Banco de México

Working Papers

N° 2022-06

The Role of Clusters in the Performance of the Mexican Economy

Miriam Juárez-Torres
Banco de México

Jonathan Puigvert-Angulo
Simon Fraser University

Francisco Zazueta-Borboa
Banco de México

April 2022

Documento de Investigación
2022-06

Working Paper
2022-06

The Role of Clusters in the Performance of the Mexican Economy*

Miriam Juárez-Torres†
Banco de México

Jonathan Puigvert-Angulo‡
Simon Fraser University

Francisco Zazueta-Borboa§
Banco de México

Abstract: This paper follows an algorithm that considers different dimensions of linkages across service and manufacturing industries to identify a cluster configuration of the Mexican economy and analyze their role in the economic performance of regions. It identifies 24 clusters and analyzes their geographical distribution, their role in regional growth, the evolution of their employment concentration, and their spillover effects. The main findings suggest that manufacturing-oriented clusters have a strong presence in the Northern states of the country, while services-oriented clusters in the Central ones. Finally, clusters such as plastic products manufacturing; retail and eating services; food and beverage manufacturing; and, automotive show relatively high direct and indirect spillover effects on the economy.
Keywords: Clusters, Agglomeration economies, Employment concentration, Economic spillovers effects
JEL Classification: L60, L80, O54

Resumen: Este documento sigue un algoritmo que considera diferentes dimensiones de los vínculos entre las industrias de servicios y manufactura para identificar una configuración de clúster para la economía mexicana y analizar su papel en el desempeño económico de las regiones. Se identifican 24 clústeres y se analiza su distribución geográfica, su papel en el crecimiento regional, la evolución de su concentración del empleo y sus efectos indirectos. Los principales hallazgos sugieren que los clústeres orientados a la manufactura tienen una fuerte presencia en los estados del norte del país, mientras que los clústeres orientados a servicios en las entidades del centro. Finalmente, clústeres tales como la fabricación de productos plásticos; servicios de comida y venta al por menor; fabricación de alimentos y bebidas; y la industria automotriz muestran efectos directos e indirectos relativamente altos en la economía.
Palabras Clave: Clústeres, Economías de aglomeración, Concentración del empleo, Efectos de derrama económica

---

*The authors are thankful to the Microdata Laboratory of the National Institute of Geography and Statistics (INEGI), in particular to Natalia Volkow, Lidia, and Liliana Martinez for the support provided to access information and for preparing this research document. Authors are grateful to Kenya Castillo and Luis Colín for their excellent research assistance, and Alejandrina Salcedo Cisneros and the two anonymous referees for their helpful and insightful comments.

† Dirección General de Investigación Económica. Email: mjuarez@banxico.org.mx.

‡ Department of Economics, Simon Fraser University. Email: jpuigvert@gmail.com.

§ Dirección General de Investigación Económica. Email: fzazuetab@banxico.org.mx.

1. Introduction

The study of clusters has been motivated by the need to identify the forces that might lead regional economic development and growth. In this context, the objective of this paper is to determine a statistically robust cluster configuration for the Mexican economy to analyze their role in regional and local growth by studying the evolution of their employment concentration and the spillover effects involved. In this context, this paper fills a gap in the empirical economic literature about clustering in Mexico by i) including a broad range of economic activities whose products and services are traded on local, regional, and international markets and, ii) by using a methodology that incorporates different linkage dimensions between activities across the Mexican economy. This comprehensive framework analysis allows conducting an enriched analysis that identifies some elements that might guide an industrial development policy to foster economic growth in local and regional economies.

Clusters are concentrations of interconnected firms and associated institutions with related economic activities that show important externalities of specialized industrial locations (Porter, 2003; Feser and Bergman, 2000). Clusters involve multiple relationships such as input-output or buyer-supplier linkages, geographical co-location, shared business, local institutions, employment, and formal or informal cooperative competition (Feser and Bergman, 2000). A proper cluster configuration is rarely represented by the standard industrial classification systems that inaccurately capture the dynamics of their industrial relationships. In terms of policy, the analysis of clusters supports the design of public policies to boost economic growth, because clusters foster existing synergies across economic activities by triggering innovation, productivity, and economic performance (Figueiredo et al., 2009).

Although clusters can be a potential tool for designing economic policy, identifying these groups remains a central matter in the empirical economic literature due to the lack of statistically robust methods that provide a unique cluster configuration; even if these approaches have been continuously improved since the early 2000s, (see Feser and Bergman, 2000; Feser et al., 2005; Delgado et al., 2014).

In this context, early seminal methodologies were based on graph theory, triangularization, factor analysis, and, later, the detection of spatial relatedness of arbitrarily defined sets of industries; or alternatively based on individual authors' own criteria (see Getis and Ord, 1995). First, the influential work of Feser and Bergman (2000) introduced factor analysis to identify specific industrial clusters based on direct and indirect linkages computed from input-output transactions of the manufacturing sector in 1987, regardless of geographical location. They defined 23 mutually exclusive manufacturing clusters of related industries in North Carolina, a key manufacturing state in the Southeastern U.S. economy.

Later, Feser et al. (2005) included the spatial dimension of so-called ‘hot spots’ in the geographical distribution of value chains by using the local *G*-statistic from Ord and Getis (1995). They systematically identified high-employment regions to conduct an interdependence analysis of sectoral employment and wages for all businesses subject to federal and state employment security law in the U.S. with data from 1989 and 1997. They applied factor analysis on industry level data from the 1992 U.S. input–output to analyze extended buyer–supplier value chains and to derive 26 value chains.

In a similar work, Argüelles et al. (2014) implemented hierarchical clustering on principal components to identify clusters based on input-output inter-industry linkages in Spain. Instead of using only factor analysis to improve the robustness of the agglomeration, they combined factor analysis and clustering methods based on a mixed algorithm: Ward’s hierarchical classification and an aggregation around mobile centers (K-means). Later, Alcácer and Zhao (2016) employed a density-based cluster identification algorithm to characterize the global semiconductor industry in some areas of the U.S. economy.

In general, methods described above might generate a non-unique cluster configuration due to the different dimensions of the possible ties. For this reason, it is advised performing several statistical analyses on different choices to check for the robustness of the selected configuration (Everitt et al., 2011). In this sense, the methodology of Delgado et al. (2016), which we follow, introduces a five-step method to generate and assess a finite set of cluster definitions to select the best quality cluster configuration in terms of its capacity to capture multiple types of inter-industry links. These authors used the Benchmark Input-Output Account of the United States (County Business Patterns – CBP) and Occupational Employment Statistics (OES) data sets to define multiple industry relatedness measures for the 778 six-digit North American Classification System (NAICS) industries in manufacturing and services in 2009 and found 51 clusters that could be mapped consistently into U.S. regions.

In the context of the empirical economic literature, most of the research focuses on developed economies and case studies within these countries, while the research for developing economies remains barely studied mainly because of the lack of disaggregated data. In particular for the Mexican economy, the application of the cluster approach has been relatively limited and few articles can be cited. For instance, Dávila (2005) identified 12 industrial clusters and evaluated their economic performance (value added and employment). In addition, he studied the evolution of the geographical location pattern of the industrial sector using the approach suggested by Feser and Bergman (2000) based on factor analysis to identify industrial clusters. Later, Villarreal and Flores (2015) used the 1996 input-output accounts and continuous spatial data techniques, such as the nearest neighbor index and hierarchical nearest-neighbor clustering to identify innovation agglomerations across industries and measure the relative specialization of each cluster. Also, Chávez and García (2015), using the standard industrial classification system, combined location coefficients and the Local Moran I index (spatial autocorrelation) to identify regional manufacturing clusters based on data from the 1994 and 2009 Economic Censuses, published by the National Institute of Statistics and Geography (INEGI by its acronym in Spanish). Their results allowed them to compare these agglomerations before and after the enactment of the North American Free Trade Agreement. Later, Villareal et al. (2017) identified and analyzed the location of manufacturing agglomerations in Mexico to design regional industrial policies by extracting relationships from the 2003 input-output matrix of INEGI and the geographical location of employment from of the 2009 Economic Census using principal component analysis. They found relevant manufacturing clusters such as electronics, automobile, chemical, apparel, and food manufacturing. However, the previously cited papers focused on the manufacturing sector, while service activities remain barely analyzed.

In this context, the main contribution of this paper consists of incorporating service industries into the cluster definition and using this framework for the analysis of the role of these agglomerations in regional economic performance. At the same time, the application of a more comprehensive methodology allows us to disclose relevant economic interdependence and externalities across traded industries, local industries, and natural resource-based industries.[1] Thus, shared natural advantages such as coastal access, mining or water availability can influence, in low extent, the colocation decision of firms, and economic interdependencies and externalities between industries. For this reason, we consider all industries with available information, including mining and oil and gas extraction, as well as secondary and tertiary activities, and only exclude primary and some retailing activities due to incomplete data. In a similar way, Delgado et al. (2016) incorporates resource-based industries into the analysis of clusters; for instance, crude petroleum and natural gas extraction, and oil and gas extraction.

To identify the clusters, we follow the methodology of Delgado et al. (2016) and consider a set of 230 four-digit NAICS industries that represent about 80% of the national gross value added according to the Economic Census 2019. For these activities, we employ information from the 2019 National Economic Census and the 2013 Input-Output Matrix, both published by INEGI.[2] We find an optimal configuration of 24 agglomerations[3] of related

[1] According to Delgado et al. (2016), "Local industries are those that serve primarily the local markets (e.g., retail), whose employment is evenly distributed across regions in proportion to regional population. Traded industries are those that are more geographically concentrated and produce goods and services that are sold across regions and countries. The set of traded industries excludes natural-resource-based industries whose location is tied to local resource availability (e.g., mining)" (page 6).
[2] The 2019 Economic Census contains basic statistical information about all establishments producing goods, merchandise marketers, and service providers to generate Mexico's economic indicators with a high level of geographical, sectoral, and thematic detail, referring to the year 2018.
[3] The 24 clusters are named according to the industry with the highest share in its gross value-added composition. These are: C1-oil and gas extraction; C2-metal mining; C3-footwear manufacturing; C4-sawmills and wood preservation; C5-medical equipment and supplies manufacturing; C6-semiconductors and other electronic components; C7-food and beverage manufacturing; C8-automotive; C9-petroleum and coal products

industries. Once the cluster configuration is defined, we study employment concentration for the clusters and the regions. In this agglomeration arrangement, we show a widespread presence of some particular agglomerations across the country: i) C7 food and beverage manufacturing; ii) C18 retail and eating services; and iii) C20 plastic products manufacturing. Clusters characterized by high technological development and located at municipalities in the North of the country such as: i) C5 medical equipment and supplies manufacturing; ii) C6 semiconductors and other electronic components; and iii) C14 steel products manufacturing show an increase in employment concentration in last decade, which could reflect the existence of strong complementarities and synergies of production with other related industries located in the U.S. In terms of policy, we find consistent results from two analytical exercises that allow us to identify the agglomerations that might influence the performance of other economic activities across the economy and regions e.g.: i) C7 food and beverage manufacturing; ii) C18 retail and eating services; iii) C20 plastic products manufacturing; and iv) C8 automotive.

This research is organized as follows. Section 2 presents a review of the methodology employed, whereas section 3 introduces the data and sources used in this paper. Section 4 describes the robust cluster configuration of related industries found for the Mexican economy and analyzes the geographical distribution of the employment for the 24 identified clusters and their growth for the period 2014-2019. Section 5 analyzes the evolution of the employment concentration at the municipality level for all the clusters of related industries across four periods: 2004, 2009, 2014, and 2019. Section 6 explores the role of the cluster in the growth of the economy at the municipality level. Section 7 provides a spillover effects analysis for the clusters in the economy. Finally, section 8 concludes.

---

manufacturing; C10-apparel manufacturing; C11-tourism and hospitality services; C12-office administrative services; C13-metal products manufacturing; C14-steel products manufacturing; C15-financial services and head offices; C16-electric power generation, transmission, and distribution, and infrastructure construction; C17-passenger transportation and communications; C18-retail and eating services; C19-employment services; C20-plastic products manufacturing; C21-freight transportation services and residential and nonresidential construction; C22-business support services; C23-education and health services; and C24-pharmaceutical and medical manufacturing and services.

2. Clustering methodology

We deploy a five-step method used to find the best cluster configuration in terms of its ability to capture the multiple existing types of linkages between the studied industries. We follow the methodology of Delgado et al. (2014) to define a statistically robust cluster configuration of the Mexican economy. Linkages between industries are captured through similarity matrices M_{ij} that measure relatedness or proximity between the industries. These matrices are composed of as many rows and columns as industries, and the element m_{ij} measures the relationship between industries i and j. In general, this methodology uses different quantitative measures of relatedness between industries to generate multiple cluster configurations and then evaluates every possible arrangement generated based on scores that assess two criteria: i) how individual clusters are different from each other, and ii) the fit of individual industries into their own cluster to find the optimal configuration. See Annex A for details on this methodology based on Delgado et al. (2016).

2.1 Similarity matrices

We use five similarity matrices based on three different relatedness measures that summarize different dimensions of the relationships across industries. For similarity matrices 1 and 2 we use the Local Correlation (LC) coefficient that captures geographical proximity and might reflect relevant economic interdependencies between a pair of industries in the location. The LC similarity matrix 1 is for employment and the LC similarity matrix 2 for establishments. For the similarity matrix 3 we employ the Co-Agglomeration Index (COI), also known as GE index (Ellison and Glaeser (1999) and Ellison et al. (2010)), quantifies the magnitude of concentration or co-agglomeration of two industries. Broadly speaking, the COI can be associated with the covariance of the area industry employment shares of the two industries, normalized to rule out the response to the size of the geographical area considered: the more positive the value, the greater the externalities from concentration. For the similarity matrix 4 at the national level, we use the input-output links account for buyer-supplier relations between a pair of industries based on their patterns of sales and purchases. Finally, for the multidimensional similarity matrix 5, we calculate it as the average of the four previously normalized similarity matrices, and it captures the four dimensions of relatedness between a pair of industries. See Annex A for additional details about definitions and calculations of similarity matrices used in this paper.

2.2 Parameter selection

Clustering techniques are part of the numerical methods applied to multivariate data to detect groups or clusters of homogeneous observations. However, these statistical methods are inconclusive with respect to identifying the 'optimal' number of agglomerations (Everitt et al., 2011) and, for this reason, parameter selection is required for the clustering function. According to Delgado et al. (2016), the selection of parameters resulting in a small number of clusters could generate a huge agglomeration including industries with low levels of relativeness between them, while, in contrast, a large number of clusters could produce groups that are not meaningfully different from each other. In particular, we must define the starting value for the number of agglomerations and the type of normalization of the underlying data for the clustering function. We set the number of cluster configurations between 10 and 40 agglomerations, a range based on previous findings from the economic literature about the Mexican economy (see Villareal et al., 2017 and INEGI, 2016). In addition, following Everitt et al. (2011), we normalize the similarity matrices using the centroid[4] for every matrix, ensuring all their elements have the same units, so larger entries indicate closer relationships between the corresponding row and column elements.

2.3 Clustering function

We select a clustering function and choose the parameters described above to create the agglomeration configurations using the similarity matrices. We have decided to use a hierarchical clustering technique combining agglomerative methods to create a series of successive fusions of individuals n into groups or divisive methods that separate the individuals n successively into finer groupings (Everitt et al., 2011). Hierarchical classifications made by both techniques can be represented by dendrograms.[5] We use the

[4] The centroid is the arithmetic mean position of all the points in a given figure or space, the center of the mass of an object of uniform density.

[5] The dendrogram is a tree-shaped diagram showing the sequential process of formation of clusters given the similarity (dissimilarity) measures between observations at each step. The level of similarity (dissimilarity)

hierarchical algorithm proposed by Ward (1963), which generates different configurations of clusters (mutually exclusive sets of industries) with the groups of industries closest to each other for every similarity matrix; the fusion of two clusters depends on the size of a sum of squared errors criterion. Thus, the clustering proceeds hierarchically at each level by merging clusters from the previous level and the objective at every stage is to minimize the increase in the total within-cluster sum of squared error (Everitt et al., 2011). In particular, the dendrogram shows the fusions made at each stage of the analysis. A problem of hierarchical clustering methods is that this technique performs very well when the data contain clusters with approximately the same numbers of points, but poorly when the clusters are of different sizes. Also, non-uniqueness of the resulting cluster configuration is an important issue, and running analyses with different choices to check for robustness is highly recommended (Everitt et al., 2011).

2.4 Performance scores

Clustering methods generate a non-unique cluster configuration due to the different dimensions of the possible ties (Everitt et al., 2011). For this reason, the methodology calculates performance scores that assess the quality of each cluster configuration $\boldsymbol{C}$ to identify the best one, the $\boldsymbol{C}^*$ configuration. To estimate these performance scores, we define a different set of similarity matrices than those used for the generation of the clusters, which allows us to compute independent validation scores that can be ranked regardless of the information matrices used to determine $\boldsymbol{C}$ (Everitt et al., 2011). Thus, for the score calculations, we use the Euclidian distance[6] to compute the dissimilarity matrices needed to estimate the performance scores.[7]

---

between observations and successive groupings are measured on the vertical axis, while observations and clusters are specified on the horizontal axis. see Annex B4.

[6] The distance between two points defined as the square root of the sum of the squares of the difference between the corresponding coordinates of the points.

[7] Delgado et al. (2016) also use a set of similarity matrices different from the initial ones that they used for the generation of the clusters but do not specify how these matrices where generated. In contrast, instead of using proximity measures to define the similarity matrices, we use Ward's distance and adjust the formulas of the validation scores in this section accordingly.

Following the methodology of Delgado et al. (2016), we assess Cluster and Industry level validation scores to assemble a Global validation score that allows us to identify the best cluster configuration $\boldsymbol{C^*}$. These scores are based on two conditions: i) how different from each other individual clusters are; and, ii) how well individual industries fit into their own cluster. First, for every possible configuration $\boldsymbol{C}$, we calculate cluster-level validation scores (VS-Cluster) as the percentage of clusters whose industries show a higher relationship with the industries within the same cluster (Within-Cluster Relatedness, WCR) than with the rest of the industries belonging to other clusters (Between-Cluster Relatedness, BCR), for the average and for the 95th percentile. Second, for every possible configuration $\boldsymbol{C}$**,** the industry-level validation scores (VS-Industry) are computed as the percentage of its industries with a WCR_{ic} higher than their average BCR_i for the average and for the 95th percentile, meaning the share of industries that show a higher relationship with the industries in the same cluster than with the rest of the industries belonging to the other clusters. See Annex A.2 for details about the calculation of the validation scores. Once we compute $Global - \boldsymbol{VS}$**,** as the average of $VS - Industry$ and $VS - Cluster$, we rank cluster configurations according to $Global - \boldsymbol{VS}$ to define a set of candidates of $\boldsymbol{C}^*$. Next, an additional robustness assessment is performed for the $\boldsymbol{C}^*$ candidates known as the $Overlap\ Score$ (OS). As the final step, we select the configuration $\boldsymbol{C}^*$ with the best performance in the $Global\ \boldsymbol{VS}$ and the $Overlap\ Score$.

*2.5 Assessing individual clusters of candidate C**

Following Delgado et al. (2016), once the $\boldsymbol{C}^*$ has been identified, we can improve this configuration by correcting the allocation of two kinds of outliers resulting from possible spurious industry groupings within clusters. We identify two types of outliers: the systematic and the marginal ones. The systematic outliers show a high overall WCR_{ic} score (based on the average of the standardized sub-scores WCR_{ic} for the five similarity matrices), indicating that they could be allocated to another cluster to improve the general allocation. On the other hand, the marginal outliers are those industries that could fit conceptually better into another cluster, even with a low WCR_{ic}. We use a rule of the thumb based on industry WCR_{ic} and an expert judgment to reassign these outliers into individual clusters in $\boldsymbol{C}^*$ by the following procedures: i) reallocation of industries between clusters; ii) combination of clusters; and iii) partition of clusters.

In particular, we identify as systematic outliers those industries with a WCR_{ic} two or more standard deviations larger than its WCR cluster mean. In this case, they are reassigned by an iterative process into the next cluster with a higher WCR_c and this process is performed as many times as necessary until no more systematic outliers are detected. For marginal outliers, expertise knowledge of industries determines the reallocation since any reallocation requires analyzing the product or service lines and the detailed definitions of the NAICS to identify the best fit. Also, some clusters can be combined or split to attain a better configuration, as Delgado et al. (2016) mention. If two clusters have a very high BCR and do not appear to be conceptually different, they can be combined. However, the reallocation process of the marginal outliers may not improve the overall VS score since some industries may be moved to clusters with high relative WCR_i scores. Once we have carried out the reallocation of the outliers, the resulting $\boldsymbol{C}^{**}$ is the final robust cluster configuration. Finally, we perform the overlap score as a robustness test of the cluster configuration $\boldsymbol{C}^{**}$.

3 Data and sources

We identify a robust 24-cluster configuration for the Mexican economy based on the information from the 2019 Economic Census, published by INEGI. The analysis is carried out at the four-digit industry level of the NAICS, considering all secondary and tertiary activities with available information, including mining and C1-oil and gas extraction and excluding primary activities due to incomplete data.[8] Thus, we focus our analysis on a final set of 230 four-digit NAICS industries that account for 71% of the total establishments of the country, 78% of national employment, 87% of gross national production, and 80% of the gross value added of the Mexican economy, according to the 2019 Economic Census. We also use the 2013 National Input-Output Account published by INEGI to obtain information

[8] Originally, the analysis considered a total of the 308 four-digit NAICS industries present in the economy. However, due to variations in the availability of information, we excluded 29 primary activities, 6 industries related to petroleum and other 41 activities from service and retailing activities with incomplete records.

about sales and purchases by industry. One limitation of this study is the data availability since we are not able to include all economic industries.

3.1 Data set up for the clustering algorithm

First, in order to have the same measuring units in the five similarity matrices, we normalize data by using the mean and standard deviation of the matrices computed in section two. Next, we set up the parameters for the clustering algorithm by defining an initial range for starting the search of the optimal cluster configuration. In the second stage, the selected range is from 10 to 40 agglomerations following previous findings of cluster analysis for the Mexican economy (see Villareal et al., 2017). With this algorithm, we generate a total of 31 possible cluster configurations $\boldsymbol{C}$ for every similarity matrix, which implies a total of 155 possible cluster configurations $\boldsymbol{C}$. In the third stage, we use a hierarchical clustering technique based on the algorithm proposed by Ward (1963) constructed on distances for creating all possible clusters configurations. In the fourth stage, we estimate the performance scores to select the best configuration based on statistical criteria by computing a different set of the five similarity matrices based on Euclidian distances. Finally, we perform the assessment of the candidate $\boldsymbol{C}^{*}$ and carry out an expert validation to obtain the final configuration $\boldsymbol{C}^{**}$, which in our case is composed of 24 clusters of related industries.

4 The robust cluster configuration of related industries for the Mexican economy

We apply the clustering methodology explained in section 2 to identify a robust cluster configuration of the Mexican economy. First, we calculate three of the five similarity matrices from the 2019 Economic Census using data for establishments and employment aggregated at the state level. Thus, we consider the 32 states of Mexico as the regional units of observation for the 230 industries to compute the similarity matrices: i) relatedness in location and distribution; ii) relatedness in employment distribution; and, iii) the Co-Agglomeration Index for employment. Next, we calculate the similarity matrix of buyer-supplier linkages using the input-output accounts for the 230 industries aggregated at the national level. Finally, we compute the multidimensional similarity matrix by averaging the four normalized matrices, as previously described. In Table 1, we present the descriptive statistics of the similarity matrices used to obtain the cluster configuration.

Table 1. Summary statistics for similarity matrices

230 industries, four-digit NAICS-2019 Mexican codes

Similarity Matrix	Mean	Standard Deviation	Median	Percentil 90	Max	Min
LC-Employment $_{ij}$	0.413	0.261	0.504	0.960	1.000	-0.403
LC-Establishments $_{ij}$	0.463	0.238	0.559	0.913	1.000	-0.476
Input-Output Max Share $_{ij}$	0.011	0.040	0.001	0.038	1.000	0.000
Coaglomeration Index $_{ij}$	0.066	0.007	0.070	0.073	0.169	-0.012
Mean Matrix$_{ij}$	0.009	0.562	0.009	0.635	6.179	-2.450

Source: Authors' own calculations based on information from INEGI.

