

# Determinación de mercados relevantes de la distribución de combustibles líquidos en algunas ciudades y regiones de Colombia

---

Estudio de FEDESARROLLO para FENDIPETRÓLEO Nacional-SOLDICOM

Directora

**Astrid Martínez Ortiz**

Investigadores

**David Forero**

**Sebastián Montaña**

*Bogotá, 20 de febrero 2019*

Estudio económico, financiero y estadístico que permite identificar los riesgos a los cuales se enfrentan las Estaciones de Servicio y determinación de los mercados relevantes de la distribución minorista de combustibles líquidos en Colombia

## **Determinación de mercados relevantes de la distribución de combustibles líquidos en algunas ciudades y regiones de Colombia**

Estudio de FEDESARROLLO para FENDIPETRÓLEO Nacional-SOLDICOM

**Directora**

**Astrid Martínez Ortiz**

**Investigadores**

**David Forero**

**Sebastián Montaña**

**Bogotá, 20 de febrero de 2019**

# Determinación de Mercados Relevantes en la Distribución Minorista de Combustibles Líquidos en Colombia

## 1. Introducción

La regulación económica, en particular en el mercado de los combustibles, debe internalizar y tener en cuenta el contexto geográfico y económico de los agentes que son objeto de ella. Así, por ejemplo, es muy diferente la regulación que debe ser diseñada para un monopolista en un mercado aislado, que aquella que se establece para un conjunto de agentes que actúan en un mercado de libre competencia. Entre los dos extremos, caben segmentos con algún poder de mercado y mercados con algún grado de competencia. Es en este contexto en el que toma relevancia el ejercicio de diferenciar e identificar, a través del análisis de las cifras económicas asociadas a las estaciones de servicio, los distintos mercados geográficos que pueden considerarse como unidades modulares de análisis, esto es, en donde sus agentes actúan bajo un principio de competencia en la distribución y venta de combustible.

Para definir cómo se delimita un mercado de un bien o producto, además del espacio físico donde es posible producir y vender dicho bien, el concepto más importante es el de sustituibilidad, referido a si un consumidor puede considerar dos productos de dicho mercado como sustitutos uno del otro. Así, en la definición de *mercado relevante*, los dos elementos más importantes a tener en cuenta son el producto y la zona geográfica. Por otro lado, la interacción entre agentes al interior de un mercado requiere hablar no solo de transacciones o intercambios sino también de la competencia, que toma forma cuando algunos oferentes buscan diferenciar su producto o ubicarse en espacios físicos más convenientes para incrementar su actividad y maximizar sus beneficios.

El Departamento de Justicia y la Comisión Federal de Comercio de Estados Unidos (1992) definen un mercado relevante como “*el producto [...] y la zona geográfica en la que se produce o vende, que siendo lo más pequeña posible, harían que un monopolística hipotético, con afán de lucro y no sometido a regulación, estuviese interesado en aumentar el precio, de forma no esporádica, en un porcentaje pequeño pero significativo, si permaneciera constantes los demás precios relativos*”.

En la actividad de distribución minorista de combustibles líquidos, debe tenerse en cuenta que para un consumidor (o vehículo) no existen productos sustitutos debido a razones tecnológicas; en otras palabras, un automóvil puede funcionar a partir de gasolina o a partir de diésel, pero no puede intercambiar entre ambos tipos de combustible a discrecionalidad (CREG, 2016). Además, el combustible ofertado por distintos agentes en Colombia es generalmente un bien homogéneo, es decir, no se presenta una marcada diferenciación de producto en la distribución



SAR<sup>2</sup>, el cual es el más apropiado para estudiar la competencia global entre distribuidores minoristas de combustibles líquidos (Lee, 2007). El objetivo principal del modelo es definir el radio (distancia o tiempo de ruta) en que los precios de las estaciones de servicio se correlacionan con los de otros distribuidores, suponiendo que observaciones cercanas tienen mayor correlación positiva<sup>3</sup>.

A pesar de que el estudio contratado por la CREG (2016) empleó la metodología propuesta para algunos casos que usó como ejemplo, el estudio citado no fue del todo claro en la metodología. Para comenzar, no se sabe si utiliza información longitudinal o un corte transversal de precios. En los escenarios analizados, tampoco es claro si se adicionan efectos fijos de estación de servicio para controlar por las particularidades de los distribuidores. Considerando la ecuación del modelo SAR del estudio, se puede señalar que, al no incluir efectos fijos u otras características observables, se está suponiendo que las estaciones de servicios son idénticas y que sus características no son relevantes a la hora de explicar la dinámica de los precios<sup>4</sup>. Pero, como se mencionó anteriormente, las características de las estaciones sí afectan los precios a los que se oferta el combustible (Lee, 2007; Neilson y Bruce, 2007). Obsérvese además que incluir características observables o efectos fijos de estación de servicio son opciones excluyentes, pero igualmente válidas para controlar por las particularidades asociadas a cada estación de servicio.

Ahora, para determinar los mercados relevantes a partir de los resultados del modelo SAR, el estudio de Econometría para la CREG usa una regla a priori: elegir como óptimo el intervalo de tiempo que se asocia al coeficiente espacial estimado con mayor magnitud, es decir, aquel intervalo de tiempo para el que se halló un mayor nivel de correlación entre precios de los competidores. No es claro por qué el mejor modelo según el estudio es aquel con un coeficiente más grande, ya que no se realizan pruebas estadísticas para determinar si existe o no diferencia con otros intervalos<sup>5</sup>. Para la solución de esta etapa del modelo en la literatura, Lee (2007) propone estimar de manera conjunta distintos intervalos de distancia (o tiempo), encontrando que solo el coeficiente a 1 milla de distancia es significativo en un modelo en que se incluyen diferentes matrices de ponderación espacial.

Pero más allá de las dudas puntuales sobre la metodología de Econometría Consultores, los resultados a los que se llegó en el estudio demuestran que no es suficiente con estimar una distancia óptima para definir los mercados relevantes. En ausencia de un ejercicio adicional de delimitación geográfica, la distancia óptima puede llevar a conclusiones indeseables, como, en

---

<sup>2</sup> Por sus siglas en inglés *Spatial Autoregressive Models* (SAR).

<sup>3</sup> En un modelo microeconómico de equilibrio parcial, con agentes racionales, la reducción del precio por parte de un oferente – que busca apropiarse de consumidores habituales de otros – será seguida por reducciones en el precio de sus competidores, quienes no desean ver afectados sus ingresos. Note que aumentos en los precios se podrían explicar por razones externas que afecten a todos los oferentes, y el mercado seguiría la senda para alcanzar un nuevo precio de equilibrio que no afecta en gran medida las ventas. Así, a partir de la teoría económica se puede esperar una correlación positiva, a lo largo del tiempo, entre los precios de los competidores que hacen parte de un mismo mercado.

<sup>4</sup> En algunos apartes, el estudio para la CREG (2016) menciona que se incluyeron cantidades vendidas de combustible por estación de servicio; sin embargo, no se presentan resultados del efecto que esta variable tiene sobre los precios.

<sup>5</sup> Obsérvese que sería posible realizar pruebas  $\chi^2$  para responder esta inquietud: si existe diferencia en las magnitudes de las correlaciones halladas.

primer lugar la de que las EDS de una ciudad capital con alta densidad poblacional, como Bogotá, pertenecen todas a un mismo mercado, o, en segundo lugar, la de que todas y cada una de las EDS ubicadas sobre una vía intermunicipal, con suficiente distancia entre ellas, pertenecen a un mercado relevante diferenciado, una conclusión que no tiene aplicabilidad alguna en términos de política pública.

En consecuencia, este trabajo desarrolla una metodología para identificar los mercados relevantes de combustibles líquidos de uso automotor en el nivel de las EDS en las principales ciudades, en las zonas de frontera y en las regiones de Colombia. Como tal, esta metodología puede ser aplicada en diferentes momentos del tiempo hacia el futuro así como a los mercados de gasolina corriente, gasolina extra y ACPM. En este documento se aplica la metodología para la gasolina corriente.

Haciendo uso de la información longitudinal de precios de venta con que cuenta el Fondo Soldicom y el Censo de Caracterización de Estaciones de Servicio (Fondo Soldicom, 2016), en este documento se estima un modelo SAR con datos panel que incluye efectos fijos de estación de servicio y efectos fijos de tiempo. Los resultados de este modelo son comparados posteriormente con los resultados de un modelo SAR que controla por características observables como la bandera de la estación, otros servicios que la estación ofrece a sus clientes, el número de accesos de cada estación, y el número de mangueras en cada surtidor, entre otras.

Acá es importante señalar que las estimaciones del modelo SAR, que definen el radio de acción óptimo de los mercados de distribución de combustible en el área analizada, son complementadas por primera vez con la aplicación de un modelo de conglomerados (*clusters*). A partir de la interacción entre estos dos modelos, es posible establecer con mayor precisión el número de mercados relevantes que se hallan al interior de una ciudad, departamento o país.

Para el presente proyecto se estimó el modelo SAR con efectos fijos de tiempo y efectos fijos de estación de servicio para los departamentos definidos como zonas de frontera, las 4 principales ciudades capitales, y el resto del país, donde cada tipo de estimación tiene algún grado de diferenciación en la especificación técnica de la metodología. Adicionalmente, se realizaron estimaciones del modelo SAR sencillo, con efectos fijos por EDS, y uno más completo en el que se introducen las diferentes características observables, lo que permite tanto comparar los resultados entre los dos modelos, como evaluar el efecto específico que tienen diferentes características de las EDS, como la bandera o la cantidad de accesos, sobre el nivel de los precios de venta de combustible que se ofrecen a los consumidores.

El énfasis del presente proyecto ha sido el desarrollo de una metodología robusta y replicable para la identificación de los mercados relevantes de distribución minorista de combustible en Colombia, con base en los estudios previos y la literatura internacional reciente. El ejercicio acá presentado tiene como uno de sus objetivos principales las zonas urbanas, en donde la gasolina corriente es el combustible que permite medir con mayor precisión el grado de competencia entre diferentes EDS. Por otra parte, teniendo en cuenta que gran parte del transporte de carga, que usa ACPM como combustible, tiene contratos directos con plantas de abastecimiento o con EDS a precios favorables –y que no son registrados en el SICOM- se considera más adecuada la información de precios disponible para gasolina corriente, con datos que son fundamentales para

la estimación de los mercados relevantes. Por estas razones, los resultados aquí presentados corresponden únicamente a la gasolina corriente. En vista de que se entregan los scripts, do-files y rutinas estadísticas, y lo más importante, las matrices de distancias entre EDS, será posible replicar de forma sencilla el ejercicio acá presentado a los datos de precios provenientes del ACPM. De hecho, la metodología planteada es lo suficientemente general para abrir la posibilidad de determinar mercados relevantes en la actividad de distribución de gasolina extra o premium y de distribución de gas natural vehicular, lo que abre nuevos caminos de investigación en este campo.

Como reflexión final, se considera que la utilidad del presente ejercicio se materializa en tres grandes dimensiones. La primera es la de ofrecer una metodología que pueda ser adoptada por las entidades de regulación y supervisión para la determinación de mercados relevantes la cual debe facilitar la interlocución entre los gremios de distribuidores y las autoridades regulatorias, al establecer un marco común para definir, por ejemplo, el régimen al que debería pertenecer ciertas estaciones de servicio y evaluar para diferentes regiones o ciudades las distancias óptimas en que se observa mayor competencia en precios. En cuanto a la supervisión de la competencia, esta metodología permite precisar el parámetro de distancia óptima, la cual está definida hoy en día como una constante para todo el país (4 km) por la Superintendencia de Industria y Comercio.

La segunda es la evaluación del efecto que tienen las diferentes características de las EDS en Colombia, incluida la bandera, sobre el precio final que se ofrece al consumidor. Por último, para los nuevos actores que quieran ingresar al mercado de distribución minorista de combustible, este documento proporciona un estudio de las condiciones del mercado circundante en la zona geográfica de interés, así como el grado de competencia al que se va a enfrentar.

## 2. Metodología de Determinación de Mercados Relevantes

La metodología para determinar los mercados relevantes se basó en el trabajo de la CREG (2016) y en Lee (2007), adaptándola al tipo de información con la que se cuenta para las estaciones de servicio (EDS) en Colombia. En particular, para determinar los mercados relevantes se siguió una estrategia en dos etapas.

En la primera, se estimó un modelo SAR usando el panel de precios de gasolina corriente entre 2013 y 2017, e incluyendo efectos fijos a nivel de estación de servicio y efectos fijos de tiempo. De manera específica, se estimó la siguiente ecuación:

$$y_{it} = \rho W^{dc} y_{it} + \alpha y_{it-1} + \sum_{i=1}^{N-1} \beta_i I(EDS_i) + \sum_{t=1}^{T-1} \phi_t I(t) + \varepsilon_{it}$$

$$i = 1, \dots, N ; t = 1, \dots, T$$

donde  $y_{it}$  es el precio de venta de la gasolina corriente para el mes  $t$  en la estación de servicio  $i$ .  $W_t^{dc}$  es una matriz cuadrada  $N$ -dimensional, definida a partir de las distancias en kilómetros

entre cada par de estaciones de servicio  $(i, j)$  y tal que:  $W^{dc}(i, j) = \frac{\omega_{ij}}{\sum_{ij} \omega_{ij}}$ . Esta matriz se restringe dependiendo de la distancia crítica  $dc$  que se defina en el ejercicio; así, si la distancia entre la estación  $i$  y la estación  $j$ ,  $d_{ij}$ , está a menos de una distancia crítica  $dc \in \{dc_0, dc_1, \dots, dc_D\}$ , entonces  $\omega_{ij} = \frac{1}{1+d_{ij}}$ , y de lo contrario es cero<sup>6</sup>.  $\rho$  es el parámetro de interés, que está asociado a los precios rezagados y ponderado espacialmente por  $W^{dc}y_{it}$  en el periodo  $t$ , y el cual es de esperar que tome valores positivos<sup>7</sup>. Observe que este parámetro permite establecer la dependencia espacio-temporal de los precios, o en otras palabras la competencia existente en el mercado. El coeficiente  $\alpha$  es el efecto marginal del precio del combustible en el periodo  $t - 1$ , lo que permite controlar por la dependencia temporal de los precios.  $\beta_i$  son parámetros que acompañan a los efectos fijos a nivel de estación de servicio, controlando así por las características de cada distribuidor minorista.  $\phi_t$  son efectos fijos de tiempo, lo que permite controlar por aquellas condiciones que podrían cambiar con el transcurso de los meses, como por ejemplo la puesta en marcha de nuevas políticas regulatorias que afectan los precios de venta del combustible para todas las estaciones de servicio.

Ahora, para determinar la distancia crítica óptima,  $dc^*$ , se realiza una serie de pruebas que permiten establecer si cada coeficiente  $\rho_{dc}$  es estadísticamente diferente de los coeficientes  $\rho_{dc'}$ , donde  $dc < dc'$ . Observe que esta prueba inicia con el coeficiente asociado a la menor distancia, denotada por  $dc_0$ , y establece bajo la hipótesis nula que éste es igual al coeficiente con mayor distancia. El proceso avanza a la siguiente distancia crítica,  $dc^1$ , si no fue posible rechazar cada una de las hipótesis nulas consideradas, es decir, avanza a  $dc^1$  si hay evidencia de que  $\rho_{dc_0} \neq \rho_{dc'}$ , para alguna distancia  $dc' > dc_0$ . Es importante mencionar que se estableció que el máximo nivel de significancia admitido como evidencia para rechazar la hipótesis nula es  $\alpha = 0.05$ .

La segunda etapa emplea una metodología de Análisis de Conglomerados (*Cluster Analysis*) que considera las características de las estaciones de servicio observadas a partir de Censo Soldicom 2016, así como el radio óptimo al cual se ejerce competencia en cada municipio o departamento considerado, determinado previamente por el modelo SAR ya descrito.

