

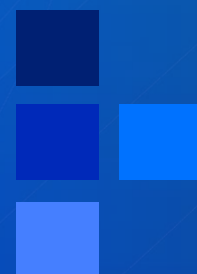
Nº 37

ENERO DE 2020

DOCUMENTOS

DEPARTAMENTO DE ECONOMÍA

UNIVERSIDAD DEL NORTE



El problema del análisis clúster en presencia de outliers: El caso de Ciénaga - Magdalena - Colombia

Nestor Garza
César Hernández Moreno

El problema del análisis clúster en presencia de outliers: El caso de Ciénaga - Magdalena - Colombia

Nestor Garza¹
César Hernández Moreno²

¹ Assistant Professor at California State University Dominguez Hills

² Egresado de economía, Universidad del Norte

Serie Documentos, 37

Enero de 2020

La serie *Documentos* del Departamento de Economía de la Universidad del Norte circula con el fin de difundir y promover las investigaciones realizadas en Uninorte, y también aquellas resultado de la colaboración con académicos e investigadores vinculados a otras instituciones. Los artículos no han sido evaluados por pares, ni están sujetos a ningún tipo de evaluación formal por parte del equipo editorial. Actualmente la serie cuenta con 36 números publicados a los cuales se puede acceder a través de la página web de la Universidad del Norte, específicamente a través del enlace:

<https://www.uninorte.edu.co/web/instituto-de-estudioseconomicos-del-caribe-ieec/publicaciones>.

Se autoriza la reproducción parcial de su contenido siempre y cuando se cite la fuente, y se solicite autorización a sus autores. Los conceptos expresados son de responsabilidad exclusiva de sus autores, y no representan la visión de la Universidad del Norte.

Comité editorial

Adolfo Meisel Roca, PhD.

Alexander Villarraga Orjuela, PhD.

Andrés Vargas Pérez, PhD.

Carlos Yanes Guerra, Mag.



Vigilada Mineducación

Universidad del Norte

Instituto de Estudios Económicos del Caribe (IEEC)

Apartado aéreo 1569

Barranquilla, Colombia

RESUMEN

Este artículo analiza la distribución espacial de establecimientos comerciales e industriales en Ciénaga, Magdalena. Apreciamos que la presencia de dos grandes empresas agro-industriales en la periferia del municipio, tiene un efecto distorsionador sobre análisis clúster tradicionales, del tipo LISA y Moran's Index. Nosotros usamos regresión cuantílica espacial en dos etapas (2SQR), para evitar el efecto distorsionador de los outliers. Los resultados de regresión nos permiten rechazar la existencia de un proceso de desarrollo productivo tipo clúster en el municipio.

Palabras clave: Regresión Espacial, Regresión Cuantílica, Ciénaga, Magdalena, Desarrollo Urbano.

Introducción

Durante los últimos años se ha popularizado el uso de técnicas de estadística o econometría espacial para detectar clústeres empresariales. Estos a su vez se han convertido en herramientas de política de desarrollo territorial, aunque la reflexión teórica de geografía económica que subyace a la idea de clústeres empresariales no ha desarrollado suficientemente su respectiva representación empírica (Martin & Sunley, 2011; Castells & Hall, 1994).

La conceptualización de clústeres se ha alimentado de contribuciones de la historia empresarial, la teoría de la organización y la ecología. Sin embargo, permanece abierta la cuestión de la aproximación empírica y las implicaciones de política. En dicho frente, y a pesar de la sofisticación relativa de las técnicas de econometría espacial, la conexión teórica aún es problemática. En particular, para poder adjudicar características espaciales a relaciones económicas estructurales (como la relación capital/trabajo, las ventas entre empresas o el compartir mercados de trabajo), se requiere definir de manera más precisa las herramientas de prueba (Maya, 2011; Simboli et al., 2012).

En el presente artículo se trabaja el municipio de Ciénaga, Magdalena, como un estudio de caso. El municipio ofrece un caso interesante al estar localizado en un área de influencia de actividades agroindustriales de gran escala, las cuales tienen un impacto territorial importante. Al mismo tiempo Ciénaga es un importante nodo regional de servicios y población que ha permanecido relativamente poco referenciado en la literatura sobre desarrollo territorial (Meisel, 2004).

Se usa una base de datos geolocalizada de 1.292 empresas, la cual incluye información sobre activos, empleo y variables de control relativas al sector económico y organización jurídica. Este tipo de trabajo no se ha realizado en Colombia para municipios como Ciénaga, nodos regionales, pero no de nivel departamental. Ahora bien, observamos una evidente correlación espacial univariada y bivariada entre empresas en términos de activos y empleo al usar

indicadores de clusterización global (I. de Moran) e indicadores locales de asociación espacial (LISA, por sus siglas en inglés), pero dicha correlación desaparece al eliminar del análisis dos grandes empresas agroindustriales (*outliers* en términos de sus dimensiones) localizadas en la periferia oriental del municipio.

Se usa regresión espacial cuantílica (*Two-Stage Quantile Regression*) ya que puede enfrentar el problema de distribuciones estadísticas fuertemente sesgadas (Koenker, 2005; Liao & Wang, 2012; Garza & Ovalle, 2019). Esta aproximación tiene en cuenta al tiempo la clusterización (que en el presente caso es del tipo SEM: *Spatial Error Model*) y la no-linealidad de la relación entre variables. El ejercicio permite rechazar la hipótesis de clústeres empresariales que existen en Ciénaga, Magdalena, con y sin las observaciones clasificadas como *outliers*. También resiste varias especificaciones y variables de control, y permanece no significativo tanto para cuantiles bajos como altos de la distribución (se rechaza la no-linealidad de la relación).

El documento está estructurado en seis secciones, la primera es esta introducción. En la segunda se presentan algunos problemas teóricos y empíricos asociados a las investigaciones tipo clúster. La tercera sección presenta las técnicas de regresión espacial, cuantílica y espacial cuantílica, incluyendo sus ventajas en presencia de observaciones extremas (*outliers*). El estudio de caso, sus particularidades territoriales y la base de datos para el análisis se presentan en la sección cuatro; y en la sección cinco se comparan los resultados derivados de la estadística espacial y de la regresión cuantil espacial, destacando las ventajas de esta última. Las conclusiones se exponen en la sexta y última sección.

1. Análisis clúster, ejemplos y problemáticas

1.1. El desafío teórico

La clusterización como fenómeno espacial es un desarrollo conceptual propio de la geografía económica, remontándose a las obras clásicas de autores como Losch, Isard y Chrystaller. Versiones recientes del tipo nueva geografía económica, que han

contribuido a una mayor precisión teórica en sus microfundamentos y su dinámica, no han aportado mucho a su entendimiento empírico (Martin & Sunley, 2011).

En la lógica de la geografía económica los procesos de “clusterización” son (o deberían ser) un resultado natural de la consolidación de economías capitalistas a nivel local y regional. En consecuencia, a pesar del poder intelectual del concepto *ventajas de aglomeración*, estas no son necesariamente un objeto factible de política económica.