We can observe in Table 1 that co-location patterns reflect the underlying economic interdependencies and externalities between industries, as well as the natural advantages and synergies between these industries. Based on these five matrices, we define our set of 155 cluster configurations (agglomerations of mutually exclusive industries, see Annex B1 for more details). To have an independent validation, we define a different set of similarity matrices based on Ward's distance (Everitt et al., 2011) that measures the closeness of an industry to all industries within the group. The closeness between two industries is measured as the Euclidean distance between a pair of characteristics of that pair of industries. We determine the $\boldsymbol{C}^*$ based on the ranking of the $Global - \boldsymbol{VS}$ and, in addition, we carry out a robustness analysis by calculating the overlap score that considers the top 29 ranking configurations (see Annex A, section A.1.3.). Table 2 shows the overlap score for the $\boldsymbol{C}^*$ configuration, resulting from the employment similarity matrices with 24 clusters, that shows a highest $Global - \boldsymbol{VS}$ of 63.75% and an overlap score of 76.1%. Table 2 presents the set of candidate configurations and their different scores and Annex B.1 shows the rest of the configurations and their respective validation scores.

Table 2. Candidate set of robust cluster configurations C*s (top-29 ranking)

Configuration	Similarity Matrix	Clustering Function	Number of Clusters	Global VS	Ranking Global VS	VS-cluster	Ranking VS-Industry	Overlap Score	Ranking Overlap
C_1	LC-Employment$_{ij}$	Herarchichal Ward	24	63.75	1	62.08	14	76.10	7
C_2	LC-Employment$_{ij}$	Herarchichal Ward	25	63.42	2	62.00	15	76.34	6
C_3	LC-Employment$_{ij}$	Herarchichal Ward	26	63.32	3	61.92	17	83.22	2
C_4	Coaglomeration Index$_{ij}$	Herarchichal Ward	21	63.30	4	75.00	1	35.97	27
C_5	LC-Employment$_{ij}$	Herarchichal Ward	23	63.14	5	61.30	28	75.48	10
C_6	LC-Employment$_{ij}$	Herarchichal Ward	22	62.76	6	61.36	25	75.05	11
C_7	LC-Employment$_{ij}$	Herarchichal Ward	27	62.67	7	61.48	20	79.10	4
C_8	Coaglomeration Index$_{ij}$	Herarchichal Ward	23	62.63	8	73.91	2	36.54	26
C_9	LC-Employment$_{ij}$	Herarchichal Ward	29	62.49	9	61.38	24	83.98	1
C_{10}	LC-Establishments$_{ij}$	Herarchichal Ward	22	62.42	10	60.91	37	71.52	17
C_{11}	LC-Employment$_{ij}$	Herarchichal Ward	28	62.36	11	61.43	21	78.86	5
C_{12}	LC-Establishments$_{ij}$	Herarchichal Ward	23	62.30	12	60.87	38	57.82	25
C_{13}	LC-Employment$_{ij}$	Herarchichal Ward	30	62.29	13	61.33	26	75.49	8
C_{14}	LC-Establishments$_{ij}$	Herarchichal Ward	24	62.26	14	60.83	39	64.62	21
C_{15}	Coaglomeration Index$_{ij}$	Herarchichal Ward	22	62.14	15	72.73	3	34.66	29
C_{16}	LC-Establishments$_{ij}$	Herarchichal Ward	29	62.01	16	60.69	45	66.68	20
C_{17}	LC-Establishments$_{ij}$	Herarchichal Ward	25	61.94	17	60.80	40	64.49	22
C_{18}	LC-Employment$_{ij}$	Herarchichal Ward	21	61.88	18	61.43	22	72.93	14
C_{19}	LC-Employment$_{ij}$	Herarchichal Ward	32	61.84	19	61.25	30	73.74	12
C_{20}	LC-Employment$_{ij}$	Herarchichal Ward	31	61.84	20	61.29	29	75.49	9
C_{21}	LC-Employment$_{ij}$	Herarchichal Ward	33	61.71	21	61.21	32	73.74	13
C_{22}	LC-Establishments$_{ij}$	Herarchichal Ward	15	61.65	22	60.00	46	63.76	23
C_{23}	LC-Establishments$_{ij}$	Herarchichal Ward	26	61.64	23	60.77	41	67.29	19
C_{24}	LC-Establishments$_{ij}$	Herarchichal Ward	14	61.58	24	60.00	47	59.05	24
C_{25}	LC-Establishments$_{ij}$	Herarchichal Ward	28	61.50	25	60.71	43	69.25	18
C_{26}	LC-Establishments$_{ij}$	Herarchichal Ward	27	61.45	26	60.74	42	71.74	16
C_{27}	LC-Establishments$_{ij}$	Herarchichal Ward	10	61.43	27	60.00	48	81.97	3
C_{28}	LC-Establishments$_{ij}$	Herarchichal Ward	21	61.36	28	60.95	36	72.67	15
C_{29}	Coaglomeration Index$_{ij}$	Herarchichal Ward	21	61.19	29	71.43	4	34.93	28

Source: Authors' own calculations based on information from INEGI.
Note: Ranking considers the 155 cluster configurations and the overlap score considers the top 29 cluster configurations based on the Global VS-score, to see the results for another's configurations see Annex B1.

In particular from configuration $\boldsymbol{C}^*$ as determined by statistical methods from the employment similarity matrix, we identify 12 systematic outliers and 50 marginal ones that are moved into a different cluster. In addition, we combine two conceptually similar clusters with a high BCR related to tourism services in a single cluster; likewise, we merge the textile and apparel clusters into one. On the other hand, we split the cluster professional services allowing industry employment services become a single cluster. Moreover, expert judgment from INEGI personnel has been extremely useful to reallocate the marginal outliers to those clusters related to services for the final consumer.[9]

[9] We are grateful for the support of Angel Fernando Pineda Solis from INEGI.

4.1 The final configuration of the 24 clusters for the Mexican economy

From the 230 industries (4-digit NAICS code) included in the analysis, we generate a unique 24-cluster configuration, with every cluster specialized in certain types of manufactured products or services. For all clusters, every industry is associated with a single cluster and agglomerations have no industries in common. Thus, every cluster can integrate several related industries and every agglomeration is named after the most important industries in terms of their gross value-added composition. For example, cluster C8-automotive is made up of seven industries that are related to the production of cars. This agglomeration includes as its main industry motor vehicle manufacturing and motor vehicle parts manufacturing, as well as other related industries like rubber product manufacturing that are part of its supply chain. On the other hand, cluster C15-financial services and head offices is composed of 17 industries where the most important industries according to gross value added are: head offices; depository credit intermediation; and wired and wireless telecommunications carriers. However, most of the other industries in this agglomeration are clearly oriented to providing financial services. For more details on cluster composition, see Annex B.2.

As can be observed in Table 3, four clusters concentrate more than a third of the gross value added recorded in the 2019 Economic Census: i) C15-financial services and head offices; ii) C8-automotive; iii) C1-oil and gas extraction; and iv) C7-food and beverage manufacturing. In the context of the gross value added per worker, a measure frequently used as a proxy for labor productivity, we can observe highly heterogenous magnitudes across clusters. On one hand, four clusters show the highest levels of productivity per worker: i) C1-oil and gas extraction; ii) C15-financial services and head offices; iii) C9-petroleum and coal product manufacturing; and iv) C2-metal mining, though they concentrate little employment nationwide (only 5%). In contrast, clusters: i) C18-retail and eating services; ii) C20-plastic products manufacturing; iii) C23-education and health services; and iv) C19-employment services[10] concentrate the highest proportion of jobs among the groups

[10] The cluster C19-Employment Services is made up of the single industry group 5613 “Employment Services”, composed of “Economic Units mainly dedicated to providing labor personnel to other economic units, according to the NAICS 2018 classification. This agglomeration groups firms dedicated to hire employees for temporary or permanent employment and the placement according to other companies’ requirements.

identified. However, these latter agglomerations show low levels of gross value added per worker. Other clusters stand out because of their share in national gross value added and their relatively well-balanced indicators in employment, production and gross value added: i) C7-food and beverage manufacturing; and, ii) C8-automotive.

Table 3. Economic statistics of the final 24 cluster configuration C according to the 2019 economic census**

No.	Name of cluster	No. Industries	Percentage of all Industries in the Economy: Employment	Establishment	Production	Gross Value Added	Gross Value-Added Per Worker (Mexican pesos)
C1	oil and gas extraction	6	0.36	0.01	4.34	8.11	8,367.31
C2	metal mining	6	0.70	0.07	2.62	2.43	1,283.19
C3	footwear manufacturing	2	0.56	0.21	0.32	0.26	169.28
C4	sawmills and wood preservation	2	0.06	0.03	0.05	0.03	170.23
C5	medical equipment and supplies manufacturing	5	1.44	0.06	1.33	1.14	291.14
C6	semiconductors and other electronic components manufacturing	4	0.96	0.02	0.65	0.72	275.11
C7	food and beverage manufacturing	10	4.71	4.79	8.75	6.33	494.69
C8	automotive	7	4.86	0.18	14.86	10.02	758.56
C9	petroleum and coal products manufacturing	4	0.32	0.02	5.84	1.17	1,355.15
C10	apparel manufacturing	7	2.01	2.04	1.07	0.87	158.82
C11	tourism and hospitality services	13	2.48	1.09	1.46	1.47	217.69
C12	office administrative services	14	3.51	0.92	1.66	2.30	241.19
C13	metal products manufacturing	11	2.06	0.38	2.73	2.10	375.47
C14	steel products manufacturing	14	1.86	0.55	4.14	3.07	608.04
C15	financial services and head offices	17	3.78	0.60	10.56	14.02	1,365.85
C16	electric power generation, transmission, and distribution and infrastructure construction	10	2.00	0.57	4.29	3.39	621.91
C17	passenger transportation and communications	13	1.12	0.11	1.47	1.33	434.27
C18	retail and eating services	6	16.62	34.86	4.33	5.42	119.97
C19	employment services	1	5.07	0.09	1.33	2.28	165.24
C20	plastic products manufacturing	21	6.91	8.34	7.51	5.51	293.29
C21	freight transportation services and residential and nonresidential construction	15	4.82	1.10	3.49	3.61	275.22
C22	business support services	9	2.76	3.87	1.24	1.36	181.10
C23	education and health services	19	7.32	11.00	1.87	2.47	124.12
C24	pharmaceutical and medical manufacturing and services	14	1.67	0.91	1.55	1.40	309.62
Industries excluded from analysis		49	22.06	28.19	12.57	19.23	320.74

Source: Authors' own calculations based on information of the 2019 Economic Census by INEGI.

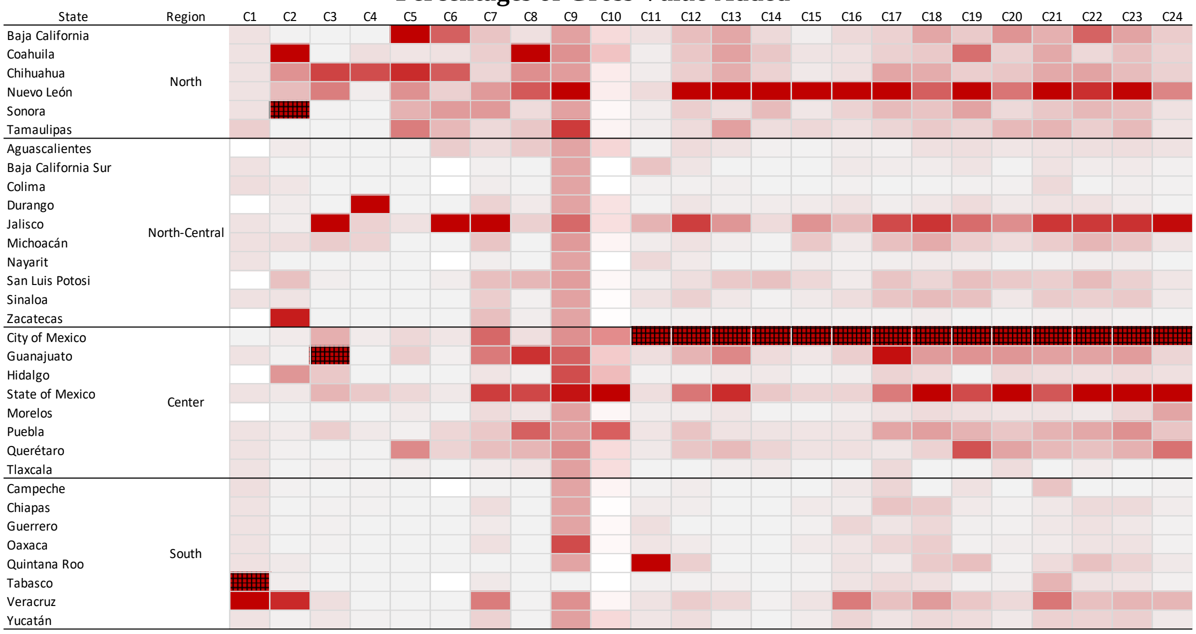
4.2 The geographical location of the 24 clusters of the Mexican economy

In this section, we show the geographical distribution of clusters based on gross value added and employment. Table 4 presents heat maps that describe the clusters' presence across states and regions. These maps can be interpreted by clusters (columns) as the share of employment and gross value added of the agglomeration at the state level, more intense color represent higher participation.

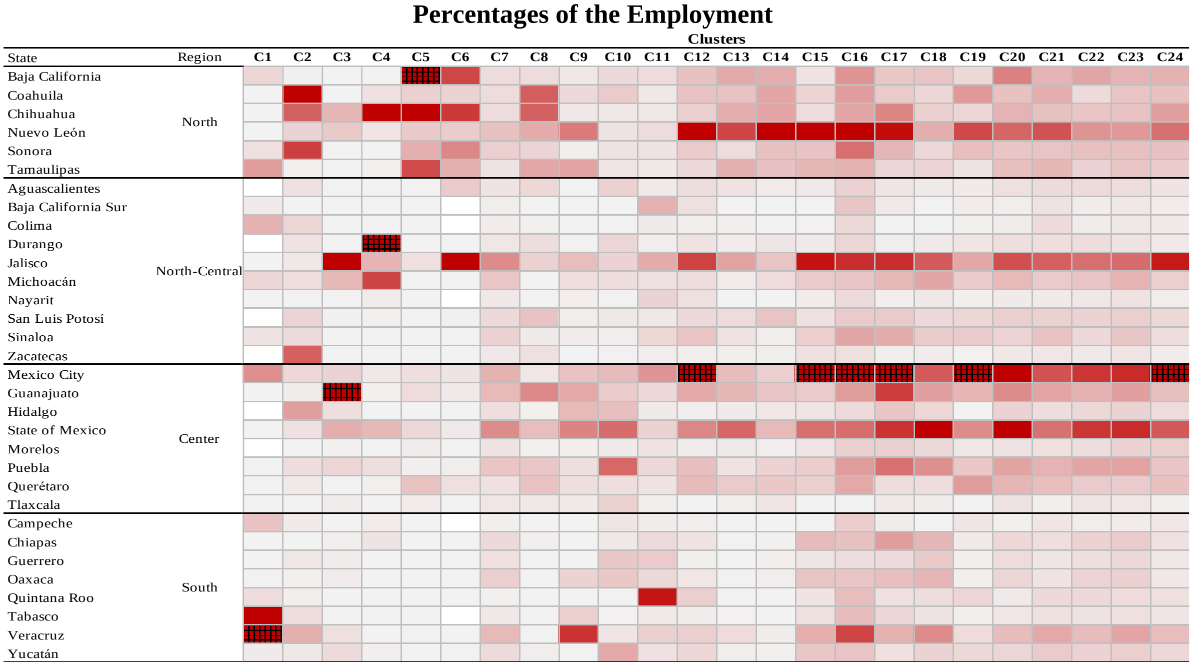
Table 4. Heat map of employment geographical distribution of the 24 clusters of related industries by region and state

Lower **Higher**

Percentages of Gross Value Added



Percentages of the Employment



Source: Authors' own calculation based on information of the 2019 Economic Census by INEGI.

Note: C1-oil and gas extraction; C2-metal mining; C3-footwear manufacturing; C4-sawmills and wood preservation; C5-medical equipment and supplies manuf.; C6-semiconductors and other electronic components; C7-food and beverage manuf.; C8-automotive; C9-petroleum and coal products manuf.; C10-apparel manufacturing; C11-tourism and hospitality serv.; C12-office administrative serv.; C13-metal products manuf.; C14-steel products manuf.; C15-financial serv. and head offices; C16-electric power generation, transmission, and distribution, and infrastructure construction; C17-passenger transportation and communications; C18-retail and eating services; C19-employment services; C20-plastic products manuf.; C21-freight transportation services and residential and nonresidential construction; C22-business support serv.; C23-education and health serv.; and C24-pharmaceutical and medical manuf. and serv. For improve visual presentation of the results, the color intensity of the cells was adjusted to a maximum value of 20%, and outliers were highlighted with a grided pattern in cells. The outlier values for employment shares are: C1(23.5), C3 (71.9), C4 (24.8), C5 (23.2), C12 (36.2), C15 (74.2), C16 (39.1), C17 (47.7), C19 (37.9), and C24 (22.8). The outlier values for gross value-added shares are: C1(49.1), C2(35.1), C3(44.1), C11 (30.3), C12(41.6), C13(20.0), C14(35.8), C15(55.1), C16(48.1), C17(62.6), C18(20.3), C19(30.2), C20(20.4), C21(21.3), C22(30.6), C23(20.8) Y C24(31.3).

Since both maps show a similar distribution, and for this reason we focus our description of the results on the employment map. In the Northern region, with its strong economic linkages with the U.S. economy, some of the most advanced technological clusters are remarkably present, especially in the border entities Baja California, Chihuahua, Sonora, Tamaulipas, and Coahuila, where the following clusters stand out: C6-semiconductors and other electronic components; C13-metal products manufacturing; and C14-steel products manufacturing. On the other hand, Nuevo León is the state with the highest level of economic development and the most populated state of the Northern region with remarkable presence of at least 8 clusters specialized in both manufacturing and services, stand out C12-office administrative services; C13-metal products manufacturing; C14-steel products manufacturing; C15-financial services and head offices; and C19-employment services.

In the North-Central region, agglomerations of mainly manufacturing: C3-footwear manufacturing; C6-semiconductors and other electronic components; C7-food and beverage manufacturing; and C15-financial services and head offices; C21-freight transportation services and residential and nonresidential construction; and, C24-pharmaceutical and medical manufacturing and services have a notable presence in Jalisco. While traditional clusters like C4-sawmills and wood preservation have presence in Durango and Michoacán and C7-food and beverage manufacturing in Michoacán, Sinaloa and San Luis Potosí.

In the Central region, both Mexico City and Estado de Mexico that jointly assemble the biggest metropolitan area in the country have a strong presence in clusters related to services e.g.: C19-employment services; C15-financial services and head offices; C12-office administrative services; C16-electric power generation, transmission, and distribution, and infrastructure construction; C17-passenger transportation and communications; and, C24-pharmaceutical and media manufacturing. Though some traditional clusters such as C7-food and beverage manufacturing; C10-apparel manufacturing; and, C13-metal products manufacturing show a remarkable presence in Estado de Mexico. In contrast, Guanajuato, Querétaro and Puebla show the presence of groups related to manufacturing, e.g. C8-automotive; C20-plastic products manufacturing; and C7-food and beverage manufacturing.

Regarding clusters related to services, C15-financial services and head offices; and C23-education and health services are present in the states of Puebla, Hidalgo, and Guanajuato.

In the Southern region, agglomeration C11-tourism and hospitality services exhibit a notorious presence in Quintana Roo, while C10-apparel manufacturing in Yucatán, Oaxaca, and Guerrero, which are states characterized by lower levels of economic growth and low specialization in secondary and tertiary activities, in comparison with the rest of the regions of the country. Likewise, cluster C1-oil and gas extraction has a notable presence in Veracruz, Campeche, and Tabasco, mostly explained by their ample natural resources.

5 The evolution of employment concentration in the clusters of the Mexican economy

In this section, using the 24-cluster configuration framework, we study the evolution of employment concentration within clusters at the national-level economy across four years: 2004, 2009, 2014, and 2019, using employment data at the municipality level from the Economic Censuses published by INEGI for these years. For this analysis, we apply the Location Quotient as a measure of relative concentration because, in this context, it is composed of ratios that compare the employment concentration of a cluster within a specific region (municipality) to the concentration of that cluster nation-wide. Following Ellison and Glaeser (1997), we define the expression of the location quotient as follows.

$$\boldsymbol{LQ_{cr}^{t}} = \frac{\boldsymbol{E_{cr}^{t}/\sum_{c=1}^{24} E_{cr}^{t}}}{\boldsymbol{\sum_{1=1}^{n} E_{cr}^{t}/\sum_{i=1}^{n}\sum_{c=1}^{24} E_{cr}^{t}}} \tag{1}$$

where E_{cr}^{t} stands for the share of employment in region r and cluster c in period t. Thus, LQ_{cr}^{t} denotes a measure of relative concentration that allows comparing the employment share in each region (municipality) of the cluster in question to the national average.[11] The closer the LQ_{cr}^{t} to 1, the closer cluster c gets the same share of its region r employment as it does of the nation. A LQ_{cr}^{t} greater than 1 indicates a cluster with a greater share of employment in the region than is the case nationwide. The change in LQ_{cr}^{t} between two years

[11] Mexico comprises 2,456 municipalities; however, economic activities are concentrated in about 400 municipalities according to the 2019 Economic Census.

may indicate the geographical expansion or contraction of employment in the clusters based on the comparison of those two years. For instance, for cluster C8-automotive, the employment concentration in 2014 in the municipality of Silao, Guanajuato, was 5.3 times than the concentration of employment for the same cluster in the whole of Mexico. This agglomeration makes up a much higher share of Silao municipality employment total than it does for the nation as a whole. For the year 2019, cluster C8-automotive in Silao, Guanajuato, recorded a LQ_{cr}^{t} of 7.6. The 2.3 increase in the LQ_{cr}^{t} between 2014 and 2019 in Silao municipality would suggest a higher concentration of employment of the cluster C8-automotive. These changes may be partially explained by the increased scale of production of important automakers plants in the central region.

Next, based on the LQ_{cr}^{t} estimations, we analyze the recent evolution of employment concentration of the clusters for the years 2014 and 2019 at the municipality level. Annex B.3 presents the spatial distribution of the clusters at the municipality level for these two points in time. At a first glance, the distribution of clusters across the country suggests highly heterogeneous employment distribution patterns. Agglomerations, e.g.: i) C7-food and beverage manufacturing; ii) C20-plastic products manufacturing; iii) C23-education and health services; iv) C11-tourism and hospitality services; v) C18-retail and eating services; and, vi) C22-business support services, show a wide-ranging geographical distribution over a high number of municipalities across the country. Another set of agglomerations, e.g.: i) C15-financial services and head offices; ii) C19-employment services; and iii) C17-passenger transportation and communications, show a high level of employment concentration in highly populated municipalities mainly located in the biggest cities of the country such as Mexico City in the Center, and Jalisco in the North center, and Nuevo León in the North. In contrast, clusters C1-oil and gas extraction and C9-petroleum and coal manufacturing have a poor presence, and are mainly located in municipalities with coastal access, showing their dependence on natural resources.

Comparing the evolution of the employment concentration of the clusters regarding their spatial distribution for municipalities between the years 2014 and 2019 in Annex B.3, we can identify agglomerations that increased their employment concentration in many municipalities across the country by raising their LQ: i) C24-pharmaceutical and medical manufacturing and services; ii) C8-automotive; iii) C18-retail and eating services; iv) C22-business support services; v) C2-metal mining; vi) C19-employment services; vii) C5-medical equipment and supplies manufacturing; and viii) C5-medical equipment and supplies manufacturing.

Another set of agglomerations, characterized by high technological development and mainly located in the Northern border municipalities and at the Center of the country, enlarged their LQ indicating an increase in employment concentration: i) C5-medical equipment and supplies manufacturing; ii) C8-automotive; iii) C6-semiconductors and other electronic components; and iv) C13-metal products manufacturing. It is important to notice that an increase in the relative measure LQ reveals an employment concentration, and not necessarily an increase of employment in absolute terms. Thus, we can confirm that employment concentration is usually experienced by high-tech clusters, whose industries show significant economies of scale and strong complementarities with external markets (see Ezcurra et al., 2006). Some of the previously mentioned agglomerations, e.g. clusters C5-medical equipment and supplies manufacturing and C8-automotive, may also reflect the existence of strong complementarities and synergies of production with other related industries located in the United States (see maps in Annex B.3).