Ahora bien, en diferentes estudios sobre los mercados relevantes en la distribución minorista de combustible se ha mostrado que la competencia en precios se ejerce en un radio de una milla, es decir, aproximadamente 1.6 kilómetros de distancia (Neilson y Bruce, 2007; Lee, 2007). Teniendo en cuenta esta evidencia previa, se estableció un conjunto de distancias críticas entre 250 metros y 4 kilómetros, con intervalos de 250 metros, como se mostrará más adelante.

---

<sup>6</sup> Observe que para cada corte transversal la matriz  $W^{dc}$  toma el mismo valor, lo que implica asumir que las distancias entre estaciones van a ser constantes en el tiempo. Sin embargo, este supuesto no es demasiado fuerte ya que es de esperar que las distancias entre estaciones de servicio permanezcan constantes entre meses, y al menos en el mediano plazo.

<sup>7</sup> Estimaciones negativas y significativas de  $\rho$  indican que, aunque existe competencia de mercado, ésta es compensada por factores cooperativos (Kao y Bera, 2016; CREG, 2016).

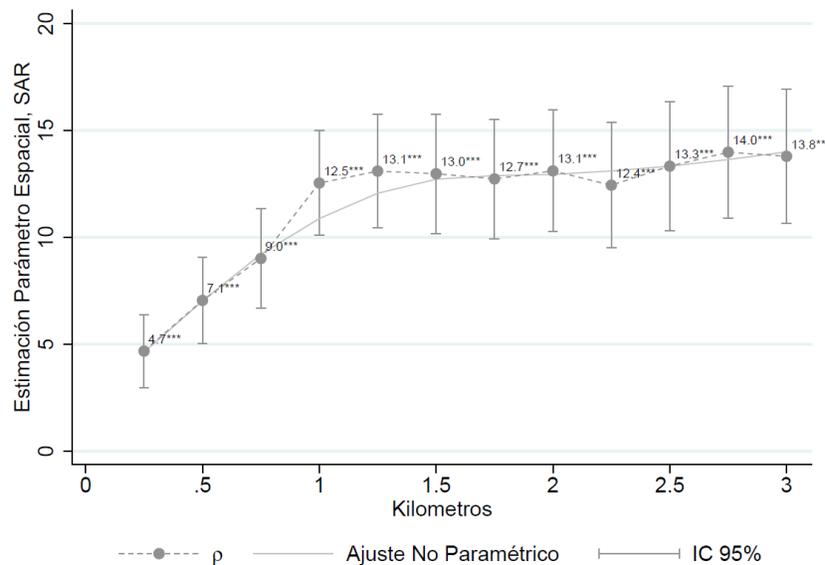
### 3. Mercados Relevantes en Zonas de Frontera

La Ley 681 de 2001 estableció como zonas de frontera a los departamentos de Amazonas, Arauca, Boyacá, Cesar, Chocó, Guainía, La Guajira, Nariño, Norte de Santander, Putumayo, Vaupés y Vichada. Para la presente entrega se definió a la frontera con Venezuela como objeto de estudio. Sin embargo, no se cuenta con información suficiente para determinar mercados relevantes en los departamentos fronterizos de Vichada y Guainía. De esta manera, se aplicará la metodología descrita para los siguientes departamentos fronterizos: La Guajira, Cesar, Norte de Santander, Boyacá y Arauca.

#### 3. 1. La Guajira

La Guajira tiene un área de aproximada de 20,848 km<sup>2</sup>. Se cuenta con georreferenciación e información longitudinal de precios para 132 estaciones de servicio en este departamento. Luego, en promedio se tiene una densidad de 1.3 estaciones de servicio por cada 200 km<sup>2</sup>. La Figura 2 presenta los resultados de estimar el modelo autorregresivo espacial para diferentes distancias críticas, que ponderan espacialmente los precios de la gasolina corriente.

Figura 2. Estimaciones del Modelo SAR, La Guajira.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Los resultados de estimar del modelo SAR para La Guajira indican que existe competencia en precios entre estaciones de servicio incluso a un radio de 250 metros, dado que a partir de esta distancia las estimaciones del parámetro  $\rho$  son estadísticamente significativas. Además, el ajuste no paramétrico presentado en la misma Figura 2 señala que la relevancia de la distancia decrece

marginalmente y permanece estable a partir de 1 kilómetro. Ahora, para determinar la distancia óptima se utiliza el proceso descrito en la sección de metodología, el cual permite comparar cada coeficiente asociado a una distancia crítica con el resto de los coeficientes, que como se mencionó están asociados a distancias críticas mayores. A continuación, se presenta el resultado de este ejercicio.

Tabla 1. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

<b>Distancia Crítica (Km)</b>	<b>0,75</b>	<b>1,00</b>	<b>1,25</b>	<b>1,50</b>	<b>1,75</b>	<b>2,00</b>	<b>2,25</b>	<b>2,50</b>	<b>2,75</b>
<b>1,00</b>	189,31								
<b>1,25</b>	100,12	3,07							
<b>1,50</b>	66,69	1,12	0,21						
<b>1,75</b>	57,65	0,21	1,79	12,80					
<b>2,00</b>	63,22	1,62	0,00	0,59	5,58				
<b>2,25</b>	37,99	0,04	3,02	3,91	1,25	10,19			
<b>2,50</b>	53,28	2,19	0,30	1,09	3,24	0,58	19,51		
<b>2,75</b>	65,50	6,68	3,72	7,15	11,39	6,83	35,06	15,61	
<b>3,00</b>	56,04	4,57	1,99	3,85	6,68	3,27	18,59	3,68	1,23

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 2. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

<b>Distancia Crítica (Km)</b>	<b>0,75</b>	<b>1,00</b>	<b>1,25</b>	<b>1,50</b>	<b>1,75</b>	<b>2,00</b>	<b>2,25</b>	<b>2,50</b>	<b>2,75</b>
<b>1,00</b>	0,000								
<b>1,25</b>	0,000	0,080							
<b>1,50</b>	0,000	0,290	0,647						
<b>1,75</b>	0,000	0,643	0,181	0,000					
<b>2,00</b>	0,000	0,203	0,970	0,444	0,018				
<b>2,25</b>	0,000	0,839	0,082	0,048	0,264	0,001			
<b>2,50</b>	0,000	0,139	0,586	0,296	0,072	0,446	0,000		
<b>2,75</b>	0,000	0,010	0,054	0,007	0,001	0,009	0,000	0,000	
<b>3,00</b>	0,000	0,033	0,158	0,050	0,010	0,071	0,000	0,055	0,267

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi^2_{gl=1})$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

En la Tabla 2 se puede observar que el menor coeficiente de correlación espacial que es estadísticamente igual a los coeficientes estimados con distancias críticas mayores, es el coeficiente asociado a una distancia de 1.25 kilómetros. Es decir, es el primer coeficiente para el

que ningún P-valor es mayor que 0.05, que es el nivel de significancia permitido para rechazar la hipótesis nula. Este resultado es coherente con las conclusiones de la Figura 1.

Tabla 3. Resultados SAR con Distancia Crítica Óptima y Controles, La Guajira.

<i>Variable Dependiente: Precio Gasolina</i>			
	MCO	SAR	
	Coef.	Coef.	Coef Wx.
1(Ayatawa)	114.24	170.39*	-45,923.05
1(Lavado Vehículos)	137.11	165.60	-13,672.34
1(Mini-mercado)	6.23	-56.13	-52,893.87
1(Otros Servicios)	91.14	62.67	40,959.80
Accesos	-114.69*	-111.99*	-7,578.37
Competidores (1Km)	6.75	133.46***	692.17
$\rho$		15.65***	
$\rho_X$		-1.60***	
$\sigma_e^2$		181,603.51***	
Observaciones	7,920	7,920	7,920
R-cuadrado		0.35	0.35
Número EDS	132	132	132

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite la variable dummy que toma 1 como valor cuando la estación de servicio pertenece a una bandera del grupo Otras. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

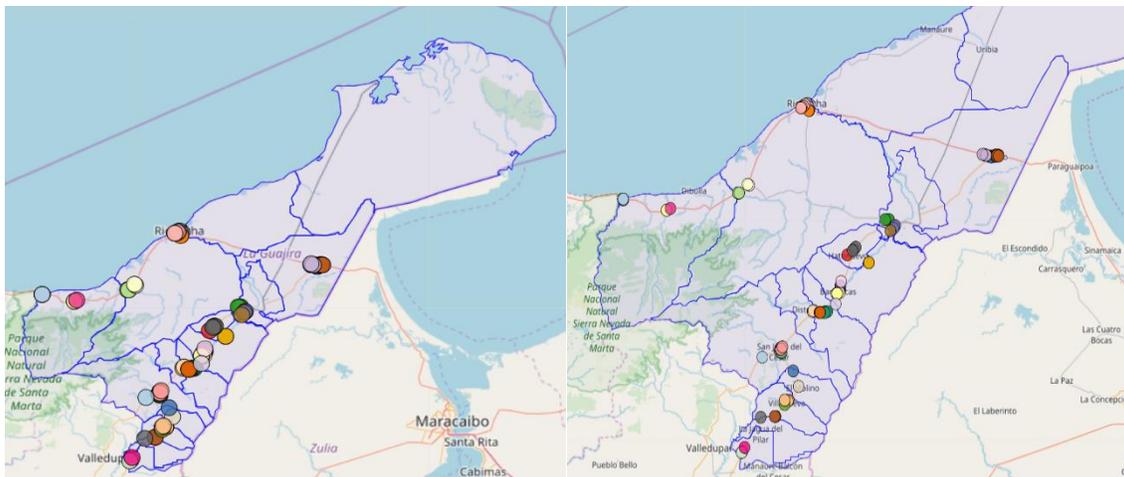
La Tabla 3 presenta los resultados del modelo SAR usando características de las estaciones de servicio, en lugar de efectos fijos<sup>8</sup>. Se observa que  $\rho$  es cercano al encontrado con el modelo alternativo, es decir, 15.6 comparado con 13.1 del modelo de efectos fijos. Luego, ambos modelos son válidos para determinar el radio en que se ejerce competencia en precios (Lee, 2007). Sin embargo, los resultados de la Tabla 3 permiten señalar que las estaciones de servicio de Ayatawa fijan en promedio un precio mayor en \$170 pesos, en comparación con las estaciones de bandera Otras. Ahora, por cada vía de acceso adicional, una estación de servicio reduce en promedio su precio en \$112 pesos, resaltando la importancia de la ubicación geográfica para competir en precios. Note también que no existe evidencia estadística para concluir que la tener lavadero, mini-mercado, y/o otros servicios influya en los precios. Se presentan estimaciones por MCO de un modelo de datos panel, esto con el fin de tener una referencia acerca del efecto marginal de cada característica observable. Note que bajo este modelo los precios de las estaciones de Ayatawa no son estadísticamente diferentes a los precios de estaciones con bandera

<sup>8</sup> Recuérdese que la diferencia entre los dos modelos es que los efectos fijos capturan el efecto idiosincrático, que no cambia en el tiempo, de cada estación sobre el nivel de precios. En contraste, el modelo de características elimina los efectos fijos, e intenta capturar esta información a través de la definición de controles específicos que introducen variables de caracterización de las EDS.

Otras. Sin embargo, estas estimaciones suponen que  $\rho$  y  $\rho_X$  son iguales a cero, y por lo tanto serán sesgadas (Ord, 1975).

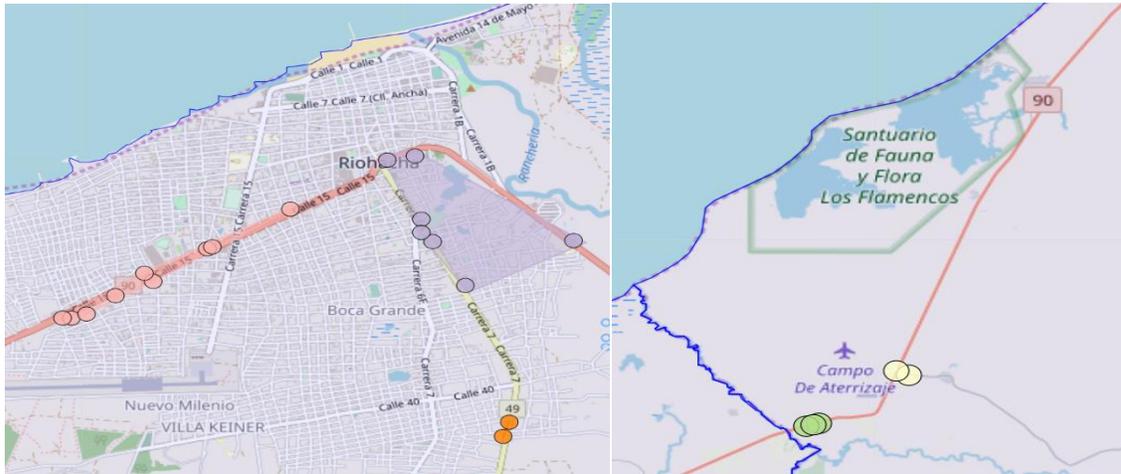
Luego de definir la distancia óptima se realiza un análisis de conglomerados (o *clusters*) que considera la distancia encontrada para determinar los distintos mercados relevantes. Como resultado de este análisis se encuentran 36 conglomerados o mercados relevantes en La Guajira, considerando como distancia crítica 1.25 kilómetros. La Figura 3 permite observar una panorámica general de los resultados. Las figuras 4 a 6 muestran los resultados para la capital del departamento, Riohacha, y para otros centros poblados o municipios. En Riohacha se identificaron 5 mercados relevantes, 3 de los cuales están en el área urbana. De manera específica, se identifica un mercado compuesto por 9 estaciones cercanas o sobre la calle 15, dos mercados con distribuidores cercanos o sobre la carrera 7 y dos mercados con estaciones cercanas a la ruta nacional 90.

