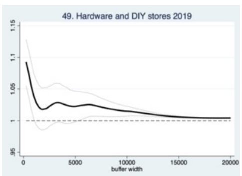
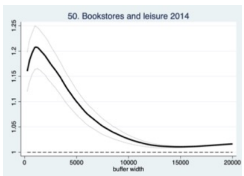
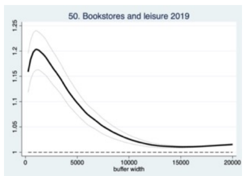
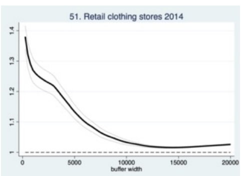
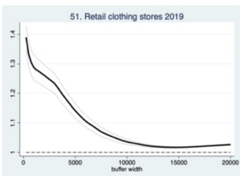
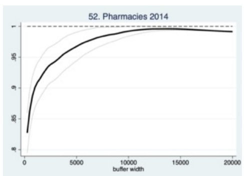
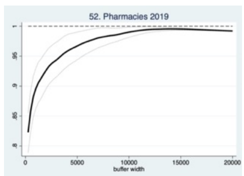
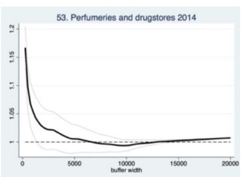
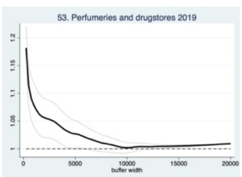
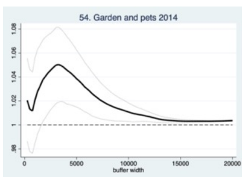
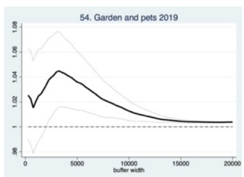
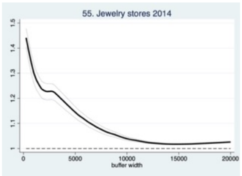
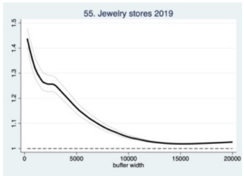
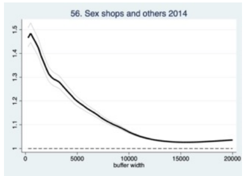
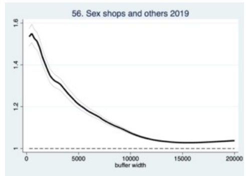
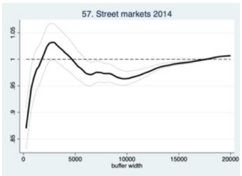
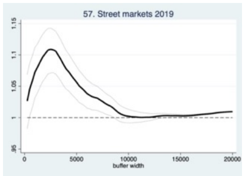
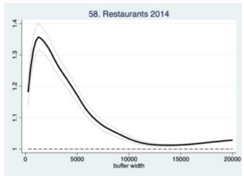
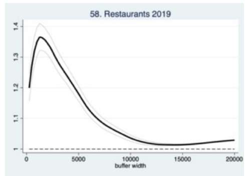
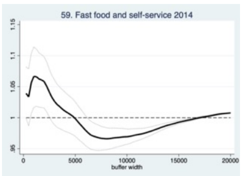
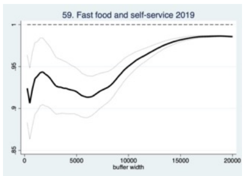
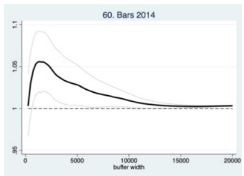
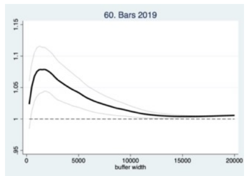
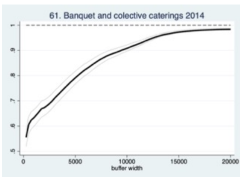
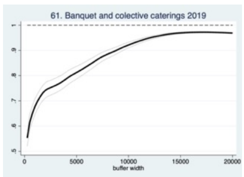
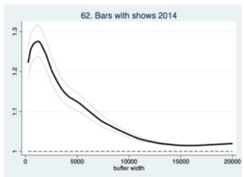
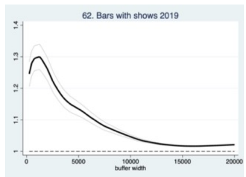
2014	2019		2014	2019
 Type 7	 Type 6		 Type 7	 Type 7
 Type 1	 Type 1		 Type 1	 Type 1
 Type 8	 Type 3		 Type 2	 Type 2
 Type 5	 Type 5		 Type 1	 Type 1
 Type 1	 Type 1		 Type 3	 Type 3
 Type 1	 Type 1		 Type 5	 Type 5
 Type 1	 Type 1		 Type 3	 Type 3

Table C.1. Sectoral global co-agglomeration curves

2014	2019		2014	2019
 Type 1	 Type 1		 Type 1	 Type 1
 Type 9	 Type 3		 Type 3	 Type 3
 Type 6	 Type 4		 Type 3	 Type 3
 Type 5	 Type 5		 Type 3	 Type 3

Note: X-axis refers to Buffer Radius; Y-axis represents Agglomeration Indicator.

Table C.2. Sectoral global co-agglomeration measures

	Synthetic Sectoral Spatial Co-agglomeration Indicators			Statistical Synthetic Sectoral Spatial Co-agglomeration Indicators	
Name	**2014**	**2019**		**2014**	**2019**
1. Food (M)	**0.89**	**0.73**		**0.99**	**0.97**
2. Textile (M)	**1.01**	**1.01**		**1.02**	**0.99**
3. Wood and paper (M)	**0.96**	**0.96**		**1.01**	**0.96**
4. Graphic arts	**0.94**	**0.93**		**1.00**	**0.97**
5. Chemicals (M)	**0.79**	**0.83**		**0.95**	**0.92**
6. Metal Products (M)	**0.81**	**0.82**		**0.99**	**0.96**
7. Electrical and electronic (M)	**0.98**	**0.97**		**1.04**	**1.00**
8. Transport Material (M)	**0.80**	**0.82**		**0.98**	**0.95**
9. Furniture (M)	**0.88**	**0.89**		**1.00**	**0.96**
10. Other Manufacturing	**0.96**	**0.96**		**1.02**	**0.98**
11. Supplies	**0.77**	**0.76**		**0.96**	**0.91**
12. Construction	**0.89**	**0.90**		**1.01**	**0.97**
13. Automobile (Trade)	**0.94**	**0.93**		**1.02**	**0.98**
14. Wholesale Trade	**0.99**	**0.90**		**1.03**	**0.98**
15. Transportation	**0.93**	**0.92**		**1.01**	**0.97**
16. Postal and delivery sector	**0.95**	**0.95**		**1.02**	**0.99**
17. Accommodation	**1.10**	**1.15**		**1.01**	**0.98**
18. Software	**0.93**	**0.93**		**0.98**	**0.95**
19. Media	**1.00**	**0.99**		**0.99**	**0.96**
20. Telecommunications	**1.01**	**1.00**		**1.04**	**1.01**
21. Insurance	**1.00**	**1.00**		**1.05**	**1.02**
22. Real estate	**0.94**	**0.95**		**1.03**	**1.00**
23. Consulting	**0.95**	**0.95**		**1.02**	**0.99**
24. Auxiliary services	**1.01**	**1.01**		**1.01**	**0.98**

Table C.2. Sectoral global co-agglomeration measures

	Synthetic Sectoral Spatial Co-agglomeration Indicators			Statistical Synthetic Sectoral Spatial Co-agglomeration Indicators	
25. Architecture	**0.95**	**0.97**		**0.97**	**0.94**
26. Research and marketing	**0.99**	**1.00**		**1.00**	**0.97**
27. Veterinarians	**0.92**	**0.92**		**1.00**	**0.97**
28. Travel Agencies	**1.08**	**1.09**		**1.05**	**1.03**
29. Building Services	**0.86**	**0.86**		**0.99**	**0.96**
30. Public Administration	**1.01**	**1.02**		**0.97**	**0.94**
31. Education	**0.90**	**0.92**		**0.99**	**0.96**
32. Health care activities	**0.93**	**0.94**		**1.02**	**0.99**
33. Residences	**0.82**	**0.81**		**0.97**	**0.93**
34. Social Services	**0.93**	**0.91**		**1.01**	**0.96**
35. Entertainment	**1.20**	**1.19**		**1.01**	**0.99**
36. Libraries and museums	**0.95**	**0.94**		**0.97**	**0.94**
37. Gaming	**1.03**	**1.01**		**1.06**	**1.03**
38. Sports	**0.94**	**0.95**		**1.00**	**0.98**
39. Associations	**0.96**	**0.96**		**1.01**	**0.98**
40. Personal Services	**0.99**	**0.99**		**1.04**	**1.01**
41. Non-special retail trade	**1.00**	**1.01**		**1.04**	**1.01**
42. Fresh food retail trade	**0.98**	**0.98**		**1.07**	**1.04**
43. Bakeries and bakeries	**1.01**	**1.05**		**1.05**	**1.02**
44. Retail beverages	**1.06**	**1.13**		**1.06**	**1.02**
45. Stores	**1.01**	**1.02**		**1.04**	**1.01**
46. Ice cream and dried fruit	**0.95**	**0.98**		**1.04**	**1.01**
47. Petrol stations	**0.70**	**0.72**		**0.93**	**0.90**
48. Electronics stores	**1.03**	**1.03**		**1.06**	**1.03**
49. Hardware and DIY stores	**1.02**	**1.03**		**1.05**	**1.02**

Table C.2. Sectoral global co-agglomeration measures

	Synthetic Sectoral Spatial Co-agglomeration Indicators			Statistical Synthetic Sectoral Spatial Co-agglomeration Indicators	
50. Bookstores and leisure	**1.10**	**1.09**		**1.04**	**1.00**
51. Retail clothing stores	**1.14**	**1.15**		**1.09**	**1.05**
52. Pharmacies	**0.95**	**0.95**		**1.01**	**0.98**
53. Perfumeries and drugstores	**1.02**	**1.04**		**1.07**	**1.03**
54. Garden and pets	**1.02**	**1.02**		**1.04**	**1.00**
55. Jewellery stores	**1.15**	**1.16**		**1.10**	**1.06**
56. Sex shops and others	**1.21**	**1.23**		**1.05**	**1.02**
57. Street markets	**0.98**	**1.04**		**1.01**	**1.00**
58. Restaurants	**1.15**	**1.16**		**1.03**	**1.00**
59. Fast food and self-service	**1.01**	**0.94**		**1.05**	**1.00**
60. Bars	**1.02**	**1.03**		**1.03**	**1.00**
61. Colective caterings	**0.80**	**0.82**		**0.99**	**0.94**
62. Bars with shows	**1.12**	**1.13**		**1.04**	**1.01**