In contrast, clusters that experience a decreasing LQ in municipalities may reflect a dispersion process, usually present in low-tech industries where economies of scale are less relevant, whose activities are more locally related, and whose largest share of products or services are sold at local markets, e.g. clusters i) C7-food and beverage manufacturing; ii) C10-apparel manufacturing; iii) C16-electric power generation, transmission, and distribution and infrastructure construction; iv) C11-tourism and hospitality services; and v) C17-passenger transportation and communications.

In the second stage of this exercise, likewise based on the LQ estimations and following the analysis of Ezcurra et al. (2006)[12], we use a nonparametric approach involving

[12] Ezcurra et al. (2006) analyze the spatial distribution of manufacturing clusters in the regions of the European Union over the period 1977-1999 by combining nonparametric techniques and spatial econometrics. They find

an adaptive kernel with flexible bandwidths to smooth the data and estimate the density probability functions of the LQ_{cr}^{t} for every cluster across the four periods of analysis (2004, 2009, 2014 and 2019).[13] We analyze the evolution of the distribution functions of employment concentration to identify, in general terms, which clusters as a whole have shown more dynamic performance in the former two decades. Figure 1 presents the distribution function for the 24 clusters of the LQ_{cr}^{t} calculated for those municipalities where the cluster is present. The horizontal axes in the graphs present the value of the location quotient LQ_{cr}^{t} for municipality *r* of cluster *c* at period *t*, and the vertical axes show their associated probability distribution. Higher LQ magnitudes indicate a stronger presence of the employment share in cluster *c* at a municipality *r* in comparison to the national level with a right-skewed or a longer right-tail distribution. Therefore, whenever probability mass associated to these curves is positively skewed and higher than *1*, it suggests a higher employment concentration conducted by a stronger presence of the cluster in the analyzed municipalities in comparison to the national level.

In chronological order, the greatest increases in municipality employment concentration during the period 2004 to 2009, reflected by a right displacement of the curve, were in clusters: i) C2-metal mining; and, ii) C9-petroleum and coal products manufacturing, for the period 2009 to 2014: i) C18-retail and eating services; ii) C15-financial services and head offices; and at a lesser magnitude: iii) C20-plastic products manufacturing; iv) C23-education and health services, and finally, for the period 2014 to 2019: i) C23-education and health services and ii) C8-automotive; and, at a lesser magnitude: iii) C6-semiconductors and other electronic components; iv) C16-electric power generation, transmission, and distribution, and infrastructure construction; and, v) C17-passenger transportation and communications. A remarkable case was the cluster C23-education and health services that notoriously showed a right displacement of the probability mass, suggesting an increasing concentration of employment across most populated cities of the country. This expansion

an increase in geographical concentration in most industrial manufacturing activities corresponding to the economic integration process underway in that period in the European Union.

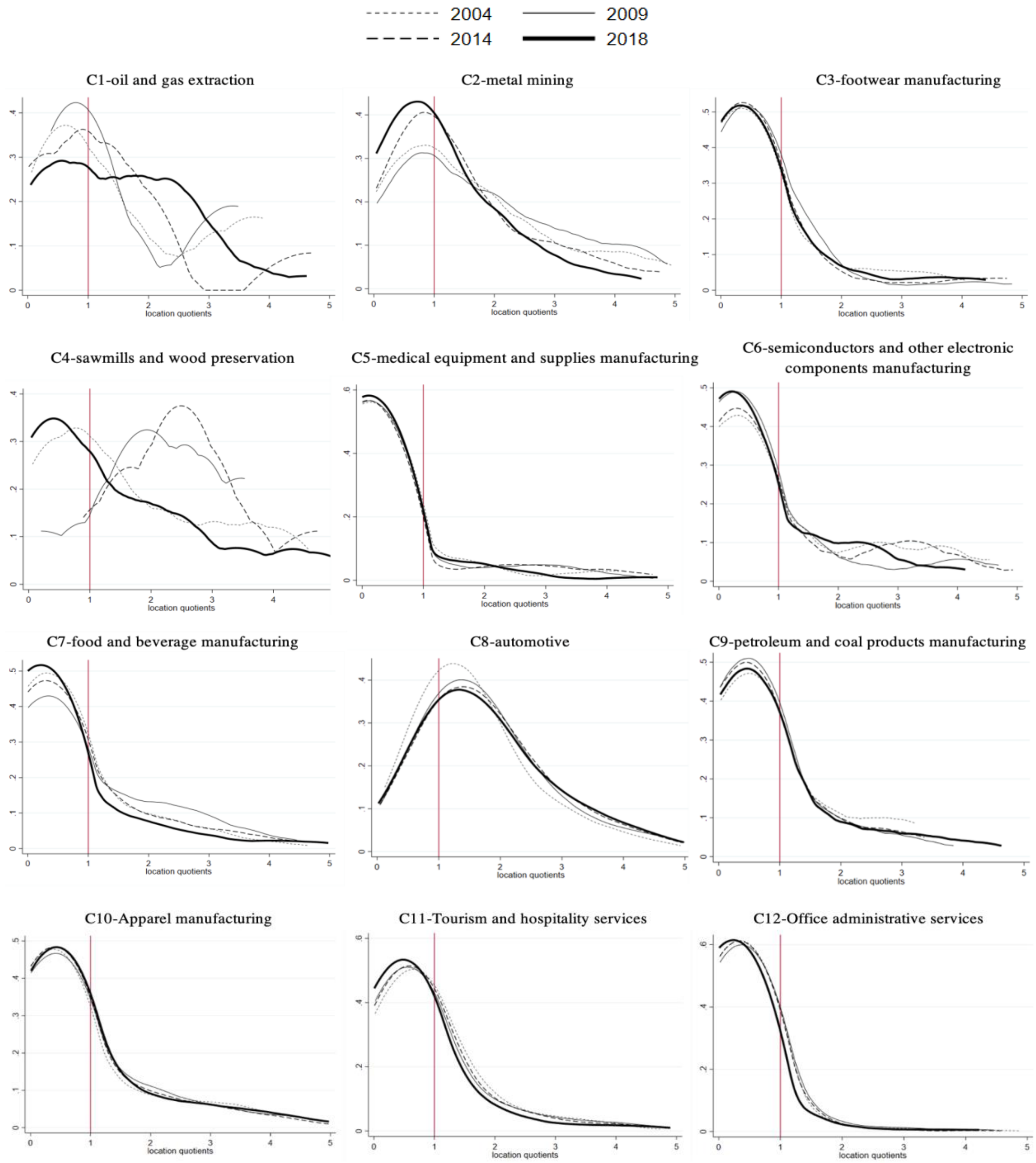
[13] We use the following kernel specification $f(LQ_{cr}^{t}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda}K\left(\frac{LQ_{cr}^{t}-LQ_{r}^{t}}{h\lambda_i}\right)$, where K is the kernel function and h is the bandwidth (Pagan and Ullah, 1999).

was mainly driven by the Mexican private health services sector that started since early 2010's with the main health care private groups in the central region. For instance, since late 2010's the first group started the construction of three new hospitals in Mexico City and updated and expand services in 8 units more, totaling 32 hospitals with more than 20,000 health specialists across Mexico.

Another relevant case is the cluster C8-automotive that shows a right displacement between 2014 and 2019, indicating an employment concentration process. The higher mass density concentrated toward right implies a recent increase of the number of municipalities with a *LQ* around 3 might signify a higher employment concentration for the most specialized municipalities. This result can be explained by the bulk arrival of new automotive manufacturing plants in many municipalities, mainly in the Northern, Central and North-Central regions of the country since the early 2010s, which has changed the national and local employment distribution in this cluster. At least a dozen of international brand automakers installed manufacturing plants, and as response, other industries in the supply chain also installed or increased scale providing inputs for these new manufacturing plants. Thus, in late 2010's automotive manufacturing plants have been continually increasing their presence in at least 12 of the 32 Mexican States.

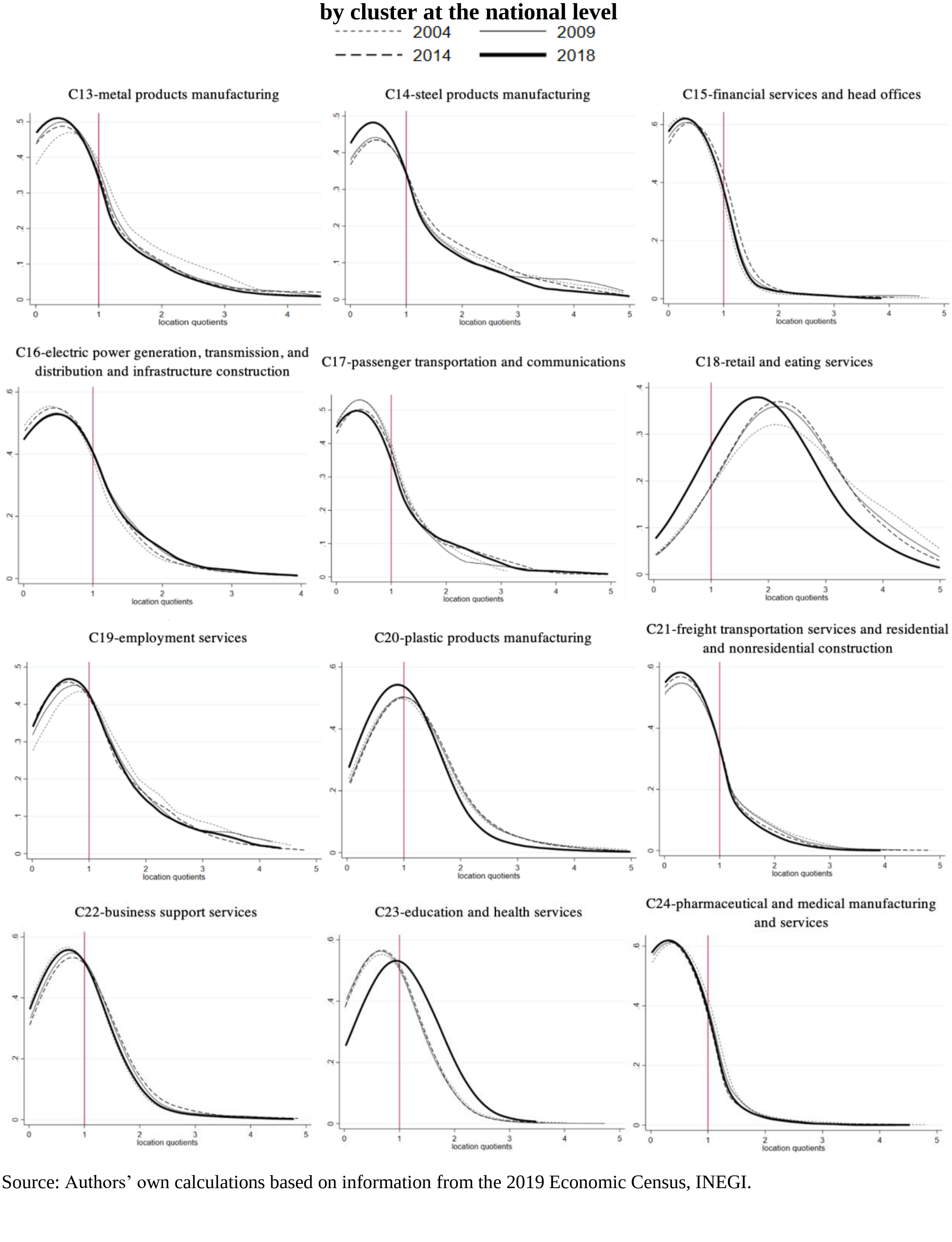
In terms of particular cases, cluster C1-oil and gas extraction stands out with a notable positively skewed distribution and an increase in the probability mass at the right part of the distribution that could be associated with industry reforms allowing higher foreign investment and partnerships, promising offshore reserves, and creating an enormous potential for shale oil and gas resources between 2009 and 2014. We also observe the contrary effect of geographical dispersion of employment in clusters: i) C2-metal mining; ii) C4 sawmill and wood preservation; iii) C7-food beverage and manufacturing; and, iii) C18-retail and eating services. On the other hand, clusters that mostly remained static at the national level in terms of employment are: i) C24-pharmaceutical and medical manufacturing and services; ii) C21-freight transportation; and, iii) C10-apparel manufacturing.

Figure 1. Distribution of the municipality employment location quotient by cluster at the national level



Source: Authors' own calculations based on information from the 2019 Economic Census, INEGI.

Figure 1 (cont.) Distributions of the municipality employment Location Quotient by cluster at the national level



Source: Authors' own calculations based on information from the 2019 Economic Census, INEGI.

6 Cluster performance and growth in the economy

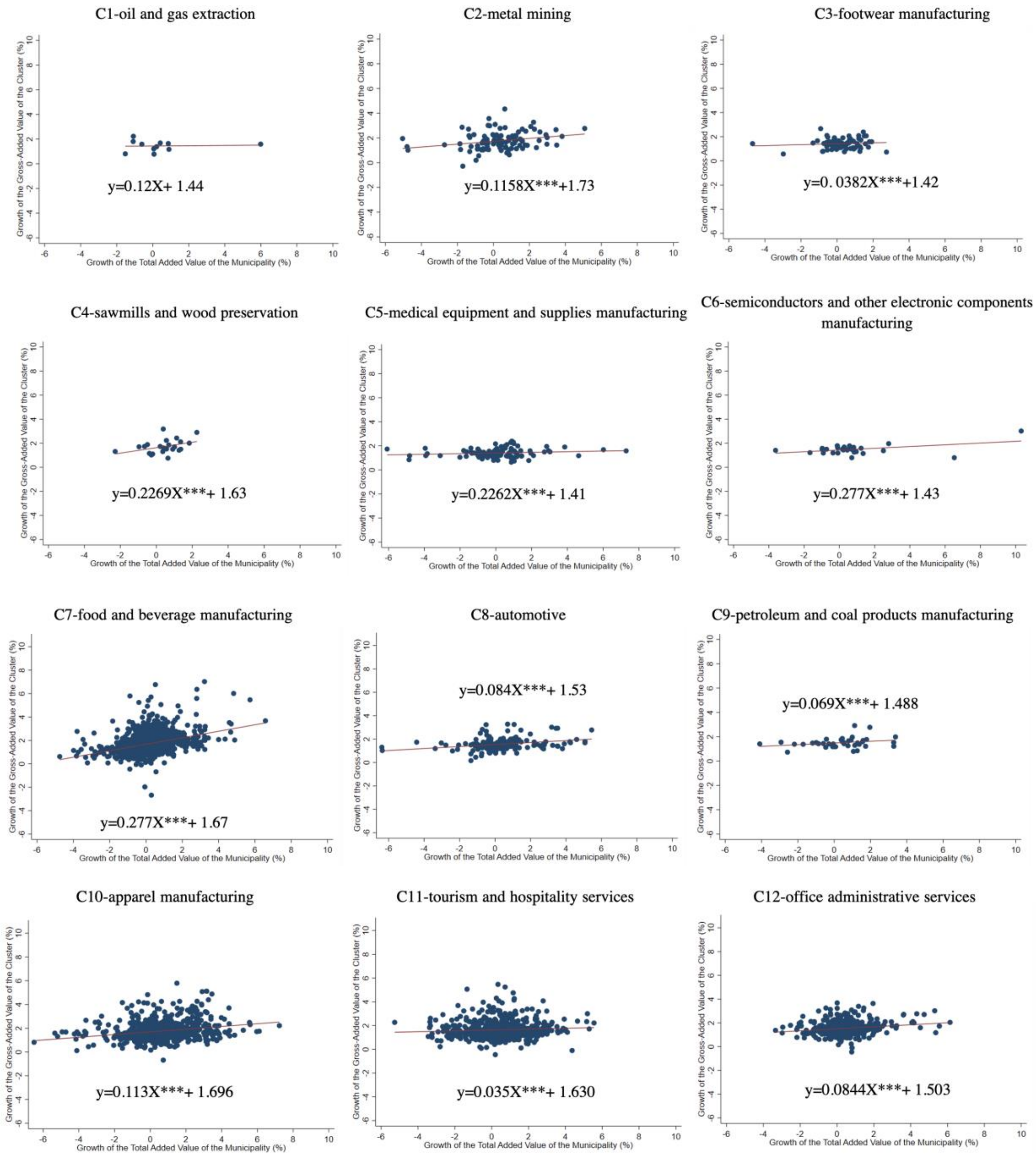
The performance of a regional economy may be influenced by the presence of certain clusters providing an extra boost to economic activities already existing in the local economy. Specific clusters can even result in an increased relatedness between different local prevailing factors like labor, expanding the possibilities of intensive knowledge sharing, and innovation spillovers. These synergies may eventually contribute to the development of more complex productive activities in these regions and induce greater growth.

To identify the kind of agglomerations that could boost synergies to foster economic performance at the municipality level, we quantify the correlation between the growth of the gross value added of the cluster present in that municipality and the growth of total added value for all economic activities in the same municipality. We used the gross value added information at the municipality level from the 2014 and 2019 Economic Censuses.

For every cluster, Figure 2 shows the correlation between regional growth and the growth exclusively from the cluster. We observe that agglomerations: i) C18-retail and eating services; ii) C7-food and beverage manufacturing; and, iii) C20-plastic products manufacturing; exhibit the higher clusters' growth and the higher positive and statistically significant relationship, suggesting that these agglomerations may be leading local economic growth in these municipalities. On the other hand, high-tech clusters, e.g. i) C5-medical equipment and supplies manufacturing; and ii) C6-semiconductors and other electronic components are present in a small number of municipalities with a relatively low growth but high correlation, suggesting a relatively more modest but important role in the growth of regional economies. In the extreme, we can identify the agglomerations displaying the lowest relationship with local growth: i) C19-employment services; ii) C11-tourism and hospitality services; iii) C16-electric power generation, transmission, and distribution, and infrastructure construction; and iv) C17-passenger transportation and communications; while C1-oil and gas extraction is non-statistically significant and most of them present in a small number of municipalities.

Figure 2. Relationship by municipality between growth of total gross value added and growth of cluster value added in the municipality

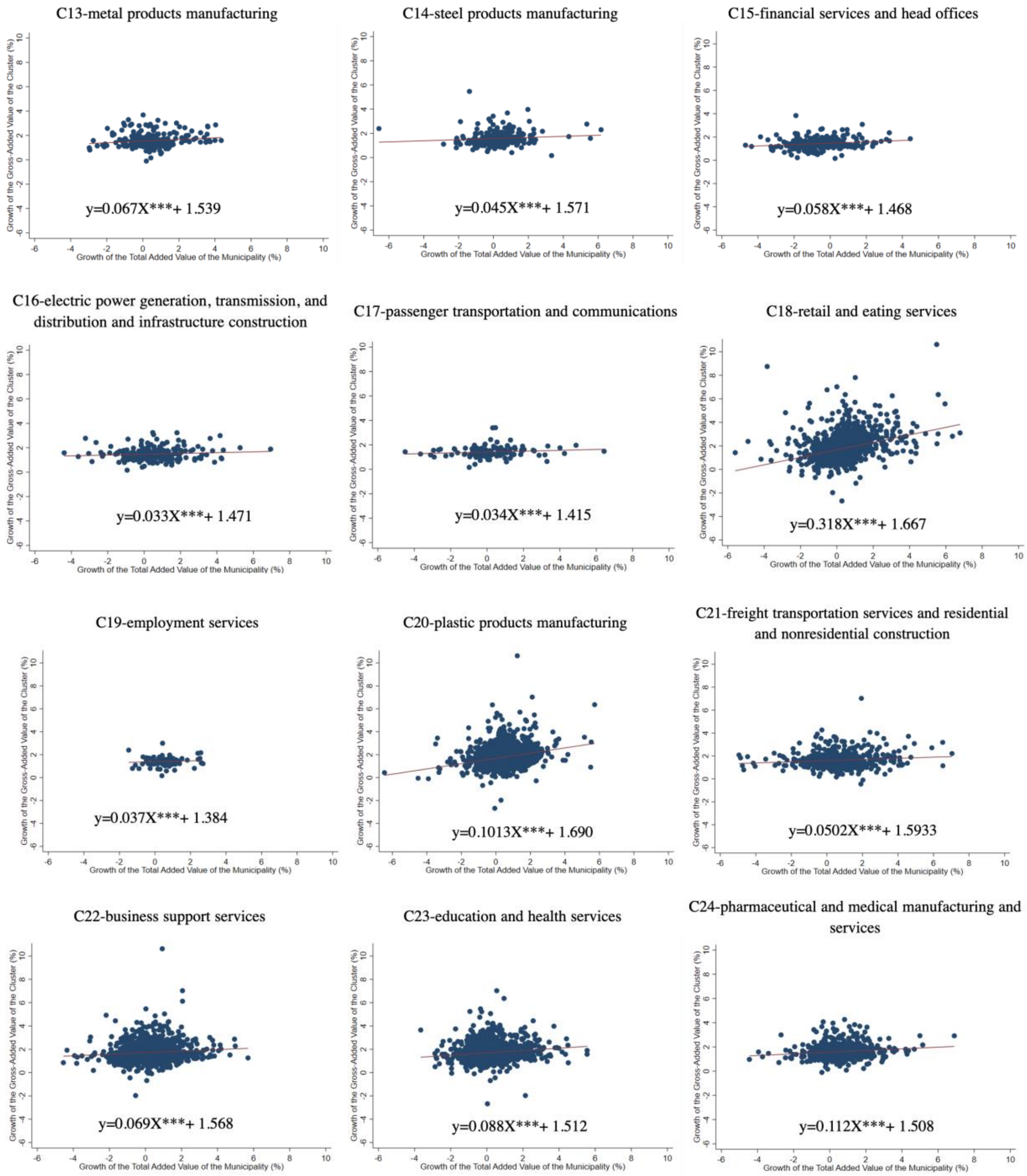
Growth of the gross value added 2014-2019, percentages



Source: Authors' own calculations based on information from INEGI.

Figure 2 (cont.) Relationship by municipality between growth of total gross value added and growth of cluster value added in the municipality

Growth of the gross value added 2014-2019, percentages



Source: Authors' own calculations based on information from INEGI.

7 Cluster spillover effects on the economy

In this section, we quantify, for every cluster, the spillover effects resulting from an exogenous demand shock in two dimensions: i) within the same cluster; and ii) toward the rest of the aglommerations in the economy. Spillover effects are estimated for the variables 'gross production', 'value added', and 'employment'. Intuitively, when a cluster experiences a positive exogenous shock, it creates a direct impact of increased demand within the cluster that receives the exogenous shock, where the magnitude of this impulse completely depends on the structure of externalities existing within the cluster. Thus, the initial demand shock triggers a virtuous circle for higher input demands and production, feeding each other after the initial shock. Also, this original shock creates an indirect effect for the rest of the agglomerations, who, in response, demand more production from the cluster where the initial shock occurred. This results in increased production in both the cluster where the initial shock occurred and in the rest of the clusters in the economy by a greater amount than the initial shock, thereby boosting production, added value and employment.

To perform this analysis, we use impact multipliers based on the input-output methodology suggested by Miller and Blair (2009), and Torre et al. (2017). According to these authors, this analysis exclusively quantifies short-run impacts because technical coefficients assume that industries within the cluster have Leontief production functions and, consequently, exhibit constant returns to scale.[14] Following this approach, for cluster i its production (x_i) equals the sum of the intermediate demand for products from other clusters j in the economy (z_{ij}) plus the demand from final consumers (f_i).

$$x_i = \sum_{j=1}^{n} z_{ij} + f_i \tag{2}$$

Thus, technical coefficients (a_{ij}) capture a fixed short-term relationship between the level of gross production of good j and the level of input of cluster i used to obtain the

[14] An important limitation of this exercise is that it assumes a perfectly elastic supply, and that the positive demand shock will therefore not generate changes in prices, which is also consistent with the short-term characteristic of the analysis.

appointed level of production, which determines the number of units that cluster i requires to produce units j demanded by cluster j:

$$a_{ij} = \frac{z_{ij}}{x_j} \tag{3}$$

For a given agglomeration i, the technical coefficient is calculated by aggregating the inputs required from all industries making up the cluster to attain the cited level of production demanded by cluster j. Considering (2) and (3), cluster production i in the Product Input model is defined as:

$$x_i = \sum_{j=1}^{n} a_{ij} x_j + f_i \tag{4}$$

For n clusters of the economy, equation (4) can be expressed in matrix notation as $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{f}$, where $\boldsymbol{I}$ stands for the identity $(n \times n)$ matrix, $\boldsymbol{A}$ is the $(n \times n)$ matrix of technical coefficients between clusters, $\boldsymbol{x}$ is the gross production vector of the economy $(n \times 1)$ and $\boldsymbol{f}$ is the vector of final demands $(n \times 1)$. In particular, the multiplier effects of an exogenous increase in the final demand of cluster j $(\Delta \boldsymbol{f}_j)$ can be computed using the following equation:

$$\Delta \boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{A})^{-1} \Delta \boldsymbol{f_j} \tag{5}$$

where the total effect on gross production relative to the increase of cluster $j's$ final demand is calculated as the sum of the array's j column elements from the matrix $\boldsymbol{L} = (\boldsymbol{I} - \boldsymbol{A})^{-1}$. This multiplier effect can be disaggregated into two effects: the direct one, which is the impact on itself (γ_{ij}), and the indirect one, obtained as the impact on the rest of the agglomerations of the economy $\sum_{i \neq j} \gamma_{ij}$.