Figura 3. Estimación de Conglomerados, La Guajira.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 1.25 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

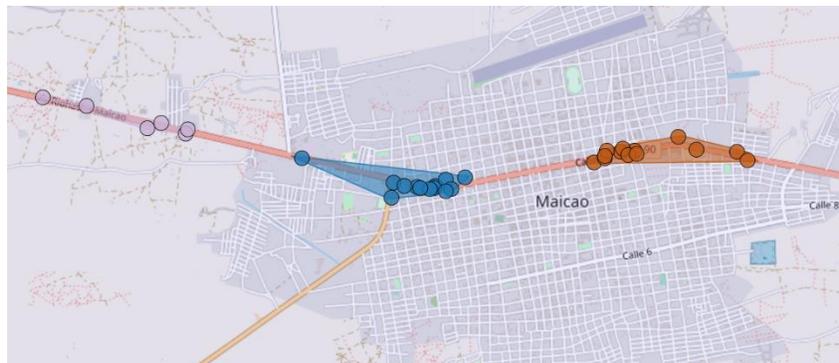
Figura 4. Estimación de Conglomerados, Riohacha.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 1.25 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

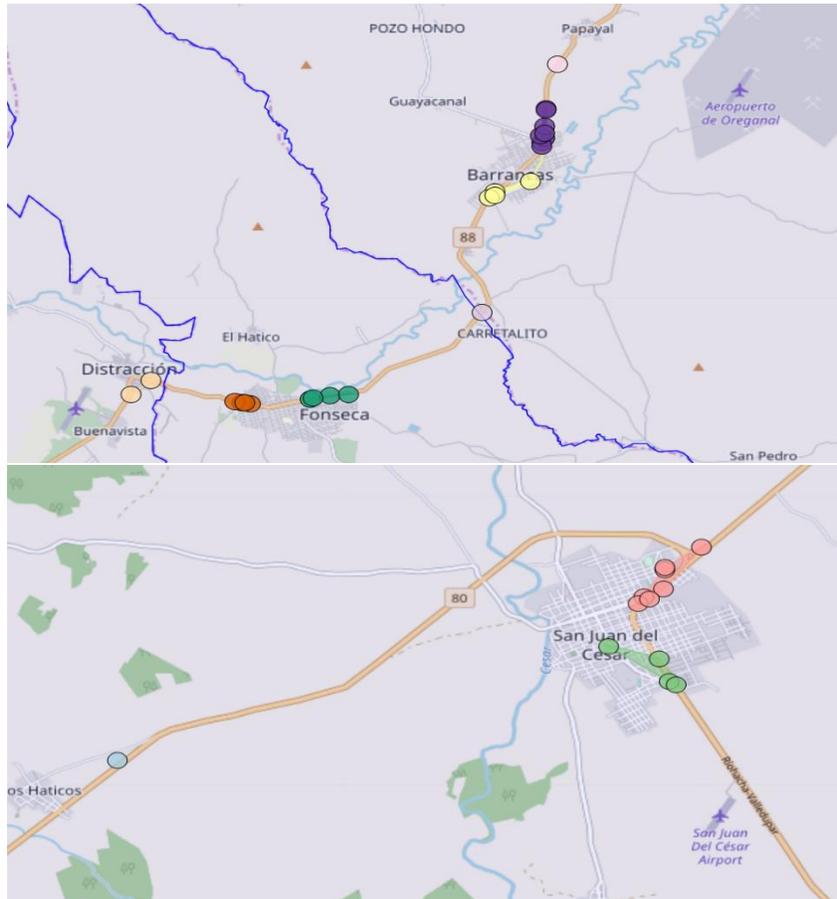
La Figura 5 presenta los resultados para Maicao, donde se identifican 3 mercados relevantes, todos cercanos a la calle 16 y la vía Riohacha-Maicao. La Figura 6 da cuenta de 10 mercados relevantes en diferentes centros poblados de La Guajira.

Figura 5. Estimación de Conglomerados, Maicao.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 1.25 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 6. Estimación de Conglomerados, Fonseca, Barrancas y San Juan de Cesar.

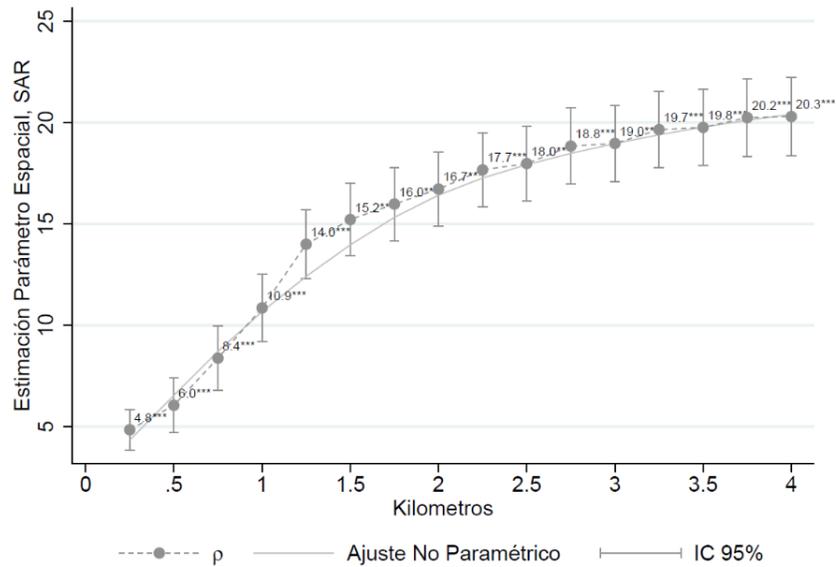


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 1.25 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

### 3. 2. Cesar

El departamento del Cesar tiene un área aproximada de 22.905 km<sup>2</sup> y cuenta con información para 147 estaciones de servicio. En este sentido, se tiene una densidad de 1.3 estaciones de servicio por cada 200 km<sup>2</sup>. La Figura 7 presenta los resultados de estimar el modelo SAR con efectos fijos de estación de servicio y efectos fijos de tiempo. Estos indican que existe competencia en precios incluso a un radio de 250 metros, al igual que en La Guajira. Asimismo, se observa que la importancia de la distancia decrece marginalmente con la distancia.

Figura 7. Estimaciones del Modelo SAR, Cesar.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

A continuación, se presentan los resultados del proceso de determinación de distancia óptima. En este caso, solo hay evidencia de que el coeficiente de correlación espacial asociado a 3.75 kilómetros es el único para el que no se rechaza la serie de pruebas. Observe que este no es estadísticamente diferente al coeficiente asociado a 4 kilómetros. Para esta distancia, el número de mercados relevantes que resulta del análisis de conglomerados es 44. La Figura 8 presenta los resultados de este método y las Figuras 9 y 10 dan cuenta de los resultados para la capital del departamento, Valledupar, y otros municipios o centros poblados. En Valledupar, se identifican dos mercados, uno de ellos compuesto por estaciones de servicio cercanas a la carrera 19, la diagonal 21 o la carrera 18D y un segundo mercado compuesto por una estación cercana a la carrera 23. Por otro lado, los municipios de Manaure, La Paz y San Diego tienen un mercado relevante cada uno, compuesto por sus únicas estaciones de servicio. Además, en Aguachica se observan 4 mercados relevantes, uno de los cuales está compuesto por 9 distribuidores cercanos al área urbana del municipio y el resto ubicados sobre la ruta nacional 45 o ruta del sol. De hecho, el análisis de conglomerados identifica 25 mercados a lo largo de la ruta nacional 45 en Cesar.

Tabla 4. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ , Cesar.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	53,96								
2,25	400,38	294,76							
2,50	172,07	121,29	5,97						
2,75	234,61	181,88	49,35	63,65					
3,00	273,03	224,15	65,99	103,20	8,04				
3,25	372,52	330,14	135,29	215,70	477,27	133,76			
3,50	359,92	312,77	134,35	192,43	173,01	89,51	3,43		
3,75	310,83	256,87	129,14	145,49	83,53	63,17	15,76	12,42	
4,00	275,36	222,39	114,10	120,90	64,82	50,53	13,24	10,26	0,30

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / [Var(\rho_{dc'}) - Var(\rho_{dc''])]$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 5. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ , Cesar.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,000								
2,25	0,000	0,000							
2,50	0,000	0,000	0,015						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,005				
3,25	0,000	0,000	0,000	0,000	0,000	0,000			
3,50	0,000	0,000	0,000	0,000	0,000	0,000	0,064		
3,75	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
4,00	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,582

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi^2_{gl=1})$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

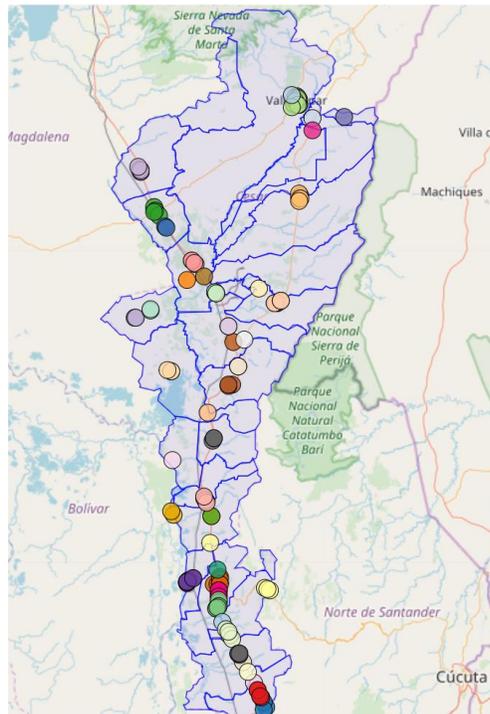
La Tabla 6, presentada a continuación, indica que en promedio las estaciones con bandera Petromil tienen un precio que es menor en 210 pesos, en comparación con las estaciones de bandera Otras en el departamento de Cesar. Por otro lado, las estaciones de bandera Zeuss y Ayatawa ofertan su gasolina a un precio mayor en 453 pesos y 500 pesos, respectivamente. Lo anterior con respecto a los precios de distribuidores de bandera Otras, que es el grupo de comparación omitido de la regresión. Del resto de características observables incluidas se observa que únicamente el número de accesos es importante en la fijación de precios, luego se puede concluir que tener un acceso adicional reduce el precio en 159 pesos.

Tabla 6. Resultados SAR con Distancia Crítica Óptima y Controles, Cesar.

	<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Terpel)	-53.60	-23.32	14,669.04
1(Zeuss)	476.01***	453.41***	262,778.40***
1(Biomax)	146.41	249.42	164,850.57*
1(Ayatawa)	718.53**	499.95*	273,514.44
1(Petromil)	-196.73**	-210.52***	47,229.80*
1(Lavado Vehículos)	-66.42	-161.65	-120,689.77
1(Mini-mercado)	22.41	-31.31	-7,424.25
1(Otros Servicios)	-179.84	-2.74	34,478.60
Accesos	-208.85***	-158.95***	2,357.93
Competidores (1Km)	-17.17	20.20	1,219.25
$\rho$			20.00***
$\rho_X$			-1.47***
$\sigma_e^2$			144,813.10***
Observaciones	8,820	8,820	8,820
R-cuadrado		0.49	0.49
Número EDS	147	147	147

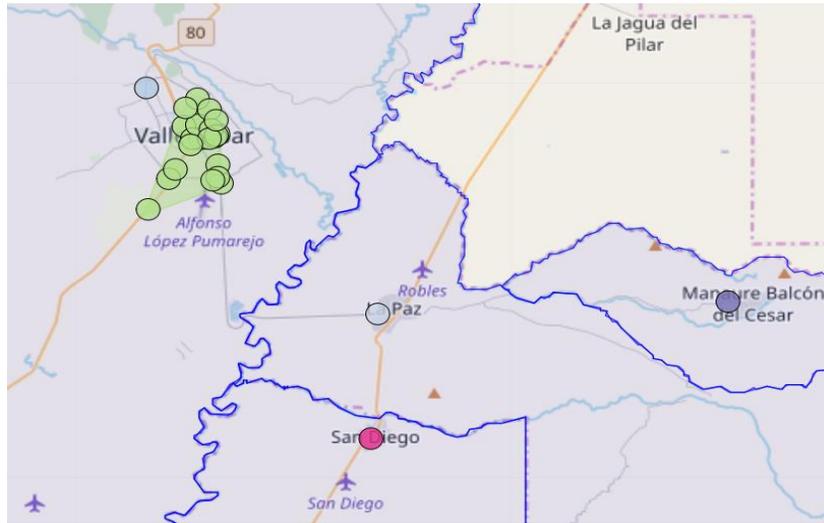
Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite la variable dummy que toma 1 como valor cuando la estación de servicio pertenece a la bandera Otras. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Figura 8. Estimación de Conglomerados, Cesar.



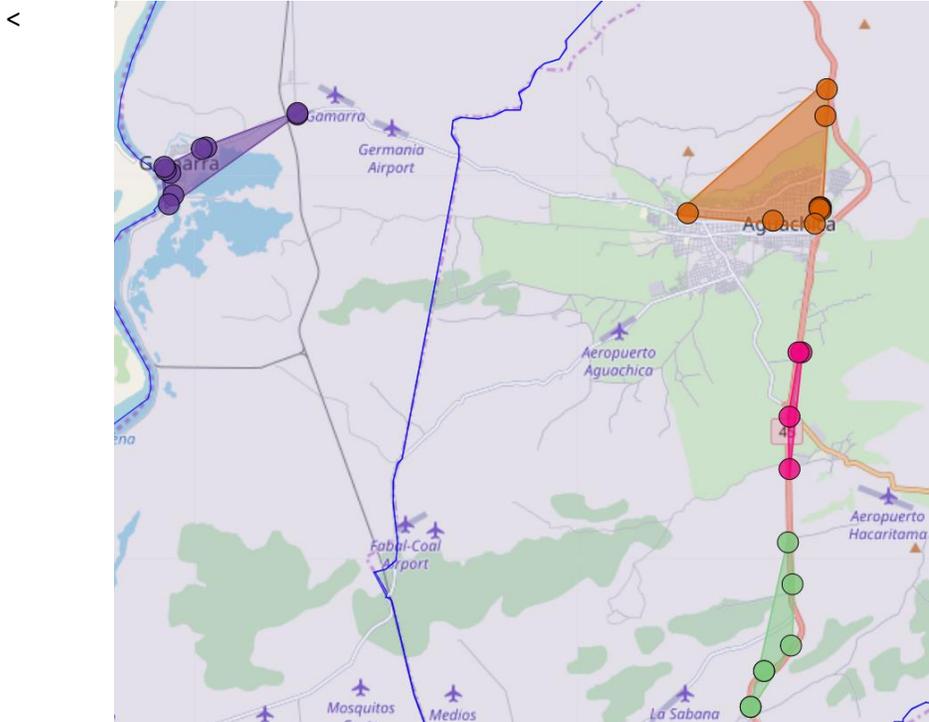
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclidianas.

Figura 9. Estimación de Conglomerados, Valledupar.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclidianas.

Figura 10. Estimación de Conglomerados, Aguachica, Gamarra y Vía Aguachica-San Martín.

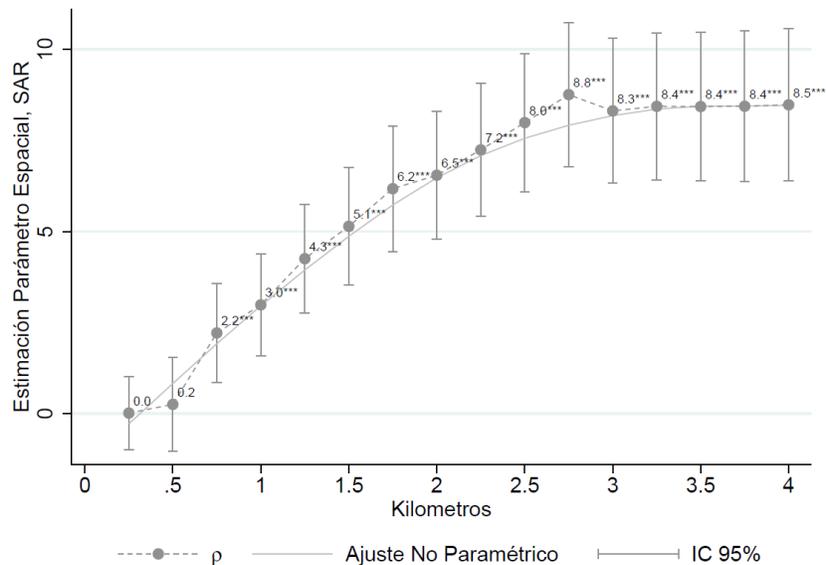


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclidianas.

### 3. 3. Norte de Santander

Norte de Santander tiene un área de aproximada de 21,658 km<sup>2</sup> y cuenta con información para 132 estaciones de servicio. Luego, en promedio se tiene una densidad de 1.2 estaciones de servicio por cada 200 km<sup>2</sup>.

Figura 11. Estimaciones del Modelo SAR, Norte de Santander.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

La Figura 11 presenta los resultados del modelo SAR. Se observa que existe evidencia de competencia en precios a partir de 750 metros, ya que a partir de esta distancia el coeficiente de correlación espacial significativo. Las Tablas 7 y 8 presentan los resultados del proceso para determinar la distancia óptima. En este caso la distancia óptima es 3.0 kilómetros, dado que a medida que se aumenta la distancia, el parámetro estimado y su intervalo de confianza aumentan lo suficiente como para no rechazar ninguna de las pruebas  $\chi^2$ . El análisis de conglomerados encuentra 42 mercados relevantes para Norte de Santander, las Figuras 12 a 14 dan cuenta de las estimaciones por conglomerados para determinar los mercados relevantes.