Table C.3. Global co-agglomeration curves for each neighbourhood

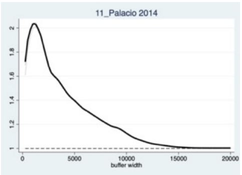
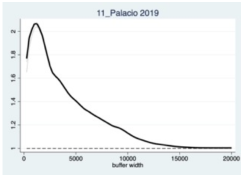
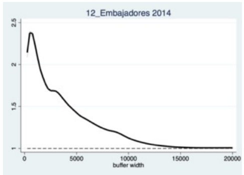
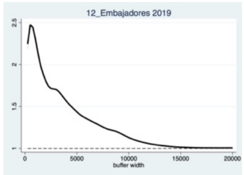
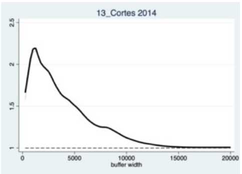
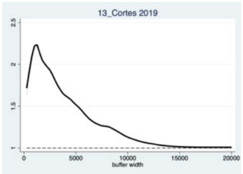
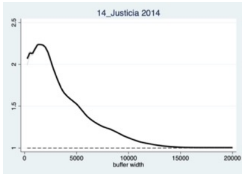
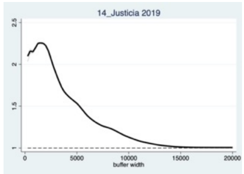
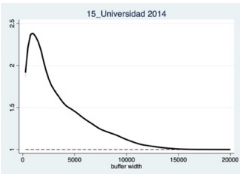
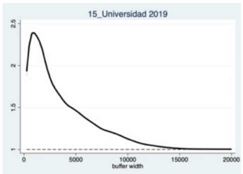
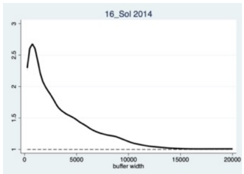
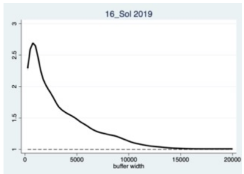
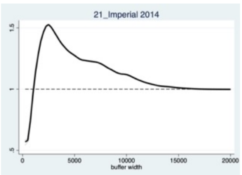
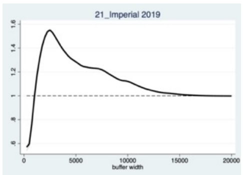
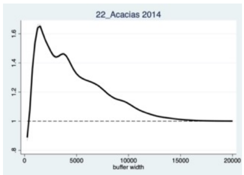
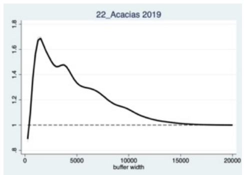
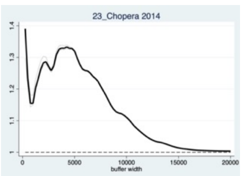
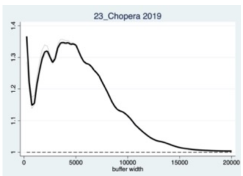
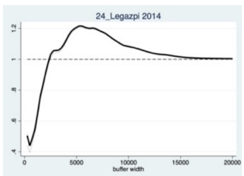
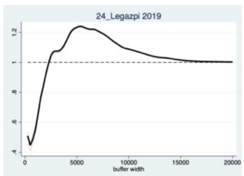
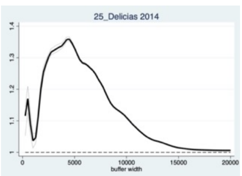
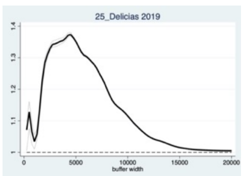
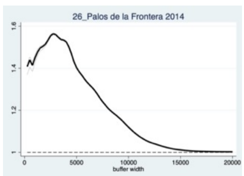
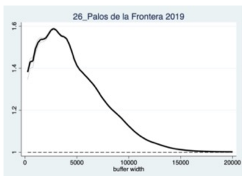
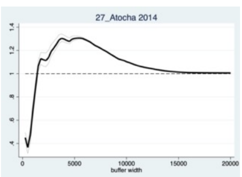
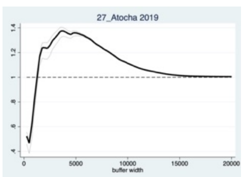
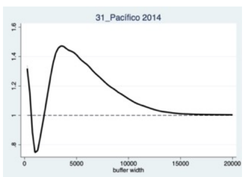
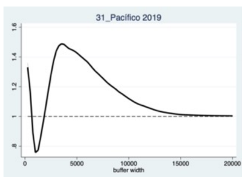
2014	2019		2014	2019
 Type 3	 Type 3		 Type 1	 Type 1
 Type 3	 Type 3		 Type 1	 Type 1
 Type 3	 Type 3		 Type 3	 Type 3
 Type 8	 Type 8		 Type 8	 Type 8
 Type 3	 Type 3		 Type 8	 Type 8
 Type 3	 Type 3		 Type 3	 Type 3
 Type 8	 Type 8		 Type 9	 Type 9

Table C.3. Global co-agglomeration curves for each neighbourhood

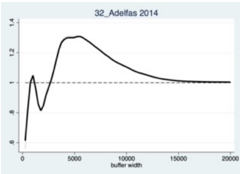
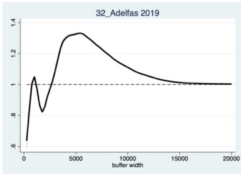
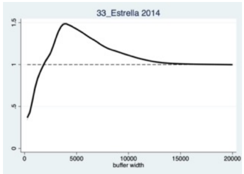
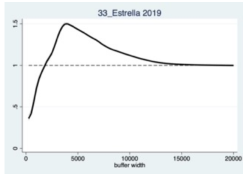
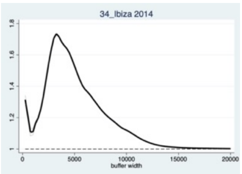
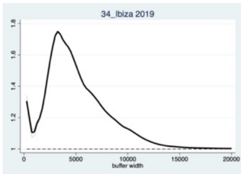
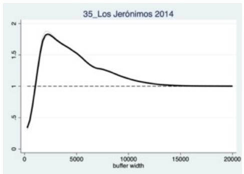
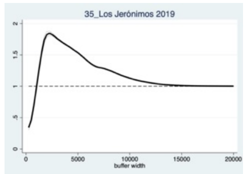
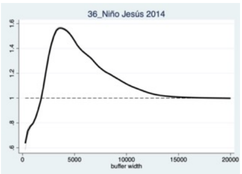
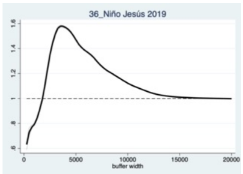
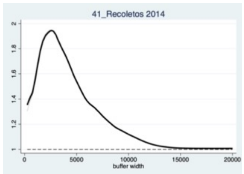
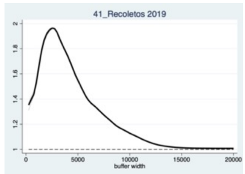
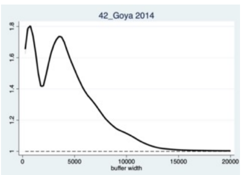
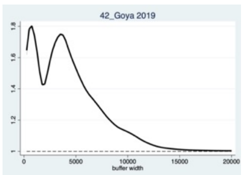
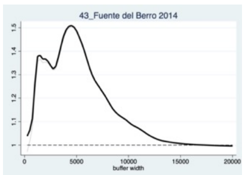
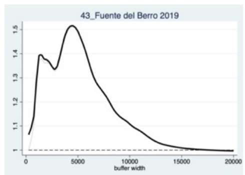
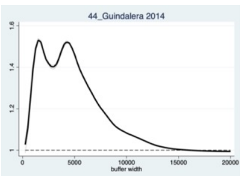
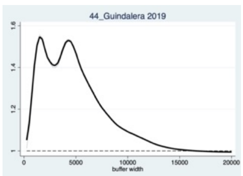
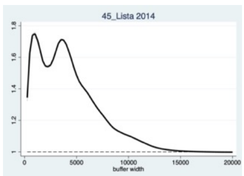
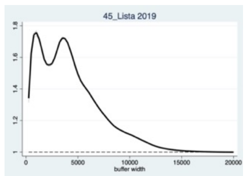
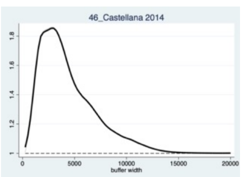
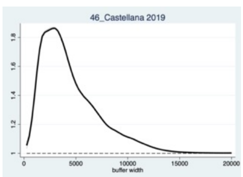
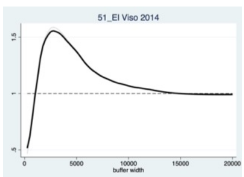
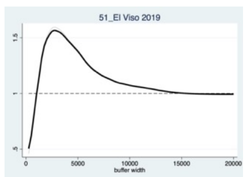
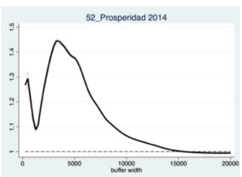
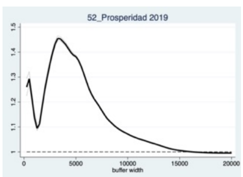
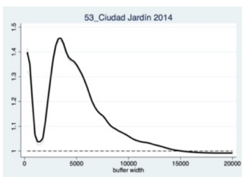
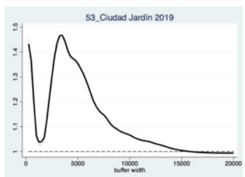
2014	2019		2014	2019
 Type 8	 Type 8		 Type 8	 Type 8
 Type 3	 Type 3		 Type 8	 Type 8
 Type 8	 Type 8		 Type 3	 Type 3
 Type 3	 Type 3		 Type 3	 Type 3
 Type 3	 Type 3		 Type 3	 Type 3
 Type 3	 Type 3		 Type 8	 Type 8
 Type 3	 Type 3		 Type 3	 Type 3