The production increase will also impact added value and employment. To calculate these effects, we multiply equation (5) by a diagonal matrix that contains the ratio value added to production ($\boldsymbol{v_j}$) and the rate employed to production ($\boldsymbol{e_j}$), respectively, for each cluster. Thus, the effects on added value ($\boldsymbol{MVA_j}$) and on employment ($\boldsymbol{ME_j}$) are defined as:[15]

$$\boldsymbol{MVA_j} = \boldsymbol{v_j}(\boldsymbol{I} - \boldsymbol{A})^{-1}\Delta\boldsymbol{f_j} \tag{6}$$

$$\boldsymbol{ME_j} = \boldsymbol{e_j}(\boldsymbol{I} - \boldsymbol{A})^{-1}\Delta\boldsymbol{f_j} \tag{7}$$

As in gross production, the effects for value added and employment can be separated into direct and indirect ones. To estimate the spillover effects of every cluster on gross production, added value and employment, we use the 2013 National Input-Product Matrix Account published by INEGI, which captures the buying and selling relationships of inputs and products between industries. To perform the analysis, we aggregate the data at the cluster level from the original data at the four-digit NAICS industry level. Thus, for every cluster, we carry out the analysis assuming an exogenous increase shock amounting to one billion Mexican pesos (MXN) on its final demand. Subsequently, we calculate the within-cluster spillover effects (direct effect) and the effect for the rest of the clusters (indirect effect). This simulation, based on standardized shocks for all agglomerations, allows a direct comparison of the differentiated effects between clusters.

Table 5 shows the equivalence of what a $1-billion-peso shock represents in gross production, gross value added and employment for each cluster, according to data from the 2019 Economic Census. Depending on the economic importance of the cluster, this amount of money could signify a substantial share of gross production in a specific agglomeration, as in the case of cluster C4-sawmills and wood preservation or, in contrast, a negligible share as in the case of cluster C8-automotive. For instance, an exogenous demand increase of one billion pesos increases production by 0.4%, raises gross value added by 1.2%, and grows employment by 2,700 jobs in C10-apparel manufacturing. In contrast, for cluster C1-oil and

15 The rate of production to value added for the cluster j is calculated as $\frac{v_j}{x_j}$. Similarly, the rate between employment and production in the same sector is $\frac{e_j}{x_j}$. These relationships are obtained from national aggregates reported in the 2014 Economic Census published by INEGI for gross production, gross value added and total employment.

gas extraction, it represents 0.1% of production, 0.12% of gross value added, and 100 thousand jobs, respectively.

Table 5.
One- billion-pesos shock in final demand for the 24 clusters of related industries [1/]

No.	Cluster	Percentage of all Industries in: Gross Production	Percentage of all Industries in: Gross Added Value	Employments Thousand of jobs
C1	oil and gas extraction	0.10	0.12	0.1
C2	metal mining	0.17	0.41	0.3
C3	footwear manufacturing	1.42	3.86	2.4
C4	sawmills and wood preservation	9.47	37.04	1.6
C5	medical equipment and supplies manufacturing	0.34	0.88	1.3
C6	semiconductors and other electronic components manufacturing	0.70	1.39	1.8
C7	food and beverage manufacturing	0.05	0.16	0.7
C8	automotive	0.03	0.10	0.4
C9	petroleum and coal products manufacturing	0.08	0.86	0.1
C10	apparel manufacturing	0.42	1.15	2.6
C11	tourism and hospitality services	0.31	0.68	2.1
C12	office administrative services	0.27	0.44	2.7
C13	metal products manufacturing	0.16	0.48	0.9
C14	steel products manufacturing	0.11	0.33	0.6
C15	financial services and head offices	0.04	0.07	0.5
C16	electric power generation, transmission, and distribution and infrastructure construction	0.11	0.30	0.6
C17	passenger transportation and communications	0.31	0.76	1.0
C18	retail and eating services	0.10	0.18	4.8
C19	employment services	0.34	0.44	4.7
C20	plastic products manufacturing	0.06	0.18	1.3
C21	freight transportation services and residential and nonresidential construction	0.13	0.28	1.7
C22	business support services	0.36	0.74	2.8
C23	education and health services	0.24	0.41	4.9
C24	pharmaceutical and medical manufacturing and services	0.29	0.71	1.4
Industries excluded from analysis		0.04	0.05	2.3

1/ Shock value in final demand of one billion pesos as a percentage of gross production or gross census value added of the cluster, as appropriate, with data from the 2019 Economic Census, INEGI. For the impact on employment, the number of jobs this initial shock would generate is estimated.
Note: Each cluster is called after the main branch that makes it up, and the cluster is identified as the one with the largest share of gross census value added in that cluster. For details on the methodology and the main branches of activity that make up the clusters, see Annex B2.
Source: Authors' own estimation based on information from the 2019 Economic Census, INEGI.

Table 6 shows the direct and the indirect effects of the $1-billion-pesos shock on the three variables ranked according to the magnitude of the total estimated impact on production. In general, the same amount in demand shock implies differentiated effects across the 24 agglomerations considered in the analysis, which precisely reflects the structure and synergies within every agglomeration. For every cluster, the indirect effects stand out in magnitude.

In the first block of Table 6, we observe a gross production effect greater than 100%, reflecting the accumulated impact of the initial exogenous demand shock within the cluster and the demand feedback effect from the other clusters. For instance, if a $1-billion-pesos demand shock occurs in cluster C20-plastic products manufacturing, it produces a 221.5% increase as the total effect on its production value, with 111.9% corresponding to the direct increased demand effect within the cluster and 109.6% linking to indirect effects from the rest of the other agglomerations which, in response, jointly demand more production from cluster C20, where the initial shock occurred. Also, this demand shock generates an 80% total increase effect on gross value added, which is equivalent to 36.9% and 43.2% increases from direct and indirect impacts on gross value added, respectively. At the same time, the total effect on employment was 3,200 jobs, corresponding to 1,300 and 1,900 jobs from direct and indirect effects, respectively. In general, the higher the indirect effect is, the higher the spillover it implies for the rest of the clusters in the economy.

In the first block of Table 6, we also observe the five agglomerations with the greatest direct effects on gross production: i) C10-apparel manufacturing; ii) C3-footwear manufacturing; iii) C20-plastic products manufacturing; iv) C7-food and beverage manufacturing; and v) C8-automotive. With respect to gross value added, two clusters stand out: i) C19-employment services; and ii) C15-financial services and head offices. In employment, the largest number of jobs resulting as a direct effect from the demand shock are in clusters: i) C23-education and health services; ii) C18-retail and eating services; and iii) C19-employment services. Regarding indirect effects, the second block of Table 6 presents gross production, value added and employment for those clusters notable for their magnitude: i) C20-plastic products manufacturing; ii) C18-retail and eating services; and, to a lower extent: iii) C7-food and beverage manufacturing; and iv) C8-automotive.

In the third block of Table 6, we can observe the total effects (direct and indirect) per cluster in terms of gross production and value added, where the clusters with the highest effects are: i) C20-plastic products manufacturing; ii) C18-retail and eating services; iii) C7-food and beverage manufacturing; and iv) C8-automotive. It is important to mention that the four clusters with the greatest indirect effects, which may exceed the direct effects in some variables, are the same as those that generate the greatest total spillover effect: i) C20-plastic products manufacturing; ii) C18-retail and eating services; iii) C7-food and beverage manufacturing; and iv) C8-automotive.

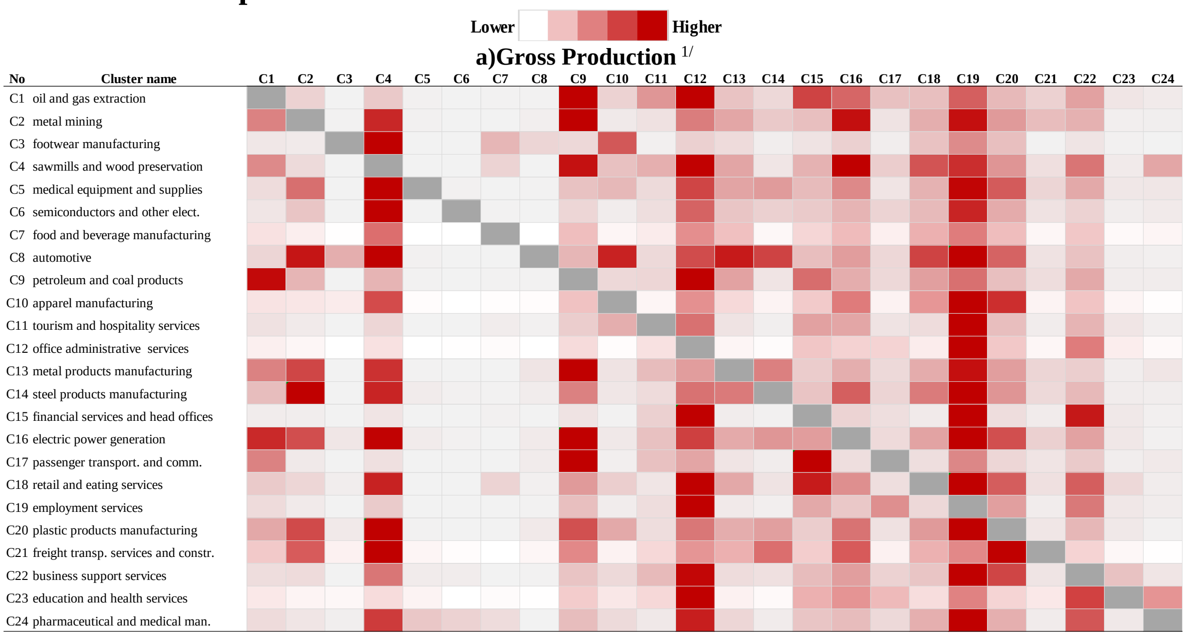
Table 6
Direct and indirect multipliers for the 24 clusters of related industries
Percentage of one billion pesos in each variable

Ranking Total Effect	Cluster	Direct Effect			Indirect Effect			Total Effect		
		Production 1/	G. Value Added 1/	Employm. 2/	Production 1/	G. Value Added 1/	Employm. 2/	Production 1/	G. Value Added 1/	Employm. 2/
1	C20-plastic products manufacturing	111.9	36.9	1.3	109.6	43.2	1.9	221.5	80.1	3.2
2	C18-retail and eating services	102.5	57.6	4.8	108.4	56.2	2.4	210.9	113.8	7.2
3	C7-food and beverage manufacturing	111.0	36.1	0.7	98.3	50.2	2.0	209.3	86.3	2.8
4	C8-automotive	109.9	33.3	0.4	93.3	42.0	1.8	203.1	75.3	2.3
5	C21-freight transportation services and residential and nonresidential construction	102.7	47.7	1.7	94.9	38.2	1.4	197.6	85.9	3.1
6	C9-petroleum and coal products manufacturing	110.1	9.9	0.1	81.0	58.5	0.7	191.1	68.4	0.8
7	C15-financial services and head offices	109.0	65.0	0.5	64.7	37.5	1.9	173.7	102.6	2.4
8	C14-steel products manufacturing	109.1	36.3	0.6	61.4	27.3	0.8	170.4	63.6	1.4
9	C16-electric power generation, transmission, and distribution and infrastructure construction	105.5	37.5	0.6	56.4	24.6	0.9	162.0	62.1	1.5
10	C13-metal products manufacturing	103.3	35.6	0.9	44.5	19.7	0.7	147.8	55.3	1.7
11	C17-passenger transportation and communications	103.4	42.0	1.0	43.1	19.5	0.6	146.6	61.5	1.6
12	C23-education and health services	101.6	60.3	4.9	40.3	20.2	0.8	141.9	80.5	5.7
13	C24-pharmaceutical and medical manufacturing and services	107.2	43.6	1.4	29.5	14.7	0.7	136.7	58.3	2.1
14	C2-metal mining	106.1	44.2	0.3	30.0	13.3	0.5	136.1	57.4	0.8
15	C10-apparel manufacturing	115.7	42.1	2.6	11.2	5.4	0.2	126.9	47.4	2.9
16	C12-office administrative services	103.7	64.5	2.7	15.8	8.6	0.4	119.4	73.1	3.1
17	C1-oil and gas extraction	101.4	85.3	0.1	17.9	8.2	0.3	119.3	93.5	0.4
18	C22-business support services	103.4	50.9	2.8	14.6	7.4	0.3	118.1	58.3	3.1
19	C5-medical equipment and supplies manufacturing	100.9	38.8	1.3	16.0	7.4	0.3	117.0	46.2	1.6
20	C3-footwear manufacturing	112.0	41.1	2.4	4.6	1.9	0.1	116.6	43.0	2.5
21	C11-tourism and hospitality services	100.2	45.3	2.1	12.2	6.6	0.3	112.4	51.9	2.4
22	C6-semiconductors and other electronic components manufacturing	100.0	50.2	1.8	11.2	5.3	0.2	111.2	55.6	2.1
23	C4-sawmills and wood preservation	104.6	26.7	1.6	2.2	1.2	0.0	106.8	27.9	1.6
24	C19-employment services	100.4	77.5	4.7	2.1	1.0	0.0	102.5	78.5	4.7
Industries excluded from analysis		105.8	72.8	2.3	86.6	41.2	1.6	192.4	113.9	3.9

1/ Percentage in the reference variable with respect to the value of the shock in final demand amounting to $1 billion pesos and thousands of jobs generated.
2/ Thousands of jobs generated from the clash in final demand amounting to $1 billion pesos.
Each cluster is called after the main branch that makes it up, and the cluster is identified as the one with the largest share of gross census value added in that cluster.
Source: Produced based on information from the Product 2013 Input Matrix and the 2019 Economic Census, INEGI.

Table 7 presents a heat map exhibiting the intensity of indirect effects for every cluster to illustrate their distribution across the rest of the agglomerations in terms of gross production, value added and employment. For each cluster (line), the intensity of the red color in cells represents the magnitude of the indirect effects. In general, we can observe that a $1-billion-pesos exogenous demand increase generates differentiated indirect effects in magnitude over the three variables. Thus, for instance, the main indirect effects of cluster C20-plastic product manufacturing on production are for agglomerations C4-sawmills and wood preservation; C9-petroleum and coal products manufacturing; and, C19-employment services. For C18-retail and eating services the highest indirect effects focus on clusters C4-sawmills and wood preservation; C19-employment services; C12-office administrative services; and, C16-electric power generation. For C8-automotive the main indirect effects are concentrated on C4-sawmills and wood preservation; C19-employment services; C2-metal mining; C13-metal products manufacturing; and, C20-plastic products manufacturing.

Table 7
Heat Map of Indirect Effects for the 24 Clusters of Related Industries

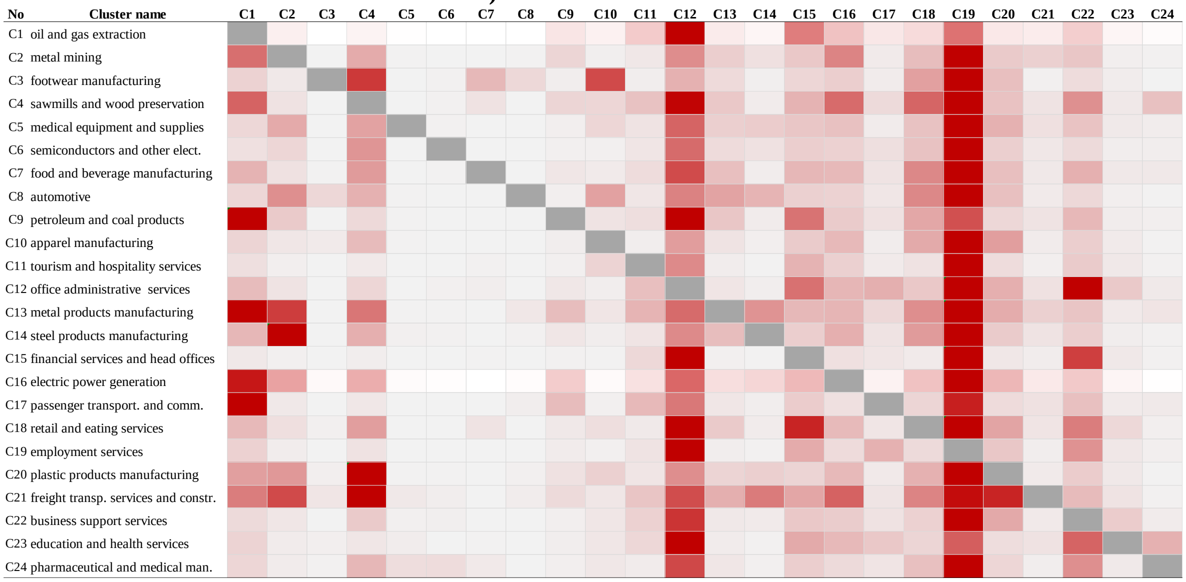


1/ Percentage relative to the value of a $1-billion-peso shock in final demand for each cluster.
Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Censuses, INEGI.

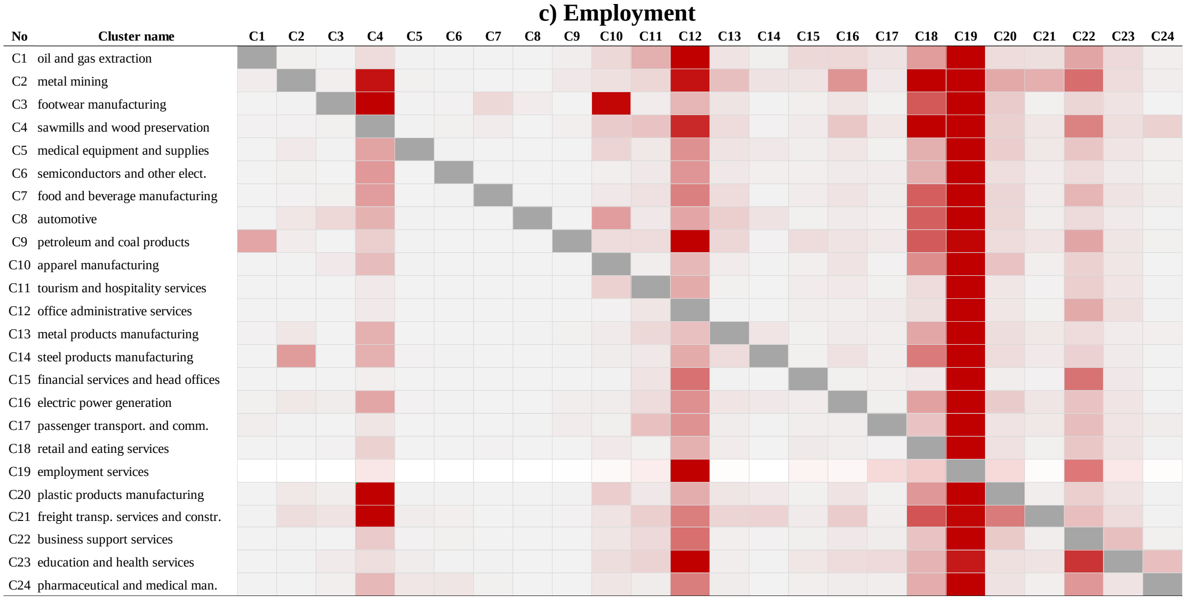
Table 7 (cont.)
Heat Map of Indirect Effects for the 24 Clusters of Related Industries

Lower Higher

b) Gross Value Added



c) Employment



1/ Percentage relative to the value of a $1-billion-peso shock in final demand for each cluster.
Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Censuses, INEGI.

Most clusters show relatively greater indirect effects on agglomerations i) C19-employment services; C4-sawmill and wood preservation; and, ii) C12-office administrative services. The common indirect effect on cluster C19-employment services could be associated with the growing use of subcontracting schemes for hiring workers. Thus, many companies across all sectors likely used more frequently these schemes. In 2017, one single company in this cluster, was the fourth largest employer in the country with more than 100,000 workers hired by outsourcing companies in almost all industries in Mexico. On the other hand, the common indirect effect of some agglomerations on cluster C4-sawmill and wood preservation is related with the manufacturing of wood containers and wood packaging materials widely used for transportation from food and perishable products to high-tech manufacturing products such as electronics, motors and steel products. At the same time, the cluster C12-office administrative services effects could be related to the demand for specialized services and business support required by clusters.

For its part, the lines that correspond to clusters C9-petroleum and coal products manufacturing and C18-retail and eating services show relatively greater indirect effects in terms of gross production and value added than other clusters. In turn, cluster C24-pharmaceutical and medical manufacturing and services shows relatively higher indirect effects in terms of gross value added and production than cluster C5-medical equipment and supplies manufacturing. Finally, generalized indirect effects of clusters on C12-office administrative services can be associated to the paper work and taxes that all companies must carry out. The C20-Plastic Products Manufacturing, has important direct and indirect effects; for example, with the following agglomerations: C4-sawmills and wood preservation, C18-retail eating services, C16-electric power generation, transmission and distribution and infrastructure construction, C9-petroleum and coal products manufacturing and C14-steel products manufacturing. These results make sense because for various processes carried out in these agglomerations, plastic supplies or packaging are required. For example, plastic resins are used for the production of plywood, while plastic is a derivative of petroleum in the same way it has an impact on industries related to the production of crude oil. In other sectors, its derivatives are the main input since they are used in packaging, auto parts, electrical appliances, having impacts on other sectors of the economy.

8 Conclusions

We study the role of clusters on regional growth starting by identifying an agglomeration configuration that incorporates different dimensions of inter-industry relationships and considers a broad range of economic activities across the Mexican economy. We find 24 clusters of related industries using data from the 2019 Economic Census.

The geographical distribution of these 24 agglomerations of related industries shows a widespread presence across the country but highly heterogenous patterns. Agglomerations C7-food and beverage manufacturing, C18-retail and eating services C23-education and health services, and C20-plastic products manufacturing show a wide-ranging geographical distribution in municipalities across the country. In contrast, some clusters, e.g. C1-oil and gas extraction and C9-petroleum and coal manufacturing, have a presence only in a few coastal municipalities, displaying their dependence on natural resources.

On the other hand, manufacturing-oriented clusters have a strong presence in the Northern States of the country, while services-oriented clusters are more frequently distributed in the Central ones. High-tech agglomerations such as i) C5-medical equipment and supplies manufacturing; ii) C6-semiconductors and other electronic components; iii) C8-automotive; and iv) C14-steel products manufacturing show an ongoing employment concentration process, suggesting strong complementarities of production with other related industries from the U.S. economy.

Our analysis of employment concentration shows that agglomerations i) C7-food and beverage manufacturing; ii) C20 plastic products manufacturing; iii) C18 retail and eating services; and iv) C22-business support services show spread wide presence across the country. At the same time, others, e.g. i) C15-financial services and head offices; ii) C19-employment services; and iii) C17-passenger transportation and communications, show a high level of employment concentration in municipalities within the biggest cities of the country, such as Mexico City in the central region, Guadalajara in the north center, and Monterrey in the north.

In addition, we estimate the spillover effects of an exogenous positive shock on each of the 24 clusters of related industries measured in terms of production, value added and employment. While, in most clusters, the effects within clusters outweigh those that extend to other agglomerations, the latter are not negligible. The following clusters stand out for showing higher spillovers effects on the economy: i) C20-plastic products manufacturing; ii) C18-retail and eating services; iii) C7-food and beverage manufacturing; and iv) C8-automotive, suggesting that they influence the economic performance of other activities in the economy. For these same clusters, we confirm that cluster growth at the municipality level shows the highest correlation with the total economic growth of these municipalities.

In terms of policy, clusters can influence the economic performance of the economy overall and of the regions by creating synergies that promote short- and long-term growth through the dissemination of knowledge, innovation, and productivity growth. The findings of this paper underline the importance of the interdependencies between different industrial agglomerations. In this context, we believe analyzing how clusters support the economic recovery of the economic regions from the COVID-19 pandemic crisis would be a fascinating opportunity for future research.