Tabla 7. Estadísticos  $\chi_{gl=1}^2$  para comparar entre coeficientes  $\rho_{dc}$ , Norte de Santander.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	9,03								
2,25	23,31	14,42							
2,50	37,18	28,43	14,08						
2,75	53,39	44,60	30,19	16,22					
3,00	34,48	26,61	13,69	2,35	27,17				
3,25	34,77	27,12	14,53	3,40	4,70	1,07			
3,50	31,69	24,46	12,65	2,70	3,02	0,51	0,00		
3,75	29,67	22,77	11,57	2,39	2,14	0,40	0,00	0,01	
4,00	29,33	22,55	11,58	2,58	1,40	0,58	0,05	0,11	0,19

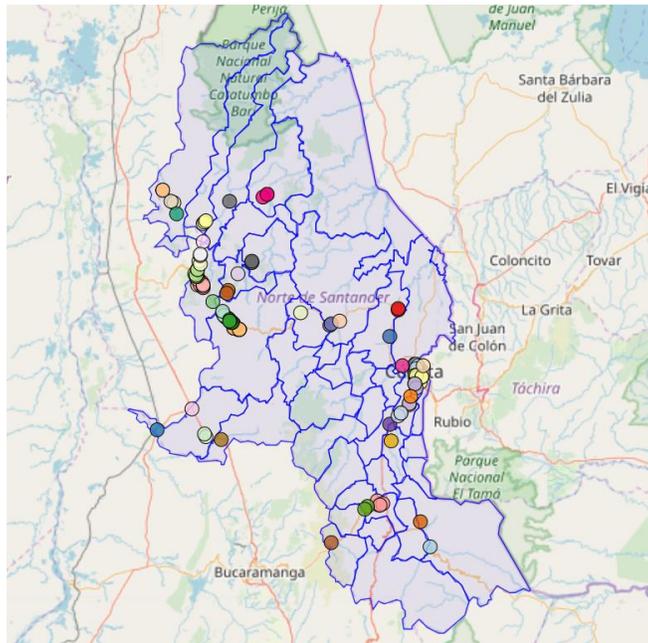
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ .  
 Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 8. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ , Norte de Santander.

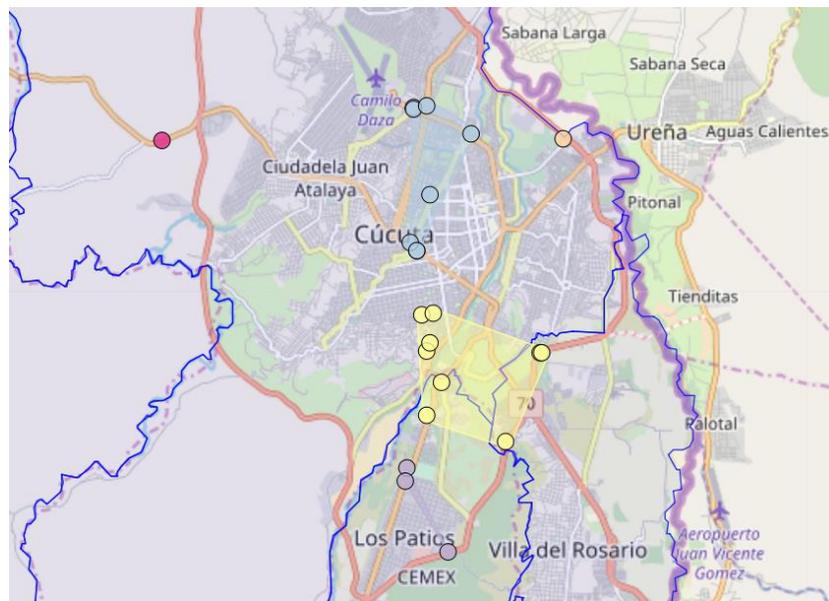
Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,003								
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,125	0,000				
3,25	0,000	0,000	0,000	0,065	0,030	0,302			
3,50	0,000	0,000	0,000	0,101	0,082	0,476	0,962		
3,75	0,000	0,000	0,001	0,122	0,144	0,525	0,987	0,941	
4,00	0,000	0,000	0,001	0,108	0,237	0,447	0,819	0,738	0,661

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 12. Estimación de Conglomerados, Norte de Santander.

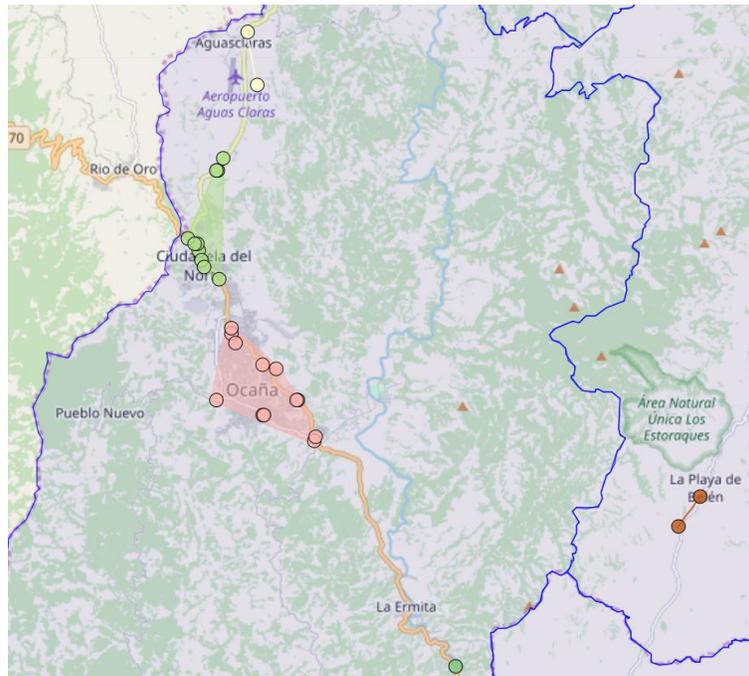


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 14. Estimación de Conglomerados, Ocaña.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Como se puede observar en la figura 13, se identificaron 4 mercados relevantes en Cúcuta, dos de los cuales están compuestos por una única estación de servicio, cercanas a la vía Ocaña-Cúcuta y al Anillo Vial. Un tercer mercado está constituido por distribuidores a lo largo de la Avenida Canal Bogotá o Avenida 5AE y un cuarto mercado con estaciones cercanas a la Avenida Libertadores y al Anillo Vial Oriental. Este último se encuentra compuesto de 4 estaciones de servicio de Cúcuta, 2 en el municipio de Los Patios y 3 en Villa del Rosario. Por otro lado, en el Municipio de Ocaña el análisis de conglomerados encuentra cinco mercados, como se puede observar en la Figura 14. Note que se identifican dos mercados sobre la vía Ocaña-Concepción, cada uno compuesto por dos estaciones de servicio, un mercado alrededor de la Ciudadela Norte y cercanos a la vía Aguachica-Ocaña, un cuarto mercado alrededor del área urbana de Ocaña y con estaciones entre la Avenida Circunvalar y Carrera 10 y, por último, un mercado con una única estación al sureste del área urbana de Ocaña y sobre la vía Ocaña-Cúcuta.

Tabla 9. Resultados SAR con Distancia Crítica Óptima y Controles, Norte de Santander.

	<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Terpel)	-49.07	-53.88	-8,492.36
1(Biomax)	-241.30	-269.18	39,108.92
1(Lavado Vehículo)	-46.57	-62.69	9,564.63
1(Mini-mercado)	-35.44	-40.88	-13,480.65
1(Otros Servicios)	-58.00	-23.86	-14,030.59
Accesos	22.09	-8.30	3,535.26
Competidores (1Km)	-7.68	-40.39	1,266.84
$\rho$		8.32***	
$\rho_x$		-1.44***	
$\sigma_e^2$		86,642.71***	
Observaciones	7,140	7,140	7,140
R-cuadrado		0.51	0.51
Número de EDS	119	119	119

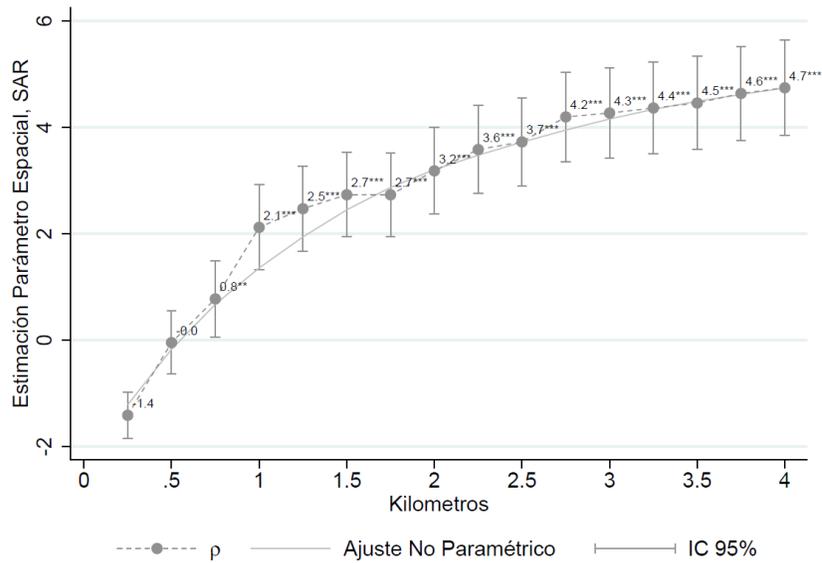
Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite la variable dummy que toma el valor 1 cuando la EDS pertenece a la bandera Otras. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

La Tabla 9 permite concluir que en Norte de Santander no existe evidencia estadística de diferencias en precios entre estaciones de servicio con diferentes banderas. Observe que a pesar de que las estaciones de Terpel y Biomax fijan precios que en promedio están 53 y 269 pesos por debajo de los precios de distribuidores con bandera Otras, estas diferencias no son significativas. Asimismo, el resto de las características no explican de manera importante los precios fijados en el departamento. Es decir, en Norte de Santander la condición más relevante es exclusivamente la geográfica, ya que el parámetro espacial es altamente significativo y positivo.

### 3. 4. Boyacá

Boyacá tiene un área aproximada de 23.189 Km<sup>2</sup> y cuenta con información para 199 estaciones de servicio, lo que implica una densidad de 1.7 distribuidores cada 200 Km<sup>2</sup>. La Figura 15 presenta los resultados del modelo SAR, permitiendo observar que a partir de 750 metros hay evidencia de correlación espacial en precios. Al igual que en el resto de los departamentos de frontera considerados, se encuentra que la importancia de la distancia decrece marginalmente.

Figura 15. Estimaciones del Modelo SAR, Boyacá.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Los resultados de la metodología para determinar la distancia crítica óptima se presentan en las Tablas 10 y 11. Toda vez que para el rango de distancias consideradas se rechaza la serie de hipótesis nulas para cada coeficiente, con excepción del valor crítico 3.75 kilómetros, se toma éste último como radio óptimo. Con la distancia crítica encontrada el modelo de conglomerados genera 90 mercados relevantes y las Figuras 16 a 18 presentan sus resultados.

La Figura 16 da una visión general de los resultados del modelo de conglomerados para Boyacá. Por otro lado, la Figura 17, revela que existen 3 mercados relevantes en Tunja, la capital del departamento. De manera específica, se puede observar un gran mercado compuesto por dieciséis estaciones servicio, aledañas a la Transversal 19, la Avenida Orienta y la Variante de Tunja. Además, se identifica un mercado con un único distribuidor al sur del área urbana de Tunja y sobre la Autopista Norte (Ruta Nacional 55), y un tercer mercado con siete estaciones al norte de la ciudad, una de ellas ubicadas en el municipio de Motavita y tres en Combita. En Duitama, se determinaron 2 mercados relevantes, uno de los cuales se compone de quince estaciones de servicio al interior de la zona urbana y el segundo tiene dos distribuidores sobre la Troncal del Norte (Ruta Nacional 55). En Paipa, se identifican 2 mercados relevantes con estaciones cercanas o sobre a la Ruta Nacional 55. En Sogamoso, se encontraron también 2 mercados, el primero tiene trece estaciones de servicio y el segundo, al norte del municipio, tiene seis estaciones, una de las cuales se ubica al interior de Nobsa.

Tabla 10. Estadísticos  $\chi_{gl=1}^2$  para comparar entre coeficientes  $\rho_{dc}$ , Boyacá.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	12,48								
2,25	30,86	22,23							
2,50	37,48	29,23	7,07						
2,75	65,98	63,29	41,70	36,09					
3,00	65,36	59,35	37,13	30,14	1,46				
3,25	64,53	55,81	34,25	27,29	3,20	1,74			
3,50	62,40	51,69	31,54	25,06	4,49	3,06	1,37		
3,75	71,19	60,94	40,38	33,71	10,51	9,12	7,68	9,88	
4,00	71,76	60,73	40,91	34,49	12,56	11,13	9,56	9,44	2,12

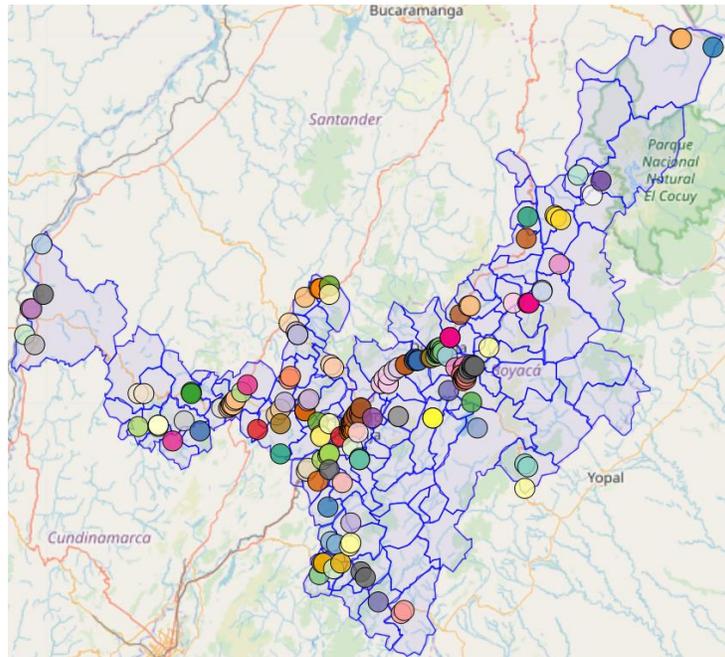
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 11. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ , Boyacá.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,000								
2,25	0,000	0,000							
2,50	0,000	0,000	0,008						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,227				
3,25	0,000	0,000	0,000	0,000	0,074	0,187			
3,50	0,000	0,000	0,000	0,000	0,034	0,080	0,242		
3,75	0,000	0,000	0,000	0,000	0,001	0,003	0,006	0,002	
4,00	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,002	0,145

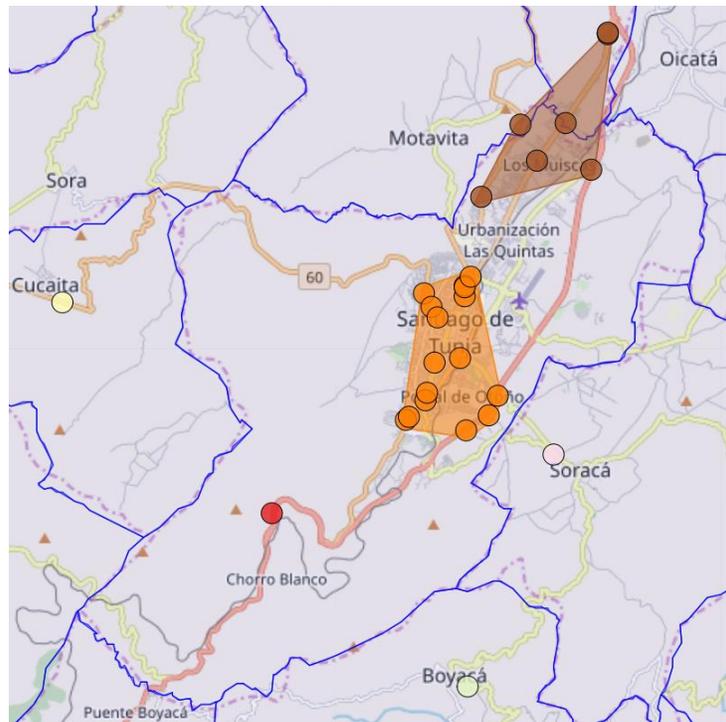
Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor < 0.05 entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 16. Estimación de Conglomerados, Boyacá.