Las lógicas voluntaristas que subyacen a la mayor parte de estudios, evaluaciones y recomendaciones sobre clústeres empresariales como determinantes de innovaciones, no reconocen su carácter aleatorio ni el sistema societal dentro del cual operan. En trabajos fundacionales sobre el tema se reconocían estas dificultades, y aunque se exponía la posibilidad de un desarrollo dirigido tipo clúster, no se veía factible tal ingeniería social en el estado de conocimiento que había en el momento sobre el tema (Castells & Hall, 1994).

El problema teórico es más serio cuando se consideran clústeres en el ámbito agrícola o agroindustrial, actividades que por definición se caracterizan por densidades relativamente bajas, mientras que la clusterización se refiere a la cercanía geográfica. Al respecto, aunque el efecto de insumos compartidos es evidente en el ámbito rural (por ejemplo, infraestructuras de transporte e irrigación), su efecto sobre la incertidumbre (fuerza económica fundamental para la aglomeración) es mínimo.

La agroindustria –especialmente la de gran escala– tiene, no obstante, un efecto espacial considerable, no solamente en su entorno rural, sino en la industria y comercio del núcleo urbano al cual se asocia. Como lo detecta Das (2011), el empleo no agrícola en las zonas rurales es el gran olvidado en los estudios y esfuerzos por consolidar clústeres agroindustriales. De hecho, estas actividades no son

industriales ni agrícolas y, sin embargo, representan la mayor parte del empleo en municipios predominantemente agrícolas.

En el caso colombiano en general, y en el Caribe en particular, son evidentes los efectos espaciales y socioeconómicos de la agroindustria de gran escala. Aunque dichos efectos son difíciles de discernir en la zona rural debido a las características ya explicadas de las actividades agrícolas, su efecto en la cabecera es un estudio de caso relevante.

En los estudios de nivel nacional las economías de aglomeración se analizan tradicionalmente en el sector manufacturero, porque cuenta con información desagregada por sectores y ciudades. En el presente trabajo nuestro énfasis en la cabecera municipal nos permite extendernos a otros sectores, especialmente el comercial que representa un papel crucial en los sistemas locales de abasto y distribución.

1.2. El problema empírico

Detectar asociaciones espaciales empresariales no es una tarea sencilla, especialmente en entornos con carencias estructurales de información, como es el caso colombiano. Al mismo tiempo, la mera detección de una asociación espacial no nos informa sobre los mecanismos por medio de los cuales la cercanía afecta el desempeño económico, la variable que en últimas deseamos entender y eventualmente modificar.

Por ejemplo, Maya (2011) muestra que, aunque en Sinaloa haya proximidad geográfica de proveedores de productos agrícolas para los Estados Unidos, estos no cumplen los criterios económicos requeridos en un sistema productivo tipo clúster. El autor denomina a esta situación “territorio desincrustado”. En el mismo orden de ideas, Simboli et al. (2012) razonan que, aunque a veces los distritos industriales se

teorizan desde los sofisticados marcos conceptuales de la ecología, sus respectivas pruebas empíricas son característicamente geografía económica.

Estos problemas se han tenido en cuenta en la literatura al realizar análisis clúster que sobrepasan la pura estructura espacial autorregresiva y usar más variables de control; es decir, que no solo se basan en un análisis de dependencia simple o generalizado, sino que hay muchos efectos espaciales de carácter local que pueden ser incorporados al estudio y que pueden detectarse, estimarse. Así mismo, se ha tenido en cuenta el problema de endogeneidad por la cercanía espacial de unidades productivas, su sector de especialización y su tamaño relativo, tal como se observa en Arbia, et al. (2012); Duranton & Overman (2005); y Marcon & Puech (2010).

El presente trabajo hace énfasis en las implicaciones que surgen de los análisis empíricos tipo clúster en un caso con grandes diferencias entre las observaciones en términos de tamaño relativo por activos o empleo. El municipio de Ciénaga presenta características ideales para la detección de *outliers* representados por grandes empresas agroindustriales. Adicionalmente, estudiamos un territorio subinvestigado y escasamente referenciado en la literatura académica.

2. Regresión espacial cuantílica

2.1. Regresión cuantílica

La regresión cuantílica ha sido desarrollada para evitar los problemas de no-linealidad de los mínimos cuadrados ordinarios (MCO). Esto porque mientras MCO trata de estimar la línea de regresión más cercana a la media de los datos, la regresión cuantílica acerca a cualquier cuantil de una distribución, lo cual es conveniente cuando la distribución de los datos está fuertemente sesgada (Koenker, 2005).

La estimación cuantílica asume que un conjunto de datos ($\mathbf{Y}$), tiene una función de distribución conocida $f(\mathbf{y}) = (\mathbf{y}_i - \bar{\mathbf{y}})$, donde cada realización $\mathbf{y}_i$ tiene una posición relativa al cuantil objetivo que se denota por τ ($0 < \tau < 1$) y que para el caso de la

mediana de la distribución este sea $\tau = 0.5$. El error relativo entre las observaciones y el cuantil objetivo se puede obtener minimizando las diferencias en el rango de toda la distribución:

$$\min(f(Y - \bar{y})) = \min(y_i - \bar{y}) \left[\int_{-\infty}^{\bar{y}} (y_i - \bar{y}) dY + \int_{\bar{y}}^{\infty} (y_i - \bar{y}) dY \right] \quad (1)$$

Una versión discreta, que se puede resolver usando un programa lineal (Simplex), se representa en (2) para la mediana de la distribución:

$$\min(f(y_i - \bar{y})) = \tau \left(\sum_{y_i < \bar{y}} (y_i - \bar{y}) + \sum_{y_i \geq \bar{y}} (y_i - \bar{y}) \right) \quad (2)$$

Una de las más interesantes características de la expresión (2) es que podemos modificar el cuantil objetivo τ , y encontrar el valor que minimiza los errores para posiciones alternativas en la distribución (por ejemplo, los quintiles más bajo o alto, respectivamente). El respectivo programa lineal Simplex que resuelve la expresión (2) tiene la forma:

$$Q_n(\tau) = \min \left\{ \sum_i [\tau - 1((y_i - \bar{y}) < 0)] (y_i - \bar{y}) \right\} \quad (3)$$

en el cual el valor que minimiza la diferencia con la distribución teórica para cada cuantil Q_i es ponderado negativamente cuando este se encuentra por debajo del cuantil objetivo $[\tau - 1((y_i - \bar{y}) < 0)]$. Esta es la mayor fortaleza de la regresión cuantílica, la posibilidad de comparar la distribución total de observaciones contra observaciones predeterminadas consideradas el objetivo de la estimación, y de esta

forma detectar diferencias en el parámetro de interés estimado en presencia de observaciones extremas o definidas como *outliers* (Koenker, 2005).

Puesto que la regresión cuantílica no usa álgebra matricial, no está sujeta a los limitantes supuestos de normalidad y no heterocedasticidad de la regresión MCO, y por definición considera que la relación entre las variables dependiente e independiente cambia para cada cuantil τ : $y = X\beta(\tau) + u$. El caso de Ciénaga, con grandes empresas agroindustriales en el borde urbano del municipio, es ideal para observar en qué medida estas influyen en la observación de patrones clúster y relaciones espaciales. Las observaciones *outliers* representan un cuantil extremo.