References

Alcácer, J., and Zhao, M. (2016). "Zooming in: A practical manual for identifying geographic clusters", *Strategic Management Journal 37*(1), 10–21.

Argüelles, M., Benavides, C., and Fernández, I. (2014). "A new approach to the identification of regional clusters: hierarchical clustering on principal components", *Applied Economics*, *46*(21), 2511–2519.

Chávez, J. C., and García, K. (2015). *Identificación de Clusters Regionales en la Industria Manufacturera Mexicana*. Banco de México Working Papers, 2015-19, 1-52.

Dávila, A. (2005). Industrial Clusters in Mexico. In E. Giuliani, R. Rabellotti, & M. P. V. Dijk (Eds.), *Clusters Facing Competition: The Importance of External Linkages* (pp. 230–257). essay, Ashgate Pub. Group.

Delgado, M., Porter, M. E., & Stern, S. (2014). "Clusters, Convergence, and Economic Performance", *Research Policy*, *43*(10), 1785–1799.

Delgado, M., Porter, M. E., and Stern, S. (2016). "Defining Clusters of Related Industries", *Journal of Economic Geography*, *16*(1), 1–38.

Ellison, G., and 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 Axplain Agglomeration?", *American Economic Review*, 89: 311 – 316

Ellison G. Glaeser E. L. Kerr W. R. (2010). "What Causes Industry Agglomeration? Evidence from coagglomeration patterns", *American Economic Review* (100): 1195 – 1213.

Everitt, B. S., S. Landau, and M. Leese. (2011). *Cluster analysis*, 4th ed. Hodder Arnold, London, United Kingdom.

Ezcurra, R., P. Pascual, and M. Rapún (2006). "The Dynamics of Industrial Concentration in the Regions of the European Union", *Growth and Change (*37)2: pp. 200–229.

Feser, E. J., and Bergman, E. M. (2000). "National Industry Cluster Templates: A Framework for Applied Regional Cluster Analysis", *Regional Studies*, *34*(1), 1–19.

Feser, E., Sweeney, S., and Renski, H. (2005). “A Descriptive Analysis of Discrete U.S. Industrial Complexes”, *Journal of Regional Science*, *45*(2), 395–419.

Feser, E., Renski, H., and Goldstein, H. (2008). Clusters and Economic Development Outcomes. *Economic Development Quarterly*, *22*(4), 324–344.

Figueiredo, O., Guimaraes, P. and Woodward, D. (2009). “Localization Economies and Establishment Size: was Marshall Right After All?” Journal of Economy Geography (9), 853-868.

Miller, R. and Blair, P. (2009). *Input-Output Analysis. Foundations and Extensions*. Cambridge University Press. 2nd Edition. Cambridge.

Ord, J. K., and Getis, A. (1995). “Local Spatial Autocorrelation Statistics: Distributional Issues and an Application”, *Geographical Analysis*, *27*(4), 286–306.

Porter, M. E. (2003). “The Economic Performance of Regions”, *Regional Studies* (37). 549-578.

Torre, L. E., Alvarado, J. A. and Quiroga, M. (2017). "Regional Input-Product Matrices: An Application to the Automotive Sector in Mexico". Research Documents, Bank of Mexico No. 2017-12.

Villarreal, A., and M. Flores. (2015). “Identificación de Clústeres espaciales y su Especialización Económica en el Sector de Innovación.” *Región y Sociedad*, *27*(62), 117–147.

Villarreal, A., M. Flores, and E. Mackb. (2017). “Industrial Complexes in Mexico: Implications for Regional Industrial Policy Based on Related Variety and Smart Specialization” *Regional Studies*, Vol. 51(4), 537–547.

Ward, J. H. (1963). “Hierarchical Groupings to Optimize an Objective Function”,. *Journal of the American Statistical Association* (58), 236–244.

World Bank (2009). *Clusters for Competitiveness: A Practical Guide and Policy Implications for Developing Cluster Initiatives*, February, Washington, D.C.

Annex A

In Annex A.1 we describe the five similarity matrices used in this paper and, in Annex A.2, we detail the calculation of the Validation Scores, both following the methodology of Delgado et al. (2016).

A.1. Similarity matrices

We use five similarity matrices that summarize different dimensions of the relationships between industries based on three different relatedness measures:

a) The proximity in the location of the establishments calculated using the measure of Local Correlation (LC).
b) The proximity in the location of employment, calculated using the LC measure.
c) The discrepancy in the distribution of employment between two industries in comparison to a situation in which such distribution was random, calculated using the measure of the Co-Agglomeration Index.
d) The buyer-supplier linkages between a pair of industries based on their total patterns of sales and purchases across multiple industries using the Input-Output accounts.
e) The multidimensional similarity matrix calculated as the average of the four previously mentioned similarity matrices that capture the four dimensions of relatedness between industries.

A.1.1. Local correlation (LC)

We use the measure of Local Correlation (LC) to calculate the first two similarity matrices: i) the location of establishments, and ii) the location of employment. For any pair of industries, LC captures location patterns between them that reflect inter-industry linkages of various types (e.g. technology, skills, supply, or demand links (Porter, 2003; Ellison and Glaeser, 1997). The LC of establishments may be an indicator of a region's current development of synergies between a pair of industries, whereas employment levels depend primarily on prior location decisions and usually tend to indicate labor market pooling, skill complementarity, and relative labor costs.

For the purposes of this paper, we define as regions the 32 states of the country, so the LC between all different pairs of industries at the State level is defined as:

$$LC - Employment_{ij} = Correlation(empoyment_{ir}, employment_{jr}) \qquad (A.1)$$

$$LC - Establishments_{ij} = Correlation(establishments_{ir}, establishments_{jr}) \quad (A.2)$$

The LC for employment is given by equation (A.1) and is obtained from calculating the correlation coefficients between the location of employment for industry *i* in State *r* and the location of establishments for industry *j* in State *r*. Equation (A.2), analogous for the location of employment, measures LC of establishments. Both measures take values between -1 and 1; positive and large values suggest relevant economic interdependences between a pair of industries. Both indicators were calculated using the aggregated information from the 2019 Economic Census at the state level of the total number of workers per industry and the aggregate number of establishments by municipality and industry.

As Delgado et al. (2016), we include two measures of the agglomeration of establishments and employment since these variables incorporate different information according to the productive profile of industries. More establishments in a location may support the interactions between companies that create spillovers in operation; while a higher concentration of employment in a given location could facilitate labor force mobility across industries, sharing the transmission of better practices and knowledge within a cluster.

A.1.2. Co-agglomeration index (COI)

This measure was initially proposed by Ellison et al. (2010) for establishments and employment that features whether two industries are more co-located in comparison with the case of employment being randomly distributed. It measures the extent of concentration or conglomeration of two industries. In other words, this measure is related to the covariance of the employment shares of two industries in the region, normalized to ruled out possible bias from the size of the geographic area considered. It captures correlations in the relative size of two industries in a region, based on employment shares, in comparison with the relative size of all other industries in the area. The index sums up all regions considered allowing us to study factors underlying the geographic concentration of industries (see, for example, Ellison and Glaeser (1999) and Ellison et al. (2010)).

$$COI_{ij} = \frac{\sum_r (s_{ri} - x_r)(s_{rj} - x_r)}{(1 - \sum_r x_r^2)} \quad \text{(A.3)}$$

Where s_{ri} is the share of industry i's employment in State r, and x_r measures the aggregate size on State r that goes from 0 to 100. This index takes a value of zero if the distribution does not differ from a random distribution., When the COI is positive and higher, it suggests greater potential for externalities between two industries.

A.1.3. Input-output linkages

To assess the linkages between pairs of industries, we use correlations based on their patterns of sales and purchases. The inter-industry transaction similarity matrix is computed at the national level following the methodology of Feser et al. (2005), where the transactions between industries are defined as the maximum proportion of purchases or sales between two industries obtained from the national input-output accounts.[16] This measure describes inter-industry links based on input-output relationships that measure how intensive the supplier and buyer flows between industries are. The input-output (IO) link between industries *i* and *j* is given by:

$$IO_{ij} = Max\{input_{i\rightarrow j}, input_{i\leftarrow j}, output_{i\rightarrow j}, output_{i\leftarrow j}\} \quad \text{(A.4)}$$

Where $input_{i\rightarrow j}$ link is the share of industry *i*'s total value of inputs coming from industry *j* and $output_{i\rightarrow j}$ link is the share of industry *i*'s total value of outputs going to industry *j*.

A.1.4. Multidimensional similarity matrix

This matrix is defined as the combination of the four previously described similarity matrices, and is computed as their average. By construction, the multidimensional similarity matrix contains numerous measures of closeness between industries and may overcome the data limitations of the single similarity matrices since they capture more types of industrial links such as demand and supply (Delgado et al., 2016).

[16] In addition to input-output to capture the national level inter-industry linkages, Delgado et al. (2016) consider similarities in labor occupations. However, for the case of Mexico, statistics of occupational employment are unavailable.

Annex A.2. Validation scores

A.2.1. Global validation score

The best configuration $\boldsymbol{C^*}$ will be the one satisfying two conditions: individual industries showing a good fit i) within their own cluster and, ii) individual clusters very different from each other. Therefore, the proposed Validation Scores assess the degree to which individual clusters and industries in a given configuration $\boldsymbol{C}$ have a high Within-Cluster Relatedness (WCR), capturing the extent in which industries fit well within their own cluster, in comparison to Between-Cluster Relatedness (BCR), evaluating (within the configuration $\boldsymbol{C}$**)** how different the individual clusters are from each other. This purpose is attained with the Global Validation Score (Global-VS) that assesses every configuration combining eight sub-scores in two dimensions: i) at the cluster level with the Validation Score for clusters $VS - Cluster;$ and, ii) at the industry level with the Validation Score for industries $VS - Industry$.

A.2.2. Cluster level

For every configuration $\boldsymbol{C}$, the $VS - Cluster$ evaluates two dimensions: i) the percentage of industries with a smaller distance in comparison to the average of the industries that belong to the same cluster ($AvgBCR_c$); and, ii) the percentage of industries with a Ward distance from their cluster's industries lower than the Ward distance of the industry's 5th percentile from other clusters ($Pc5BCR_c$).[17]

$$VS - ClusterAvg_C^M = \left(\frac{100}{N_c}\right).\sum_c I\left[WCR_c\left(M_{ij}\right) < AvgBCR_c\left(M_{ij}\right)\right] \quad \text{(A.5)}$$

$$VS - ClusterPc5_C^M = \left(\frac{100}{N_c}\right).\sum_c I\left[WCR_c\left(M_{ij}\right) < Pc5BCR_c(M_{ij})\right] \quad \text{(A.6)}$$

where N_c is the number of clusters in configuration $\boldsymbol{C}$ and I is an indicator function equal to 1 for a given cluster c if $WCR_c < AvgBCR_c$ in equation (A.5) and, $WCR_c < Pc5BCR_c$ in equation (A.6). We calculate both scores for every similarity matrix. At the cluster level, we calculate WCR_c as the average closeness between a pair of industries within a cluster, while

[17] The expressions for equations (A.5) and (A.6) are based on Ward distance and, for this reason, the inequality sign was adjusted to indicate that a lower magnitude indicates a higher level of relatedness.

BCR_c is the average closeness between industries in cluster c and those in another cluster. Thus, $VS - Cluster$ is obtained as the average of the percentage in a cluster configuration with a WCR lower than the average WCR and lower than the fifth percentile of BCR. That is, we are calculating the percentage of clusters that show a smaller average distance to the rest of the industries inside of the cluster in comparison to the overall distance with other clusters. Since we are using Ward distance as an input for validation scores, a higher WCR_{ic} indicates lower linkage. Accordingly, we adjust equations (A.5) to (A.8) to reflect this difference with respect to the notation of the paper by Delgado et al. (2016). In consequence, a high $VS - cluster$ score suggests that individual clusters in a configuration $\boldsymbol{C}$ are meaningfully different.

A.2.3. Industry Level

For every configuration $\boldsymbol{C}$, $VS - Industry$ evaluates the fit of the individual industries within their own cluster.[18]

$$VS - IndustryAvg_C^M = \left(\frac{100}{N_i}\right).\sum_i I\left[WCR_{ic}\left(M_{ij}\right) < AvgBCR_i\left(M_{ij}\right)\right] \tag{A.7}$$

$$VS - IndustryPc5_C^M = \left(\frac{100}{N_i}\right).\sum_i I\left[WCR_{ic}\left(M_{ij}\right) < Pc5BCR_i(M_{ij})\right] \tag{A.8}$$

where N_i is the number of industries in C, and we calculate both scores for every similarity matrix. At the industry level, the WCR_{ic} is calculated as the average pairwise relatedness between the focal industry and the other industries within the cluster, and BCR_i is the average relatedness between the focal industry and the industries in different clusters. In this score, we measure the percentage of industries with a WCR_{ic} lower than their average BCR_i and lower than their fifthpercentile of BCR_i. A score of 100 suggests that all individual clusters in C contain industries that show a high level of proximity based on multiple linkages. Finally, we obtain $VS - Industry$ by averaging $VS - IndustryAvg_C^M$ and $VS - IndustryPc5_C^M$.

[18] As above, the expressions for equations (A.7) and (A.8) are based on Ward distance and, for this reason, the inequality sign was adjusted to indicate that a lower magnitude indicates a higher level of relatedness.

As a last step, to obtain the final validation score, we compute $Global - \boldsymbol{VS}$ by averaging $VS - Industry$ and $VS - Cluster$. Once we have computed $Global - \boldsymbol{VS}$, we can proceed to rank cluster configurations according to $Global - \boldsymbol{VS}$ to define a set of candidates for $\boldsymbol{C}^{*}$. This set of possible candidates is used as a framework of reference to estimate an additional robustness score to corroborate the best configuration $\boldsymbol{C}^{*}$. Intuitively, the $Overlap\,Score$ (OS) over a given set of the candidates of $\boldsymbol{C}^{*}$ captures the overlap between a pair of clusters across the set of candidates of configurations to ensure the robustness of the results.

$$Overlap_{c,b} = 100.\left[\frac{Share\ Industries_{c,b}}{\sqrt{Industries_{c} * Industries_{b}}}\right] \tag{A.9}$$

$$Overlap\ Score_{C_1 - C_2} = \frac{1}{N} \sum_{c \epsilon C_1} Overlap_{c,b} \tag{A.10}$$

where equations (A.9) and (A.10) compare two given cluster configurations $C_1 - C_2$, and we calculate the overlap between a pair of clusters $\boldsymbol{c}$ and $\boldsymbol{b}$. For every individual cluster $\boldsymbol{c}$ in C_1 we find a matching cluster $\boldsymbol{b}$ in C_2. Finally, we select the configuration $\boldsymbol{C}^{*}$ with the best performance in $Global\ \ \boldsymbol{VS}$ and the $Overlap\ Score$.

Annex B.1.

Table B.1. Cluster configurations and valuation scores

Configuration	Similarity Matrix	Clustering Function	Number of Clusters	Global VS	Ranking Global VS	VS-cluster	Ranking VS-Industry	Overlap Score	Ranking Overlap
C_{30}	LC-Employment$_{ij}$	Herarchichal Ward	15	61.19	30	61.33	27	61	73
C_{31}	LC-Employment$_{ij}$	Herarchichal Ward	14	61.08	31	61.43	23	61	80
C_{32}	LC-Establishments$_{ij}$	Herarchichal Ward	11	60.57	32	50	49	62	64
C_{33}	LC-Employment$_{ij}$	Herarchichal Ward	36	60.83	33	59.44	67	62	58
C_{34}	LC-Employment$_{ij}$	Herarchichal Ward	12	60.76	34	60	50	62	66
C_{35}	LC-Employment$_{ij}$	Herarchichal Ward	20	60.75	35	61.5	19	60	85
C_{36}	LC-Employment$_{ij}$	Herarchichal Ward	35	60.73	36	59.43	68	62	60
C_{37}	LC-Employment$_{ij}$	Herarchichal Ward	34	60.7	37	59.41	69	62	61
C_{38}	LC-Establishments$_{ij}$	Herarchichal Ward	30	60.65	38	58.67	80	63	49
C_{39}	LC-Establishments$_{ij}$	Herarchichal Ward	37	60.61	39	58.92	73	62	56
C_{40}	LC-Employment$_{ij}$	Herarchichal Ward	11	60.55	40	60	51	61	71
C_{41}	LC-Employment$_{ij}$	Herarchichal Ward	37	60.57	41	59.46	66	62	65
C_{42}	LC-Employment$_{ij}$	Herarchichal Ward	19	60.55	42	61.05	35	60	84
C_{43}	LC-Establishments$_{ij}$	Herarchichal Ward	31	60.55	43	58.71	79	62	53
C_{44}	LC-Establishments$_{ij}$	Herarchichal Ward	12	60.45	44	60	52	61	*77*
C_{45}	LC-Establishments$_{ij}$	Herarchichal Ward	38	60.43	45	58.95	72	62	62
C_{46}	LC-Employment$_{ij}$	Herarchichal Ward	13	60.35	46	60	53	61	81
C_{47}	LC-Employment$_{ij}$	Herarchichal Ward	18	60.34	47	61.11	34	59.57	87
C_{48}	LC-Employment$_{ij}$	Herarchichal Ward	38	60.32	48	59.47	65	61	*72*
C_{49}	LC-Establishments$_{ij}$	Herarchichal Ward	13	60.3	49	60	54	61	82
C_{50}	LC-Employment$_{ij}$	Herarchichal Ward	40	60.29	50	59.75	63	60.82	79
C_{51}	LC-Employment$_{ij}$	Herarchichal Ward	39	60.26	51	59.49	64	61	74
C_{52}	LC-Establishments$_{ij}$	Herarchichal Ward	16	60.22	52	60	55	60	83
C_{53}	LC-Establishments$_{ij}$	Herarchichal Ward	39	60.2	53	58.97	71	61	68
C_{54}	LC-Establishments$_{ij}$	Herarchichal Ward	34	60.17	54	58.82	76	62	67
C_{55}	LC-Establishments$_{ij}$	Herarchichal Ward	33	60.11	55	58.79	77	61	69
C_{56}	LC-Establishments$_{ij}$	Herarchichal Ward	32	60.09	56	58.75	78	61	70
C_{57}	Coaglomeration Index$_{ij}$	Herarchichal Ward	30	60.02	57	70	5	50.04	151
C_{58}	LC-Establishments$_{ij}$	Herarchichal Ward	40	60.02	58	59	70	61	75
C_{59}	LC-Establishments$_{ij}$	Herarchichal Ward	35	59.95	59	58.86	75	61	76
C_{60}	LC-Establishments$_{ij}$	Herarchichal Ward	36	59.88	60	58.89	74	58.48	78
C_{61}	LC-Employment$_{ij}$	Herarchichal Ward	17	59.83	61	61.18	33	59.65	97
C_{62}	LC-Establishments$_{ij}$	Herarchichal Ward	17	59.83	62	60	56	58.4	86
C_{63}	LC-Employment$_{ij}$	Herarchichal Ward	16	59.82	63	61.25	31	58.4	98
C_{64}	LC-Employment$_{ij}$	Herarchichal Ward	10	59.68	64	60	57	59.35	89
C_{65}	Coaglomeration Index$_{ij}$	Herarchichal Ward	25	59.67	65	68	7	51.34	137
C_{66}	Coaglomeration Index$_{ij}$	Herarchichal Ward	19	59.28	66	68.42	6	50.13	150
C_{67}	LC-Establishments$_{ij}$	Herarchichal Ward	18	59.20	67	60	58	58,4	99
C_{68}	Coaglomeration Index$_{ij}$	Herarchichal Ward	27	55.96	68	66.67	8	51.26	140
C_{69}	LC-Establishments$_{ij}$	Herarchichal Ward	19	55.57	69	60	59	57.14	104
C_{70}	Coaglomeration Index$_{ij}$	Herarchichal Ward	18	55.57	70	66.67	9	50.48	147
C_{71}	LC-Establishments$_{ij}$	Herarchichal Ward	20	55.51	71	60	60	57.01	105
C_{72}	Coaglomeration Index$_{ij}$	Herarchichal Ward	26	55.34	72	65.38	10	51.3	138
C_{73}	Coaglomeration Index$_{ij}$	Herarchichal Ward	17	57.76	73	64.71	11	50.82	143
C_{74}	Coaglomeration Index$_{ij}$	Herarchichal Ward	14	57.16	74	64.29	12	50.04	152
C_{75}	Coaglomeration Index$_{ij}$	Herarchichal Ward	31	56.41	75	58.06	83	55	125
C_{76}	Coaglomeration Index$_{ij}$	Herarchichal Ward	16	56.40	76	62.5	13	50.3	148
C_{77}	Input-Output Max Share$_{ij}$	Herarchichal Ward	10	56.35	77	62	16	50.69	145
C_{78}	Coaglomeration Index$_{ij}$	Herarchichal Ward	28	56.07	78	60.71	44	51.43	135
C_{79}	Coaglomeration Index$_{ij}$	Herarchichal Ward	35	56.00	79	57.14	84	54.85	124
C_{80}	Coaglomeration Index$_{ij}$	Herarchichal Ward	30	55.97	80	56.67	85	55.28	115
C_{81}	Coaglomeration Index$_{ij}$	Herarchichal Ward	29	55.83	81	58.62	81	53.03	130
C_{82}	Coaglomeration Index$_{ij}$	Herarchichal Ward	13	55.97	82	61.54	18	50	153
C_{83}	Coaglomeration Index$_{ij}$	Herarchichal Ward	10	55.52	83	60	61	51.04	141
C_{84}	Coaglomeration Index$_{ij}$	Herarchichal Ward	34	55.47	84	55.88	86	55.06	119
C_{85}	Coaglomeration Index$_{ij}$	Herarchichal Ward	36	55.35	85	55.56	87	55.15	118
C_{86}	Coaglomeration Index$_{ij}$	Herarchichal Ward	15	55.26	86	60	62	50.52	146
C_{87}	Coaglomeration Index$_{ij}$	Herarchichal Ward	40	55.10	87	55	88	55.19	117
C_{88}	Coaglomeration Index$_{ij}$	Herarchichal Ward	12	54.82	88	58.33	82	51.3	121
C_{89}	Coaglomeration Index$_{ij}$	Herarchichal Ward	33	54.78	89	54.55	89	55.02	122
C_{90}	Coaglomeration Index$_{ij}$	Herarchichal Ward	39	54.43	90	53.85	91	55.02	122

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.