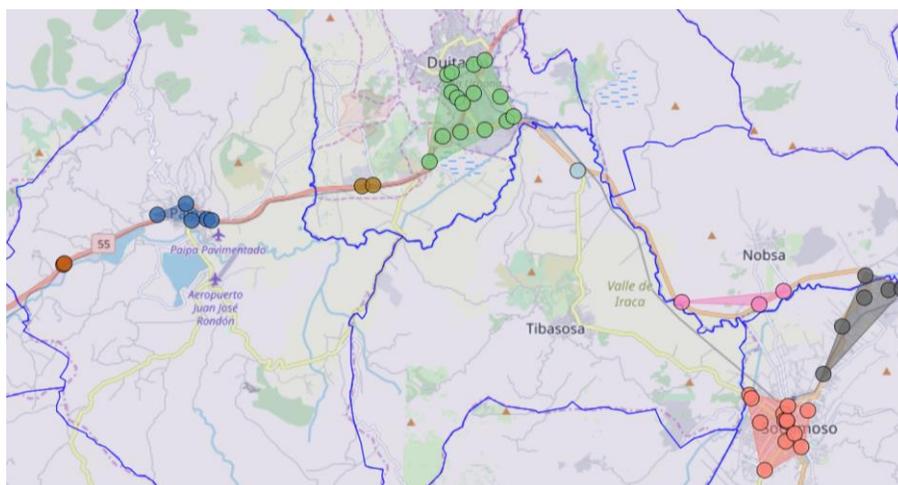
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 17. Estimación de Conglomerados, Tunja.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 18. Estimación de Conglomerados, Duitama Sogamoso y Paipa.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Tabla 12. Resultados SAR con Distancia Crítica Óptima y Controles, Boyacá.

	<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Terpel)	-141.83***	-147.97***	-12,660.17
1(Petromil)	-140.38	-140.25	11,924.87
1(Texaco)	-151.22	-156.20*	-17,458.97
1(Petrobras)	-226.27***	-243.39***	-18,903.36
1(Mobil)	-103.55	-125.79*	-6,706.26
1(Biomax)	-67.35	-62.99	15,913.15
1(Lavado Vehículos)	104.48**	75.66	47,795.92**
1(Mini-mercado)	3.65	-6.28	-18,233.82
1(Servicios)	-10.66	9.34	-9,010.93
Competidores (1Km)	-11.74	-11.31	5,252.48
$\rho$		4.68***	
$\rho_X$		-2.40***	
$\sigma_e^2$		10,990.99***	
Observaciones	11,940	11,940	11,940
R-cuadrado		0.76	0.76
Número EDS	199	199	199

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite la variable dummy que toma 1 como valor cuando la estación de servicio pertenece a una bandera del grupo Otras. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

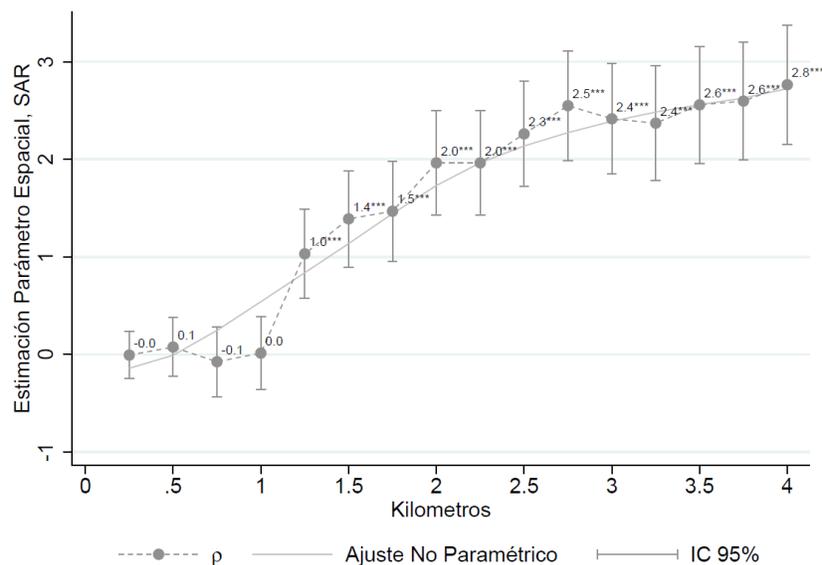
La Tabla 12 permite observar que Petrobras fija precios menores en 243 pesos, en comparación con los precios de distribuidores con bandera Otras. Asimismo, se puede señalar que Texaco,

Mobil y Terpel reaccionan a los precios bajos de Petrobras, que es la bandera con precios más bajos. Luego, en comparación con estaciones de bandera Otras, Terpel fija precios 148 pesos más bajos, Texaco tiene precios menores en 156 pesos y Mobil tiene precios menores en 126 pesos. Note que el modelo de datos panel, que supone irrelevante la dimensión espacial de la competencia en precios, señala que las estaciones de servicio con lavado de vehículo fijan precios más altos en 104 pesos. Sin embargo, lo anterior no es cierto a partir de los resultados en el modelo SAR, que corrige el sesgo de las estimaciones cuando  $\rho$  es diferente de cero. Luego, según del modelo autorregresivo espacial, tener lavadero de vehículos no es una característica relevante para fijar precios, ya que el coeficiente no es estadísticamente significativo. Por último, ninguna otra característica es relevante para explicar los precios.

### 3. 5. Arauca

El departamento de Arauca tiene un área aproximada de 23.818 Km<sup>2</sup> y se cuenta con información para 25 estaciones de servicio, luego la densidad es de 1 estación de servicio por cada 1.000 Km<sup>2</sup>. La Figura 19 presenta los resultados del modelo SAR para Arauca, e indica que existe correlación espacial a partir de 1.25 kilómetros. Los resultados del proceso de determinación de distancia óptima, que se presentan en las Tablas 13 y 14, señalan que 3.25 kilómetros es el radio con el cual se debe determinar los mercados relevantes, ya que se rechaza la hipótesis nula de la serie de pruebas para las distancias críticas menores a este valor.

Figura 19. Estimaciones del Modelo SAR, Arauca



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabla 13. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ , Arauca.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,70								
2,25	17,66	24,65							
2,50	17,66	24,65							
2,75	33,59	45,22	22,45	22,45					
3,00	39,65	46,34	22,48	22,48	7,36				
3,25	30,81	35,26	13,18	13,18	2,07	86,04			
3,50	20,73	21,60	5,97	5,97	0,51	2,58	0,16		
3,75	25,97	27,08	10,42	10,42	2,95	0,00	1,11	5,69	
4,00	27,47	28,73	11,66	11,66	3,71	0,12	1,74	7,71	3,38

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 14. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ , Arauca.

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,70								
2,25	17,66	24,65							
2,50	17,66	24,65							
2,75	33,59	45,22	22,45	22,45					
3,00	39,65	46,34	22,48	22,48	7,36				
3,25	30,81	35,26	13,18	13,18	2,07	86,04			
3,50	20,73	21,60	5,97	5,97	0,51	2,58	0,16		
3,75	25,97	27,08	10,42	10,42	2,95	0,00	1,11	5,69	
4,00	27,47	28,73	11,66	11,66	3,71	0,12	1,74	7,71	3,38

Fuente: Cada entrada se computa de la siguiente manera  $1 - F^{-1}(\chi^2_{gl=1})$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

La Tabla 15 presenta los resultados del modelo SAR considerando características observables en lugar de efectos fijos de estación de servicio. Observe que en el departamento de Arauca las 25 estaciones de servicio consideradas son de bandera Terpel, luego es imposible considerar esta dimensión dentro de la regresión presentada a continuación. Sin embargo, se puede observar que las estaciones de servicio que ofrecen servicio de lavado de vehículos tienen un precio promedio superior en \$613 pesos y que aumentar el número de accesos viales en una unidad implica que se ofrezca la gasolina a un precio \$107 pesos mayor, aunque esta última característica no es estadísticamente significativa. Por último, note que el parámetro espacial es muy significativo en los resultados de la Tabla 15, resaltando la importancia que tiene esta dimensión en la determinación de los precios de los minoristas en Arauca.

Tabla 15. Resultados SAR con Distancia Crítica Óptima y Controles, Arauca.

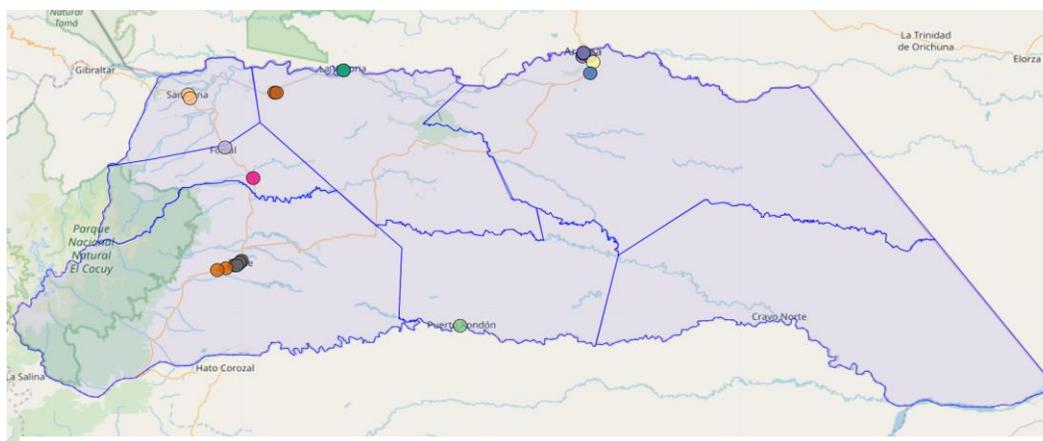
	<i>Variables Dependientes: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Lavado Vehículos)	682.44	613.40*	91,231.02***
1(Mini-mercado)	-176.03	-119.33	-23,200.18
Accesos	32.58	107.75	5,366.23
Competidores (1Km)	35.69	37.62	651.19
$\rho$		4.21***	
$\rho_X$		-1.67***	
$\sigma_e^2$		122,506.44***	
Observaciones	2,400	2,400	2,400
R-cuadrado		0.58	0.58
Número EDS	25	25	25

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar robustos.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

El modelo de conglomerados con la distancia óptima considerada, es decir, 3.25 kilómetros, genera 11 mercados al interior del departamento de Arauca. Las Figuras 20 a 22 dan cuenta de estas estimaciones. De esta manera, la Figura 20 presenta un panorama general de los resultados. La Figura 21 indica que en el municipio de Arauca se identificaron 3 mercados relevantes, el primero de los cuales está constituido por seis estaciones de servicio y aledañas a la Carrera 19 y la Calle 1 Sur. Los dos mercados restantes están compuestos por una estación de servicio cada uno y se ubican sobre la Ruta Nacional 66, al sur del área urbana de Arauca. En Tame se identificaron 2 mercados relevantes, el primero compuesto por seis estaciones de servicio al interior de la zona urbana de Tame y el segundo con dos distribuidores sobre la Ruta 66 al suroeste de la zona urbana.

Figura 20. Estimación de Conglomerados, Arauca.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.25 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclidianas.



de sus resultados concluyen que los distribuidores minoristas ubicados a lo largo de diferentes corredores principales de una ciudad forman diferentes mercados.

Por ende, la nueva metodología implica desarrollar las siguientes tareas: 1. Ubicar las estaciones de servicio cercanas a las salidas de la ciudad<sup>9</sup>; 2. Implementar la metodología inicial y correr el análisis de conglomerados excluyendo las estaciones identificadas en el primer paso; 3. Estimar el modelo SAR con controles y efectos fijos de mercado relevante, es decir, aquellos identificados por el análisis de conglomerados y los construidos considerando las salidas de la ciudad. De esta forma, se espera que, si los coeficientes asociados a los efectos fijos de cada mercado son significativos, se demostraría que las estaciones de servicio que se ubican sobre las principales vías de entrada y de salida de las ciudades conforman un mercado independiente de aquellos determinados a través del ejercicio tradicional de *cluster*.

#### 4. 1. Bogotá

La ciudad de Bogotá tiene un área de aproximada de 1,775 km<sup>2</sup> y cuenta con información para 319 estaciones de servicio. Así, en promedio se tiene una densidad de 1.8 estaciones de servicio por cada 10 km<sup>2</sup>. Para Bogotá se identificaron 7 mercados relevantes a partir de las vías de salida e ingreso. La Figura 23 presenta los mercados viales identificados, que a continuación se listan en detalle:

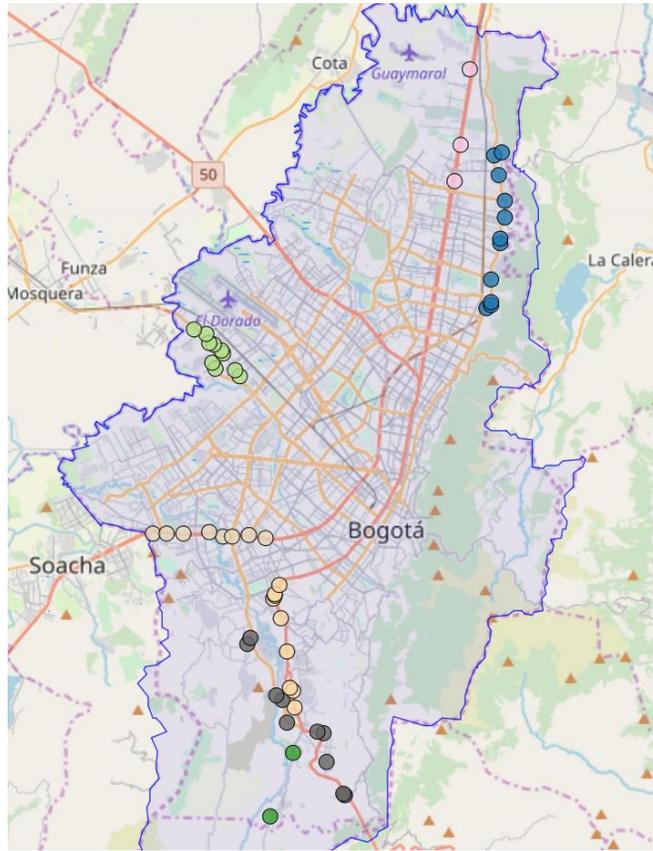
1. Avenida Boyacá, Vía Bogotá-Villavicencio: 10 EDS
2. Avenida Carrera 1, Vía Usme: 2 EDS
3. Avenida Caracas, Vía Bogotá-Villavicencio: 9 EDS
4. Autopista Sur, Vía Soacha: 8 EDS
5. Avenida Calle 17 y Diagonal 16, Vía Bogotá-Funza: 10 EDS
6. Avenida Carrera 7, a partir de la Calle 120: 11 EDS
7. Autopista Norte, a partir de la Calle 170: 3 EDS

La Figura 24 presenta los resultados de estimar el modelo SAR con diferentes distancias críticas. Por otro lado, las Tablas 16 y 17 presentan los resultados del algoritmo para determinar la distancia óptima. Se puede observar que la distancia óptima en este caso es 1.5 kilómetros, ya que con esta distancia no es posible rechazar la serie de pruebas estadísticas que se describen en la metodología inicial. Luego, con esta distancia se identifican 42 mercados relevantes, 35 de los cuales resultan de la estimación por el método de conglomerados. En la Tabla 18 y la Figura 25 se presenta las estimaciones del Modelo SAR que incluye efectos fijos de conglomerados o mercados relevantes. La Figura 26 permite observar los resultados de la estimación por conglomerados, al igual que visualizar los mercados viales.