Table C.3. Global co-agglomeration curves for each neighbourhood

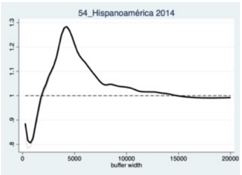
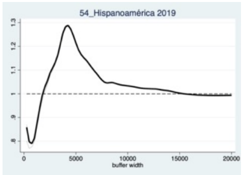
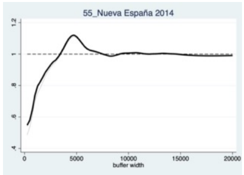
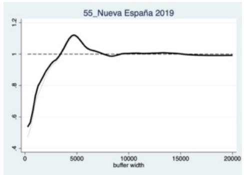
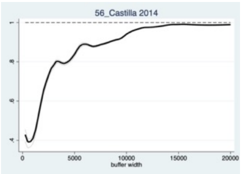
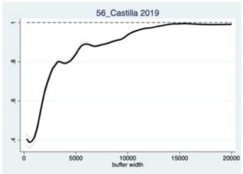
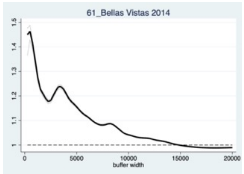
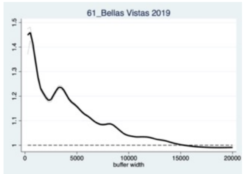
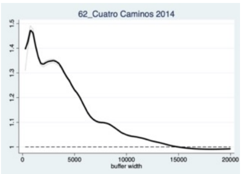
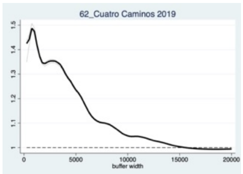
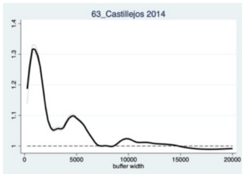
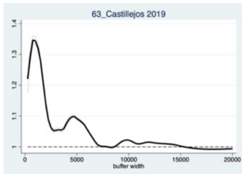
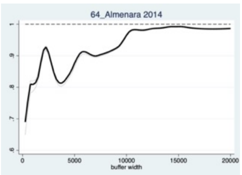
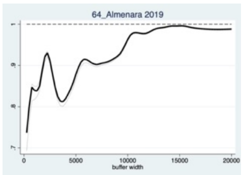
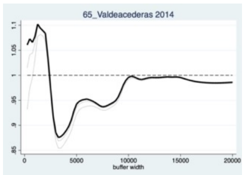
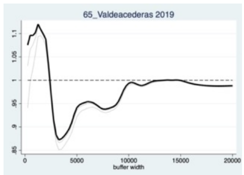
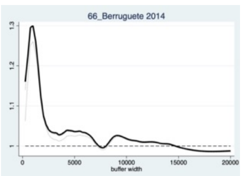
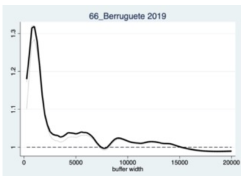
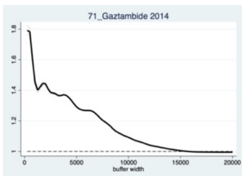
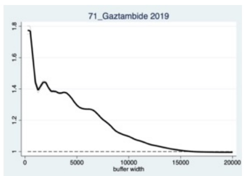
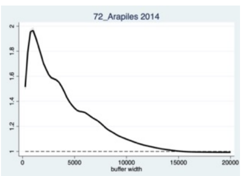
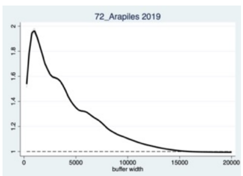
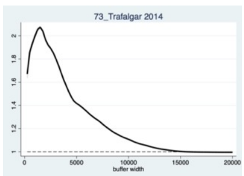
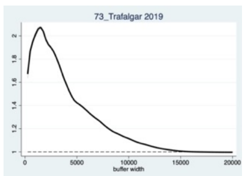
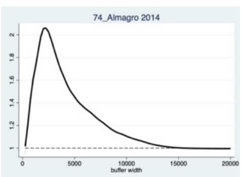
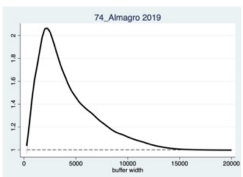
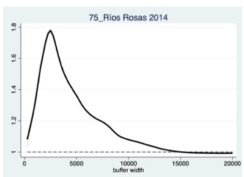
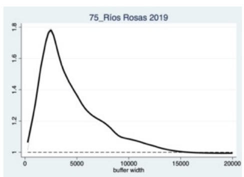
2014	2019		2014	2019
 Type 8	 Type 8		 Type 10	 Type 10
 Type 5	 Type 5		 Type 1	 Type 1
 Type 1	 Type 1		 Type 3	 Type 3
 Type 4	 Type 4		 Type 7	 Type 7
 Type 3	 Type 3		 Type 1	 Type 1
 Type 3	 Type 3		 Type 3	 Type 3
 Type 3	 Type 3		 Type 3	 Type 3

Table C.3. Global co-agglomeration curves for each neighbourhood

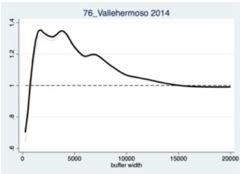
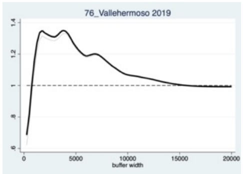
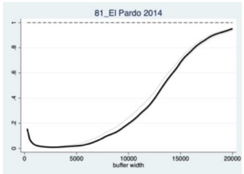
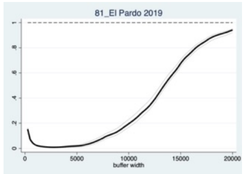
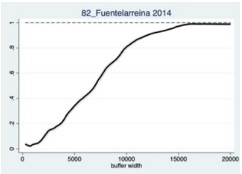
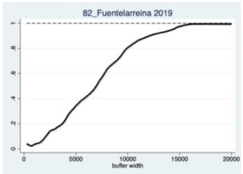
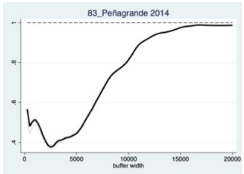
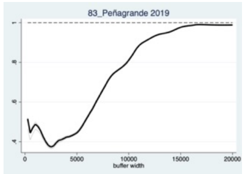
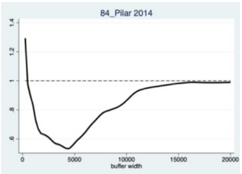
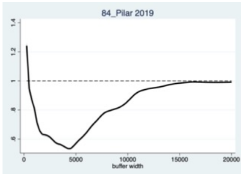
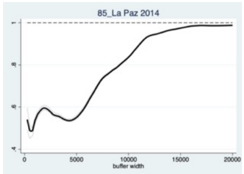
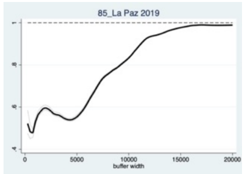
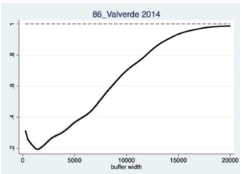
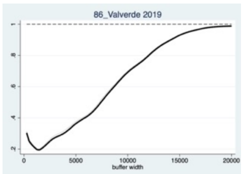
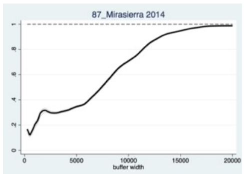
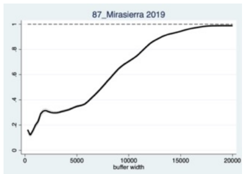
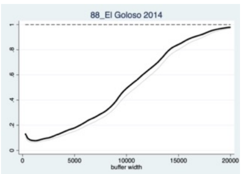
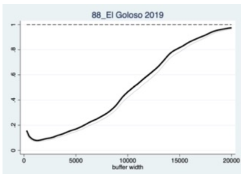
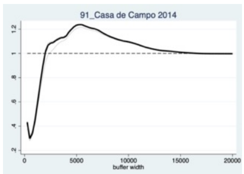
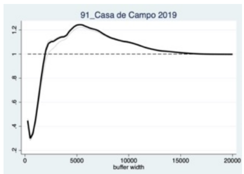
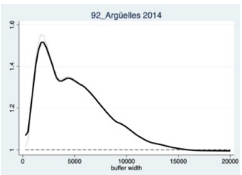
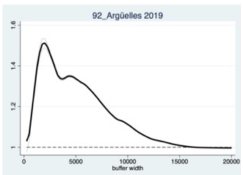
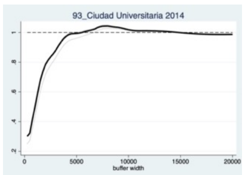
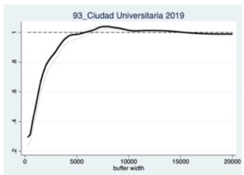
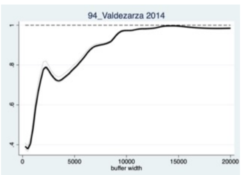
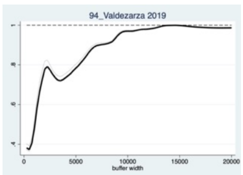
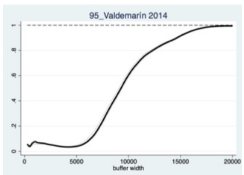
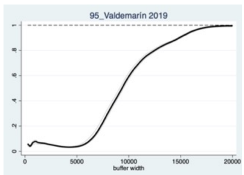
2014	2019		2014	2019
 Type 8	 Type 8		 Type 2	 Type 2
 Type 5	 Type 5		 Type 2	 Type 2
 Type 7	 Type 7		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 10	 Type 10
 Type 3	 Type 3		 Type 10	 Type 10
 Type 5	 Type 5		 Type 5	 Type 5

Table C.3. Global co-agglomeration curves for each neighbourhood

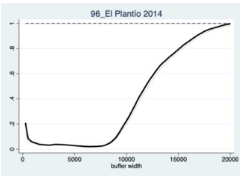
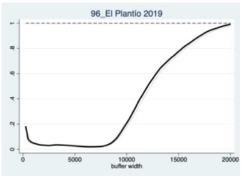
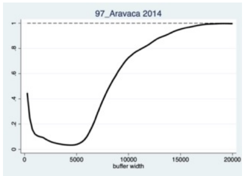
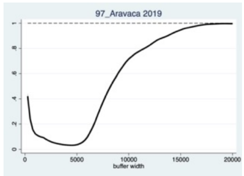
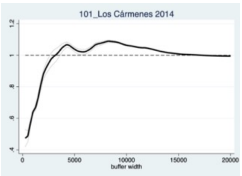
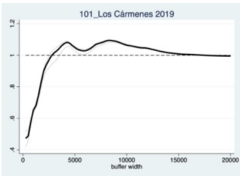
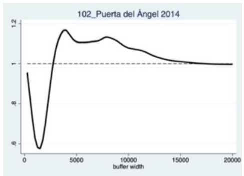
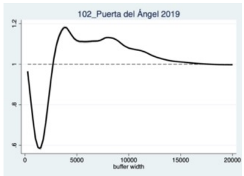
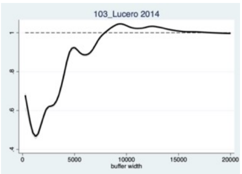
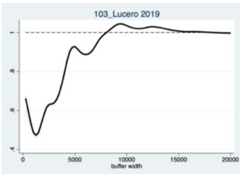
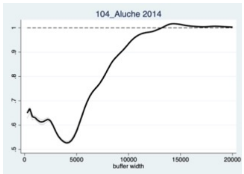
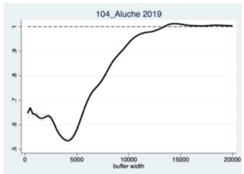
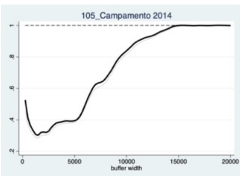
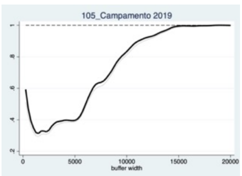
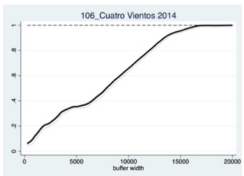
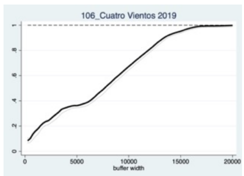
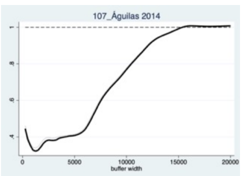
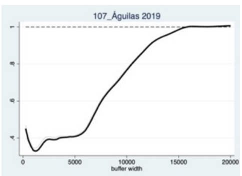
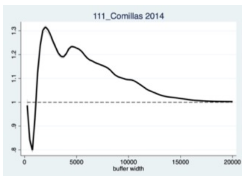
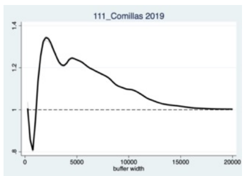
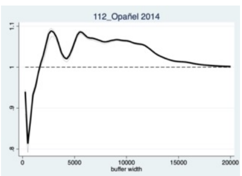
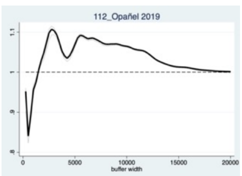
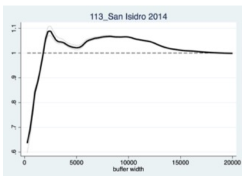
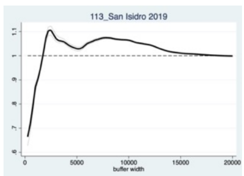
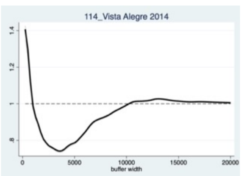
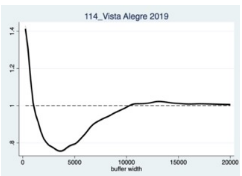
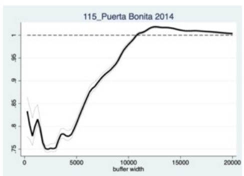
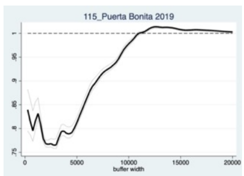
2014	2019		2014	2019
 Type 2	 Type 2		 Type 2	 Type 2
 Type 10	 Type 10		 Type 9	 Type 9
 Type 7	 Type 7		 Type 2	 Type 2
 Type 2	 Type 2		 Type 5	 Type 5
 Type 2	 Type 2		 Type 8	 Type 8
 Type 10	 Type 10		 Type 10	 Type 10
 Type 7	 Type 7		 Type 7	 Type 7