2.2. Regresión espacial

Según la primera ley de la geografía de Tobler, observaciones espaciales cercanas se parecen, pero su similitud disminuye con la distancia. En regresión espacial una variable se asocia con su versión espacialmente rezagada, usando la denominada matriz de pesos espaciales (W) en la que las relaciones w_i entre todas las observaciones i se determinan según la distancia inversa: en la versión estandarizada la relación uno (1) representa la máxima distancia factible entre observaciones, mientras que cero (0) es la distancia mínima; este caso se da en la diagonal donde relacionamos cada observación consigo misma (La Sage & Pace, 2009).

Existen dos especificaciones básicas en regresión espacial: el modelo espacial autorregresivo (SAR: *Spatial Autoregressive*): $y_i = \beta_1 + \beta_2 w_i y_i + u_i$; y el modelo de error espacial (SEM: *Spatial Error Model*): $y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i e_i + u_i$. En el primer caso el error o residuo (u_i) cumple con la condición de ser independiente e idénticamente distribuido (iid), mientras que en el segundo caso los errores (e_i) de una regresión MCO entre las variables (y) y (x) no cumplen con tal condición.

Por definición, tanto el SAR como el SEM no cumplen la propiedad de no autocorrelación, por lo que su estimación por MCO no es eficiente y requieren un

estimador de máxima verosimilitud completa (FML: *Full Maximum Likelihood*). Esta particularidad se puede revelar como una complicación al utilizar estimación cuantílica para determinar los parámetros espaciales (SAR y SEM), ya que esta técnica no usa FML, sino que resuelve un problema de optimización lineal Simplex. Podemos, sin embargo, agregar que a diferencia de la regresión MCO que es paramétrica, la regresión cuantil es no-lineal y se acerca a la verdadera distribución de los datos como lo hace la FML. Esto hace que la diferencia entre las dos estimaciones se haga progresivamente insignificante en muestras grandes.

2.3. Regresión espacial cuantílica

La regresión cuantílica para datos espaciales no requiere hipótesis previas sobre las perturbaciones aleatorias en la relación entre regresores y la variable endógena, por lo que la hace semiparamétrica. Su conveniencia en este caso alude a que genera buenos resultados cuando se dispone de una gran cantidad de datos de corte transversal. Una ventaja estructural de este tipo de estimación frente a MCO es que cuando tenemos elementos atípicos en la muestra (*outliers*) en mínimos cuadrados las observaciones intervienen en igual medida y los puntos atípicos del plano medio tiran de este, ya que el objetivo sería minimizar la suma de los residuos cuadrados; esto dificulta un análisis segmentado. En caso contrario, los puntos extraños en la estimación cuantílica (mediana por ejemplo) no cambiarían la solución. Otra ventaja es que muestra el comportamiento de los parámetros según cambia de cuantil; es como analizar la relación de las variables para diferentes valores o tamaños de la endógena estimada.

Existen dos formas de trabajar conjuntamente la regresión espacial y la regresión cuantílica: usar regresión en dos etapas 2SQR (*Two-Stages Quantile Regression*), o usar

variables instrumentales IVQR (*Instrumental Variables Quantile Regression*) (Liao & Wang, 2012). La expresión (4) representa estas estrategias en un SAR:

$$\mathbf{y} = \lambda_{\tau} \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta}_{\tau} + \mathbf{u}_{\tau} \quad (4)$$

En el SAR (4), los parámetros λ_{τ} y $\boldsymbol{\beta}_{\tau}$ son matrices de tamaño $n \cdot k$ variables $\mathbf{y}$ con $\mathbf{u}_{\tau} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$. En el caso 2SQR es necesario estimar primero $\mathbf{W} \mathbf{y}$ como una función de las $\mathbf{X}$ variables independientes, además de sus respectivos rezagos espaciales $\mathbf{W} \mathbf{X}$, y después obtener el λ_q vector de parámetros que asocia $\mathbf{y}$ con esa $\widehat{\mathbf{W} \mathbf{y}}$ estimada. El caso IVQR usa como variable dependiente los residuos de la regresión espacial no cuantil, mientras que las variables independientes se introducen instrumentadas³: $(\mathbf{y} - \lambda \mathbf{W} \mathbf{y}) = \mathbf{X} \boldsymbol{\beta}_{\tau} + \mathbf{u}_{\tau}$. En el presente trabajo, puesto que no tenemos muchas variables para escoger instrumentos, nos decantamos por la estrategia 2SQR, siguiendo la lógica de Zhang & Wang (2016).

3. Caso de estudio e información disponible

3.1. Caso de estudio

Ciénaga es un municipio de Colombia en el departamento del Magdalena, localizado a orillas del mar Caribe en el extremo nororiental de la Ciénaga Grande de Santa Marta. La población se encuentra a una altitud de 3 msnm promedio en la zona urbana. Dista 35 km de Santa Marta (capital del departamento) y tiene un área aproximada de 1.242,68 km². El municipio tiene 104.319 habitantes de los cuales 98.694 se ubican en la zona urbana. Esta última observación coincide con lo

³ La regresión espacial contemporánea usa solamente FML, pero la econometría espacial original trabajaba con el estimador de variables instrumentales, particularmente por la baja capacidad de los computadores de la época (Anselin, 1988; La Sage & Pace, 2009). En este sentido, la IVQR retorna al problema de estimación espacial original, aunque la razón no sea la falta de capacidad de los computadores actuales.

referenciado en nuestra sección conceptual, en la que se expresa la gran importancia del empleo no agrícola en una región fuerte en actividades agroindustriales.

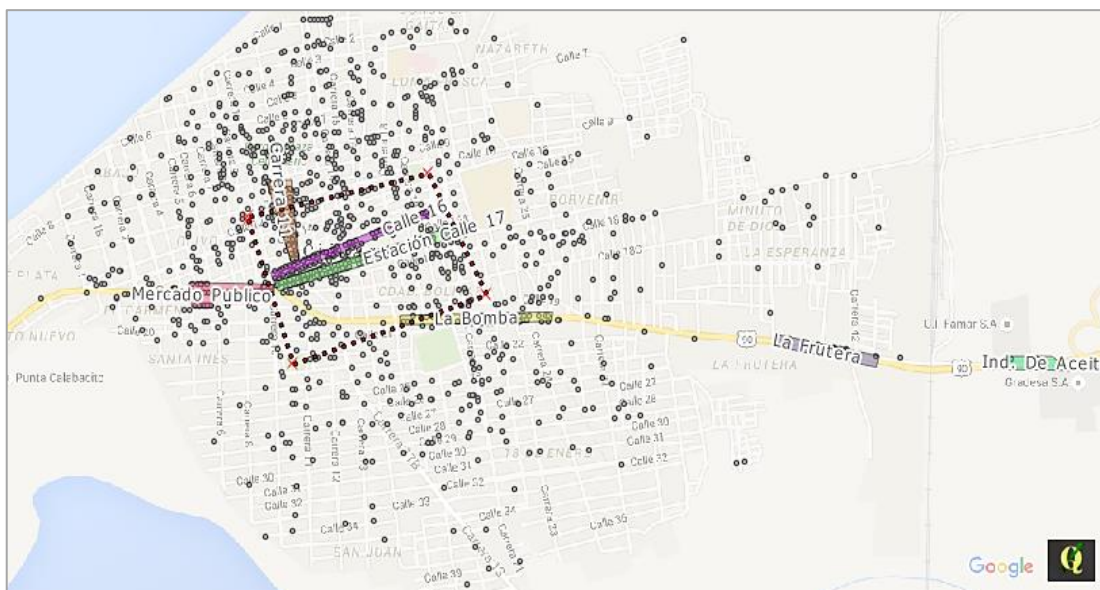
Desde finales del siglo XIX el cultivo y exportación de banano han influenciado la economía de Ciénaga, haciéndola una de las más dinámicas de la Región Caribe colombiana. En el periodo 1966 - 1993 la industria ha tenido muchos altibajos debido al retiro de la United Fruit Company en 1965. En 1990 el municipio se convirtió en puerto marítimo para la exportación carbonífera de la multinacional Drummond, lo que ha implicado la llegada de recursos adicionales por regalías (Meisel, 2004).