Table B.1 (Cont.) Cluster Configurations and Valuation Scores

Configuration	Similarity Matrix	Clustering Function	Number of Clusters	Global VS	Ranking Global VS	VS-cluster	Ranking VS-Industry	Overlap Score	Ranking Overlap
C_{91}	Coaglomeration Index$_{ij}$	Herarchichal Ward	32	54.07	91	53.13	92	55.02	123
C_{92}	Coaglomeration Index$_{ij}$	Herarchichal Ward	38	54.04	92	53.63	93	55.45	114
C_{93}	Coaglomeration Index$_{ij}$	Herarchichal Ward	37	53.49	93	51.35	95	55.63	113
C_{94}	Input-Output Max Share$_{ij}$	Herarchichal Ward	25	52.83	94	46.4	107	59.26	91
C_{95}	Input-Output Max Share$_{ij}$	Herarchichal Ward	26	52.77	95	47.31	104	58.23	100
C_{96}	Input-Output Max Share$_{ij}$	Herarchichal Ward	23	52.72	96	48.26	102	57.19	103
C_{97}	Coaglomeration Index$_{ij}$	Herarchichal Ward	11	52.68	97	54.55	90	50.82	144
C_{98}	Input-Output Max Share$_{ij}$	Herarchichal Ward	28	52.56	98	46.43	106	58.7	95
C_{99}	Input-Output Max Share$_{ij}$	Herarchichal Ward	14	52.53	99	50	97	55.06	120
C_{100}	Input-Output Max Share$_{ij}$	Herarchichal Ward	24	52.52	100	48.33	101	56.71	107
C_{101}	Input-Output Max Share$_{ij}$	Herarchichal Ward	15	52.49	101	52	94	52.99	131
C_{102}	Input-Output Max Share$_{ij}$	Herarchichal Ward	29	52	102	45.17	112	58.83	94
C_{103}	Input-Output Max Share$_{ij}$	Herarchichal Ward	22	51.89	103	46.82	105	56.97	106
C_{104}	Input-Output Max Share$_{ij}$	Herarchichal Ward	27	51.89	104	45.56	109	58.23	101
C_{105}	Input-Output Max Share$_{ij}$	Herarchichal Ward	17	51.72	105	50.59	96	52.86	132
C_{106}	Input-Output Max Share$_{ij}$	Herarchichal Ward	16	50.93	106	48.75	99	53.12	129
C_{107}	Input-Output Max Share$_{ij}$	Herarchichal Ward	18	50.71	107	47.78	103	53.64	126
C_{108}	Mean Matrix$_{ij}$	Herarchichal Ward	22	50.42	108	24.09	125	76.75	1
C_{109}	Input-Output Max Share$_{ij}$	Herarchichal Ward	30	50.29	109	41.67	114	58.92	93
C_{110}	Input-Output Max Share$_{ij}$	Herarchichal Ward	21	50.26	110	45.24	111	55.28	116
C_{111}	Input-Output Max Share$_{ij}$	Herarchichal Ward	11	50.15	111	50	98	50.3	149
C_{112}	Input-Output Max Share$_{ij}$	Herarchichal Ward	31	49.84	112	40.32	116	59.35	90
C_{113}	Mean Matrix$_{ij}$	Herarchichal Ward	23	49.75	113	23.04	128	76.45	2
C_{114}	Input-Output Max Share$_{ij}$	Herarchichal Ward	34	49.71	114	40.88	115	58.53	96
C_{115}	Input-Output Max Share$_{ij}$	Herarchichal Ward	33	49.46	115	39.8	120	59.22	92
C_{116}	Input-Output Max Share$_{ij}$	Herarchichal Ward	19	49.28	116	45.26	110	53.29	128
C_{117}	Input-Output Max Share$_{ij}$	Herarchichal Ward	13	49.21	117	48.46	100	49.96	154
C_{118}	Input-Output Max Share$_{ij}$	Herarchichal Ward	32	49.07	118	38.75	123	59.39	88
C_{119}	Mean Matrix$_{ij}$	Herarchichal Ward	24	48.99	119	22.08	136	75.89	3
C_{120}	Input-Output Max Share$_{ij}$	Herarchichal Ward	35	48.73	120	39.71	119	57.75	102
C_{121}	Mean Matrix$_{ij}$	Herarchichal Ward	21	48.46	121	23.33	126	73.59	5
C_{122}	Input-Output Max Share$_{ij}$	Herarchichal Ward	28	48.29	122	40.26	117	56.32	111
C_{123}	Input-Output Max Share$_{ij}$	Herarchichal Ward	40	48.26	123	20.25	118	56.28	112
C_{124}	Input-Output Max Share$_{ij}$	Herarchichal Ward	20	48.19	124	43	113	53.38	127
C_{125}	Input-Output Max Share$_{ij}$	Herarchichal Ward	37	48.08	125	39.46	121	56.71	108
C_{126}	Input-Output Max Share$_{ij}$	Herarchichal Ward	39	47.84	126	39.23	122	56.45	110
C_{127}	Mean Matrix$_{ij}$	Herarchichal Ward	25	47.81	127	21.2	137	74.42	4
C_{128}	Input-Output Max Share$_{ij}$	Herarchichal Ward	12	47.79	128	45.83	108	49.74	155
C_{129}	Input-Output Max Share$_{ij}$	Herarchichal Ward	36	47.64	129	38.61	124	56.67	109
C_{130}	Mean Matrix$_{ij}$	Herarchichal Ward	26	47.37	130	21.15	138	73.59	6
C_{131}	Mean Matrix$_{ij}$	Herarchichal Ward	28	47.27	131	21.07	140	73.46	7
C_{132}	Mean Matrix$_{ij}$	Herarchichal Ward	27	47.2	132	21.11	139	73.29	8
C_{133}	Mean Matrix$_{ij}$	Herarchichal Ward	18	46.99	133	23.33	127	70.65	24
C_{134}	Mean Matrix$_{ij}$	Herarchichal Ward	14	46.73	134	22.86	129	70.61	25
C_{135}	Mean Matrix$_{ij}$	Herarchichal Ward	30	46.7	135	20.33	129	73.07	9
C_{136}	Mean Matrix$_{ij}$	Herarchichal Ward	29	46.61	136	20.34	143	72.99	10
C_{137}	Mean Matrix$_{ij}$	Herarchichal Ward	13	46.55	137	22.31	142	70.91	23
C_{138}	Mean Matrix$_{ij}$	Herarchichal Ward	15	46.47	138	22.67	135	70.43	28
C_{139}	Mean Matrix$_{ij}$	Herarchichal Ward	16	46.44	139	22.5	130	70.43	29
C_{140}	Mean Matrix$_{ij}$	Herarchichal Ward	17	46.36	140	22.35	132	70.52	27
C_{141}	Mean Matrix$_{ij}$	Herarchichal Ward	20	46.32	141	22.5	134	70.22	30
C_{142}	Mean Matrix$_{ij}$	Herarchichal Ward	19	46.03	142	22.63	133	70	31
C_{143}	Mean Matrix$_{ij}$	Herarchichal Ward	34	46.01	143	19.41	131	72.64	12
C_{144}	Mean Matrix$_{ij}$	Herarchichal Ward	32	45.98	144	19.38	147	71.13	13
C_{145}	Mean Matrix$_{ij}$	Herarchichal Ward	12	45.93	145	20.83	148	72.77	22
C_{146}	Mean Matrix$_{ij}$	Herarchichal Ward	33	45.9	146	19.09	141	72.38	11
C_{147}	Mean Matrix$_{ij}$	Herarchichal Ward	35	45.89	147	19.43	149	72.03	15
C_{148}	Mean Matrix$_{ij}$	Herarchichal Ward	40	45.84	148	19.75	146	72.65	16
C_{149}	Mean Matrix$_{ij}$	Herarchichal Ward	31	45.61	149	19.03	145	71.21	14
C_{150}	Mean Matrix$_{ij}$	Herarchichal Ward	11	45.44	150	20	150	71.9	21
C_{151}	Mean Matrix$_{ij}$	Herarchichal Ward	39	45.29	151	18.97	144	71.9	17
C_{152}	Mean Matrix$_{ij}$	Herarchichal Ward	38	45.15	152	18.68	152	71.69	18
C_{153}	Mean Matrix$_{ij}$	Herarchichal Ward	36	45.08	153	18.61	153	71.77	20
C_{154}	Mean Matrix$_{ij}$	Herarchichal Ward	37	44.8	154	18.38	154	71.69	19
C_{155}	Mean Matrix$_{ij}$	Herarchichal Ward	10	44.08	155	19	151	71.77	26

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.

Annex B.2. Cluster classification

C1-oil and gas extraction

Number of indsutries: 6

4-digit NAICS	**Oil and gas extraction**	WCR_k	Outlier	Percentages wthin the Cluster: Employment	Establishments	Production	G. value added
2111	Oil and Gas Extraction	2.64		5.53	50.55	95.97	93.67
3366	Ship and Boat Building	2.41		16.60	12.39	0.47	1.21
4831	Deep Sea, Coastal, and Great Lakes Water Transportation	2.27		8.30	10.60	1.33	1.69
4862	Pipeline Transportation of Natural Gas	3.17		6.17	5.57	0.87	1.27
4869	Other Pipeline Transportation	2.51		1.28	1.62	0.41	0.65
4883	Support Activities for Water Transportation	2.47		62.13	19.27	0.95	1.51

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This is the original configuration of the cluster.

C2-metal mining

Number of indsutries: 6

4-digit NAICS	**Metal minning**	WCR_k	Outlier	Percentages wthin the Cluster: Employment	Establishments	Production	G. value added
2122	Metal Ore Mining	3.52	**	9.54	42.35	59.01	43.36
3311	Iron and Steel Mills and Ferroalloy Manufacturing	3.27		0.79	14.26	24.54	37.28
3365	Railroad Rolling Stock Manufacturing	3.25		1.39	10.69	6.48	8.19
2123	Nonmetallic Mineral Mining and Quarrying	3.36		80.95	16.28	4.95	4.66
2131	Support Activities for Mining	3.33	**	5.27	11.89	3.26	4.62
2121	Coal Mining	2.86		2.05	4.53	1.75	1.89

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration.

C3-footwear manufacturing

Number of indsutries: 2

4-digit NAICS	**Footwear manufacturing**	WCR_k	Outlier	Percentages wthin the Cluster: Employment	Establishments	Production	G. value added
3162	Footwear Manufacturing	1.86		90.32	85.22	68.91	70.21
3161	Leather and Hide Tanning and Finishing	1.86		9.68	14.78	31.09	29.79

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This is the original configuration of the cluster.

C4-sawmills and wood preservation

Number of indsutries: 2

4-digit NAICS	**Sawmills and wood preservation**	WCR_k	Outlier	Percentages wthin the Cluster: Employment	Establishments	Production	G. value added
3212	Veneer, Plywood, and Engineered Wood Product Manufacturir	3.38		9.11	34.71	38.78	55.70
3211	Sawmills and Wood Preservation	3.38		90.89	65.29	61.22	44.30

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This is the original configuration of the cluster.

C5-medical equipment and supplies manufacturing

Number of indsutries: 5

4-digit NAICS	**Medical equipment and supplies manufacturing**	WCR_k	Outlier	Percentages wthin the Cluster: Employment	Establishments	Production	G. value added
3352	Household Appliance Manufacturing	3.79	**	8.41	16.82	25.02	46.40
3391	Medical Equipment and Supplies Manufacturing	3.25		79.43	44.09	33.69	22.65
3364	Aerospace Product and Parts Manufacturing	4.23		4.34	10.53	16.38	12.37
3342	Communications Equipment Manufacturing	3.02	**	4.55	13.43	13.26	10.33
3343	Audio and Video Equipment Manufacturing	3.24		3.28	15.12	11.65	8.26

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration.

C6-semiconductors and other electronic components manufacturing

Number of indsutries: 4

4-digit NAICS	Semiconductor and other electronic component manufacturi	WCR_k	Outlier	Employment	Establishments	Production	G. value added
				Percentages wthin the Cluster			
3344	Semiconductor and Other Electronic Component Manufacturin	3.23		59.04	75.00	70.81	63.82
3341	Computer and Peripheral Equipment Manufacturing	3.08		7.05	15.64	15.00	17.45
3325	Hardware Manufacturing	3.64		32.05	7.03	11.72	14.51
3346	Manufacturing and Reproducing Magnetic and Optical Media	3.18		1.86	2.33	2.47	4.22

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This is the original configuration of the cluster.

C7-food and beverage manufacturing

Number of indsutries: 10

4-digit NAICS	Food and beverage manufacturing	WCR_k	Outlier	Employment	Establishments	Production	G. value added
				Percentages wthin the Cluster			
3121	Beverage Manufacturing	3.87	**	10.96	14.51	28.06	21.06
3112	Grain and Oilseed Milling	3.80		0.63	3.88	8.29	12.28
3116	Animal Slaughtering and Processing	4.43	*	1.79	9.22	10.55	11.56
3118	Bakeries and Tortilla Manufacturing	4.09	**	73.80	39.79	13.40	10.94
3115	Dairy Product Manufacturing	3.68		6.85	8.31	8.04	10.81
3119	Other Food Manufacturing	3.89	**	3.06	7.05	12.86	10.49
3111	Animal Food Manufacturing	3.91		0.30	2.46	6.49	8.75
3113	Sugar and Confectionery Product Manufacturing	3.78		1.63	7.26	6.53	7.89
3114	Fruit and Vegetable Preserving and Specialty Food Manufactur	3.84	**	0.88	6.27	5.06	5.14
3117	Seafood Product Preparation and Packaging	6.03		0.09	1.25	0.71	1.09

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These four systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C8-automotive

Number of indsutries: 7

4-digit NAICS	Automotive	WCR_k	Outlier	Employment	Establishments	Production	G. value added
				Percentages wthin the Cluster			
3361	Motor Vehicle Manufacturing	4.63	*	0.65	8.11	45.08	50.78
3363	Motor Vehicle Parts Manufacturing	5.31	**	23.03	80.96	49.04	44.16
3262	Rubber Product Manufacturing	3.84		11.13	5.00	2.60	2.47
3362	Motor Vehicle Body and Trailer Manufacturing	4.30		10.29	2.45	1.72	1.27
3336	Engine, Turbine, and Power Transmission Equipment Manufac	4.50	**	0.72	1.83	1.26	1.12
3169	Other Leather and Allied Product Manufacturing	4.25		35.03	1.10	0.16	0.12
3159	Apparel Accessories and Other Apparel Manufacturing	4.25		19.15	0.56	0.13	0.07

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C9-petroleum and coal products manufacturing

Number of indsutries: 4

4-digit NAICS	Petroleum and coal products manufacturing	WCR_k	Outlier	Employment	Establishments	Production	G. value added
				Percentages wthin the Cluster			
3241	Petroleum and Coal Products Manufacturing	5.44		21.64	36.03	62.42	70.59
3251	Basic Chemical Manufacturing	4.67		44.72	44.24	18.12	24.07
3253	Pesticide, Fertilizer, and Other Agricultural Chemical Manufact	4.47		29.92	16.14	12.44	4.23
3122	Tobacco Manufacturing	5.26		3.72	3.58	7.03	1.11

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This is the original configuration of the cluster.

C10-apparel manufacturing

Number of indsutries: 7				Percentages wthin the Cluster			
4-digit NAICS	**Apparel manufacturing**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
3152	Cut and Sew Apparel Manufacturing	3.91		38.69	55.04	50.07	43.09
3132	Fabric Mills	3.58		1.33	9.52	14.34	20.06
3131	Fiber, Yarn, and Thread Mills	4.32		16.48	7.78	8.92	8.28
3133	Textile and Fabric Finishing and Fabric Coating Mills	3.75	**	0.61	3.80	8.19	7.74
3149	Other Textile Product Mills	4.56	**	37.17	14.09	6.43	7.53
3151	Apparel Knitting Mills	4.04		2.57	5.34	6.59	7.52
3141	Textile Furnishings Mills	3.67		3.15	4.43	5.45	5.78

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration.

C11-tourism and hospitality services

Number of indsutries: 13				Percentages wthin the Cluster			
4-digit NAICS	**Tourism and hospitality services**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
7211	Traveler Accommodation	4.77		40.40	72.04	69.56	71.23
5615	Travel Arrangement and Reservation Services	5.22		14.96	8.71	14.95	13.94
5321	Automotive Equipment Rental and Leasing	4.88		3.87	2.59	4.83	5.52
7131	Amusement Parks and Arcades	5.06		23.83	5.88	5.42	4.18
7121	Museums, Historical Sites, and Similar Institutions	4.97		1.33	1.65	1.24	1.42
4871	Scenic and Sightseeing Transportation, Land	4.51		0.81	1.27	0.95	0.94
4859	Other Transit and Ground Passenger Transportation	5.28		0.62	1.64	0.84	0.77
7213	Rooming and Boarding Houses, Dormitories, and Workers' Car	4.59		6.75	1.91	0.84	0.76
7111	Performing Arts Companies	4.63		5.58	2.98	0.62	0.51
4872	Scenic and Sightseeing Transportation, Water	5.23		0.97	0.70	0.32	0.33
4832	Inland Water Transportation	5.11		0.58	0.33	0.28	0.25
7212	RV (Recreational Vehicle) Parks and Recreational Camps	5.06	**	0.29	0.28	0.13	0.15
4879	Scenic and Sightseeing Transportation, Other	5.03	**	0.01	0.03	0.02	0.01

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration, and this cluster is the result of merging three different clusters.

C12-office administrative services

Number of indsutries : 14				Percentages wthin the Cluster			
4-digit NAICS	**Office administrative services**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
5611	Office Administrative Services	3.96		4.83	29.58	36.18	33.63
5416	Management, Scientific, and Technical Consulting Services	3.58		17.37	19.67	19.50	19.07
5415	Computer Systems Design and Related Services	3.64		7.59	11.07	10.79	11.91
5412	Accounting, Tax Preparation, Bookkeeping, and Payroll Servic	3.86		39.19	14.26	11.24	10.09
5418	Advertising, Public Relations, and Related Services	3.62		17.48	9.27	8.05	8.87
5617	Services to Buildings and Dwellings	3.69		5.09	9.99	5.77	5.47
7112	Spectator Sports	5.51	*	0.33	0.83	3.15	4.24
5612	Facilities Support Services	3.59		0.84	2.00	2.34	2.66
7113	Promoters of Performing Arts, Sports, and Similar Events	3.65		2.80	1.35	1.37	1.98
5417	Scientific Research and Development Services	3.51		0.93	1.06	1.00	1.05
5122	Sound Recording Industries	4.21	**	0.54	0.47	0.42	0.68
7114	Agents and Managers for Artists, Athletes, Entertainers, and Otl	4.00	**	0.14	0.07	0.10	0.23
7115	Independent Artists, Writers, and Performers	4.38	**	2.54	0.29	0.06	0.09
6243	Vocational Rehabilitation Services	4.16		0.33	0.10	0.03	0.03

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These three systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C13-metal products manufacturing

Number of indsutries: 11				Percentages wthin the Cluster			
4-digit NAICS	**Metal product manufacturing**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
4841	General Freight Trucking	4.11		20.93	30.85	20.53	18.85
3272	Glass and Glass Product Manufacturing	3.85		5.15	9.93	13.14	14.74
3329	Other Fabricated Metal Product Manufacturing	3.64		6.10	12.16	13.93	11.91
3339	Other General Purpose Machinery Manufacturing	3.54		4.92	9.52	12.00	10.89
3313	Alumina and Aluminum Production and Processing	3.62		0.70	3.79	5.66	8.18
3328	Coating, Engraving, Heat Treating, and Allied Activities	3.67		4.12	5.88	6.97	7.94
3324	Boiler, Tank, and Shipping Container Manufacturing	3.76		2.28	4.63	6.52	7.85
3327	Machine Shops; Turned Product; and Screw, Nut, and Bolt Mai	3.66		48.77	12.85	8.80	7.14
3322	Cutlery and Handtool Manufacturing	4.29	*	3.53	4.19	6.11	5.89
3321	Forging and Stamping	4.12		2.79	4.03	3.90	4.02
3369	Other Transportation Equipment Manufacturing	3.91		0.71	2.18	2.45	2.59

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This marginal outlier was reallocated from the initial configuration.

C14-steel products manufacturing

Number of indsutries : 14				Percentages wthin the Cluster			
4-digit NAICS	**Steel product manufacturing**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
3312	Steel Product Manufacturing from Purchased Steel	4.45		2.15	9.15	20.40	23.88
3314	Nonferrous Metal (except Aluminum) Production and Processii	4.92		0.52	2.43	23.76	21.22
3359	Other Electrical Equipment and Component Manufacturing	4.31		1.27	16.10	11.37	12.02
3334	Ventilation, Heating, Air-Conditioning, and Commercial Refrig	4.37		1.93	11.10	9.13	9.88
3353	Electrical Equipment Manufacturing	4.33		1.32	14.83	9.41	7.69
3331	Agriculture, Construction, and Mining Machinery Manufacturi	4.34		1.61	5.82	4.74	6.34
3271	Clay Product and Refractory Manufacturing	5.26		77.84	19.96	7.57	5.84
3315	Foundries	4.48		1.94	6.13	5.32	5.65
3345	Navigational, Measuring, Electromedical, and Control Instrume	4.57		0.51	4.54	2.74	2.51
3274	Lime and Gypsum Product Manufacturing	5.21		7.67	2.39	1.12	1.59
3351	Electric Lighting Equipment Manufacturing	4.31		0.98	3.94	2.02	1.44
3335	Metalworking Machinery Manufacturing	4.19		0.98	1.60	1.40	1.17
3333	Commercial and Service Industry Machinery Manufacturing	4.38		1.18	1.69	0.82	0.67
4882	Support Activities for Rail Transportation	5.19		0.09	0.31	0.21	0.10

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This configuration was the initial configuration.

C15-financial services and head offices

Number of indsutries : 17				Percentages wthin the Cluster			
4-digit NAICS	**Financial services and head offices**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
5221	Depository Credit Intermediation	4.78		0.18	29.63	27.54	26.06
5511	Head offices	4.25		1.27	13.57	28.69	21.25
5173	Wired and Wireless Telecommunications Carriers	4.36		0.96	16.79	7.84	19.34
5241	Insurance Carriers	4.43		0.74	5.75	12.14	8.51
5224	Other credit and financial intermediation institutions, non-stock	5.31		46.61	13.23	3.85	6.97
5222	Nondepository Credit Intermediation	4.82		0.06	1.89	7.80	5.69
5211	Monetary Authorities-Central Bank	4.22		0.00	0.33	4.82	3.41
5151	Radio and Television Broadcasting	5.06	*	4.79	4.21	1.92	2.27
5223	Activities Related to Credit Intermediation	6.24	**	18.14	4.47	0.96	1.88
5242	Agencies, Brokerages, and Other Insurance Related Activities	5.39	**	15.77	6.61	1.70	1.85
5231	Securities and Commodity Contracts Intermediation and Brokei	5.31	**	10.32	1.24	1.24	1.18
5225	Services related to credit intermediation, non-stock exchange	4.22		0.11	1.03	0.96	1.00
5112	Software Publishers	4.29		0.42	0.76	0.17	0.19
5232	Securities and Commodity Exchanges	4.31		0.02	0.08	0.22	0.17
5174	Satellite Telecommunications	4.82		0.30	0.16	0.06	0.07
5152	Cable and Other Subscription Programming	4.66		0.11	0.17	0.05	0.07
5331	Lessors of Nonfinancial Intangible Assets (except Copyrighted	4.27		0.19	0.07	0.06	0.06

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C16-electric power generation, transmission, and distribution and infrastructure construction

Number of indsutries : 10				Percentages wthin the Cluster			
4-digit NAICS	**Electric power generation, transmission and distribution and infraestructure construction**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
2211	Electric Power Generation, Transmission and Distribution	4.98	**	0.45	16.86	47.40	59.03
2373	Highway, Street, and Bridge Construction	4.88		4.93	13.49	9.00	8.95
2371	Utility System Construction	4.73		4.73	12.33	7.22	7.73
5312	Offices of Real Estate Agents and Brokers	4.97		22.86	13.07	10.87	6.99
5413	Architectural, Engineering, and Related Services	4.49		26.47	19.70	8.95	5.23
5313	Activities Related to Real Estate	4.44		23.26	9.50	7.04	4.36
2382	Building Equipment Contractors	4.45		14.69	8.58	4.95	3.75
2389	Other Specialty Trade Contractors	4.74		1.80	4.04	2.31	2.13
2379	Other Heavy and Civil Engineering Construction	4.53		0.77	2.14	1.06	1.18
2212	Natural Gas Distribution	4.78	**	0.04	0.27	1.20	0.65

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These two systematic outliers were reallocated from the initial configuration.

C17-passenger transportation and communications

Number of indsutries : 13				Percentages wthin the Cluster			
4-digit NAICS	**Passenger transportation and communications**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
4811	Scheduled Air Transportation	5.41		0.68	9.95	26.19	37.04
4852	Interurban and Rural Bus Transportation	6.18	*	21.98	29.03	22.54	20.03
4821	Rail Transportation	5.39		0.17	5.14	15.35	13.61
5111	Newspaper, Periodical, Book, and Directory Publishers	4.99		25.10	12.90	8.14	6.84
4921	Couriers and Express Delivery Services	4.87		10.86	14.00	6.90	6.33
4881	Support Activities for Air Transportation	5.26		6.01	12.33	8.22	6.32
5182	Data Processing, Hosting, and Related Services	5.11		4.83	7.23	6.14	4.03
4812	Nonscheduled Air Transportation	5.37	**	1.44	1.49	2.22	2.41
4855	Charter Bus Industry	5.44	**	11.41	3.06	1.56	1.41
4889	Other Support Activities for Transportation	5.37	**	4.40	2.87	1.55	1.02
5191	Other Information Services	4.88		8.44	1.30	0.98	0.77
4853	Taxi and Limousine Service	5.16		1.07	0.29	0.09	0.09
4922	Local Messengers and Local Delivery	5.47	**	3.61	0.41	0.10	0.09

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These four systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C18-retail and eating services

Number of indsutries : 6				Percentages wthin the Cluster			
4-digit NAICS	**Retail and eating servicies**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
7225	Restaurants and Other Eating Places	7.36	*	34.75	44.31	33.36	45.71
4311	Wholesale trade of groceries and food	7.16	*	1.31	9.60	29.86	25.47
4611	Retail trade of groceries and food	6.91	*	55.54	39.11	32.26	23.22
7224	Drinking Places (Alcoholic Beverages)	6.68		1.69	2.73	2.09	2.77
8114	Personal and Household Goods Repair and Maintenance	5.86		5.79	3.40	1.89	2.30
5322	Consumer Goods Rental	5.95		0.92	0.86	0.54	0.53

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These three marginal outliers were reallocated from the initial configuration.