Table C.3. Global co-agglomeration curves for each neighbourhood

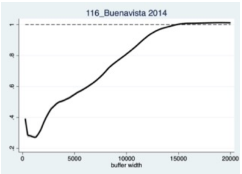
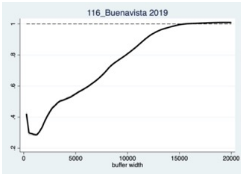
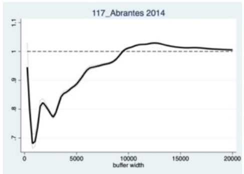
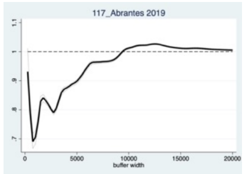
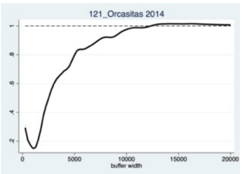
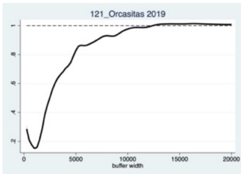
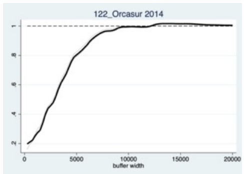
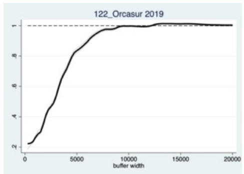
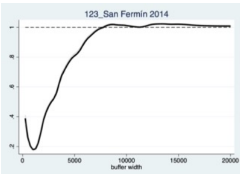
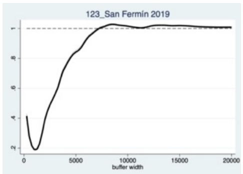
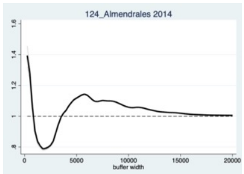
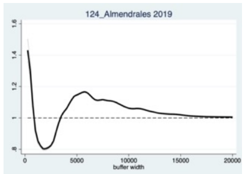
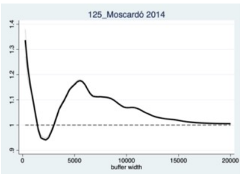
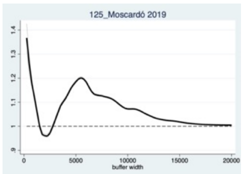
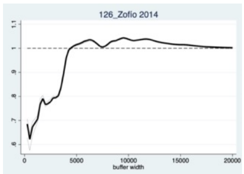
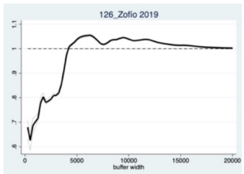
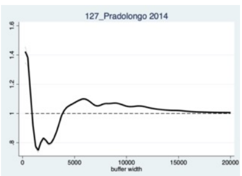
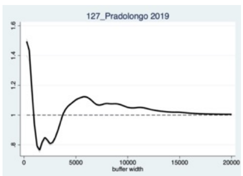
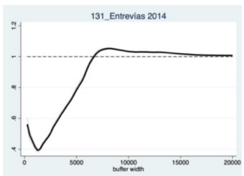
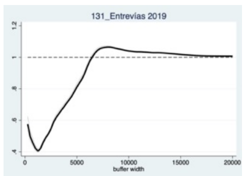
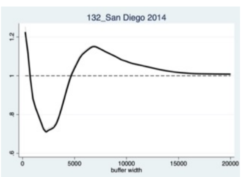
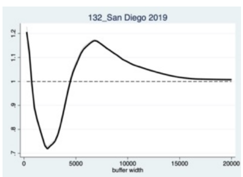
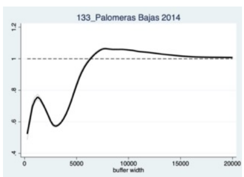
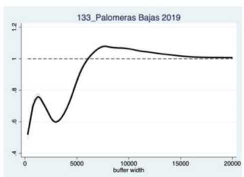
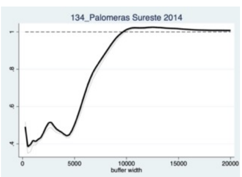
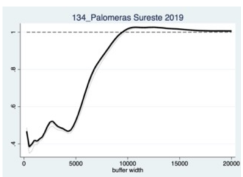
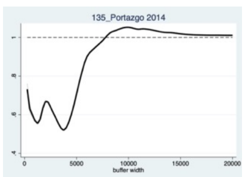
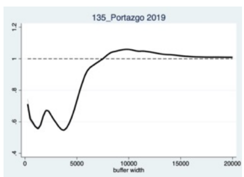
2014	2019		2014	2019
 Type 2	 Type 2		 Type 7	 Type 7
 Type 2	 Type 2		 Type 5	 Type 5
 Type 2	 Type 2		 Type 9	 Type 9
 Type 9	 Type 9		 Type 10	 Type 10
 Type 9	 Type 9		 Type 10	 Type 10
 Type 9	 Type 9		 Type 7	 Type 7
 Type 7	 Type 7		 Type 7	 Type 7

Table C.3. Global co-agglomeration curves for each neighbourhood

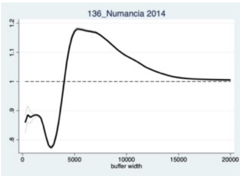
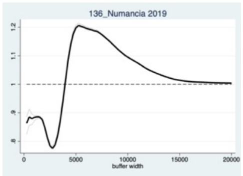
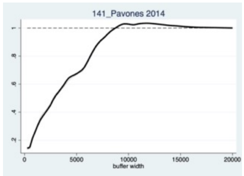
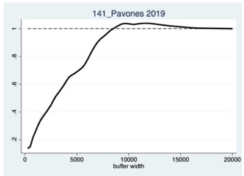
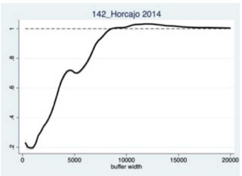
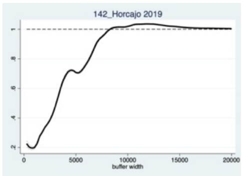
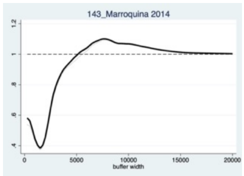
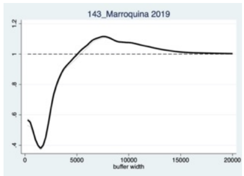
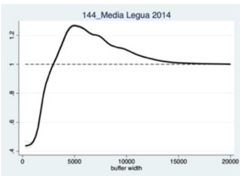
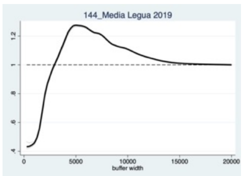
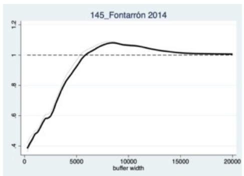
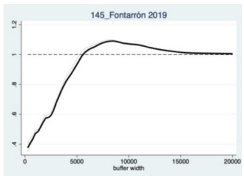
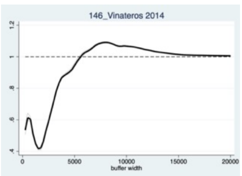
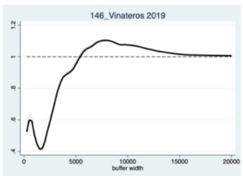
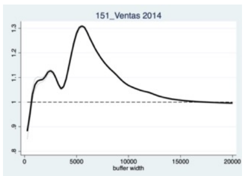
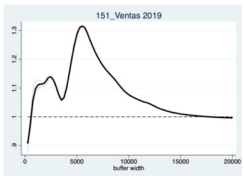
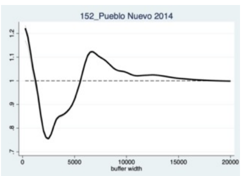
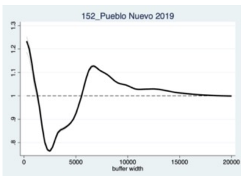
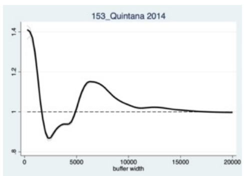
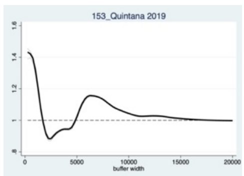
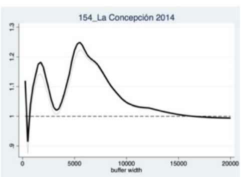
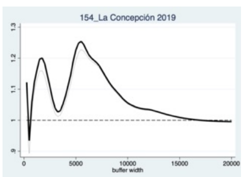
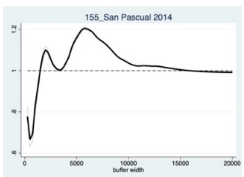
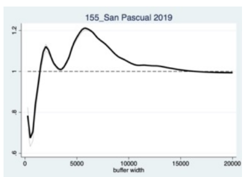
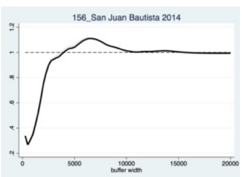
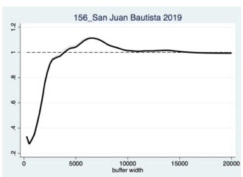
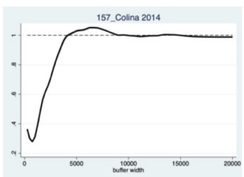
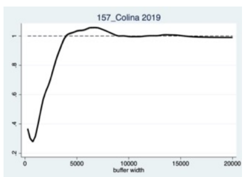
2014	2019		2014	2019
 Type 9	 Type 9		 Type 10	 Type 10
 Type 10	 Type 10		 Type 10	 Type 10
 Type 8	 Type 8		 Type 10	 Type 10
 Type 10	 Type 10		 Type 8	 Type 8
 Type 9	 Type 9		 Type 9	 Type 9
 Type 9	 Type 9		 Type 9	 Type 9
 Type 10	 Type 10		 Type 10	 Type 10