En consonancia con los requisitos del caso de estudio, la actividad agrícola es la que tiene mayor participación en el territorio municipal, con el 93% de sus 122.160 hectáreas. Al mismo tiempo, no obstante, la actividad comercial equivale al 52% del ingreso municipal. El sector industrial lo comprenden mayormente dos empresas localizadas en las afueras de la cabecera. Estas dos empresas tienen un efecto importante tanto en el paisaje periurbano como en la economía de la ciudad. Dichas empresas se especializan en la elaboración de aceites vegetales, margarinas y derivados, cuyos productos se corresponden con la agroindustria de la palma aceitera, la cual ha tenido efectos socioespaciales notorios en el departamento del Magdalena (Rey, 2013), y especialmente en los municipios cercanos a Ciénaga (Goebertus, 2008).

En el mapa 1 presentamos una aproximación a la estructura urbana del municipio. Este mapa ha sido producido usando el *software* libre Quantum-GIS, geolocalizando las observaciones por dirección a partir de una capa ráster producida desde GoogleMaps. En el mapa se aprecia un trazado donde la carretera que une a Santa Marta, Barranquilla y Bogotá, determina la estructura urbana, esperable en un municipio relativamente pequeño y orientado al comercio. Sin embargo, el centro histórico de la ciudad se encuentra fuera de esta área de influencia, reflejando la estructura económica pasada, con una economía más endógena y especializada en el banano. El ordenamiento territorial del municipio condesciende con esta aspiración territorial al determinar que el centro tradicional, de trazado modernista y arquitectura fusión, constituye el núcleo urbanístico del municipio, mientras que

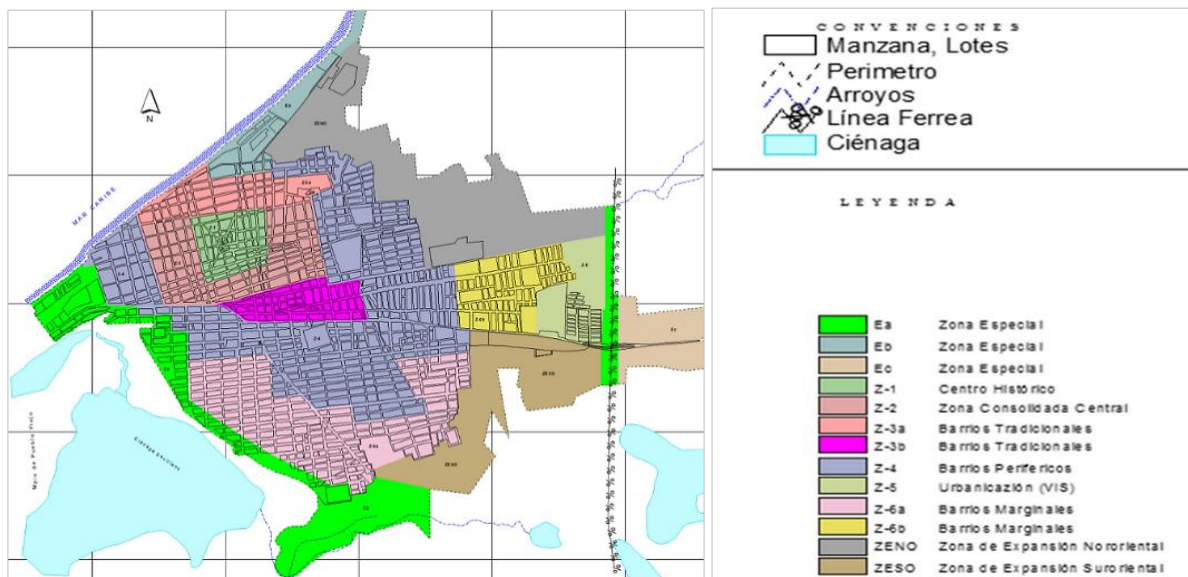
los alrededores de la carretera principal no tienen consideraciones especiales de ordenamiento. Lo anterior puede apreciarse en los mapas 2 y 3. No es lugar aquí para discutir los determinantes de dichas decisiones de planeación urbanística, pero podemos especular que se trata de un refuerzo a la identidad local en contraposición

las fuerzas dominantes de Barranquilla, Santa Marta e incluso Bogotá, representadas en la carretera central que cruza el casco urbano.



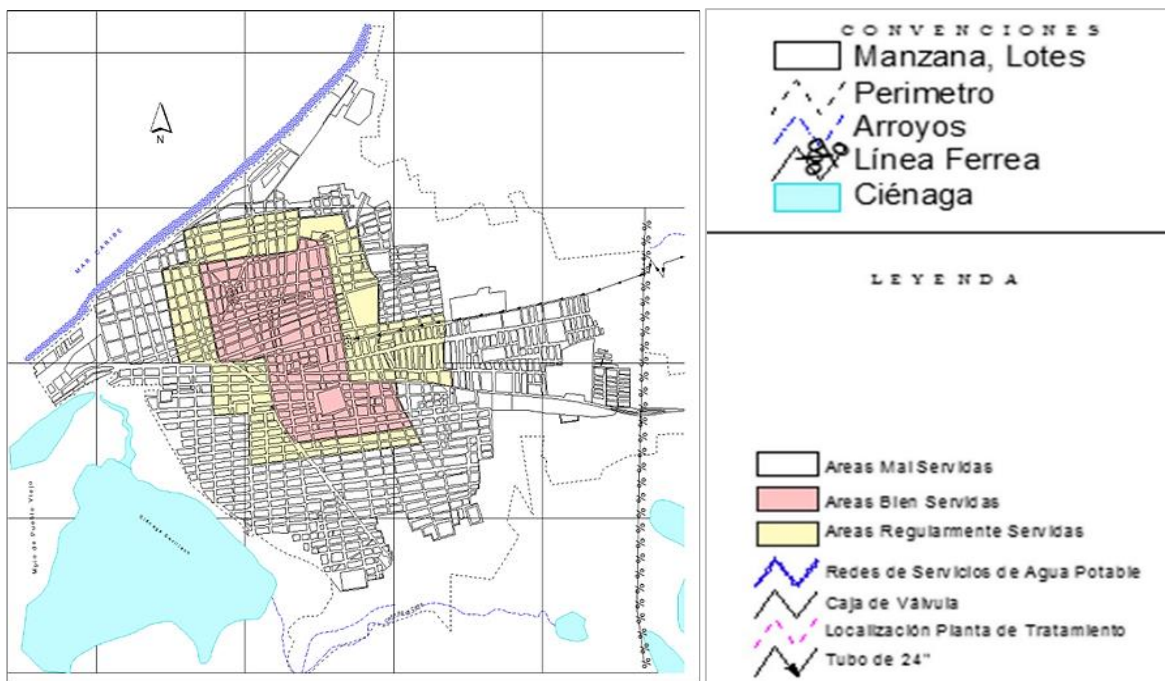
Fuente: elaboración propia, usando un ráster de imagen con GoogleMaps.