C19-employment services

Number of indsutries : 1				Percentages wthin the Cluster			
4-digit NAICS	**Employment services**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
5613	Employment Services	3.17	*	100.00	100.00	100.00	100.00

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This industry is a marginal outlier that becomes a single industry cluster.

C20-plastic products manufacturing

Number of indsutries : 21

4-digit NAICS	Plastics product manufacturing	WCR_k	Outlier	Percentages wthin the Cluster Employment	Establishments	Production	G. value added
3261	Plastics Product Manufacturing	5.35		1.30	16.84	19.49	23.32
3222	Converted Paper Product Manufacturing	5.23		1.44	6.25	9.65	11.21
3256	Soap, Cleaning Compound, and Toilet Preparation Manufactur	5.74	**	0.41	2.96	10.93	9.39
3252	Resin, Synthetic Rubber, and Artificial and Synthetic Fibers anc	5.51	**	0.07	0.99	6.74	9.15
3273	Cement and Concrete Product Manufacturing	5.72		1.68	3.23	8.32	7.84
3221	Pulp, Paper, and Paperboard Mills	5.84	**	0.08	1.60	4.93	5.60
8111	Automotive Repair and Maintenance	5.64		57.87	28.93	8.62	5.59
3399	Other Miscellaneous Manufacturing	5.76		4.56	6.46	4.90	4.34
3255	Paint, Coating, and Adhesive Manufacturing	5.21		0.15	1.30	4.01	4.29
3323	Architectural and Structural Metals Manufacturing	5.28		14.55	9.23	4.55	3.93
3371	Household and Institutional Furniture and Kitchen Cabinet Mar	5.68		7.12	7.14	3.86	2.97
3259	Other Chemical Product and Preparation Manufacturing	5.19		0.29	1.54	3.10	2.92
3219	Other Wood Product Manufacturing	5.38		7.99	4.82	2.18	1.92
3326	Spring and Wire Product Manufacturing	5.28		0.18	1.00	1.79	1.90
4931	Warehousing and Storage	5.25		0.24	1.78	1.67	1.25
3279	Other Nonmetallic Mineral Product Manufacturing	5.54		1.11	1.33	1.21	1.17
3379	Other Furniture Related Product Manufacturing	5.64	**	0.17	1.23	1.00	0.93
3332	Industrial Machinery Manufacturing	5.31	**	0.20	0.73	1.09	0.92
5621	Waste Collection	5.29		0.15	0.76	0.97	0.69
4854	School and Employee Bus Transportation	6.01	*	0.10	1.66	0.85	0.59
5323	General Rental Centers	5.29		0.34	0.21	0.13	0.08

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These five systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

C21-freight transportation services and residential and nonresidential construction

Number of indsutries : 15

4-digit NAICS	Freight transportation services and residencial and nonresidential construction	WCR_k	Outlier	Percentages wthin the Cluster Employment	Establishments	Production	G. value added
4885	Freight Transportation Arrangement	5.84		8.14	10.79	23.12	19.44
2362	Nonresidential Building Construction	5.78		8.27	16.14	13.42	18.84
2361	Residential Building Construction	6.08	**	7.24	13.14	9.89	13.90
5616	Investigation and Security Services	5.79		7.84	20.55	15.45	10.97
4851	Urban Transit Systems	6.33		3.71	12.04	6.90	7.34
8113	Commercial and Industrial Machinery and Equipment (except	5.69	**	35.32	8.22	7.59	6.78
5324	Commercial and Industrial Machinery and Equipment Rental ar	5.47		11.65	3.60	5.50	4.79
2372	Land Subdivision	6.66	**	2.49	3.30	3.55	4.64
5619	Other Support Services	5.60		3.25	4.81	5.77	4.60
4884	Support Activities for Road Transportation	5.74		3.66	1.66	3.86	3.24
3372	Office Furniture (including Fixtures) Manufacturing	5.52		2.53	2.32	1.92	2.32
2381	Foundation, Structure, and Building Exterior Contractors	5.48		2.42	1.52	1.03	1.32
5622	Waste Treatment and Disposal	5.58		0.27	0.49	0.85	0.70
2383	Building Finishing Contractors	5.47		2.59	0.72	0.50	0.58
5629	Remediation and Other Waste Management Services	5.70	**	0.62	0.69	0.65	0.53

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These four systematic outliers were reallocated from the initial configuration.

C22-business support services

Number of indsutries : 9

4-digit NAICS	Business support services	WCR_k	Outlier	Percentages wthin the Cluster Employment	Establishments	Production	G. value added
3231	Printing and Related Support Activities	5.14		10.24	17.07	18.97	27.68
5121	Motion Picture and Video Industries	5.37		0.72	6.41	21.32	20.83
5614	Business Support Services	6.03		28.13	28.80	20.30	16.94
5411	Legal Services	5.17		15.45	17.45	17.05	13.03
5419	Other Professional, Scientific, and Technical Services	5.67		12.35	8.69	7.45	6.47
8122	Death Care Services	5.34		19.94	10.72	4.34	4.48
5179	Other Telecommunications	5.36		1.76	2.64	4.26	4.19
8124	Parking lots and garages for motor vehicles	5.22		8.77	5.57	3.83	3.88
5414	Specialized Design Services	5.18		2.62	2.65	2.48	2.49

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
This cluster corresponds to the original configuration of the cluster.

C23-education and health services

Number of indsutries : 19				Percentages wthin the Cluster			
4-digit NAICS	**Education and health services**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
6113	General secondary education schools	5.25		0.78	12.70	25.15	22.83
6111	Basic, middle and special needs education schools	5.49		4.69	23.01	26.26	21.45
8121	Personal Care Services	5.41		40.98	16.84	7.70	8.69
5311	Lessors of Real Estate	5.78		5.76	4.85	8.33	8.66
7139	Other Amusement and Recreation Industries	5.36		5.20	6.49	6.43	8.33
6211	Offices of Physicians	5.33		11.92	6.98	5.84	6.36
6215	Medical and Diagnostic Laboratories	5.54	**	2.67	3.17	5.11	6.28
6212	Offices of Dentists	5.18		11.34	5.60	3.79	4.48
6116	Higher middle education schools	5.40		3.83	4.05	2.78	2.85
6213	Offices of Other Health Practitioners	5.19		5.16	3.11	2.11	2.28
8123	Drycleaning and Laundry Services	5.48		1.16	1.61	1.47	2.10
6244	Child Day Care Services	5.84	**	1.26	3.40	2.11	1.83
6214	Outpatient Care Centers	5.57	**	0.22	0.59	0.89	1.42
8129	Other Personal Services	5.36	**	1.08	0.76	0.56	0.72
6115	Terminal technical middle education schools	5.54		0.56	0.67	0.43	0.47
6114	Technical secondary education schools	5.89		0.19	0.43	0.42	0.47
6241	Individual and Family Services	5.95	**	3.12	5.41	0.22	0.38
6112	Higher technical education schoolsT	5.66		0.06	0.24	0.25	0.26
6117	Educational support servicesT	6.91	**	0.02	0.09	0.15	0.14

Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These six systematic outliers were reallocated from the initial configuration.

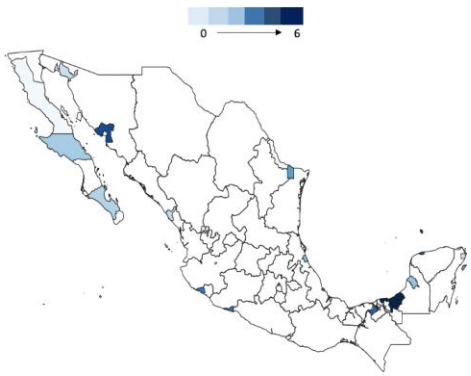
C24-pharmaceutical and medical manufacturing and services

Number of indsutries : 14				Percentages wthin the Cluster			
4-digit NAICS	**Pharmaceutical and medical manufacturing and services**	WCR_k	Outlier	Employment	Establishments	Production	G. value added
3254	Pharmaceutical and Medicine Manufacturing	6.51		1.92	19.61	55.17	55.03
6221	General Medical and Surgical Hospitals	5.92		4.14	25.12	18.46	20.09
7132	Gambling Industries	5.93		11.93	9.78	10.29	10.49
8112	Electronic and Precision Equipment Repair and Maintenance	6.02		63.52	15.07	5.93	5.04
6223	Specialty (except Psychiatric and Substance Abuse) Hospitals	5.83		1.89	6.39	4.16	3.91
7223	Special Food Services	5.84		5.71	10.93	4.24	3.61
6219	Other Ambulatory Health Care Services	6.87	**	1.90	4.24	0.36	0.65
6233	Continuing Care Retirement Communities and Assisted Living I	5.82		1.94	2.64	0.49	0.41
6239	Other Residential Care Facilities	5.93		2.50	2.60	0.29	0.29
6216	Home Health Care Services	7.36	**	0.49	0.50	0.27	0.16
6242	Community Food and Housing, and Emergency and Other Reli	6.36	**	1.83	1.34	0.13	0.13
6232	Residential Intellectual and Developmental Disability, Mental H	6.12	**	1.66	1.20	0.10	0.09
6222	Psychiatric and Substance Abuse Hospitals	6.40	*	0.15	0.20	0.07	0.06
6231	Nursing Care Facilities (Skilled Nursing Facilities)	6.18		0.42	0.37	0.06	0.05

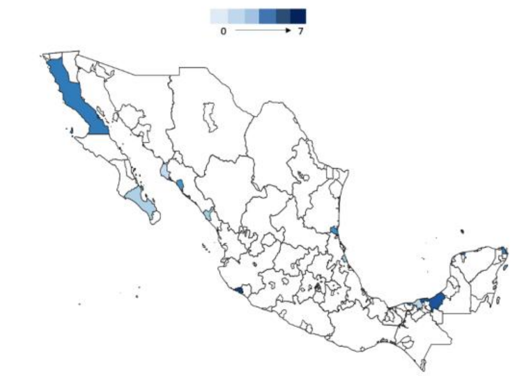
Source: Authors' own estimation based on information from the 2013 Input-Output Account Matrix and the 2019 Economic Census, INEGI.
Note: * indicates marginal outliers and ** stands for systematic outliers.
These four systematic outliers were reallocated from the initial configuration, as well as the marginal outlier.

Annex B.3.
Employment Location Quotient by Cluster at Municipality Level
C1-oil and gas extraction

Distribution at the Municipality Level of Location Quotient, 2014

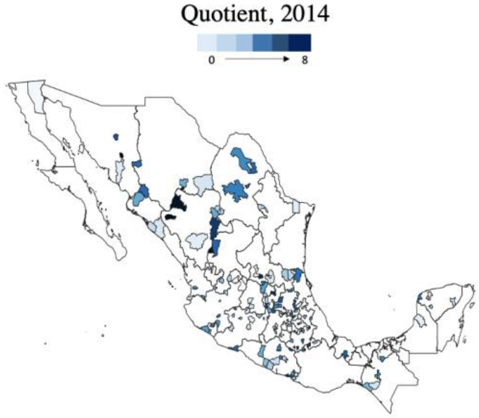


Distribution at the Municipality Level of Location Quotient, 2018

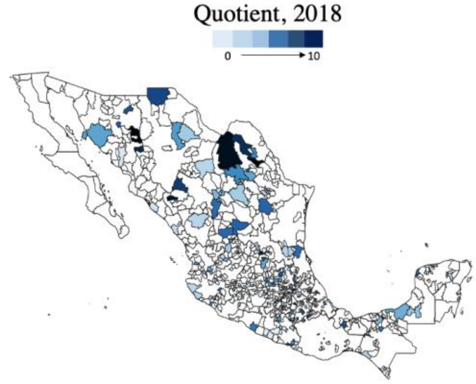


C2-metal mining

Distribution at the Municipality Level of Location Quotient, 2014

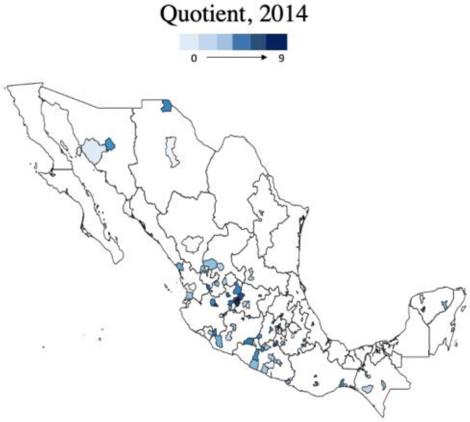


Distribution at the Municipality Level of Location Quotient, 2018



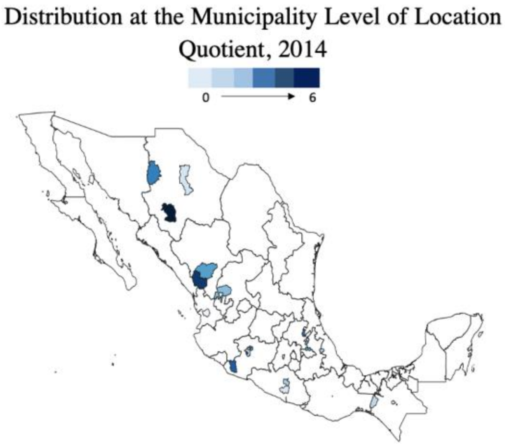
C3-footwear manufacturing

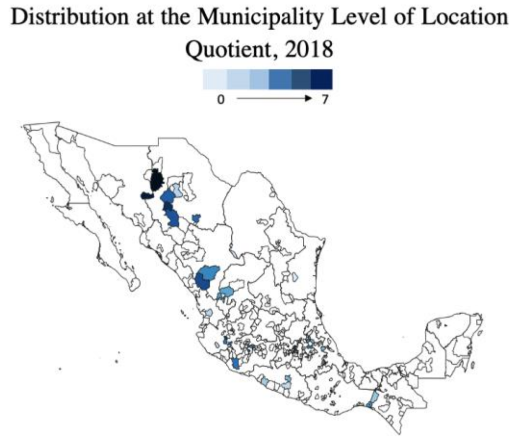
Distribution at the Municipality Level of Location Quotient, 2014



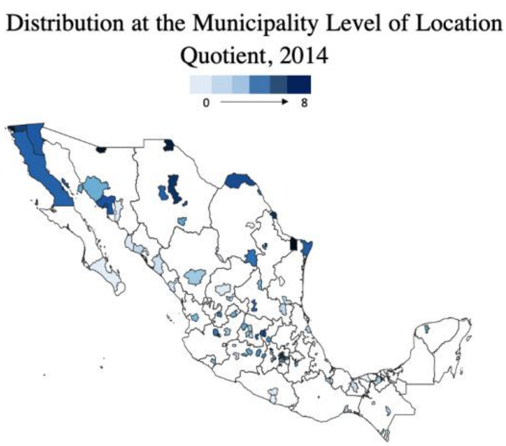
Distribution at the Municipality Level of Location Quotient, 2018

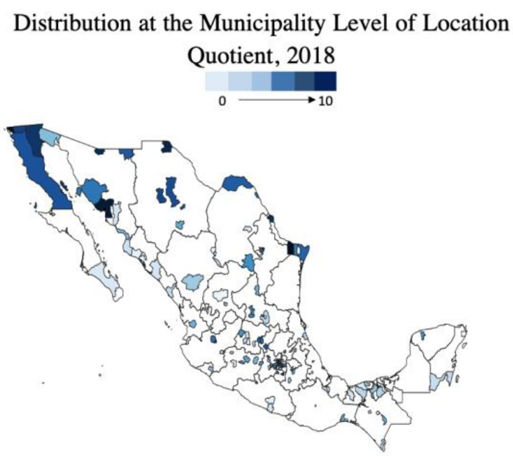
C4-sawmills and wood preservation



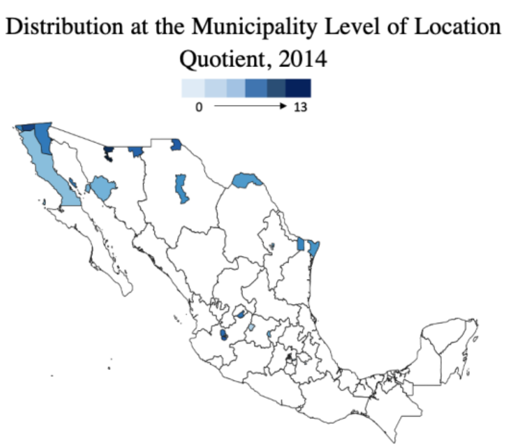


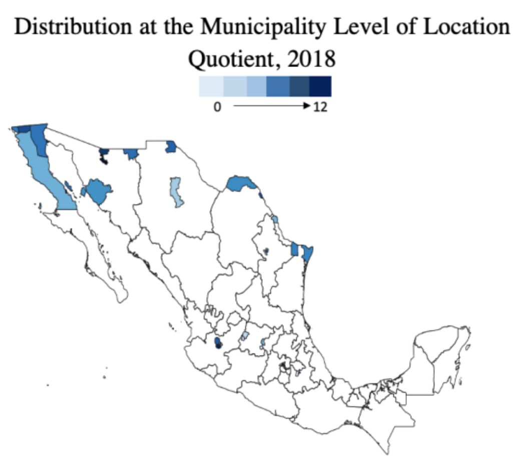
C5-medical equipment and supplies manufacturing



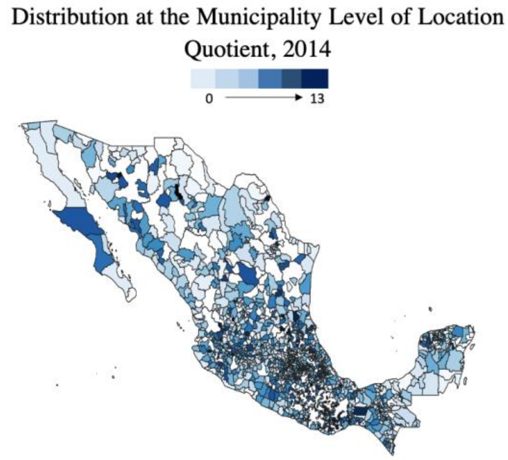


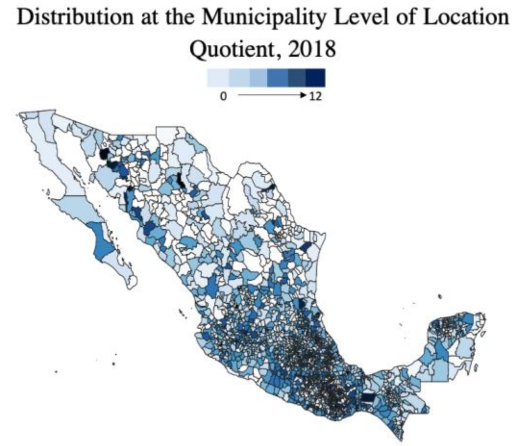
C6-semiconductors and other electronic components



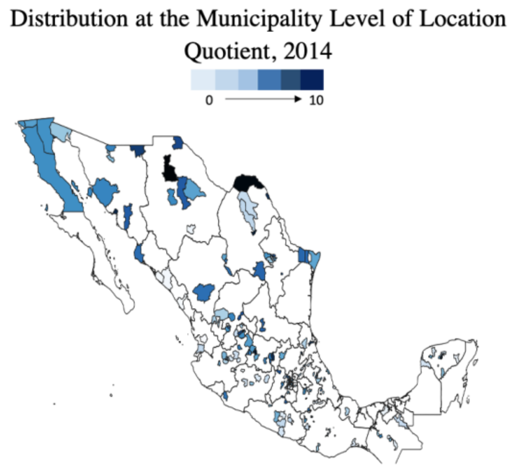


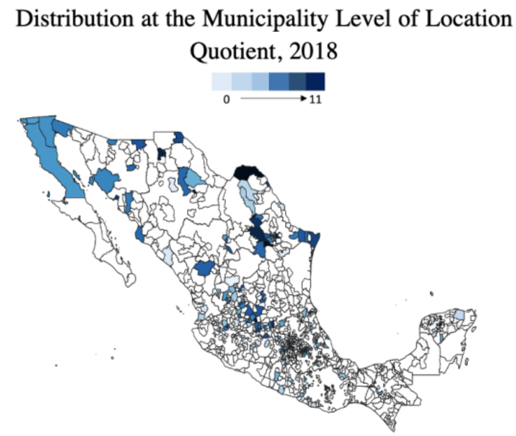
C7-food and beverage manufacturing



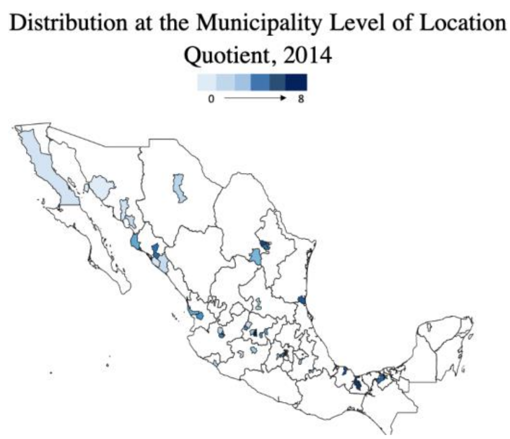


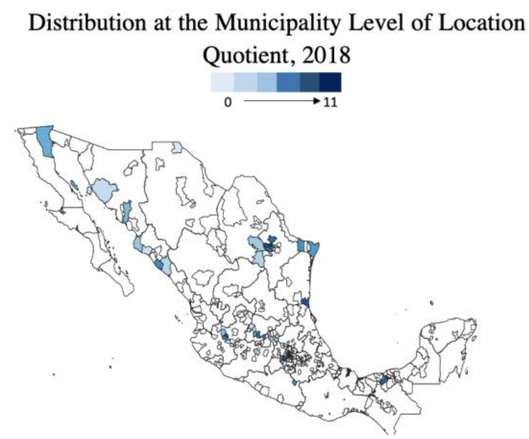
C8-automotive



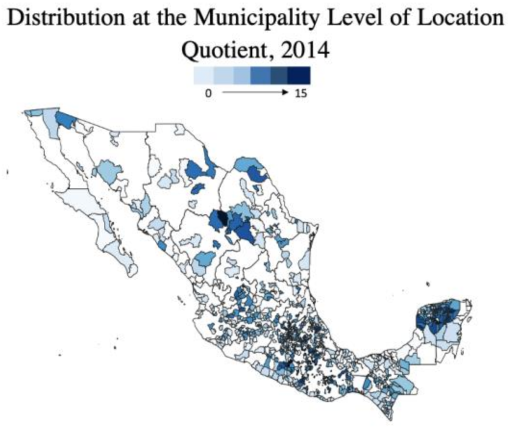


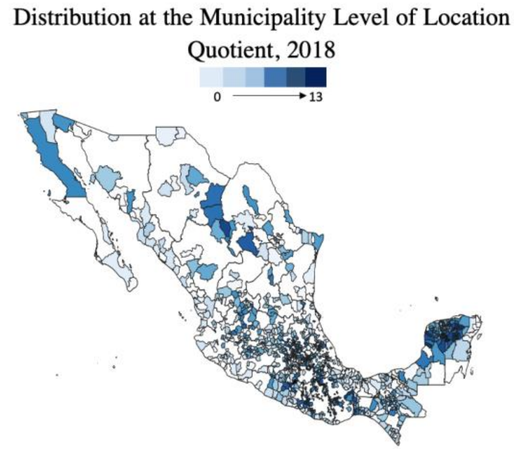
C9-petroleum and coal products manufacturing