Table C.3. Global co-agglomeration curves for each neighbourhood

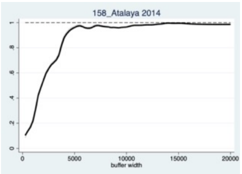
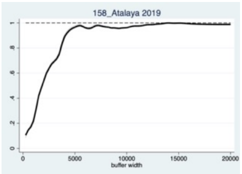
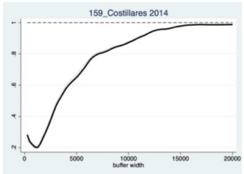
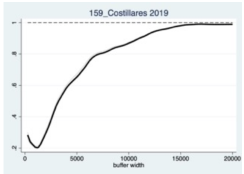
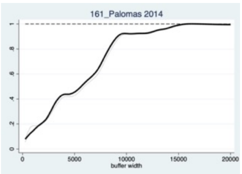
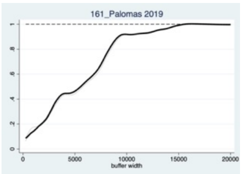
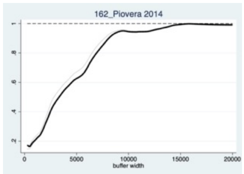
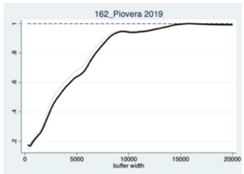
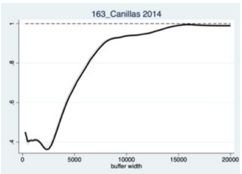
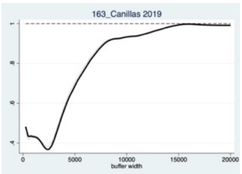
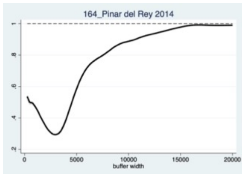
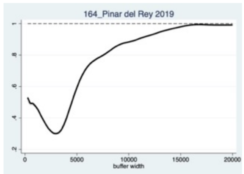
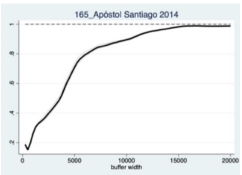
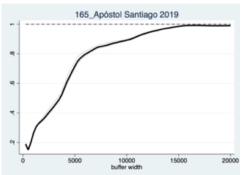
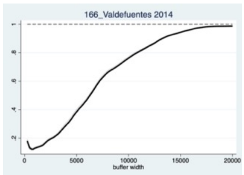
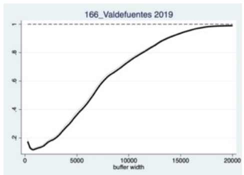
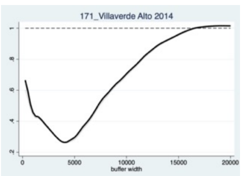
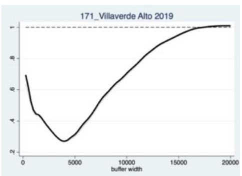
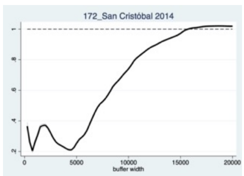
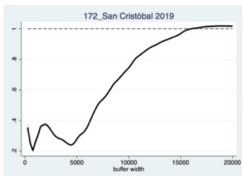
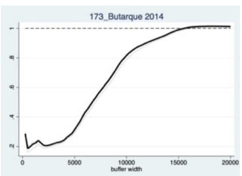
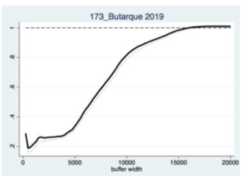
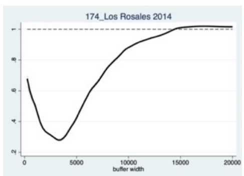
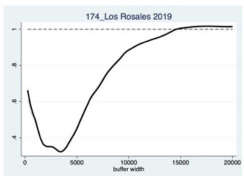
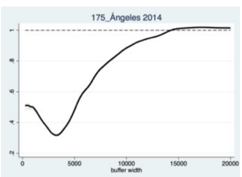
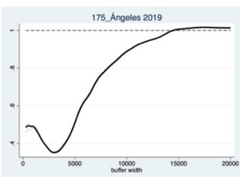
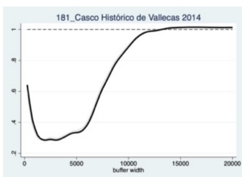
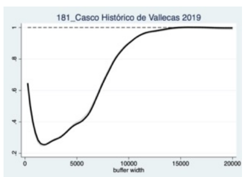
2014	2019		2014	2019
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 2	 Type 2		 Type 2	 Type 2
 Type 5	 Type 5		 Type 5	 Type 5
 Type 2	 Type 2		 Type 4	 Type 4
 Type 5	 Type 5		 Type 2	 Type 2
 Type 2	 Type 2		 Type 2	 Type 2

Table C.3. Global co-agglomeration curves for each neighbourhood

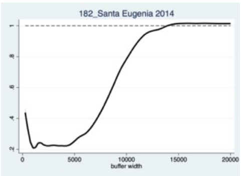
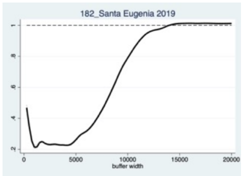
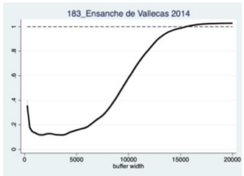
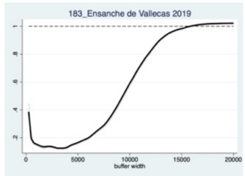
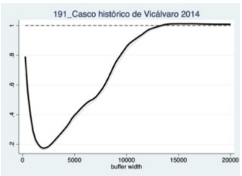
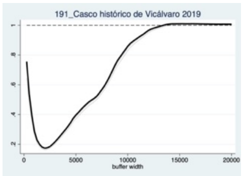
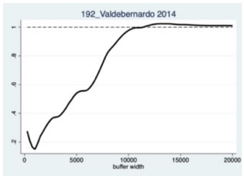
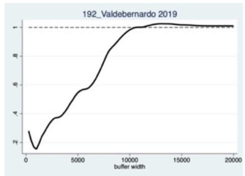
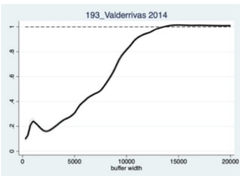
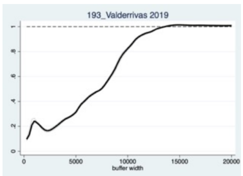
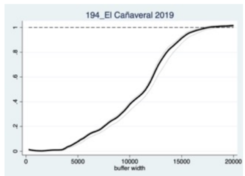
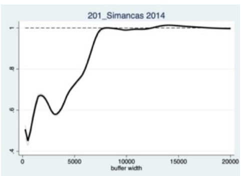
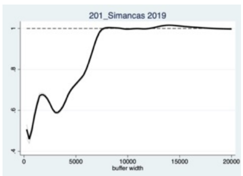
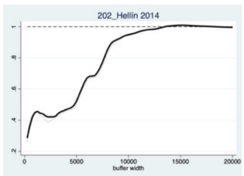
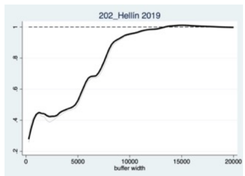
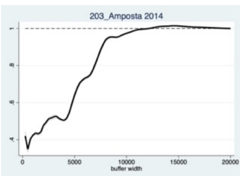
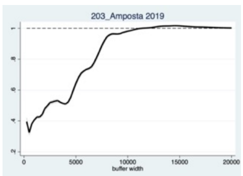
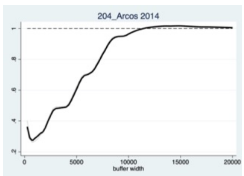
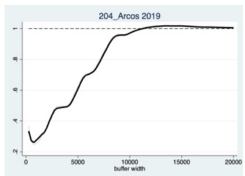
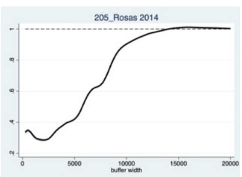
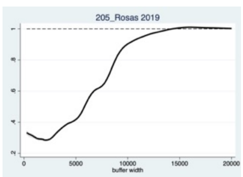
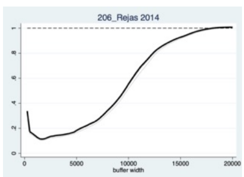
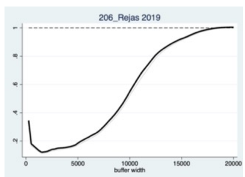
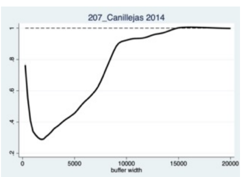
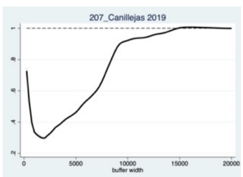
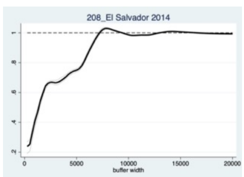
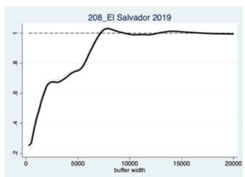
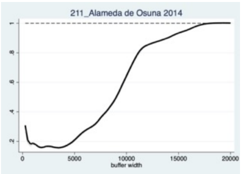
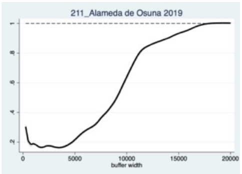
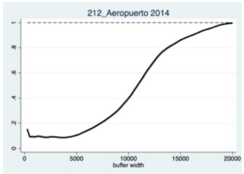
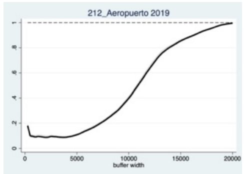
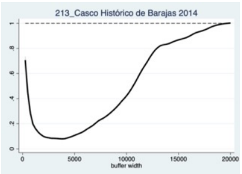
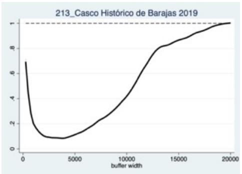
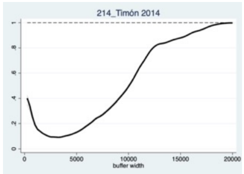
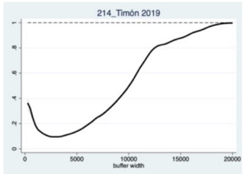
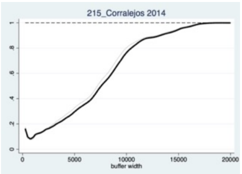
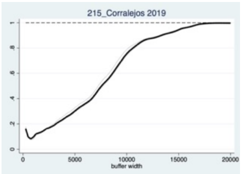
2014	2019		2014	2019
 Type 2	 Type 2		 Type 2	 Type 2
 Type 2	 Type 2		 Type 5	 Type 5
 Type 5	 Type 5			 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 5	 Type 5
 Type 5	 Type 5		 Type 2	 Type 2
 Type 2	 Type 2		 Type 5	 Type 5

Table C.3. Global co-agglomeration curves for each neighbourhood

2014	2019		2014	2019
 Type 2	 Type 2		 Type 5	 Type 5
 Type 2	 Type 2		 Type 2	 Type 2
 Type 5	 Type 5			