Mapa 1. Municipio de Ciénaga (amanzanado, vías principales y observaciones en la base de datos).



Fuente: POT municipal.

Mapa 2. Zonas urbanas según el POT 2001-2010.



Fuente: POT municipal.

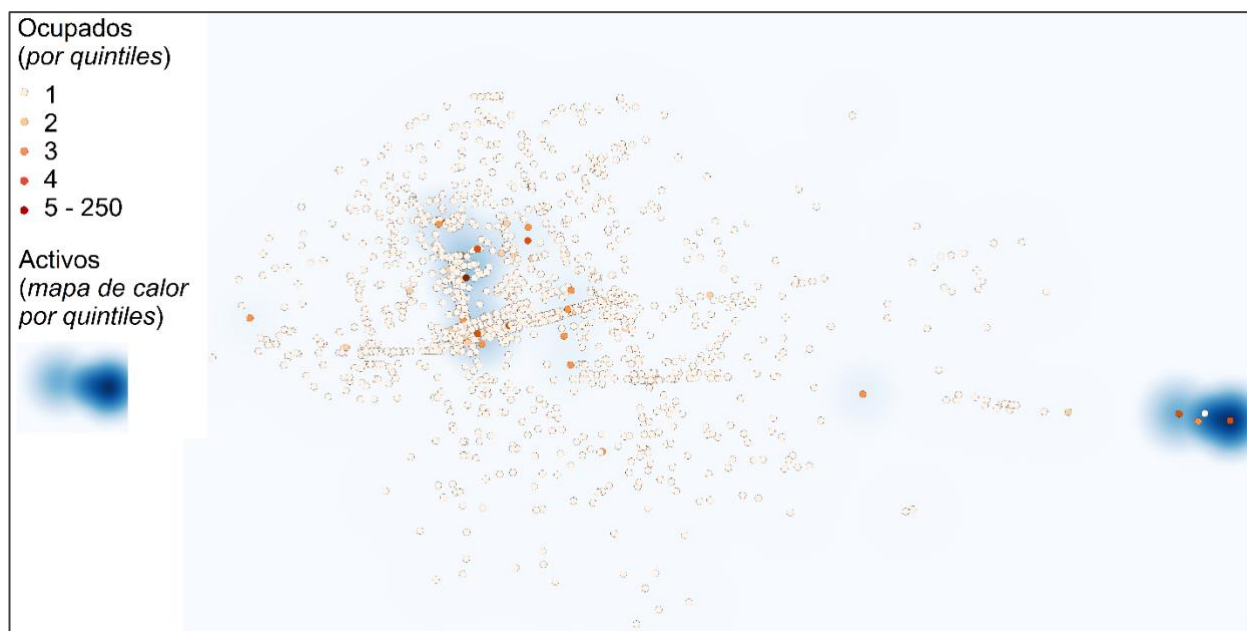
Mapa 3. Prestación de servicios según el POT 2001-2010.

3.2. Fuentes de información y construcción de la base de datos

Para esta investigación hemos obtenido información sobre 1.292 observaciones a nivel de establecimiento que incluyen localización exacta (dirección), activos, empleo, código CIIU y tipo de organización jurídica. Los datos utilizados provienen de la base de datos de la Cámara de Comercio (2012), facilitados por la coordinación de impuestos municipales de la Secretaría de Hacienda del municipio de Ciénaga, e incluyen zonas urbanas y periurbanas del municipio.

Como podemos apreciar en el mapa 4, las dos empresas con el mayor tamaño en activos se encuentran en la periferia oriente de la ciudad, representando notorios *outliers* que pueden imponer grandes distorsiones a cualquier análisis clúster no controlado por esta característica. En el mismo mapa se aprecia que, aunque estas dos empresas también tienen el mayor tamaño en empleo, las diferencias en esta variable no son tan marcadas. Esta es una particularidad coherente con el hecho de

que las actividades agroindustriales de gran escala son por definición intensivas en capital.



Fuente: elaboración propia con base en información de la Cámara de Comercio y GoogleMaps.

Mapa 4. Localización de activos y empleo en 2012 (1.292 observaciones).

Nuestra versión espacial de regresión cuantil evaluará el signo y valor absoluto del parámetro de asociación espacial en diferentes grupos (percentiles) de la distribución de la muestra por tamaños de activos o empleo. En otras palabras,

evaluará la existencia de no-linealidades, especialmente en la zona de la distribución de los datos cercana a los dos *outliers* descritos anteriormente.

4. Resultados

4.1. Resultados clúster

En los estudios clústeres tradicionales se usan indicadores espaciales de asociación espacial tipo Moran, tanto del nivel local como del global. La razón para la popularidad de estas herramientas analíticas es que reflejan de forma sencilla la idea de economías de aglomeración. En el presente documento se hace un barrido con la información disponible en la intención de detectar clústeres globales (nivel municipal) y locales (nivel barrial) usando tanto el Índice de Moran como los índices locales de asociación espacial (LISA), tal y como los calcula el *software* libre GEODA. Es esta una de las más populares estrategias de investigación observadas en la literatura, con abundantes ejemplos en el caso colombiano (Galvis, L. & Meisel, A., 2010; Garza, et al., 2009).

En el caso del presente artículo, detectamos estas asociaciones simples para después proceder a probar su verosimilitud usando regresión espacial cuantílica. Las dos variables de mayor interés de las que tenemos información son activos y ocupados, las cuales por definición están asociadas teóricamente, como se discutió arriba. La tabla 1 resume todos los resultados de asociación espacial global y local de las variables de interés, así como en relación con otras variables de interés obtenidas en la base de datos de la investigación.

La tabla 1 muestra que hay correlación espacial tanto univariada como bivariada. Con respecto al análisis univariado la autocorrelación es positiva en todos los casos para las variables de interés: activos y ocupados. En el caso de las variables *dummy*: tipo de organización jurídica y sector de actividad económica, encontramos que, tanto para los ocupados como para los activos existe correlación negativa con la

empresa unipersonal. Esto es debido a su abundancia en la base de datos, lo que coincide con clústeres *Low-High* de esta variable en la tabla 1.