---

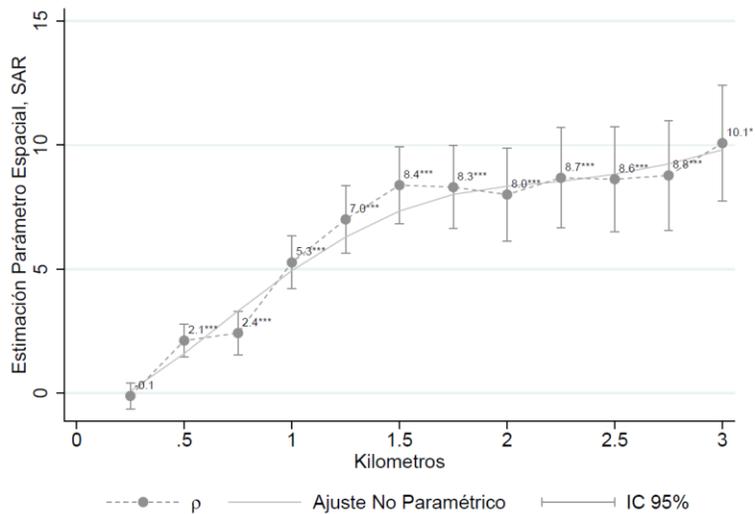
<sup>9</sup> Para ubicar las estaciones de servicio cercanas a las salidas de cada ciudad se usó la georreferenciación del Censo de Caracterización de Estaciones de Servicio del Fondo Soldicom y el mapa de carreteras de INVIAS, que se puede visualizar en el siguiente link: <https://hermes.invias.gov.co/carreteras/>.

Figura 23. Mercados Viales, Bogotá



Fuente: Cálculos propios con datos de georreferenciación del Fondo Soldicom.

Figura 24. Estimaciones del Modelo SAR, Bogotá.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabla 16. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50
0,75	0,86								
1,00	48,79	85,16							
1,25	68,55	87,46	20,80						
1,50	87,65	104,61	39,70	19,11					
1,75	75,24	86,41	30,20	10,55	0,11				
2,00	57,11	62,50	18,57	3,88	0,91	0,91			
2,25	62,95	68,00	24,23	8,38	0,37	0,81	5,90		
2,50	57,91	61,76	21,36	6,87	0,20	0,46	3,09	0,07	
2,75	56,94	60,22	21,40	7,31	0,46	0,83	3,47	0,10	0,49

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / [Var(\rho_{dc'}) - Var(\rho_{dc''])]$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 17. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50
0,75	0,354								
1,00	0,000	0,000							
1,25	0,000	0,000	0,000						
1,50	0,000	0,000	0,000	0,000					
1,75	0,000	0,000	0,000	0,001	0,738				
2,00	0,000	0,000	0,000	0,049	0,339	0,341			
2,25	0,000	0,000	0,000	0,004	0,544	0,368	0,015		
2,50	0,000	0,000	0,000	0,009	0,653	0,496	0,079	0,797	
2,75	0,000	0,000	0,000	0,007	0,496	0,362	0,063	0,754	0,486

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi^2_{gl=1})$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

La Tabla 18 presenta las estimaciones del Modelo SAR con características observables de las estaciones de servicio y efectos fijos de mercados relevantes, tanto viales como los estimados por el método de conglomerados. Sin embargo, las estimaciones de los coeficientes de los mercados relevantes se presentan en la Figura 25 por brevedad. Se puede observar que, en Bogotá, las estaciones con banderas de Terpel, Mobil, Petrobras y Biomax, tiene precios menores a los de estaciones con bandera Otras. Más aún, la diferencia con este último tipo de distribuidores varía entre 221 pesos y 78 pesos. Ahora bien, tener lavado de vehículos, tener mini-mercado y el número de accesos no son variables importantes para fijar los precios de la gasolina. Sin embargo, las estaciones con otros servicios - que incluye cambio de aceite, monta

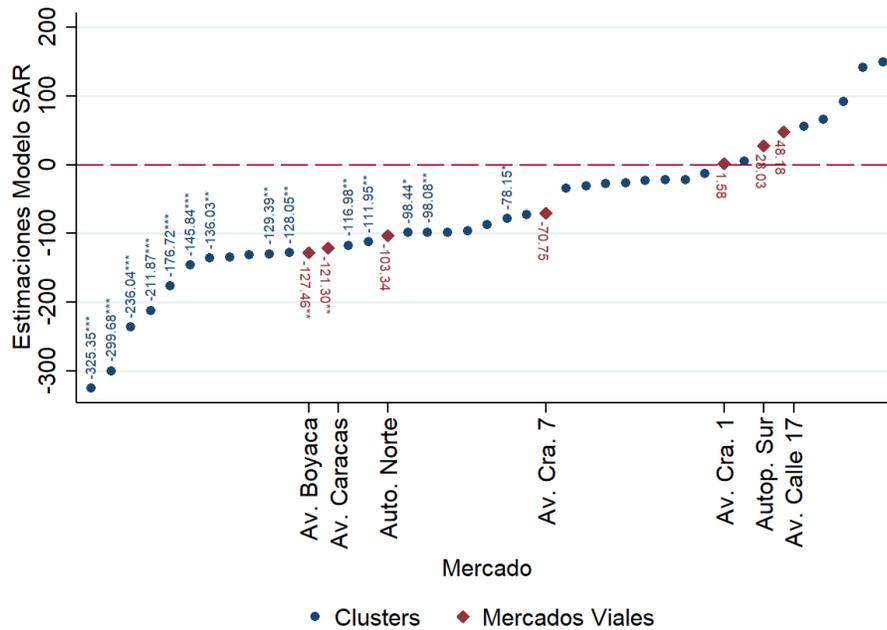
llantas y alineación y balanceo – si tienen precios mayores en 42 pesos, en comparación con las estaciones que no cuentan con ninguno de estos servicios.

Tabla 18. Estimación modelo SAR incluyendo Efectos Fijos de Mercados Relevantes.

	<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Texaco)	-59.45	-60.30	-12,885.84
1(Mobil)	-110.47***	-105.41***	3,092.75
1(Terpel)	-202.71***	-221.23***	-11,998.91
1(Zeuss)	-120.72	78.78	79,559.51
1(Biomax)	-105.70***	-77.70**	63,690.64***
1(Petrobras)	-95.21**	-78.45**	27,730.61
1(Lavado Vehículos)	-8.83	-23.61	-26,445.44
1(Mini-mercado)	-26.29	-24.32	-5,304.98
1(Otros Servicios)	38.37	41.70*	-3,293.60
Accesos	5.16	4.52	1,753.06
Competidores (1Km)	-13.45**	-13.17	-207.20
$\rho$		8.40***	
$\rho_x$		-1.88***	
$\sigma_e^2$		13,639.38***	
Observaciones	19,140	19,140	19,140
R-cuadrado		0.81	0.81
Número EDS	319	319	319

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite las estaciones de servicio con bandera Otras. Las regresiones incluyeron efectos fijos de mercado relevante, los cuales se presentan en la Figura 24 por brevedad. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Figura 25. Estimación de Efectos Fijos de Mercados Relevantes, Bogotá.

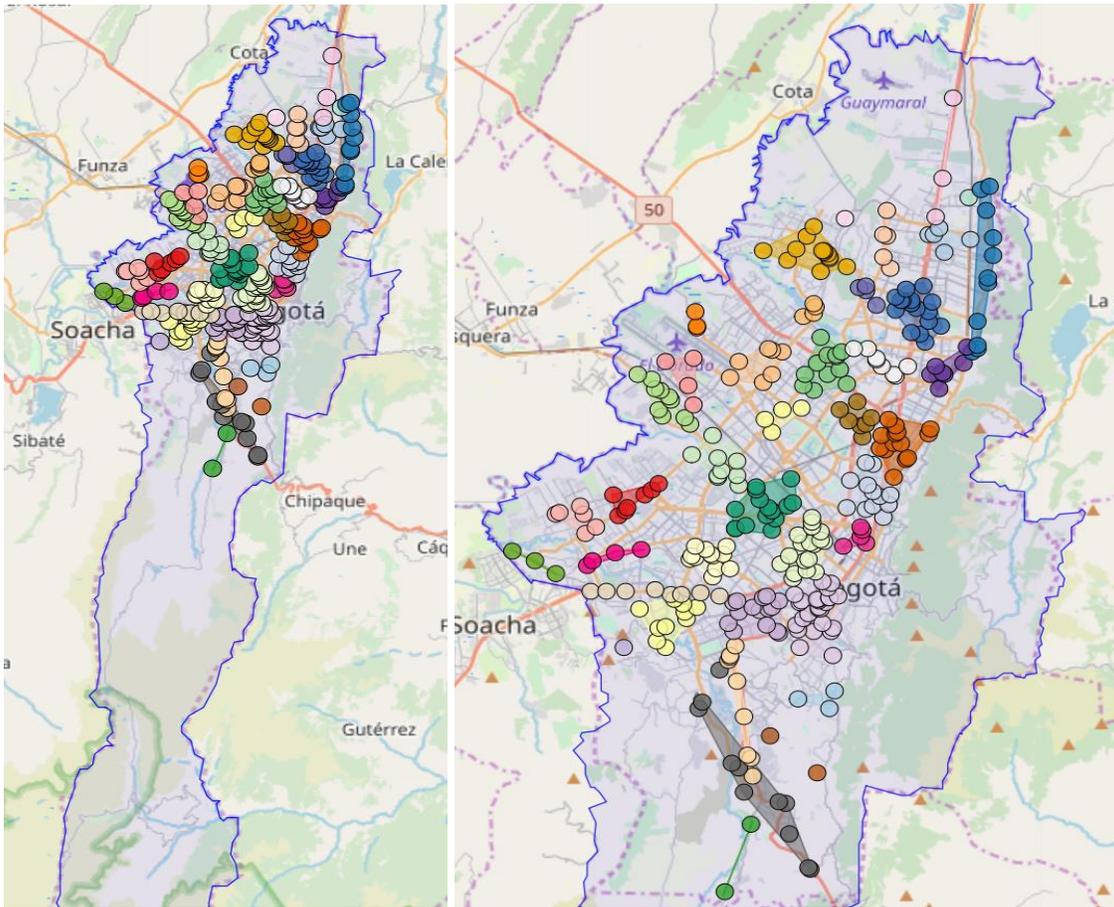


Fuente: Cada punto corresponde a la estimación máxima verosímil de la variable dummy que representa cada mercado. Se omite uno de los mercados para realizar comparaciones. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

La Figura 25 presenta las estimaciones de los efectos fijos para los mercados relevantes, a partir del modelo SAR con controles. Es importante señalar que se omite un mercado relevante de aquellos estimados por el método de conglomerados. Luego, de los 42 mercados relevantes que se identificaron para Bogotá, se presentan 41, el restante se usa como base de comparación.<sup>10</sup> Se observa que los precios de los mercados de la Avenida Caracas y la Avenida Boyacá son estadísticamente diferentes a los precios del mercado omitido. Lo anterior ofrece evidencia de la importancia de los corredores viales, aunque para el resto de los mercados viales la diferencia no sea estadísticamente significativa.

<sup>10</sup> Note además que es irrelevante cual es el mercado omitido, ya que la significancia estadística de las diferencias no se ve alterada, como si se ve alterada la magnitud de la diferencia en precios.

Figura 26. Estimación de Conglomerados y Mercados Viales, Bogotá



Fuente: Estimaciones por clústeres jerárquicos usando la metodología promedio y una distancia crítica de 1.5 kilómetros e identificación de estaciones en corredores viales. Las distancias usadas para generar los clústeres fueron computadas como distancias esféricas, también conocidas como geodésicas.

## 4. 2. Cali

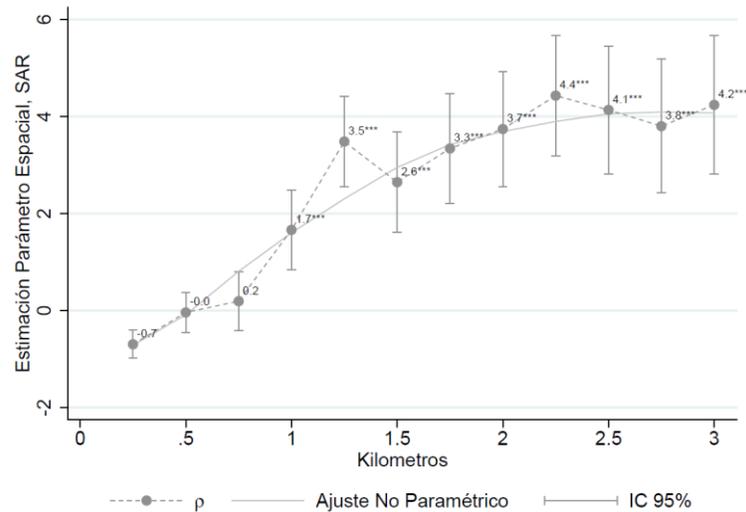
La ciudad de Cali tiene un área de aproximada de 619 km<sup>2</sup> y cuenta con información para 125 estaciones de servicio<sup>11</sup>. Es decir, en promedio se puede encontrar 2 estaciones de servicio por cada 10 km<sup>2</sup>. Para la ciudad de Cali se identificaron 7 mercados relevantes a partir de las vías de salida, los cuales se detallan a continuación:

1. Calle 25 y Carrera 100, vía Jamundí – Cali (Red Vial 2504): 8 Estaciones de Servicio.
2. Calle 5, vía Jamundí – Cali: 8 EDS

<sup>11</sup> Es importante resaltar que aquí se está considerando el área total del municipio y no solo el área urbana, caso en el cual aumentaría la densidad promedio.

3. Calle 18, vía Jamundí – Cali: 3 EDS
4. Calle 70, vía Cali - Yumbo (Red Vial 2301): 14 EDS
5. Carrera 1, vía Cali - Palmira (Red Vial 2505): 9 EDS
6. Avenida 4 Oeste, Cali – Loboguerrero (Red Vial 1901): 2 EDS
7. Carrera 25 y Carrera 27: 4 EDS

Figura 27. Estimaciones del Modelo SAR, Cali



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

La Figura 27 presenta los resultados de estimar el modelo SAR para Cali, usando diferentes distancias críticas. Se observa que existe una importante correlación espacial en los precios de la gasolina corriente, puesto que con una distancia mínima de 1 kilómetro hay evidencia de competencia en precios entre distribuidores minoristas. Aunque el parámetro espacial estimado aumenta en magnitud a medida que se incrementa la distancia crítica del ejercicio, se resalta también que a medida que incrementamos la distancia crítica, aumenta también el intervalo de confianza del estimador puntual. En otras palabras, la predicción para una estación de servicio es más precisa si se consideran sus vecinos más cercanos.