Note: X-axis refers to Buffer radius; Y-axis represents Agglomeration Indicator

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
Name	**2014**	**2019**		**2014**	**2019**
11. Palacio	**1.41**	**1.43**		**1.65**	**1.64**
12. Embajadores	**1.49**	**1.51**		**1.82**	**1.83**
13. Cortes	**1.47**	**1.48**		**1.70**	**1.69**
14. Justicia	**1.53**	**1.54**		**1.84**	**1.82**
15. Universidad	**1.51**	**1.52**		**1.82**	**1.79**
16. Sol	**1.57**	**1.58**		**1.96**	**1.93**
21. Imperial	**1.11**	**1.12**		**1.03**	**1.01**
22. Acacias	**1.23**	**1.24**		**1.27**	**1.25**
23. Chopera	**1.18**	**1.19**		**1.27**	**1.24**
24. Legazpi	**0.96**	**0.96**		**0.84**	**0.82**
25. Delicias	**1.15**	**1.16**		**1.19**	**1.15**
26. Palos de la Frontera	**1.29**	**1.30**		**1.41**	**1.38**
27. Atocha	**1.02**	**1.07**		**0.90**	**0.93**
31. Pacífico	**1.14**	**1.15**		**1.18**	**1.16**
32. Adelfas	**1.05**	**1.05**		**0.99**	**0.97**
33. Estrella	**1.04**	**1.04**		**0.90**	**0.87**
34. Ibiza	**1.25**	**1.25**		**1.31**	**1.27**
35. Los Jerónimos	**1.18**	**1.19**		**1.07**	**1.04**
36. Niño Jesús	**1.10**	**1.11**		**1.03**	**1.00**
41. Recoletos	**1.38**	**1.38**		**1.48**	**1.44**
42. Goya	**1.36**	**1.37**		**1.55**	**1.51**
43. Fuente del Berro	**1.20**	**1.21**		**1.23**	**1.21**
44. Guindalera	**1.22**	**1.23**		**1.27**	**1.25**
45. Lista	**1.34**	**1.35**		**1.48**	**1.45**

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
46. Castellana	1.30	1.31		1.34	1.31
51. El Viso	1.12	1.12		1.05	1.01
52. Prosperidad	1.18	1.18		1.26	1.22
53. Ciudad Jardín	1.17	1.17		1.27	1.24
54. Hispanoamérica	1.02	1.02		1.01	0.96
55. Nueva España	0.92	0.92		0.85	0.81
56. Castilla	0.78	0.78		0.68	0.64
61. Bellas Vistas	1.16	1.16		1.30	1.26
62. Cuatro Caminos	1.20	1.20		1.33	1.30
63. Castillejos	1.08	1.09		1.18	1.16
64. Almenara	0.89	0.90		0.88	0.86
65. Valdeacederas	0.99	1.00		1.06	1.03
66. Berruguete	1.07	1.07		1.16	1.13
71. Gaztambide	1.28	1.28		1.48	1.44
72. Arapiles	1.36	1.37		1.56	1.54
73. Trafalgar	1.44	1.44		1.67	1.64
74. Almagro	1.34	1.35		1.41	1.38
75. Ríos Rosas	1.24	1.25		1.30	1.27
76. Vallehermoso	1.11	1.11		1.10	1.05
81. El Pardo	0.23	0.23		0.19	0.15
82. Fuentelarreina	0.48	0.48		0.31	0.27
83. Peñagrande	0.66	0.65		0.61	0.55
84. Pilar	0.82	0.82		0.91	0.86
85. La Paz	0.71	0.71		0.66	0.62
86. Valverde	0.52	0.52		0.43	0.39

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
87. Mirasierra	**0.52**	**0.52**		**0.39**	**0.36**
88. El Goloso	**0.37**	**0.36**		**0.27**	**0.24**
91. Casa de Campo	**0.95**	**0.95**		**0.79**	**0.77**
92. Argüelles	**1.20**	**1.20**		**1.25**	**1.20**
93. Ciudad Universitaria	**0.85**	**0.85**		**0.73**	**0.69**
94. Valdezarza	**0.79**	**0.79**		**0.69**	**0.65**
95. Valdemarín	**0.36**	**0.36**		**0.25**	**0.21**
96. El Plantío	**0.27**	**0.26**		**0.20**	**0.16**
97. Aravaca	**0.44**	**0.43**		**0.36**	**0.32**
101. Los Cármenes	**0.92**	**0.93**		**0.82**	**0.79**
102. Puerta del Ángel	**0.98**	**0.98**		**0.95**	**0.92**
103. Lucero	**0.83**	**0.83**		**0.76**	**0.72**
104. Aluche	**0.77**	**0.77**		**0.73**	**0.70**
105. Campamento	**0.62**	**0.63**		**0.55**	**0.54**
106. Cuatro Vientos	**0.48**	**0.50**		**0.35**	**0.34**
107. Águilas	**0.61**	**0.61**		**0.52**	**0.49**
111. Comillas	**1.10**	**1.11**		**1.10**	**1.08**
112. Opañel	**1.01**	**1.03**		**1.02**	**1.00**
113. San Isidro	**0.98**	**0.99**		**0.93**	**0.91**
114. Vista Alegre	**0.97**	**0.97**		**1.08**	**1.06**
115. Puerta Bonita	**0.88**	**0.89**		**0.88**	**0.85**
116. Buenavista	**0.63**	**0.64**		**0.52**	**0.50**
117. Abrantes	**0.91**	**0.92**		**0.90**	**0.87**
121. Orcasitas	**0.70**	**0.71**		**0.54**	**0.50**
122. Orcasur	**0.71**	**0.72**		**0.54**	**0.52**

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
123. San Fermín	**0.73**	**0.74**		**0.58**	**0.56**
124. Almendrales	**1.04**	**1.05**		**1.13**	**1.11**
125. Moscardó	**1.08**	**1.10**		**1.17**	**1.15**
126. Zofío	**0.91**	**0.92**		**0.85**	**0.83**
127. Pradolongo	**1.03**	**1.05**		**1.14**	**1.13**
131. Entrevías	**0.79**	**0.80**		**0.69**	**0.67**
132. San Diego	**0.99**	**1.00**		**1.04**	**1.01**
133. Palomeras Bajas	**0.85**	**0.86**		**0.77**	**0.75**
134. Palomeras Sureste	**0.71**	**0.72**		**0.62**	**0.58**
135. Portazgo	**0.82**	**0.83**		**0.77**	**0.73**
136. Numancia	**0.99**	**1.00**		**0.97**	**0.94**
141. Pavones	**0.69**	**0.70**		**0.52**	**0.49**
142. Horcajo	**0.69**	**0.70**		**0.52**	**0.49**
143. Marroquina	**0.85**	**0.86**		**0.75**	**0.71**
144. Media Legua	**0.93**	**0.94**		**0.80**	**0.76**
145. Fontarrón	**0.83**	**0.83**		**0.69**	**0.65**
146. Vinateros	**0.85**	**0.85**		**0.74**	**0.70**
151. Ventas	**1.08**	**1.09**		**1.08**	**1.06**
152. Pueblo Nuevo	**1.00**	**1.01**		**1.08**	**1.05**
153. Quintana	**1.08**	**1.09**		**1.20**	**1.18**
154. La Concepción	**1.08**	**1.09**		**1.12**	**1.10**
155. San Pascual	**1.00**	**1.01**		**0.96**	**0.94**
156. San Juan Bautista	**0.85**	**0.86**		**0.69**	**0.66**
157. Colina	**0.82**	**0.82**		**0.68**	**0.65**
158. Atalaya	**0.75**	**0.75**		**0.57**	**0.53**

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
159. Costillares	**0.64**	**0.65**		**0.51**	**0.47**
161. Palomas	**0.57**	**0.57**		**0.41**	**0.37**
162. Piovera	**0.65**	**0.65**		**0.49**	**0.46**
163. Canillas	**0.71**	**0.71**		**0.61**	**0.58**
164. Pinar del Rey	**0.68**	**0.68**		**0.60**	**0.57**
165. Apóstol Santiago	**0.66**	**0.66**		**0.50**	**0.46**
166. Valdefuentes	**0.51**	**0.50**		**0.38**	**0.33**
171. Villaverde Alto	**0.60**	**0.61**		**0.60**	**0.58**
172. San Cristóbal	**0.54**	**0.55**		**0.45**	**0.41**
173. Butarque	**0.54**	**0.55**		**0.42**	**0.40**
174. Los Rosales	**0.67**	**0.67**		**0.63**	**0.59**
175. Ángeles	**0.67**	**0.68**		**0.60**	**0.56**
181. Casco Histórico de Vallecas	**0.62**	**0.63**		**0.58**	**0.55**
182. Santa Eugenia	**0.54**	**0.54**		**0.45**	**0.43**
183. Ensanche de Vallecas	**0.44**	**0.45**		**0.33**	**0.32**
191. Casco histórico de Vicálvaro	**0.61**	**0.61**		**0.58**	**0.54**
192. Valdebernardo	**0.63**	**0.63**		**0.47**	**0.44**
193. Valderrivas	**0.52**	**0.52**		**0.37**	**0.34**
194. El Cañaveral	**-**	**0.31**		**-**	**0.15**
201. Simancas	**0.79**	**0.80**		**0.71**	**0.68**
202. Hellín	**0.68**	**0.68**		**0.57**	**0.52**
203. Amposta	**0.72**	**0.72**		**0.60**	**0.56**
204. Arcos	**0.68**	**0.68**		**0.54**	**0.49**
205. Rosas	**0.62**	**0.61**		**0.51**	**0.46**
206. Rejas	**0.43**	**0.43**		**0.34**	**0.31**

Table C.4. Neighbourhood global co-agglomeration measures

	Synthetic Geographical Spatial Co-agglomeration Indicators			Statistical Synthetic Geographical Spatial Co-agglomeration Indicators	
207. Canillejas	**0.66**	**0.66**		**0.62**	**0.58**
208. El Salvador	**0.76**	**0.77**		**0.62**	**0.60**
211. Alameda de Osuna	**0.46**	**0.46**		**0.37**	**0.33**
212. Aeropuerto	**0.35**	**0.35**		**0.25**	**0.23**
213. Casco Histórico de Barajas	**0.42**	**0.42**		**0.43**	**0.39**
214. Timón	**0.42**	**0.42**		**0.38**	**0.33**
215. Corralejos	**0.48**	**0.48**		**0.34**	**0.31**

GENERAL CONCLUSIONS

Each chapter provides specific conclusions related to its proposals and applications, however, the following are the overarching conclusions of this doctoral thesis. The thesis introduces a new family of distance-based indicators designed to address gaps in the literature highlighted by previous research and further emphasized in this research. This family of indicators stands out due to four key characteristics that distinguish it from existing distance-based measures:

1. **Point-Level Basis**: The indicators are constructed at the individual level, departing from the traditional focus of current indicators on sectoral aggregates. This approach enhances the granularity and specificity of the analysis.
2. **Geographical Perspective**: The methodology reinterprets both individual and aggregate indicators from a geographical standpoint. It allows for the creation of indicators for sub-areas (e.g., neighbourhoods, prefectures) that reflect the ability of businesses operating within these territories to attract productive activity.
3. **Relaxation of Restrictive Assumptions**: The indicators aim to relax or revise the restrictive assumptions underlying previous approaches. For instance, edge-effect corrections are replaced by a data-driven solution, and the assumptions regarding the baseline distribution for testing the null hypothesis of no concentration are revisited to align more closely with empirical realities.
4. **Analytical Flexibility**: This family of indicators provides a wide array of tools for analysis. It is highly adaptable, capable of measuring various agglomeration phenomena, including intra-activity and inter-activity dynamics, as well as global and bilateral relationships. This adaptability addresses limitations commonly encountered in existing indicators.

These innovative features position this family of indicators as a significant advancement in the study and measurement of spatial agglomeration and co-agglomeration phenomena.