Cuando se usó el tipo de organización colectiva no hubo en general asociación espacial significativa, incluso en los sectores financiero y de servicios. En este caso pensamos que es el escaso número de observaciones el que produce dicho resultado. Cuando se usan los sectores agropecuario, industrial y transporte tienen correlación positiva, mientras que los sectores comercial y otros tienen correlación negativa. Como ya sabemos que esta evaluación de clústeres está expuesta al sesgo de los *outliers*, a continuación repetimos los mismos análisis eliminando las dos observaciones extremas. Los resultados se presentan en la tabla 2.

La primera y más importante diferencia que se puede apreciar entre las tablas 1 y 2, es que la asociación espacial entre activos y ocupados desaparece. Así mismo, no hay asociación con la empresa accionaria y colectiva respectivamente. En general, comparados con los resultados de la tabla 1, hay menos casos de correlación espacial en la tabla 2, incluyendo sectores que previamente la reportaban, como agropecuario, industrial y construcción. En otras palabras, el mero análisis de clusterización es vulnerable a observaciones extremas, y no es una forma confiable de abordar esta clase de investigación. La regresión espacial cuantílica es una mejor

herramienta, ya que tiene en cuenta simultáneamente la distribución estadística y espacial de las observaciones.

ANÁLISIS DEL ÍNDICE DE MORAN GLOBAL					ANÁLISIS DEL ÍNDICE DE MORAN LOCAL					
MORAN UNIVARIADO					MORAN UNIVARIADO (Filtro de Significancia 0,05)					
VARIABLES DE INTERES (X); (Y) SU REZAGO		I. DE MORAN	P-VALOR	CORRELACIÓN ESPACIAL P<0,05	VARIABLES DE INTERES		HIGH- HIGH	LOW- LOW	LOW- HIGH	HIGH- LOW
OCUPADOS		0,1643	0,001	S+	OCUPADOS		3	218	2	2
ACTIVOS		0,2301	0,001	S+	ACTIVOS		7	121	18	23
MORAN BIVARIADO					MORAN BIVARIADO (Filtro de Significancia 0,05)					
(X)	(Y) REZAGO	I. DE MORAN	P-VALOR	CORRELACIÓN ESPACIAL P<0,05	(X)	(Y) REZAGO	HIGH- HIGH	LOW- LOW	LOW- HIGH	HIGH- LOW
OCUPADOS	ACTIVOS	0,1952	0,001	S+	OCUPADOS	ACTIVOS	5	179	0	41
"	UNIPERSONAL	-0,0595	0,001	S-	"	UNIPERSONAL	15	272	137	108
"	LTDA	-0,0015	0,301	N	"	LTDA	108	47	267	29
"	ACCIONARIA	0,0872	0,001	S+	"	ACCIONARIA	4	24	31	10
"	COLECTIVA	-0,0017	0,251	N	"	COLECTIVA	47	11	128	76
"	AGROPECUARIO	0,0553	0,001	S	"	AGROPECUARIO	3	126	47	71
"	INDUSTRIAL	0,0388	0,001	S+	"	INDUSTRIAL	15	16	81	11
"	COMERCIAL	-0,0456	0,001	S-	"	COMERCIAL	125	218	289	62
"	FINANCIERO	-0,0022	0,169	N	"	FINANCIERO	53	2	160	51
"	CONSTRUCCION	-0,0007	0,393	N	"	CONSTRUCCION	37	242	146	110
"	SERVICIOS	-0,0016	0,27	N	"	SERVICIOS	84	106	277	52
"	TRANSPORTE	0,0302	0,001	S+	"	TRANSPORTE	25	8	73	11
"	OTRAS	-0,0122	0,001	S-	"	OTRAS	19	369	120	166
ACTIVOS	OCUPADOS	0,1952	0,001	S+	ACTIVOS	OCUPADOS	3	140	23	3
"	UNIPERSONAL	-0,0685	0,001	S-	"	UNIPERSONAL	1	368	152	13
"	LTDA	-0,0057	0,014	S-	"	LTDA	9	55	366	6
"	ACCIONARIA	0,1047	0,001	S+	"	ACCIONARIA	3	28	31	1
"	COLECTIVA	-0,0026	0,085	N	"	COLECTIVA	4	13	184	7
"	AGROPECUARIO	0,0648	0,001	S+	"	AGROPECUARIO	1	177	49	6
"	INDUSTRIAL	0,0382	0,001	S+	"	INDUSTRIAL	2	22	95	0
"	COMERCIAL	-0,0453	0,001	S-	"	COMERCIAL	11	269	407	10
"	FINANCIERO	-0,0028	0,12	N	"	FINANCIERO	5	4	211	7
"	CONSTRUCCION	-0,0107	0,001	S-	"	CONSTRUCCION	8	336	177	13
"	SERVICIOS	-0,0043	0,055	N	"	SERVICIOS	11	134	349	6
"	TRANSPORTE	0,038	0,001	S+	"	TRANSPORTE	0	8	99	2
"	OTRAS	-0,0114	0,003	S-	"	OTRAS	3	518	134	17
999 RANDOMIZATIONS (con empresas grandes)					999 RANDOMIZATIONS (con empresas grandes)					

Nota: Es común considerar significativos los p-valores inferiores a 0.05.

Tabla 1. Índice de Moran global y local (univariado y bivariado), todas las observaciones.

ANALISIS DEL INDICE DE MORAN GLOBAL				
MORAN UNIVARIADO				
VARIABLES DE INTERES (X); (Y) SU REZAGO	I. DE MORAN	P-VALOR	CORRRELACION ESPACIAL P<0,05	
OCUPADOS	-0,0017	0,4	N	
ACTIVOS	0,0026	0,71	N	
MORAN BIVARIADO				
(X)	(Y) REZAGO	I. DE MORAN	P- VALOR	CORRELACION ESPACIAL P<0,05
OCUPADOS	ACTIVOS	0,0021	0,129	N
"	UNIPERSONAL	-0,006	0,021	S-
"	LTDA	0,0079	0,015	S+
"	ACCIONARIA	0,0006	0,322	N
"	COLECTIVA	0,0008	0,328	N
"	AGROPECUARIO	0,0041	0,047	S+
"	INDUSTRIAL	0,0105	0,007	S+
"	COMERCIAL	-0,014	0,007	S-
"	FINANCIERO	0,0014	0,22	N
"	CONSTRUCCION	0,014	0,002	S+
"	SERVICIOS	0,0069	0,021	S+
"	TRANSPORTE	-0,001	0,315	N
"	OTRAS	-0,007	0,003	S-
ACTIVOS	OCUPADOS	0,0013	0,171	N
"	UNIPERSONAL	-0,005	0,032	S-
"	LTDA	0,0055	0,028	S+
"	ACCIONARIA	0,0013	0,26	N
"	COLECTIVA	-1E-04	0,463	N
"	AGROPECUARIO	0,0004	0,436	N
"	INDUSTRIAL	0,0015	0,224	N
"	COMERCIAL	-0,003	0,087	N
"	FINANCIERO	0,0024	0,094	N
"	CONSTRUCCION	0,0028	0,128	N
"	SERVICIOS	0,0075	0,013	S+
"	TRANSPORTE	0,0005	0,371	N
"	OTRAS	-0,006	0,012	S-
999 RANDOMIZATIONS (sin empresas grandes)				