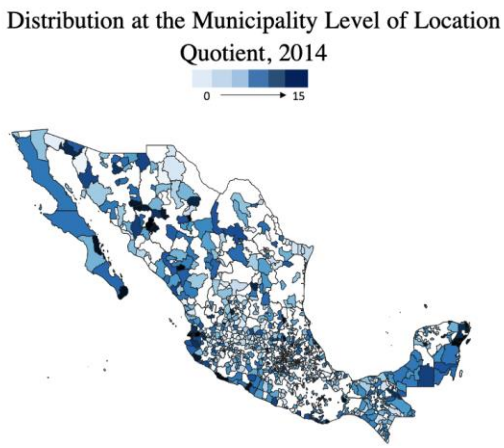


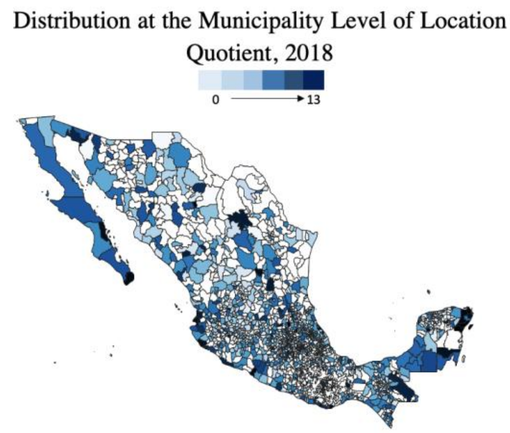
C10-apparel manufacturing



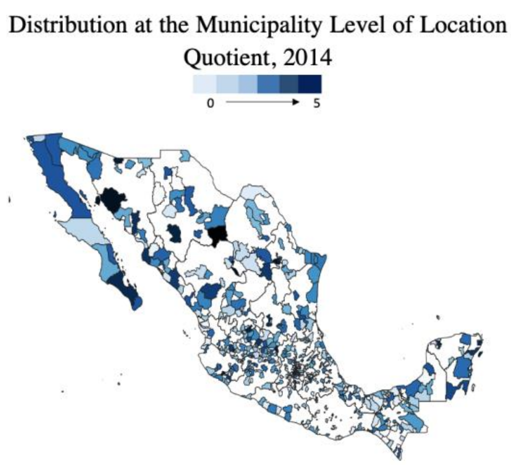


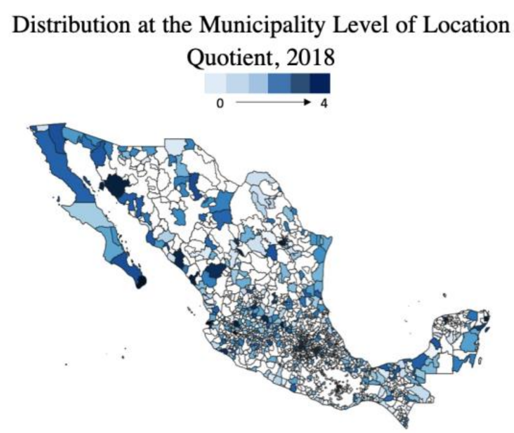
C11-tourism and hospitality services



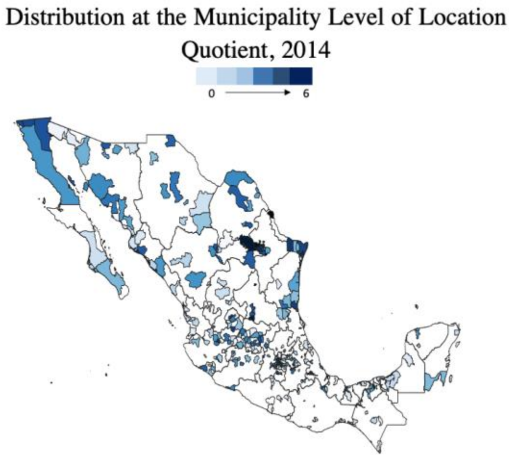


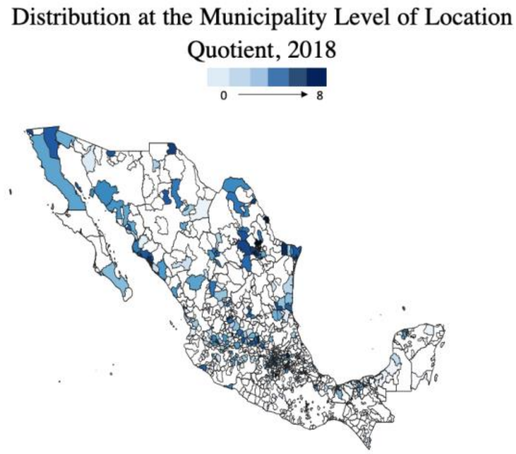
C12-office administrative services



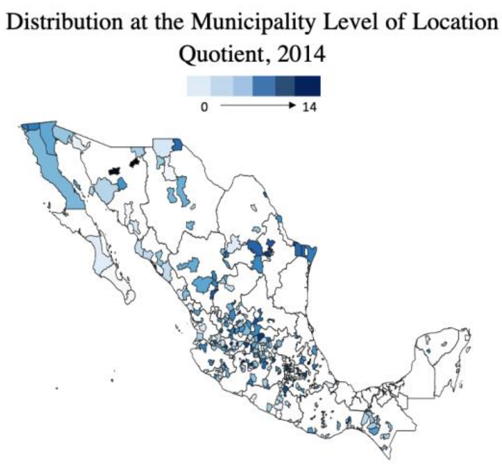


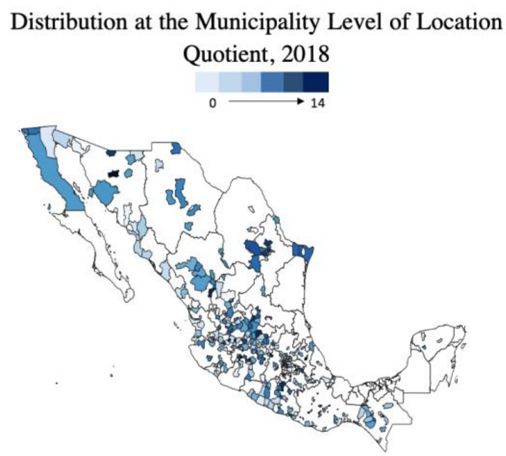
C13-metal product manufacturing



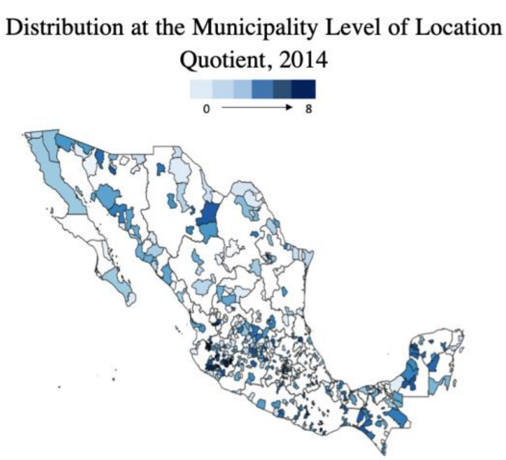


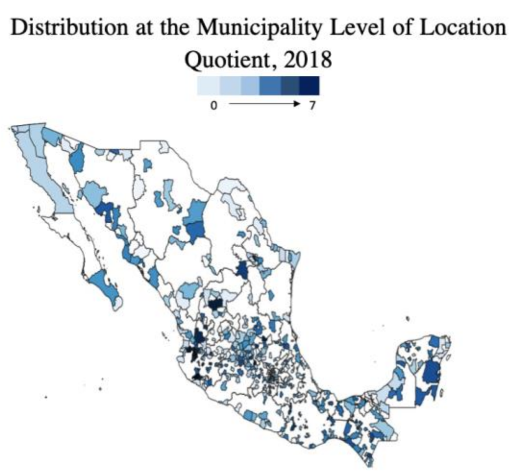
C14-steel products manufacturing



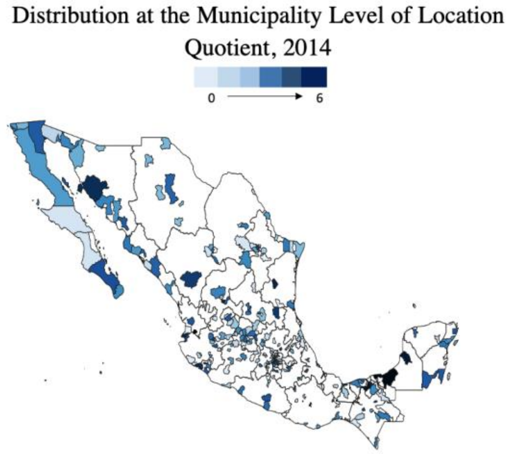


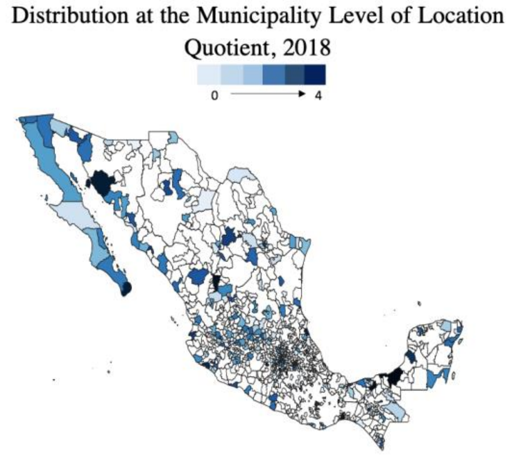
C15-financial services and head offices



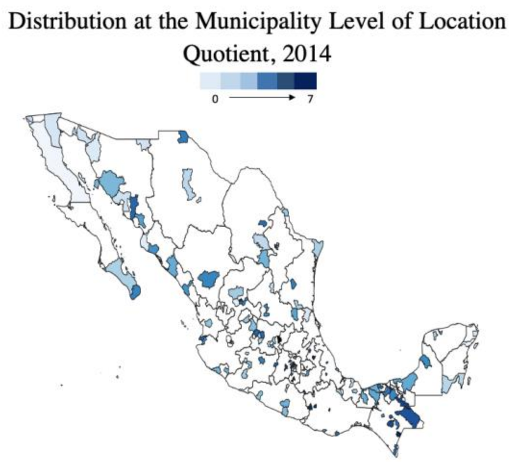


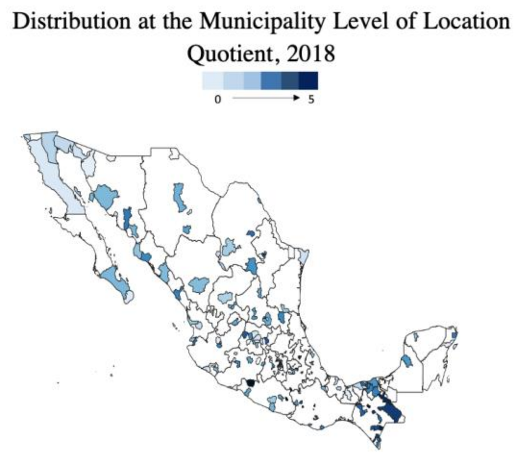
C16-electric power generation, transmission and distribution and infrast. construction



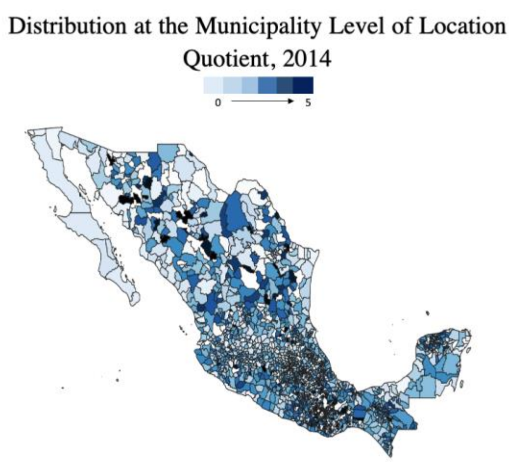


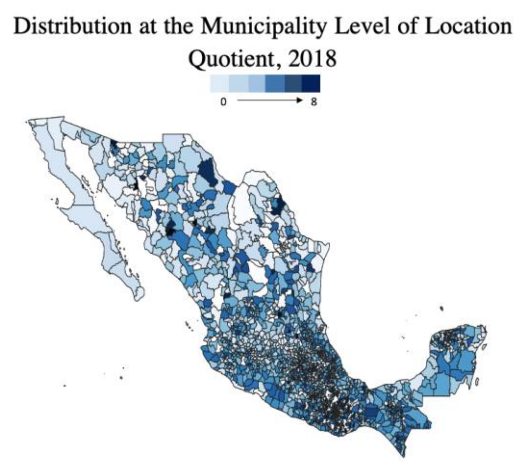
C17-passenger transportation and communications



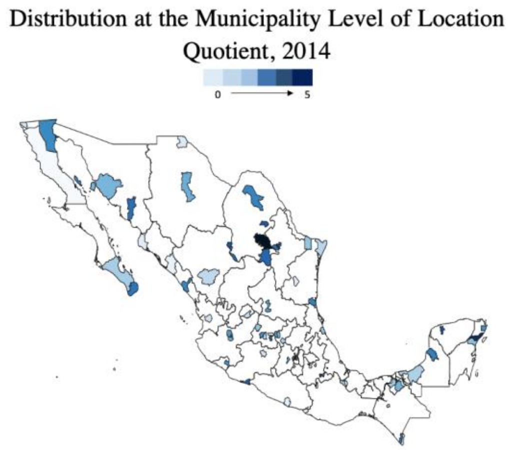


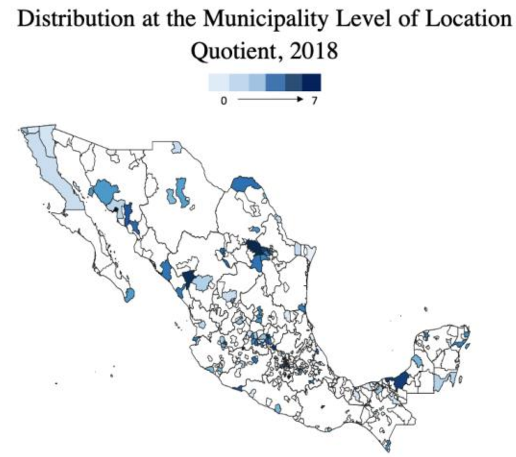
C18-retail and eating services



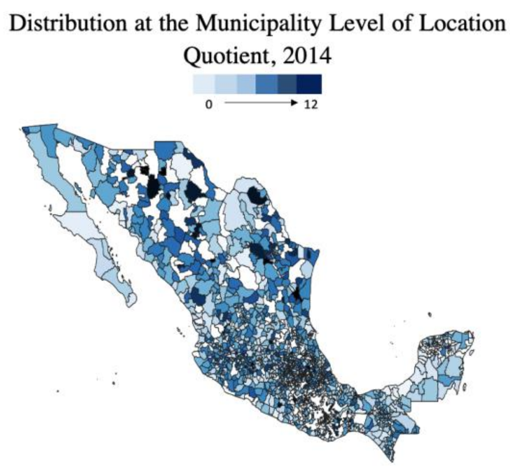


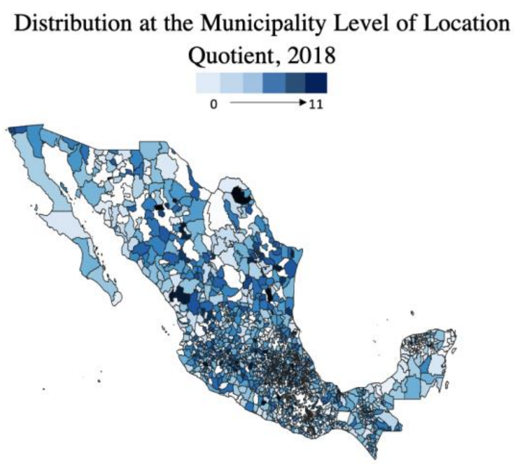
C19-employment services



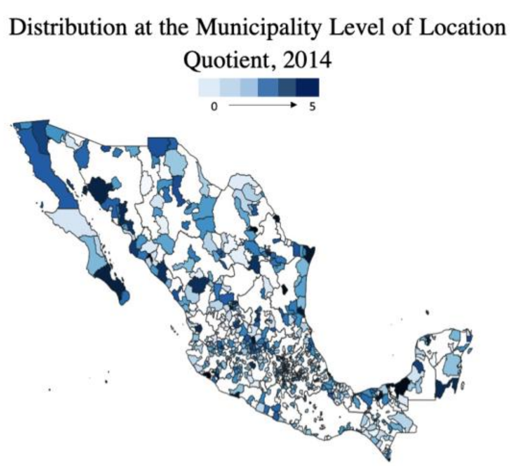


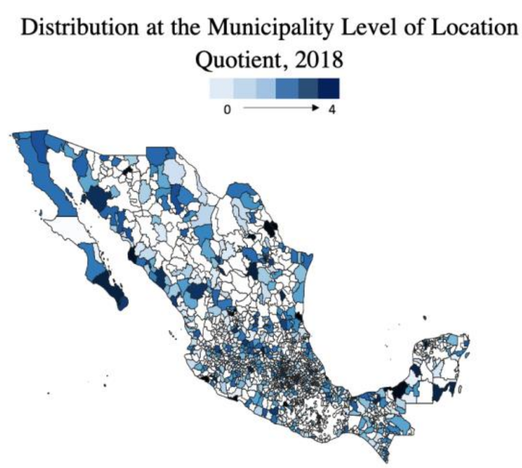
C20-plastic products manufacturing



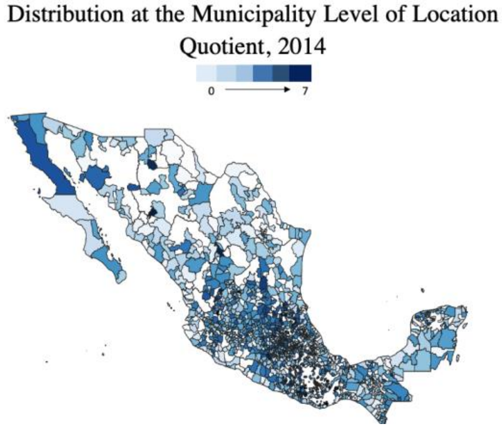


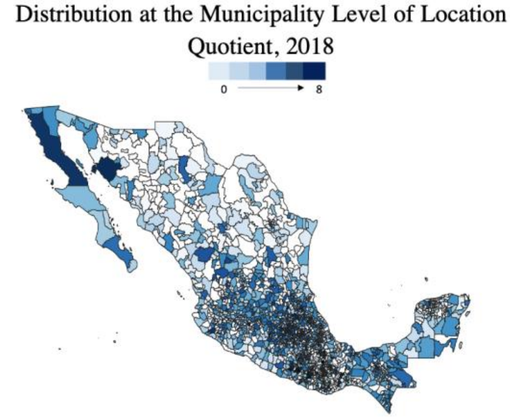
C21-freight transportation services and residential and nonresidential construction



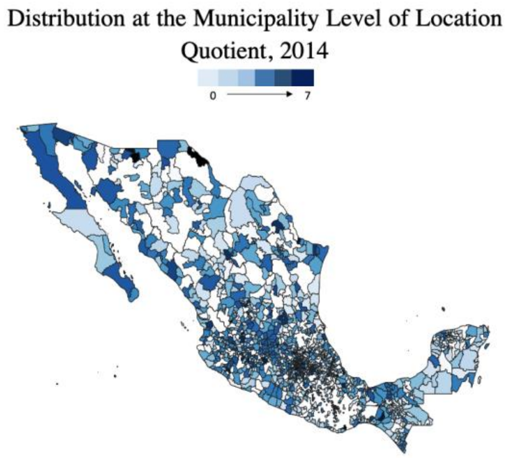


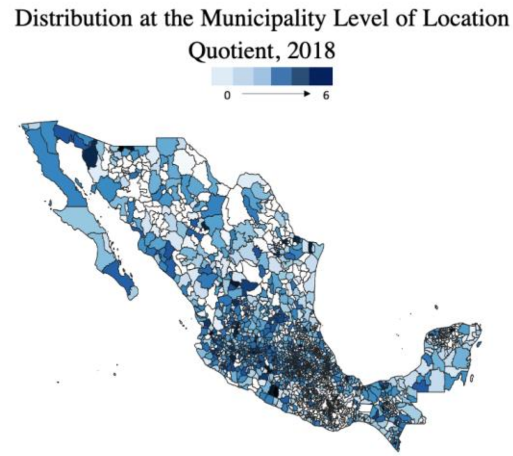
C22-business support services



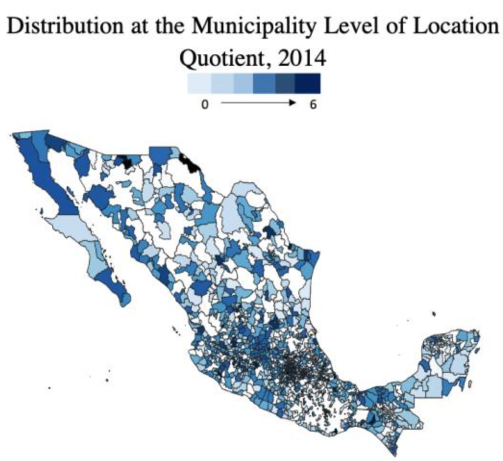


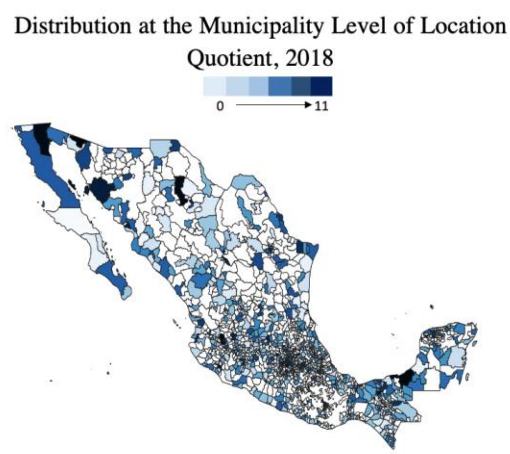
C23-education and health services





C24-pharmaceutical and medical manufacturing and services

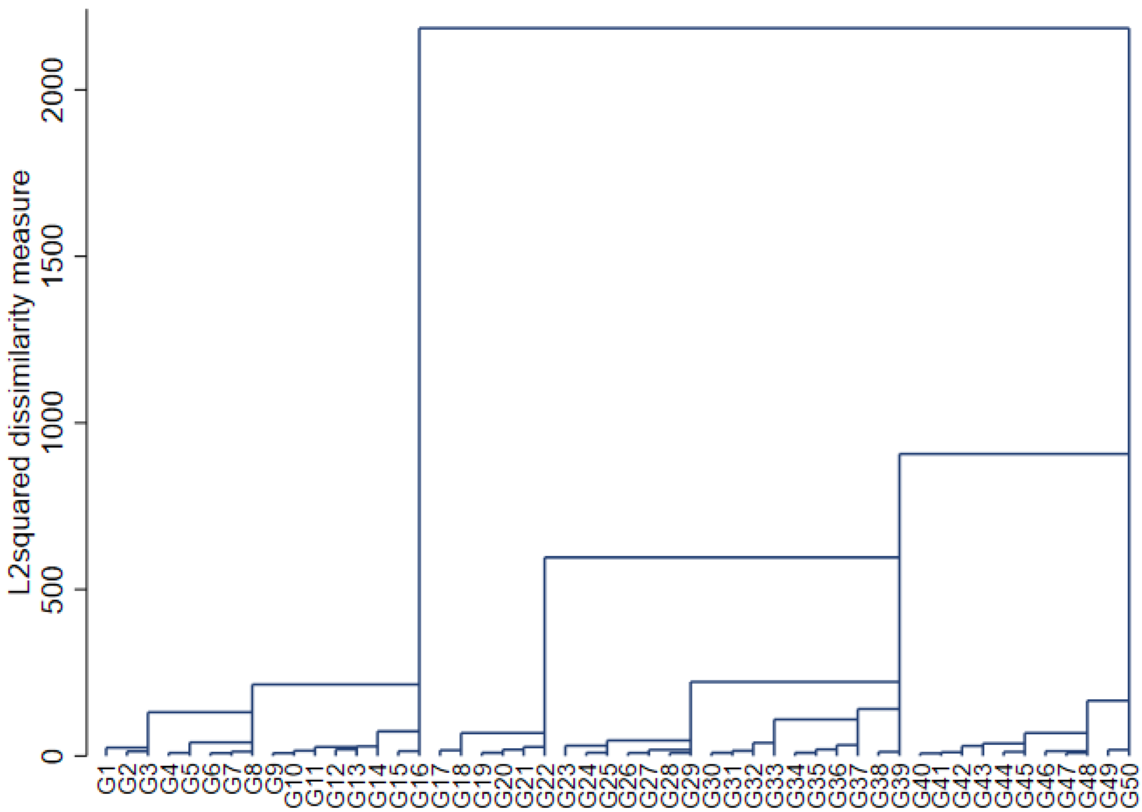




Annex B.4

Dendrogram

Dendrogram is a tree diagram that shows the groups that are formed by creating clusters of observations at each step and their levels of similarity. The level of similarity is measured on the vertical axis (alternatively, the level of distance can be displayed) and the measure of the dissimilarity (similarity) of the units from the observations are specified on the horizontal axis.



Source: Authors' own calculation based on data from INEGI.