All these features lead to remarkable versatility and significant computational time savings compared to other existing approaches. Furthermore, this efficiency enables the measurement of agglomeration in cases involving a very large number of production points or firms, as exemplified by the case of Chinese manufacturing.

From an empirical perspective, the proposed indicators are applied to two entirely different realities: all productive activities in the city of Madrid, i.e., a monocentric urban context, and manufacturing across China, i.e., a country with a notable population concentration in the east.

In the case of Madrid, significant intra-industry concentration (agglomeration) is observed in activities with substantial location restrictions, such as manufacturing, those that are land-intensive, high-value-added sectors, and activities related to tourism. The first two are concentrated in specific locations on the periphery, while the latter two are primarily located in the city centre. Conversely, there is significant dispersion in daily consumption activities that require proximity to consumers. From a geographical perspective, there is notable persistence in the high concentration found in the central neighbourhoods of the city and in some areas in the southern and eastern periphery where industrial estates were developed during the analysed years.

Conversely, in terms of inter-industry agglomeration (global co-agglomeration), a high level of stability is observed at the individual, sectoral, and geographic levels during the analysed period. This suggests that global co-agglomeration is a structural geographic feature that would only change over the very long term. Furthermore, greater variability is detected among neighbourhoods than among sectors, highlighting its higher relevance from a geographic perspective rather than an activity-based one. From a sectoral perspective, significant co-location attractiveness is observed in activities related to daily consumption or leisure, which require proximity to consumers and exhibit co-consumption patterns. In contrast, some land-intensive industries (e.g., manufacturing, accommodation) and creative industries are less likely to exhibit co-located attractiveness. Comparing intra-sectoral and inter-sectoral agglomeration reveals that while some activities exhibit similar relative behaviours, others invert their positions due to restrictions, requirements, or complementarities with other activities related to co-consumption. From a geographic perspective, these activities have been highly co-agglomerated in the central neighbourhoods of the city, showing significant stability across the two analysed years.

In the case of China's manufacturing industry, this is the first study to analyse the entire sample without imposing restrictions, offering new insights or refining previous findings based on geographically, sectoral, or firm-type restricted subsamples. e.g., focusing solely on medium and large firms. The empirical analysis reveals that agglomeration predominantly occurs along the eastern coast of the country, gradually diminishing as it moves westward and northward, though significant heterogeneity exists within all regions. Agglomeration is concentrated in a limited number of sectors and geographic areas, particularly in prefectures with a strong urban profile. Manufacturing activities related to natural resources, capital-intensive sectors, and some technology-driven industries exhibit the highest levels of spatial concentration. Conversely, labour-intensive industries show a relative dispersion, a novel finding compared to previous literature but consistent with theoretical frameworks. Lastly, Economic Development Zones (EDZs) appear to have no significant distortive impact on the spatial concentration process of China's manufacturing sector.

As is likely the case with most research works, and particularly with doctoral theses, many aspects must be set aside, at least temporarily, to bring the project to completion. This thesis is no exception, as the initial scope of the research was more ambitious than what has been achieved. These are issues that will likely be addressed in the coming years. Three main areas have been deferred:

1. **Additional measurement aspects**: While the proposed family of indicators is adaptable, there are potential extensions that remain unexplored. These include bilateral co-agglomeration and density-based (rather than cumulative) measurements, which would allow for a more comprehensive analysis and

comparison of results. Additionally, with the same dataset for Madrid, it would be valuable to assess how business dynamics (entries, exits, and turnover) are influenced by the levels of agglomeration or co-agglomeration, an area of interest for evaluating the geographical characteristics of the city.

2. **Theoretical limitations**: A notable constraint lies in the interpretation and application of individual indicators. The relative position of production points within a territory, especially their proximity to others, means that the measurement of spatial concentration reflects both active agglomeration, caused by individual production points attracting others, and passive agglomeration, resulting from proximity to areas dominated by active agglomeration. A potential avenue for refinement could involve hybridising the distance-based methodology with clustering techniques and Gibbs models to improve and perfect these indicators.

3. **Determinants and effects**: A notable area of research, primarily relying on first and second-generation measures, has focused on explaining the causes or determinants of productive concentration processes, as well as their effects. These effects span business outcomes (e.g., productivity, profitability), factors of production (e.g., wages), distributive elements (e.g., per capita income), and urban structures. From this perspective, a key objective for future research is to establish both theoretical and empirical links between sectors and areas with higher levels of agglomeration and co-agglomeration and business dynamics. The underlying hypothesis is that greater productive agglomeration is associated with higher business turnover.

A second limitation, common to all research, stems from the availability of data and computational resources. Data limitations restrict the estimation of desirable metrics (e.g., weighted estimators, firm-level rather than establishment-level data, or inter-firm relationships that could produce more realistic indicators). While the methods proposed in this thesis significantly reduce the computational burden of these indicators, current technological constraints still limit their application to contexts with a very large number of production points.

Despite these limitations and challenges, adapting a sentiment from the author Anne McCaffrey, I hope that this thesis has provided solutions rather than dwelling excessively on the problems that could not yet be resolved.

REFERENCES

Asian Development Bank. (2021). *The 14th Five-Year Plan of the People's Republic of China - Fostering High-Quality Development.*

Albert, J. M., Casanova, M. R., & Orts, V. (2012). Spatial location patterns of Spanish manufacturing firms. *Papers in Regional Science*, *91*(1), 107-137.

Alder, S., Shao, L., & Zilibotti, F. (2016). Economic reforms and industrial policy in a panel of Chinese cities. *Journal of Economic Growth*, *21*(4), 305–349.

Andersson, M., & Lööf, H. (2011). Agglomeration and productivity: evidence from firm-level data. *The Annals of Regional Science*, *46*, 601-620.

Arbia, G. (2001). Modelling the geography of economic activities on a continuous space. *Papers in Regional Science*, *80*(4), 411–424.

Arbia, G., Espa, G., Giuliani, D., & Mazzitelli, A. (2012). Clusters of firms in an inhomogeneous space: The high-tech industries in Milan. *Economic Modelling*, *29*(1), 3-11.

Audretsch, D.B., & Feldman, M.P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, *86*(3), 630–640.

Baddeley, A.J., Møller, J., & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, *54*(3), 329–350.

Bai, C. E., Du, Y., Tao, Z., & Tong, S. Y. (2004). Local protectionism and regional specialization: evidence from China's industries. *Journal of International Economics*, *63*(2), 397-417.

Barrios, S., Bertinelli, L., & Strobl, E. (2006). Coagglomeration and spillovers. *Regional Science and Urban Economics*, *36*(4), 467-481.

Belt and road forum. (2023). *Third Belt and Road Forum for International Cooperation.* http://www.beltandroadforum.org/english/index.html

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, *122*(3), 507-553.

Bertinelli, L., & Decrop, J. (2005). Geographical agglomeration: Ellison and Glaeser's index applied to the case of Belgian manufacturing industry. *Regional Studies*, *39*(5), 567–583.

Bonneu, F., & Thomas-Agnan, C. (2015). Measuring and testing spatial mass concentration with micro-geographic data. *Spatial Economic Analysis*, *10*(3), 289-316.

Brakman, S., Garretsen, H., & Zhao, Z., (2017). Spatial concentration of manufacturing firms in China. *Papers in Regional Science*, *96*, S179–S206.

Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). WTO accession and performance of Chinese manufacturing firms. *American Economic Review*, *107*(9), 2784–2820.

Briant, A., Combes, P. P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?. *Journal of Urban Economics*, *67*(3), 287-302.

Brülhart, M., & Mathys, N. (2008). Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics*, *38*(4), 348–362.

Brülhart, M., & Traeger, R. (2005). An account of geographic concentration patterns in Europe. *Regional Science and Urban Economics*, *35*(6), 597–624.

Carlino, G. A. (1978). Agglomeration economies: a survey of the literature. *Economies of scale in manufacturing location.* (vol.12, pp. 6-29). Springer.

Casanova, M. R., Orts, V., & Albert, J. M. (2017). Sectoral scope and colocalisation of Spanish manufacturing industries. *Journal of Geographical Systems*, *19*, 65-92.

Chain, C. P., Santos, A. C. D., Castro, L. G. D., & Prado, J. W. D. (2019). Bibliometric Analysis of the Quantitative Methods Applied to the Measurement of Industrial Clusters. *Journal of Economic Surveys*, *33*(1), 60–84.

Ciccone A., & Hall R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, *86*(1), 54–70.

Coll-Martínez, E., Moreno-Monroy, A. I., & Arauzo-Carod, J. M. (2019). Agglomeration of creative industries: An intra-metropolitan analysis for Barcelona. *Papers in Regional Science*, *98*(1), 409-432.

Combes, P. P., & Duranton, G. (2006). Labour pooling, labour poaching, and spatial clustering. *Regional Science and Urban Economics*, 36(1), 1-28.

Combes, P. P., & Overman, H. G. (2004). The spatial distribution of economic activities in the European Union. In *Handbook of regional and urban economics* (Vol. 4, pp. 2845-2909). Elsevier.

Combes, P. P., Thisse, J. F., & Mayer, T. (2008). *Economic geography: The integration of regions and nations*. Princeton University Press.

Cressie, N. (1993). Spatial point patterns, *Statistics for Spatial Data* (pp. 575–723). John Wiley & Sons.

Dahl, M. S., & Klepper, S. (2015). Whom do new firms hire?. *Industrial and corporate change*, *24*(4), 819-836.

De Dominicis, L., Arbia, G., & De Groot, H. L. (2013). Concentration of manufacturing and service sector activities in Italy: Accounting for spatial dependence and firm size distribution. *Regional Studies*, *47*(3), 405-418.

Devereux, M. P., Griffith, R., & Simpson, H. (2004). The geographic distribution of production activity in the UK. *Regional science and urban economics*, 34(5), 533-564.

Diggle, P.J., & Chetwynd, A.G. (1991). Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, 47(3), 1155–1163.

Diggle, P. J., Gómez-Rubio, V., Brown, P. E., Chetwynd, A. G., & Gooding, S. (2007). Second-order analysis of inhomogeneous spatial point processes using case–control data. *Biometrics*, *63*(2), 550-557.

Diodato, D., Neffke, F., & O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, *106*, 1-26.

Dumais, G., Ellison, G., & Glaeser, E. L. (2002). Geographic Concentration as a Dynamic Process. *The Review of Economics and Statistics*, *84*(2), 193–204.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics* (Vol. 4, pp. 2063-2117). Elsevier.

Duranton, G., & Overman, H. G. (2002). *Localisation in UK manufacturing industries*. CEPR Discussion Paper 3379. Centre for Economic Policy Research.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077-1106.

Duranton, G., & Overman, H. G. (2008). Exploring the detailed location patterns of UK manufacturing industries using microgeographic data. *Journal of Regional Science*, *48*(1), 213-243.

Eberts, R. W., & McMillen, D. P. (1999). Agglomeration economies and urban public infrastructure. *Handbook of regional and urban economics* (Vol. 3, pp.1455-1495). Elsevier.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of Political Economy*, *105*(5), 889-927.

Ellison, G., & Glaeser, E. L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration?. *American Economic Review*, *89*(2), 311-316.

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), 1195-1213.

Faggio, G., Silva, O., & Strange, W. C. (2017). Heterogeneous agglomeration. *Review of Economics and Statistics*, *99*(1), 80-94.

Faggio, G., Silva, O., & Strange, W. C. (2020). Tales of the city: what do agglomeration cases tell us about agglomeration in general?. *Journal of Economic Geography*, *20*(5), 1117-1143.