ANALISIS DEL INDICE DE MORAN LOCAL					
MORAN UNIVARIADO (Filtro de Significancia 0,05)					
VARIABLES DE INTERES		HIGH- HIGH	LOW- LOW	LOW- HIGH	HIGH- LOW
OCUPADOS		19	115	40	19
ACTIVOS		19	195	450	3
MORAN BIVARIADO (Filtro de Significancia 0,05)					
(X)	(Y) REZAGO	HIGH- HIGH	LOW- LOW	LOW- HIGH	HIGH- LOW
OCUPADOS	ACTIVOS	142	165	327	32
"	UNIPERSONAL	15	279	124	108
"	LTDA	110	64	264	26
"	ACCIONARIA	5	21	49	11
"	COLECTIVA	32	12	94	74
"	AGROPECUARIO	5	126	34	74
"	INDUSTRIAL	14	10	88	10
"	COMERCIAL	123	228	291	60
"	FINANCIERO	72	3	192	46
"	CONSTRUCCION	39	254	154	106
"	SERVICIOS	85	92	277	53
"	TRANSPORTE	29	8	77	11
"	OTRAS	24	372	126	165
ACTIVOS	OCUPADOS	1	131	57	4
"	UNIPERSONAL	2	375	143	14
"	LTDA	12	77	366	4
"	ACCIONARIA	1	28	51	2
"	COLECTIVA	4	14	126	7
"	AGROPECUARIO	1	182	42	11
"	INDUSTRIAL	2	13	103	2
"	COMERCIAL	17	278	396	10
"	FINANCIERO	11	5	254	6
"	CONSTRUCCION	8	340	181	17
"	SERVICIOS	12	128	351	5
"	TRANSPORTE	3	8	102	3
"	OTRAS	3	513	142	24
999 RANDOMIZATIONS (sin empresas grandes)					

Nota: Es común considerar significativos los p-valores inferiores a 0.05.

Tabla 2. Índice de Moran global y local (univariado y bivariado), excluyendo outliers.

4.2. Resultados de regresión

La tabla 3 nos permite determinar si la especificación SAR o SEM ofrece la mejor calidad de la estimación, con y sin variables de control. Estas fueron escogidas usando un proceso *stepwise* que eligió cuatro variables para sectores y dos para tipo de organización jurídica. Las estimaciones SAR muestran además del efecto total, el

efecto indirecto (*feedback*) hasta la potencia siete (7) de la matriz espacial, en la cual nuevas contribuciones al valor absoluto del parámetro se acercan a cero⁴.

Como podemos ver en la tabla 3, la especificación SAR no es significativa en ninguna de sus estimaciones, aunque su efecto indirecto lo es a pesar de su pequeño tamaño absoluto (no logra cambiar el signo). Cuando se le usó simultáneamente con ocupados o las variables de control, se repitieron los resultados.

	FML1	FML2	FML3	FML4	FML5	FML6	FML7
Constante	21.878 **	14.650 *	12.874 ***	14.274 ***	20.436 **	23.429 ***	11.829 *
SAR Directo	-0.310	-0.074			-0.256	-0.414	0.093
SAR Total	-0.310 ***	-0.074 ***			-0.256 ***	-0.414 ***	0.093 ***
Ocupados		1.904 ***	1.918 ***	1.611 ***		1.919 ***	1.617 ***
SEM			-0.018 ***	-0.013 **	0.006	-0.020 ***	
Tipo Unipersonal				-1.309 ***			-1.323 ***
Tipo Colectiva				-1.124			-1.173 *
Sector Financiero				2.057 ***			2.054 ***
Sector Otras				-0.386 ***			-0.369 ***
Sector Industrial				-0.388 ***			-0.356 ***
Sector Servicios				-0.459 ***			-0.497 ***
Log Likelihood	-2,492.3	-1,903.4	-1,898.7	-1,777.7	-2,491.7	-1,897.8	-1,786.0
Akaike	3.861	2.951	2.944	2.770	3.862	2.944	2.779

Nota: Lectura del P-Valor: *** significativo al 99%; ** significativo al 95; *significativo al 90%.

Tabla 3. Estimaciones FML (Full Maximum Likelihood). Variable dependiente: activos.

En consecuencia, nos decantamos por las especificaciones SEM, las cuales no tienen el problema del efecto indirecto y se comportaron bien en presencia de ocupados y de las variables de control (modelos FML3, FML4 y FML6). La especificación SEM es usada en todos los casos de regresión IVQR de la tabla 4, pero allí pierde significancia en presencia de los controles y cuando nos alejamos de la mediana para estimar un cuantil objetivo superior (90%) o inferior (10%). En otras palabras, no hay

⁴ Este procedimiento es necesario porque en el SAR el valor del parámetro incluye su efecto indirecto; es decir, cómo los vecinos de una observación i afectan su valor después que el valor de la misma los afectó a ellos primero.

evidencia de clusterización entre los activos y los ocupados en un amplio conjunto de especificaciones cuantil objetivo, con y sin controles.

	Median	90%	10%	Median	90%	10%
Constant	12.996 ***	13.607 ***	12.410 ***	14.833 ***	15.036 ***	13.065 ***
Ocupados	1.397 ***	2.597 ***	0.799 ***	1.135 ***	2.124 ***	0.696 ***
SEM	-0.010 **	-0.003	-0.001	-0.004	-0.003	0.002
Tipo Unipersonal				-1.772 ***	-1.325 ***	-0.603 ***
Colectiva				-1.520 ***	-0.906	-0.134
Sector Financiero				0.770 ***	3.208 ***	0.655 **
Otras				-0.328 ***	-0.360 ***	-0.419 ***
Industrial				-0.177 **	-0.351 **	-0.011
Servicios				-0.250 **	-0.523 ***	-0.075
R ²	0.237	0.542	0.120	0.306	0.588	0.136
Error Estándar	1.121	1.705	1.680	1.022	1.546	1.610
LR (p-value)	0.000	0.000	0.000	0.000	0.000	0.000

Nota: Lectura del P-Valor: *** significativo al 99%; ** significativo al 95; *significativo al 90%.

Tabla 4. SEM Cuantil 2SQR (todas las observaciones). Variable dependiente: activos.

	Median	90%	10%	Median	90%	10%
Constant	12.997 ***	13.607 ***	12.409 ***	14.848 ***	14.982 ***	13.139 ***
Ocupados	1.384 ***	2.603 ***	0.784 ***	1.123 ***	2.157 ***	0.661 ***
SEM	-0.009 *	-0.003	0.000	-0.004	-0.003	0.002
Tipo Unipersonal				-1.784 ***	-1.271 ***	-0.652 ***
Colectiva				-1.572 ***	-0.852	-0.183
Sector Financiero				0.778 ***	3.185 ***	0.630 **
Otras				-0.334 ***	-0.360 ***	-0.419 ***
Industrial				-0.177 **	-0.351 **	-0.011
Servicios				-0.233 **	-0.523 **	-0.100
R ²	0.237	0.536	0.118	0.303	0.582	0.134
Error Estándar	1.121	1.707	1.672	1.018	1.561	1.598
LR (p-value)	0.000	0.000	0.000	0.000	0.000	0.000

Nota: Lectura del P-Valor: *** significativo al 99%; ** significativo al 95; *significativo al 90%.