Las Tablas 19 y 20 presentan los resultados del procedimiento para determinar la distancia crítica óptima. Se puede observar que este algoritmo indica que la distancia óptima para la ciudad de Cali es 2.5 kilómetros, ya que esta es la menor distancia para la cual no se rechazan ninguna de las pruebas de hipótesis que permiten comparar el coeficiente de correlación con el coeficiente asociado a distancias mayores. Ahora el análisis de conglomerados usando la distancia crítica óptima genera como resultado 7 mercados relevantes, adicionales a los 7 identificados a partir de las rutas de salida de la ciudad. La Figura 28 presenta los resultados de implementar la metodología descrita.

Tabla 19. Estadísticos  $\chi_{gl=1}^2$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75
1,00	19,64								
1,25	64,71	57,78							
1,50	28,03	9,24	14,84						
1,75	37,29	18,09	0,21	9,46					
2,00	43,25	23,86	0,54	15,70	6,31				
2,25	56,25	36,62	5,90	30,47	22,20	16,95			
2,50	43,41	24,65	2,23	15,48	6,84	2,31	2,25		
2,75	33,47	16,39	0,45	7,66	1,72	0,04	5,69	3,62	
3,00	39,87	22,08	2,35	12,96	5,57	2,07	0,40	0,21	8,80

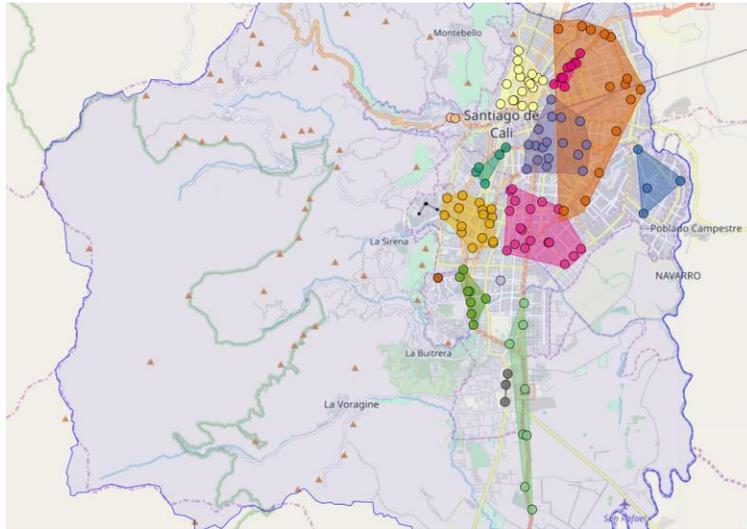
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 20. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75
1,00	0,000								
1,25	0,000	0,000							
1,50	0,000	0,002	0,000						
1,75	0,000	0,000	0,648	0,002					
2,00	0,000	0,000	0,463	0,000	0,012				
2,25	0,000	0,000	0,015	0,000	0,000	0,000			
2,50	0,000	0,000	0,136	0,000	0,009	0,128	0,134		
2,75	0,000	0,000	0,501	0,006	0,190	0,851	0,017	0,057	
3,00	0,000	0,000	0,125	0,000	0,018	0,150	0,529	0,648	0,003

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 28. Estimación de Conglomerados y Mercados Viales, Cali



Fuente: Estimaciones por clústeres jerárquicos usando la metodología promedio y una distancia crítica de 2.5 kilómetros e identificación de estaciones en corredores viales. Las distancias usadas para generar los clústeres fueron computadas como distancias esféricas, también conocidas como geodésicas.

Tabla 21. Estimación modelo SAR incluyendo Efectos Fijos de Mercados Viales.

Variable Dependiente: Precio Gasolina		
	MCO	SAR
<i>Mercados por Conglomerados:</i>		
Mercado 2	-67.78	-101.70*
Mercado 3	-107.21*	-110.56
Mercado 4	-94.98	-149.14**
Mercado 5	-115.16*	-165.85***
Mercado 6	-10.43	-124.37
Mercado 7	-231.64*	-305.87**
<i>Mercados Viales:</i>		
Mercado 1	-36.69	-147.54*
Mercado 2	-48.21	-113.53
Mercado 3	21.68	-118.28
Mercado 4	-205.77***	-307.51***
Mercado 5	-129.04*	-192.69**
Mercado 6	-21.43	-127.98
Mercado 7	15.37	-104.81
<i>Otros Controles:</i>		
1(Terpel)	8.47	15.06
1(Lavadero Vehículos)	-29.25	-29.60
1(Mini-mercado)	27.73	49.89**
1(Otros Servicios)	45.89	42.59
Accessos	-28.32	-33.76**
Competidores (1Km)	-27.47**	-11.17
$\rho$		4.21***
$\sigma_e^2$		11,066.24***
Observaciones	7,500	7,500
R-cuadrado		0.86
Número EDS	125	125

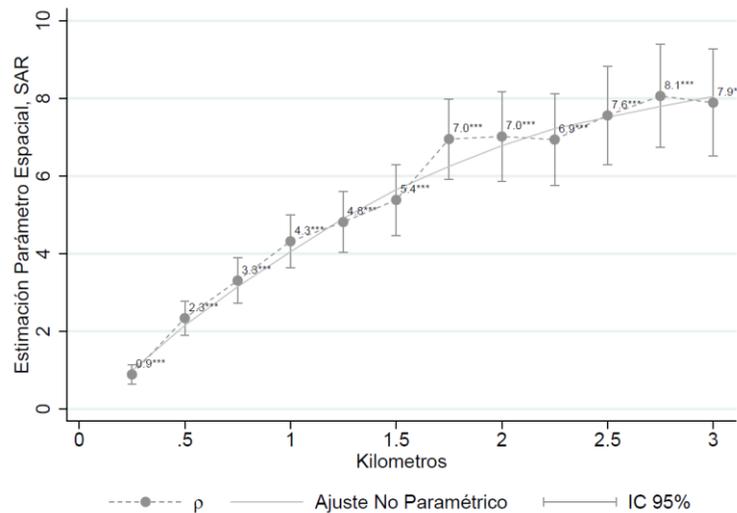
Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite el Mercado 1 como referencia para realizar comparaciones. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Por último, la Tabla 21 presenta las estimaciones de la nueva metodología que permite determinar mercados relevantes, separando aquellos que se ubican en las vías de entrada y de salida de las ciudades. Se observa que, como era esperado, el comportamiento de los precios es diferente entre mercados. Por ejemplo, los mercados por conglomerados 2, 4, 5 y 7 y los mercados viales 1, 4 y 5, son estadísticamente diferentes del mercado omitido, que se elige de entre los mercados por conglomerados. Sin embargo, se observa que no todas las diferencias entre mercados son significativas, lo que podría explicarse por la proximidad entre algunos conglomerados y las salidas de la ciudad.

### 4. 3. Medellín

La ciudad de Medellín tiene un área de aproximada de 381 km<sup>2</sup> y cuenta con información para 105 estaciones de servicio. Luego, en promedio se tiene una densidad de 2.8 estaciones de servicio por cada 10 km<sup>2</sup>. La Figura 29 presenta las estimaciones del modelo SAR para Medellín.

Figura 29. Estimaciones del Modelo SAR, Medellín



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Como se puede observar, en Medellín hay evidencia de competencia en precios incluso a una distancia de 250 metros. Esto puede explicarse por la mayor densidad de las estaciones de servicio en la ciudad, en comparación con Cali. Además, como revelan las Tablas 22 y 23, la distancia óptima para Medellín es de 1.75 kilómetros. En esta ciudad se identificaron 6 mercados relevantes a partir de las vías de salida. A continuación, se describen estos mercados viales:

1. Carrera 64, sector Medellín – Hoyo Rico (Red vial 2510): 5 EDS
2. Vía Las Palmas, sector Medellín – Las Palmas (Red vial 5601): 3 EDS
3. Vía Santa Elena: 1 EDS
4. Calle 49, vía a Santa Elena: 3 EDS
5. Conexión Vial Guillermo Gaviria Correa, sector San Cristóbal: 1 EDS
6. Carrera 55 y Calle 43 A Sur, San Antonio de Prado: 1 EDS

El análisis de conglomerados usando la distancia óptima obtuvo como resultado 15 mercados, adicionales a los 6 mercados identificados sobre las vías de salida de la ciudad. Luego, en Medellín se encuentran 21 mercados relevantes. Dichos resultados se presentan en la Figura 30.

Finalmente, los resultados del último paso de la metodología modificada se presentan en la Tabla 24. Al igual que en el resto de las ciudades se omitió un mercado determinado por el método de conglomerados para realizar comparaciones. Se observa que solo dos mercados relevantes definidos por la metodología de conglomerados tienen un comportamiento de precios diferente, estos son los mercados 7 y 11. Adicionalmente, los mercados viales 2 y 3 tienen un comportamiento de precios diferente. Observe que la densidad de los mercados identificados a través de las rutas de ingreso y salida de la ciudad es baja, lo que podría explicar las razones por las que algunos mercados no sean estadísticamente diferente del resto.

Tabla 22. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25
0,50	21,82								
0,75	33,73	12,18							
1,00	52,66	30,86	20,45						
1,25	57,26	35,48	23,73	5,38					
1,50	60,16	38,72	26,55	10,08	4,85				
1,75	92,79	70,99	59,63	40,13	35,97	40,73			
2,00	81,95	60,48	48,32	31,00	25,65	21,79	0,07		
2,25	78,59	57,29	45,12	28,32	22,95	18,62	0,00	0,89	
2,50	86,48	65,18	53,01	36,08	30,70	26,49	3,12	5,23	7,89

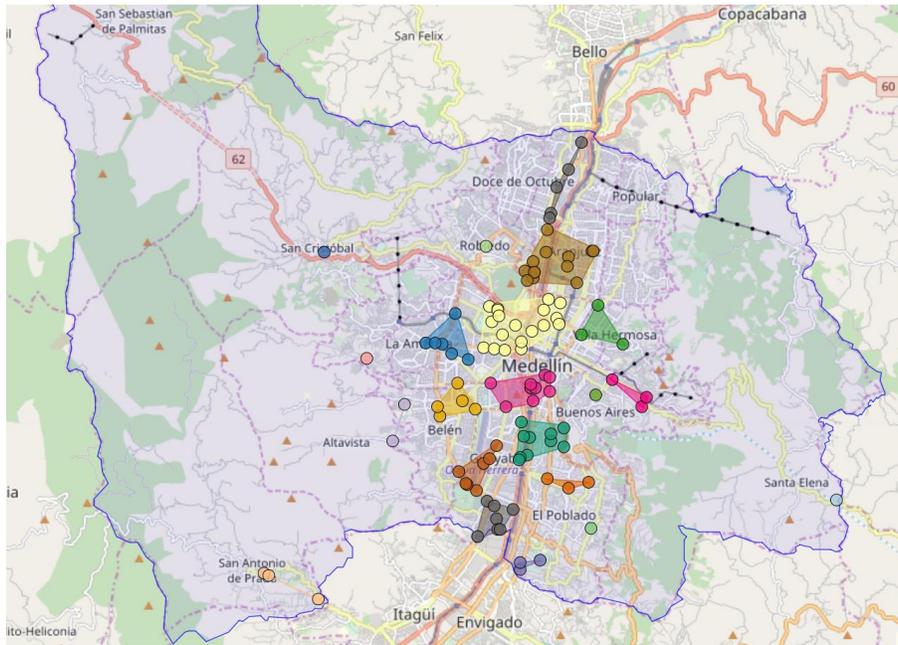
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 23. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25
0,50	0,000								
0,75	0,000	0,000							
1,00	0,000	0,000	0,000						
1,25	0,000	0,000	0,000	0,020					
1,50	0,000	0,000	0,000	0,001	0,028				
1,75	0,000	0,000	0,000	0,000	0,000	0,000			
2,00	0,000	0,000	0,000	0,000	0,000	0,000	0,791		
2,25	0,000	0,000	0,000	0,000	0,000	0,000	0,958	0,345	
2,50	0,000	0,000	0,000	0,000	0,000	0,000	0,077	0,022	0,005

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi^2_{gl=1})$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 30. Estimación de Conglomerados y Mercados Viales, Medellín



Fuente: Estimaciones por clústeres jerárquicos usando la metodología promedio y una distancia crítica de 1.75 kilómetros e identificación de estaciones en corredores viales. Las distancias usadas para generar los clústeres fueron computadas como distancias esféricas, también conocidas como geodésicas.

Tabla 24. Estimación modelo SAR incluyendo Efectos Fijos de Mercados Viales.

<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR
<i>Mercados por Conglomerados:</i>		
Mercado 2	21.52	-74.43
Mercado 3	162.34	-97.46
Mercado 4	6.83	50.41
Mercado 5	-3.01	16.87
Mercado 6	-86.32	-99.76
Mercado 7	17.70	-197.86*
Mercado 8	-22.64	-14.75
Mercado 9	71.84	-25.35
Mercado 10	42.69	-87.65
Mercado 11	-111.49	-203.80*
Mercado 12	-31.09	-94.78
Mercado 13	79.32	-5.36
Mercado 14	107.41	-21.76
Mercado 15	45.29	561.90
<i>Mercados Viales:</i>		
Mercado 1	-54.01	-23.57
Mercado 2	133.61	-421.87**
Mercado 3	311.41**	268.06**
Mercado 4	-53.72	972.73
Mercado 5	69.69	-19.68
Mercado 6	120.73	-8.45
<i>Otros Controles:</i>		
1(Terpel)	-101.34***	-80.35**
1(Lavado Vehículos)	-32.42	18.03
1(Mini-mercado)	4.84	-27.14
1(Otros Servicios)	6.89	89.42*
Accesos	-2.15	-46.64*
Competidores (1Km)	-21.23*	-0.53
$\rho$		6.90***
$\sigma_e^2$		11,461.04***
Observaciones	6,300	6,300
R-cuadrado		0.86
Número de Observaciones	105	105

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite el Mercado 1 como referencia para realizar comparaciones. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

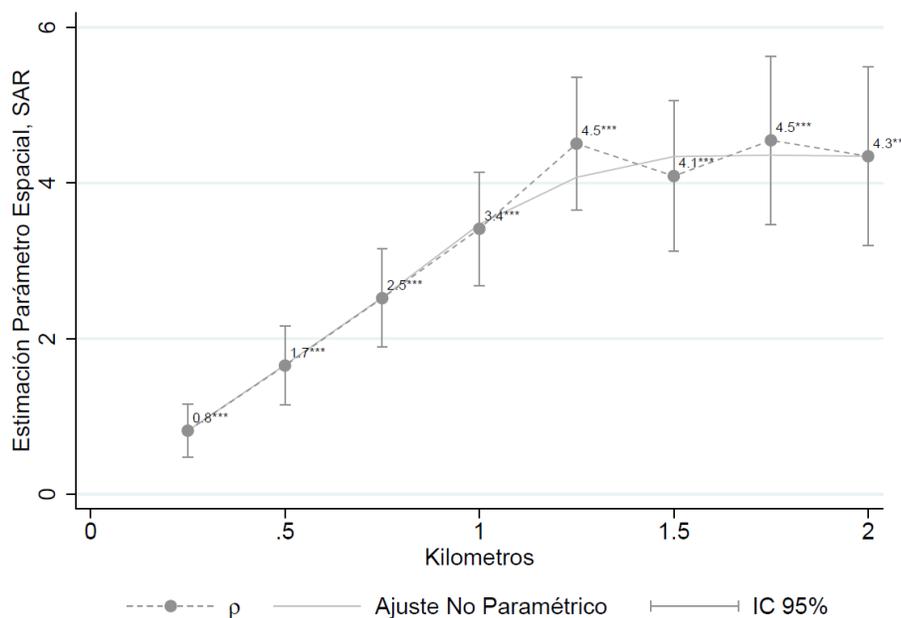
#### 4. 4. Barranquilla

La ciudad de Barranquilla tiene un área aproximada de 154 km<sup>2</sup> y cuenta con información para 88 estaciones de servicio. Es decir, en promedio se puede encontrar 5.7 estaciones de servicio por cada 10 km<sup>2</sup>. Para Barranquilla se identificaron 8 mercados relevantes a partir de las vías de salida, los cuales se detallan a continuación:

1. Vía Barranquilla – Tubara: 1 EDS
2. Avenida Carrera 38: 3 EDS
3. Avenida Calle 110: 6 EDS
4. Avenida Calle 109: 4 EDS
5. Avenida Calle 30, Vía Soledad: 8 EDS
6. Avenida Carrera 46: 5 EDS
7. Avenida Calle 17, Vía Malambo: 5 EDS
8. Avenida Carrera 9, Puente Laureano Gómez: 2 EDS

La Figura 31 presenta las estimaciones del modelo SAR para Barranquilla. Se observa que a partir de 250 metros existe competencia en precios, ya que el coeficiente de correlación espacial es significativo al considerar esta distancia crítica. Al igual que en el resto de ciudades se tiene que la importancia de la dimensión espacial es marginalmente decreciente. Las Tablas 25 y 26 señalan que la distancia crítica óptima es de 1.25 Km.