Fan, C. C., & Scott, A. J. (2003). Industrial agglomeration and development: a survey of spatial economic issues in East Asia and a statistical analysis of Chinese regions. *Economic geography*, *79*(3), 295-319.

Fujita, M., & Thisse, J. F. (2002). *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press.

Fujita, M., Krugman, P., & Venables, A.J. (1999). *The Spatial Economy: Cities, Regions, and International Trade.* MIT Press.

Giuliano, G., Kang, S. & Yuan, Q. (2019). Agglomeration economies and evolving urban form. *The Annals of Regional Science, 63*, 377–398.

Glaeser, E. L. (1999). Learning in cities. *Journal of urban Economics*, *46*(2), 254-277.

Glaeser, E. L., & Kahn, M. E. (2001). *Decentralized employment and the transformation of the American city* (NBER Working Paper No. w8117). National Bureau of Economic Research.

Goldstein, G. S., & Gronberg, T. J. (1984). Economies of scope and economies of agglomeration. *Journal of Urban Economics*, *16*(1), 91-104.

Gómez-Antonio, M., & Alañón-Pardo, Á. (2020). Point pattern methods for analyzing industrial location. *Investigación económica*, *79*(314), 51-74.

He, C., & Wang, J. (2010). Geographical agglomeration and co-agglomeration of foreign and domestic enterprises: A case study of Chinese manufacturing industries. *Post-Communist Economies*, *22*(3), 323-343.

He, C., & Wang, J. (2012). Regional and sectoral differences in the spatial restructuring of Chinese manufacturing industries during the post-WTO period. *GeoJournal*, *77*, 361-381.

He, C., Wei, Y. D., & Pan, F. (2007). Geographical concentration of manufacturing industries in China: The importance of spatial and industrial scales. *Eurasian geography and economics*, *48*(5), 603-625.

He, C. F., & Zhu, S. J. (2009). Industrial agglomeration and labour productivity in transition: An empirical study of Chinese manufacturing industries. *Post-Communist Economies*, *21*(1), 103-115.

Helsley, R. W., & Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*, *20*(2), 189-212.

Helsley, R. W., & Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, *122*(5), 1064-1093.

Henderson, V., & Thisse, J. F. (2004). *Handbook of regional and urban economics: cities and geography* (Vol. 4). Elsevier.

Hong, J., & Fu, S. (2011). Information and communication technologies and the geographical concentration of manufacturing industries: Evidence from China. *Urban studies*, *48*(11), 2339-2354.

Howard, E., Newman, C., & Tarp, F. (2016). Measuring industry coagglomeration and identifying the driving forces. *Journal of Economic Geography*, *16*(5), 1055-1078..

Huang, Y., Sheng, K., & Sun, W. (2022). Influencing factors of manufacturing agglomeration in the Beijing-Tianjin-Hebei region based on enterprise big data. *Journal of Geographical Sciences*, *32*(10), 2105-2128..

Huang, Y., Zhuo, L., & Cao, J. (2024). An Integrated Duranton and Overman Index and Local Duranton and Overman Index Framework for Industrial Spatial Agglomeration Pattern Analysis. *ISPRS International Journal of Geo-Information*, *13*(4), 116..

Kominers, S. D. (2008). Measuring agglomeration. In *Harvard Urban and Social Economics Seminar (Economics 2800b) http://www. scottkom. com/articles/measure_agglomeration. pdf*.

Kopczewska, K. (2018). Cluster-based measures of regional concentration. Critical overview. *Spatial statistics*, *27*, 31-57.

Kopczewska, K., Churski, P., Ochojski, A., & Polko, A. (2019). SPAG: Index of spatial agglomeration, *Papers in Regional Science*, *98*(6), 2391-2425.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, *99*(3), 483–499.

Krugman, P. (1992). *Geography and Trade*. MIT Press.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79-86.

Lagache, T., Lang, G., Sauvonnet, N., & Olivo-Marin, J. C. (2013). Analysis of the spatial organization of molecules with robust statistics. *PLoS One*, *8*(12), e80914.

Lang, G., Marcon, E., & Puech, F. (2020). Distance-based measures of spatial concentration: introducing a relative density function. *The Annals of Regional Science*, *64*, 243-265.

Li, J., Zhang, W., Yu, J., & Chen, H. (2015). Industrial spatial agglomeration using distance-based approach in Beijing, China. *Chinese Geographical Science*, *25*, 698-712.

Li, Z., Ding, C., & Niu, Y. (2019). Industrial structure and urban agglomeration: evidence from Chinese cities. *The Annals of Regional Science*, *63*, 191-218.

Lin, H.L., Li, H.Y., & Yang, C.H. (2011). Agglomeration and productivity: Firm-level evidence from China's textile industry. *China Economic Review*, *22*(3), 313–329.

Liu, Y., Wu, Y., & Zhu, X. (2023). Development zones and firms' performance: the impact of development zones on firms' performance for a Chinese industrial cluster. *Regional Studies*, *57*(5), 868-879.

Liu, Y., Zhang, X., Pan, X., Ma, X., & Tang, M. (2020). The spatial integration and coordinated industrial development of urban agglomerations in the Yangtze River Economic Belt, China. *Cities*, *104*, 102801.

Long, C., & Zhang, X. (2012). Patterns of China's industrialization: Concentration, specialization, and clustering. *China Economic Review*, *23*(3), 593-612.

Loop, M. S., & McClure, L. A. (2015). Testing for clustering at many ranges inflates family-wise error rate (FWE). *international Journal of health geographics*, *14*(1), 1-6.

Loosmore, N. B., & Ford, E. D. (2006). Statistical inference using the G or K point pattern spatial statistics. *Ecology*, *87*(8), 1925-1931.

Lu, J. (2010). Agglomeration of economic activities in China: Evidence from establishment censuses. *Regional Studies*, *44*(3), 281-297.

Lu, J., & Tao, Z. (2009). Trends and determinants of China's industrial agglomeration. *Journal of urban economics*, *65*(2), 167-180.

Lu, Y., Wang, J., & Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from China’s economic zone program. *American Economic Journal: Economic Policy*, *11*(3), 325-360.

Mao, Q., Dong, S., Wang, F., & Li, J. (2013). Evolving spatial distribution of manufacturing industries in China. *Acta Geographica Sinica*, *68*(4), 435–448.

Marcon, E., & Puech, F. (2003). Evaluating the geographic concentration of industries using distance-based methods. *Journal of economic geography*, *3*(4), 409-428.

Marcon, E., & Puech, F. (2010). Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, *10*(5), 745-762.

Marcon, E., & Puech, F. (2017). A typology of distance-based measures of spatial concentration. *Regional Science and Urban Economics*, *62*, 56-67.

Marcon, E., Puech, F., & Traissac, S. (2012). Characterizing the relative spatial structure of point patterns. *International Journal of Ecology*, *2012*(1), 619281.

Marshall, A. (1890). *Principles of Economics*. Springer.

Maurel, F., & Sédillot, B. (1999). A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, *29*(5), 575-604.

Mori, T., & Smith, T. E. (2014). A probabilistic modeling approach to the detection of industrial agglomerations. *Journal of Economic Geography*, *14*(3), 547-588.

Mori, T., Nishikimi, K., & Smith, T. E. (2005). A divergence statistic for industrial localization. *Review of Economics and Statistics*, *87*(4), 635-651.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(2), 381-404.

Nathan, M., & Overman, H. G. (2013). Agglomeration, clusters, and industrial policy. *Oxford review of economic policy*, *29*(2), 383-404.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlated coefficients: Three experiments on the modifiable areal unit problem. In *Statistical applications in the spatial sciences* (pp. 127–144). Pion.

Ottaviano, G., & Thisse, J. F. (2004). Agglomeration and economic geography. *Handbook of regional and urban economics* (Vol. 4, pp. 2563-2608). Elsevier.

Pélissier, R., & Goreaud, F. (2001). A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. *Journal of Vegetation Science*, *12*(1), 99-108.

Penttinen, A. (2006). Statistics for marked point patterns. *The Yearbook of the Finnish Statistical Society* (pp. 70-91). Statistics Finland.

Penttinen, A., Stoyan, D., & Henttonen, H. M. (1992). Marked point processes in forest statistics. *Forest science*, *38*(4), 806-824.

Porter, M. E. (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic development quarterly*, *14*(1), 15-34.

Quigley, J. M. (1998). Urban diversity and economic growth. *Journal of Economic Perspectives*, *12*(2), 127-138.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, *13*(2), 255-266.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(2), 172–192.

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics*, *50*(2), 191-229.

Rysman, M., & Greenstein, S. (2005). Testing for agglomeration and dispersion. *Economics Letters*, *86*(3), 405-411.

Saxenian, A. (1996). *Regional advantage: Culture and competition in Silicon Valley and route 128, with a new preface by the author*. Harvard University Press.

Scholl, T., & Brenner, T. (2016). Detecting spatial clustering using a firm-level cluster index. *Regional Studies*, *50*(6), 1054-1068.

Steijn, M. P., Koster, H. R., & Van Oort, F. G. (2022). The dynamics of industry agglomeration: Evidence from 44 years of coagglomeration patterns. *Journal of Urban Economics*, *130*, 103456.

Sweeney, S., & Gómez-Antonio, M. (2016). Localization and industry clustering econometrics: An assessment of Gibbs models for spatial point processes. *Journal of Regional Science*, *56*(2), 257-287.

Tabuchi, T. (1998). Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics, 44*(3), 333–351.

Tabuchi, T., & Yoshida, A. (2000). Separating urban agglomeration economies in consumption and production. *Journal of Urban Economics*, *48*(1), 70-84.

Tian, S., Wang, J., Gui, Z., Wu, H., & Wang, Y. (2017). A case study: Exploring industrial agglomeration of manufacturing industries in Shanghai using Duranton and Overman's K-density function. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 149-154.

Venables, A. J. (1996). Equilibrium locations of vertically linked industries. *International economic review*, 341-359.

Wang, J. (2013). The economic impact of special economic zones: Evidence from Chinese municipalities. *Journal of development economics*, *101*, 133-147.

Weber, A. (1909). *Über den Standort der Industrien.* Tübingen, J.C.B. Mohr (Paul Siebeck); translated by Friedrich, C. J. (1929) as *Alfred Weber's Theory of the Location of Industries*. University of Chicago Press.

Wei, Y. D. (1999). Regional inequality in China. *Progress in human geography*, *23*(1), 49-59.

Wei, Y. D. (2001). Decentralization, marketization, and globalization: The triple processes underlying regional development in China. *Asian Geographer*, *20*(1-2), 7-23.

Wen, M. (2004). Relocation and agglomeration of Chinese industry. *Journal of Development Economics*, *73*(1), 329–347.

Wu, X., Huang, Y., & Gao, J. (2022). Impact of industrial agglomeration on new-type urbanization: Evidence from Pearl River Delta urban agglomeration of China. *International Review of Economics & Finance*, *77*, 312-325.

Xi, J. P. (2022). Hold high the great banner of socialism with Chinese characteristics and strive in unity to build a modern socialist country in all respects. *Report to the 20th National Congress of the Communist Party of China.*

Zhao, X. B., & Zhang, L. (1999). Decentralization reforms and regionalism in China: a review. *International regional science review*, *22*(3), 251-281.

Zheng, S., Sun, W., Wu, J., & Kahn, M. E. (2017). The birth of edge cities in China: Measuring the effects of industrial parks policy. *Journal of Urban Economics*, *100*, 80-103.