Tabla 5. SEM Cuantil 2SQR (excluyendo outliers). Variable dependiente: activos.

Replicando el ejercicio I. de Moran y LISA que realizamos en la sección anterior, eliminamos de la población las dos grandes empresas agroindustriales del extremo oriente del municipio, y repetimos en la tabla 5 las especificaciones ofrecidas en la tabla 4. Como se puede apreciar en la tabla 5, ningún resultado se ha modificado al eliminar los *outliers*; el SEM fue significativo solo para la mediana y sin controles, mientras que no fue significativo en ninguno de los otros casos. Los controles repitieron los mismos signos y fueron significativos en los mismos casos, mientras que la relación entre activos y ocupados fue siempre positiva y significativa.

En el municipio de Ciénaga hay una asociación teóricamente esperable y de sentido común entre activos y ocupados, pero dicha asociación no tiene un carácter o representación espacial explícita. Ese resultado se sostiene no importando la exclusión o no de *outliers* y de variables de control. Tampoco se encontró dicha evidencia al enfatizar cuantiles extremos (10% y 90%). No hay evidencia de desarrollo empresarial tipo clúster.

Tenemos, entonces que los análisis LISA y de I. de Moran pueden ser “engañados” por los *outliers*, mostrando asociaciones espaciales positivas cuando las dos grandes empresas *outliers* se incluyen en el análisis. Por el contrario, la regresión espacial cuantil ha sido inmune a la presencia o no de dichos *outliers*, rechazando la existencia de clústeres empresariales en todas las especificaciones. Este resultado reconoce que las grandes empresas agroindustriales de la periferia del casco urbano son extrañas a la realidad socioeconómica del municipio, donde el comercio de pequeña escala es la actividad predominante, y con el cual estas no tienen casi interrelaciones.

Conclusiones

Aunque la detección de clústeres espaciales representa empíricamente algunos argumentos sobre aglomeración, presentes en teorías de geografía económica, no se trata en sentido estricto de pruebas de hipótesis a tales teorías y la complejidad de la estructura socioeconómica que estas representan. Sin embargo, el uso de herramientas estandarizadas de búsqueda y clasificación de clústeres empresariales

es ahora una práctica de consultoría e investigación, con sus correspondientes implicaciones en términos de política: la decisión de impulsar desarrollos territoriales tipo clúster.

El análisis espacial por sí solo no sirve para determinar con claridad la presencia de clústeres, o su relevancia e impacto sobre el desarrollo local. Una aproximación que combina los ejercicios espaciales con regresión cuantílica, por el contrario, puede permitir observar el efecto de observaciones extremas (*outliers*), y la no-linealidad del impacto de la clusterización espacial sobre variables económicas de interés.

En el presente artículo usamos 1.292 observaciones de activos, empleo, sector económico y tipo de organización jurídica en el municipio de Ciénaga, Magdalena. Se trata de un municipio con una importante tradición y un valioso rol como nodo regional, el cual, no obstante, casi no tiene referencias en la literatura sobre desarrollo territorial.

Con la información disponible podemos realizar tanto análisis cluster, incluyendo o no las grandes empresas agroindustriales de su periferia oriental, como análisis de econometría espacial cuantílica tipo 2SQR (*Two-Stages Quantile Regression*). Los ejercicios realizados permiten observar el efecto de las observaciones extremas sobre los resultados, encontrando que los ejercicios clúster son muy susceptibles a dicha información, mientras que los ejercicios de econometría espacial cuantílica no lo son.

Si bien, por un lado, los ejercicios de clusterización parecen sugerir la presencia de aglomeraciones económicas de activos y ocupados, pero son alterados por la presencia de *outliers*, de otra parte, los ejercicios econométricos rechazan la hipótesis de clusterización y no son afectados por la inclusión o no de las empresas *outliers*. El resultado tiene sentido, pues la actividad económica predominante en el municipio

es el comercio de pequeña escala (52% según el DANE, 2005), y no la agroindustria intensiva en capital.

Referencias

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Arbia, G. et al. (2012). Clusters of firms in an inhomogeneous space: The high-tech industries in Milan. *Economic Modelling* 29(1), 3-11

Castells, M. & Hall, P. (1994). *Las tecnopolis del mundo: la formación de los complejos industriales del siglo XXI*. Madrid: Alianza Editorial.

Das, K. (2011). Rural industrialization in India: enhancing reach and returns. En Das, K. (ed), *Micro and small enterprises in India: the era of reforms* (pp. 208-224). Nueva Delhi: Routledge – Taylor & Francis.

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

Galvis, L. & Meisel, A. (2010). Persistencia de las desigualdades regionales en Colombia: un análisis espacial. *Documentos de Trabajo sobre Economía Regional* 120. Cartagena de Indias: Banco de la Republica.

Garza, N., Gutierrez, M. & Nieto, L. (2009). El homicidio en Barranquilla: una lectura espacial. *Revista de Economía del Caribe* 3, 176-229.

Garza, N. & C. Ovalle (2019). Tourism and housing prices in Santa Marta, Colombia: Spatial determinants and interactions. *Habitat International* 87, 36-43,

Goebertus, J. (2008). Palma de aceite y desplazamiento forzado en Zona Bananera: “trayectorias” entre recursos naturales y conflicto. *Colombia Internacional* 67, 152-175.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

La Sage, J. & Pace, K. (2009). *Introduction to Spatial Econometrics*. Cheltenham UK: CRC Press – Taylor & Francis..

Liao, W. & Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics* 21(1) 16-27.

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

Martin, R. & Sunley, M. (2011). The new economic geography and policy relevance. *Journal of Economic Geography* 11, 357-369.

Maya, C. (2011). Sinaloa: ¿Clúster agroindustrial o territorio desincrustado? *Análisis* 14(41), 127-160.

Meisel, A (2004). La economía de Ciénaga después del banano. *Documento de Trabajo sobre Economía Regional* 50. Cartagena de Indias: Banco de la república..

Rey, C. (2013), Análisis espacial de la correlación entre cultivo de palma de aceite y desplazamiento forzado en Colombia. *Cuadernos de Economía* 32(61), 683-718

Simboli, A., Taddeo, R. & Morgante, A. (2012). Place-based approaches to industrial ecology. Considerations about the application in long-standing industrial clusters. *Environment and Energy. First Issue* (pp.177-189). Milano: FrancoAngeli.



El Instituto de Estudios Económicos del Caribe (IEEC) fue creado en 2003 para adelantar en la Universidad del Norte las actividades de docencia e investigación en el campo de la Economía. En consecuencia, el programa de pregrado en Economía, así como cursos especiales y de posgrado. Además, el IEEC lleva a cabo proyectos de investigación, con énfasis en el estudio de temas y problemas atinentes a la Costa Caribe colombiana.

El IEEC tiene como una de sus prioridades la difusión de sus investigaciones y ensayos mediante la publicación de libros y muy especialmente, a través de su serie *Documentos*. De esta manera, el IEEC aspira a contribuir a la discusión pública de los más significativos problemas que afectan a nuestra sociedad.