La Figura 31. Estimaciones del Modelo SAR, Barranquilla



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabla 25. Estadísticos  $\chi_{gl=1}^2$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	0,25	0,50	0,75	1,00	1,25	1,50	1,75
0,50	8,26						
0,75	19,88	12,29					
1,00	34,59	28,14	16,30				
1,25	53,23	47,62	35,93	19,63			
1,50	34,08	25,86	14,63	3,84	2,96		
1,75	37,70	29,45	18,42	7,40	0,02	3,83	
2,00	30,62	22,51	12,78	4,11	0,17	0,71	1,13

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 26. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

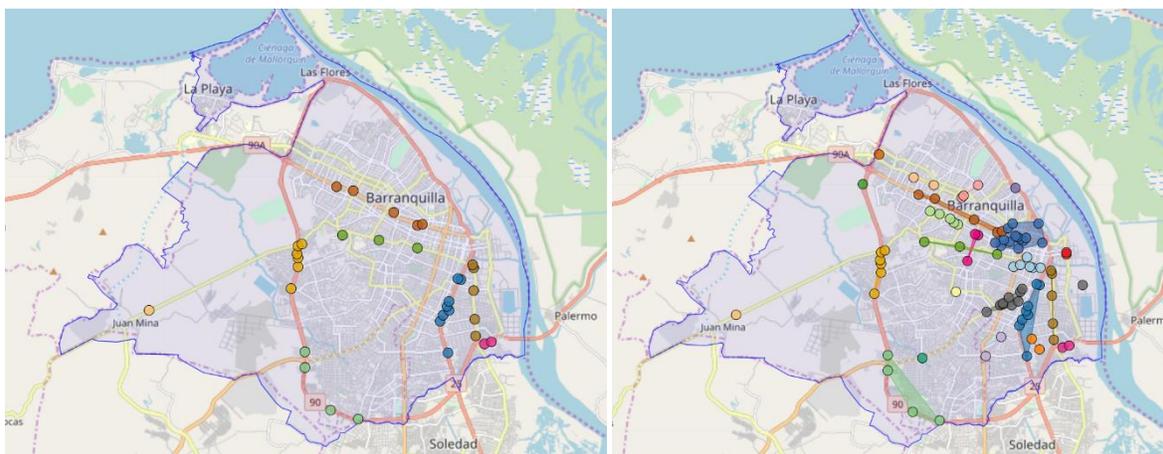
Distancia Crítica (Km)	0,25	0,50	0,75	1,00	1,25	1,50	1,75
0,50	0,004						
0,75	0,000	0,000					
1,00	0,000	0,000	0,000				
1,25	0,000	0,000	0,000	0,000			
1,50	0,000	0,000	0,000	0,050	0,086		
1,75	0,000	0,000	0,000	0,007	0,899	0,050	
2,00	0,000	0,000	0,000	0,043	0,677	0,401	0,288

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor < 0.05 entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

La Figura 32 presenta los mercados viables y los resultados de determinar los mercados por el método de conglomerados, usando la distancia óptima. Además de los 8 mercados identificados a través de las vías de entrada y salida de la ciudad (Panel a), se identificaron a través del método de clústers 16 conglomerados adicionales en la ciudad de Barranquilla, los cuales se muestran en el Panel b. Una vez identificados los mercados relevantes de la ciudad, la Tabla 27 presenta las estimaciones del Modelo SAR con características observables de las estaciones de servicio y efectos fijos de mercados relevantes; las estimaciones de los coeficientes de los mercados relevantes se presentan en la Figura 33. Como se puede observar los coeficientes de siete de los mercados viales son significativos, lo que indica que estos mercados fijan sus precios con una importante diferencia respecto al mercado omitido. Entre los mercados viales, el que fija precios menores para la gasolina es el mercado de la Carrera 38, mientras que el que fija precios más altos es el de la Calle 109. En términos de banderas de las EDS, los resultados también permiten concluir que, en Barranquilla, las estaciones de servicio de Terpel fijan un precio de la gasolina 67 pesos mayor que el precio de las estaciones de bandera Otras. Además, las estaciones de servicio con mini-mercado tienen precios menores en 49 pesos respecto a las estaciones que no

tienen este servicio. Se observa además que aumentar los accesos viales en una unidad permite a las estaciones de servicio ser más competitivas y reducir su precio con un promedio de 89 pesos.

Figura 32. Estimación de Conglomerados y Mercados Viales, Barranquilla



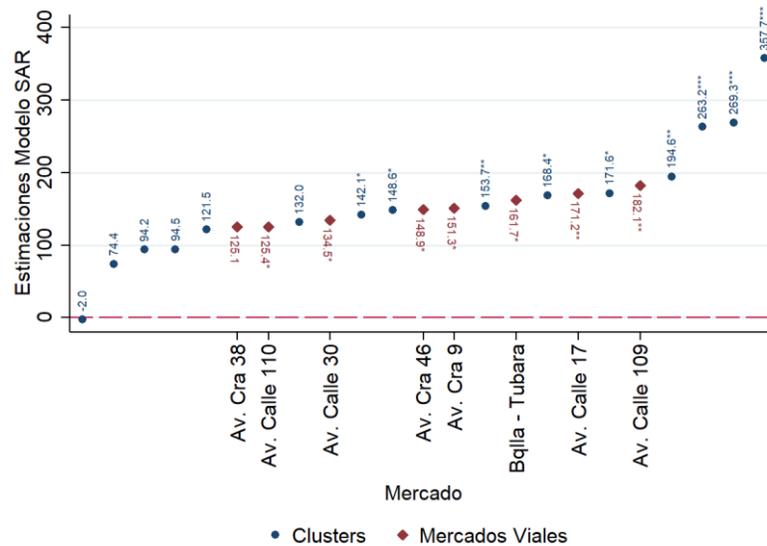
Fuente: Estimaciones por clústeres jerárquicos usando la metodología promedio y una distancia crítica de 1.25 kilómetros e identificación de estaciones en corredores viales. Las distancias usadas para generar los clústeres fueron computadas como distancias esféricas, también conocidas como geodésicas.

Tabla 27. Estimación modelo SAR incluyendo Efectos Fijos de Mercados Relevantes.

	<i>Variable Dependiente: Precio Gasolina</i>		
	MCO	SAR	
	Coef.	Coef.	Coef. Wx
1(Texaco)	33.24	39.74	999.78
1(Mobil)	51.69	10.68	-17,642.76
1(Terpel)	72.86**	67.49**	5,781.11
1(Biomax)	36.76	34.15	-392.76
1(Petromil)	56.94	22.11	-6,081.97
1(Lavado Vehículos)	23.54	6.52	-8,144.82
1(Mini-mercado)	-43.85	-49.19**	-5,482.78
1(Otros Servicios)	40.39	22.66	6,922.64
Accesos	-108.96***	-89.06***	15,418.69*
Competidores (1Km)	-10.58**	-1.70	-1,040.95
$\rho$		21.39***	
$\rho_x$		-0.42***	
$\sigma_e^2$		53,285.21***	
Observaciones	8,448	8,448	8,448
R-cuadrado		0.58	0.58
Número EDS	88	88	88

Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Errores estándar clusterizados a nivel de estación de servicio. Se omite las estaciones de servicio con bandera Otras. Las regresiones incluyeron efectos fijos de mercado relevante, los cuales se presentan en la Figura 33 por brevedad. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Figura 33. Estimación de Efectos Fijos de Mercados Relevantes, Barranquilla.



Fuente: Cada punto corresponde a la estimación máxima verosímil de la variable dummy que representa cada mercado. Se omite uno de los mercados para realizar comparaciones. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

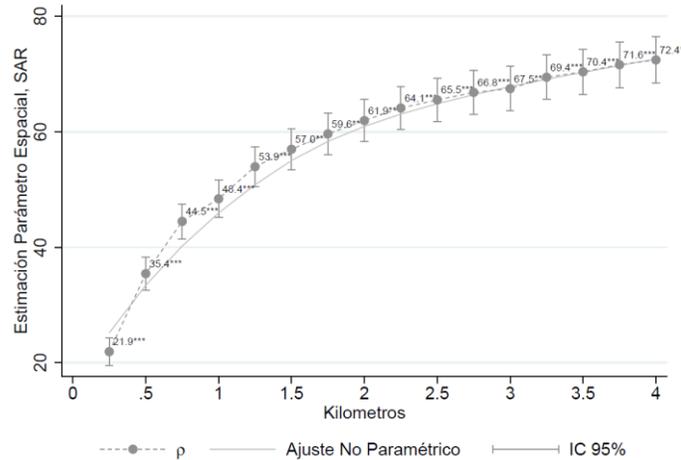
## 5. Mercados Relevantes por Regiones

Para determinar los mercados relevantes a nivel nacional se definieron siete regiones, las cuales siguen la división propuesta para el Sistema General de Regalías, pero sustrayendo los dos departamentos de la frontera con Ecuador, Nariño y Putumayo, como una región aparte por su especial interés como zona de frontera. Para poder procesar la ingente cantidad de datos asociados a todo el territorio nacional, fue necesario modificar la forma de medición de las distancias usadas para ponderar los precios espacialmente, las cuales fueron computadas como distancias geodésicas y no como distancias de ruta. De los ejercicios por región fueron excluidos los departamentos de la frontera con Venezuela y las ciudades principales, para los que se aplicó previamente la metodología de determinación de mercados relevantes a través de distancias de ruta.

### 5.1 Frontera con Ecuador: Nariño y Putumayo

Nariño tiene un área aproximada de 33,268 km<sup>2</sup> y Putumayo tiene un área aproximada de 24,885 km<sup>2</sup>; ambos departamentos cuentan con 302 y 97 estaciones de servicio respectivamente. Luego, esta región tiene una densidad aproximada de 1.4 estaciones de servicio, por cada 200 km<sup>2</sup>. La Figura 34 presenta los resultados del modelo SAR.

La Figura 34. Estimaciones del Modelo SAR, Frontera con Ecuador



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Las Tablas 28 y 29 presentan los resultados del procedimiento de determinación de distancia óptima. Ya que para el rango de distancias críticas consideradas no se rechaza ninguna hipótesis nula, entonces se toma 4 kilómetros como distancia crítica para determinar los mercados relevantes de la región fronteriza con Ecuador. El análisis de conglomerados da como resultado 122 mercados relevantes. La Figura 35 presenta un panorama general de estos resultados y la Figura 36 permite observar el número de mercados relevantes en Pasto y Mocoa, las capitales de los departamentos. Por su cuenta, en Pasto se identifican 3 mercados relevantes y en Mocoa 2.

Tabla 28. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,25	379,76	157,74							
2,50	447,14	234,91	80,33						
2,75	535,17	323,79	168,00	88,98					
3,00	511,27	314,88	167,64	89,33	17,78				
3,25	735,41	521,77	364,41	288,50	200,31	356,40			
3,50	754,48	547,59	392,14	312,74	224,16	256,39	39,50		
3,75	804,49	602,09	448,24	368,12	280,57	294,52	98,81	59,83	
4,00	786,14	594,14	445,53	365,61	282,00	279,29	115,66	76,57	23,08

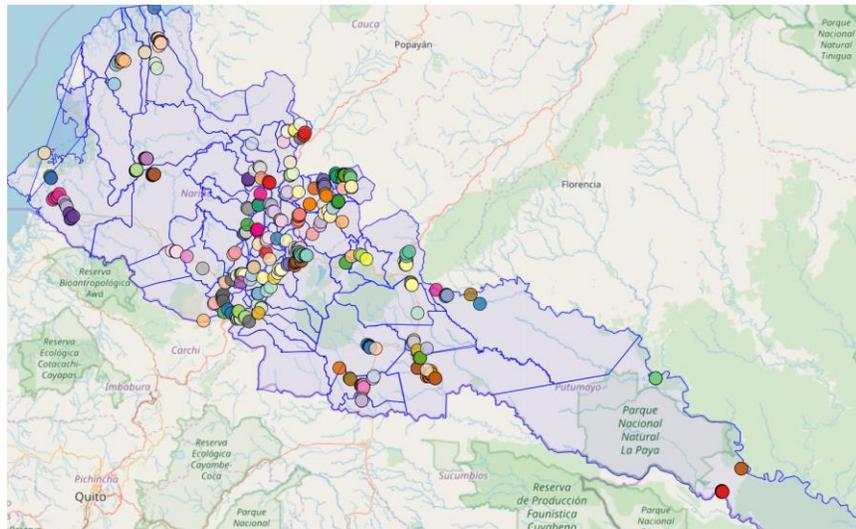
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 29. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,000				
3,25	0,000	0,000	0,000	0,000	0,000	0,000			
3,50	0,000	0,000	0,000	0,000	0,000	0,000	0,000		
3,75	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
4,00	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

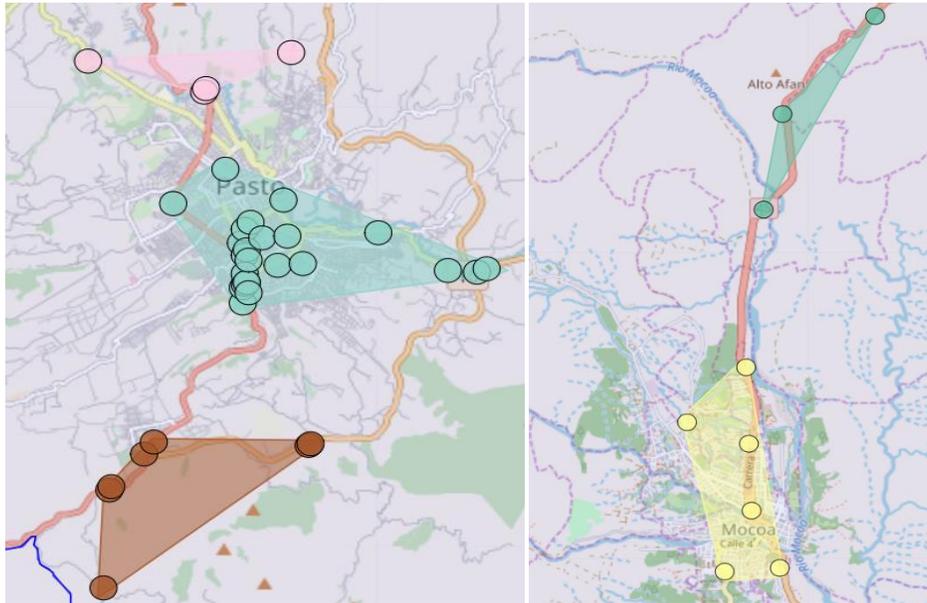
Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl}^2 = 1)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 35. Estimación de Conglomerados, Frontera con Ecuador.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 36. Estimación de Conglomerados, Pasto y Mocoa.

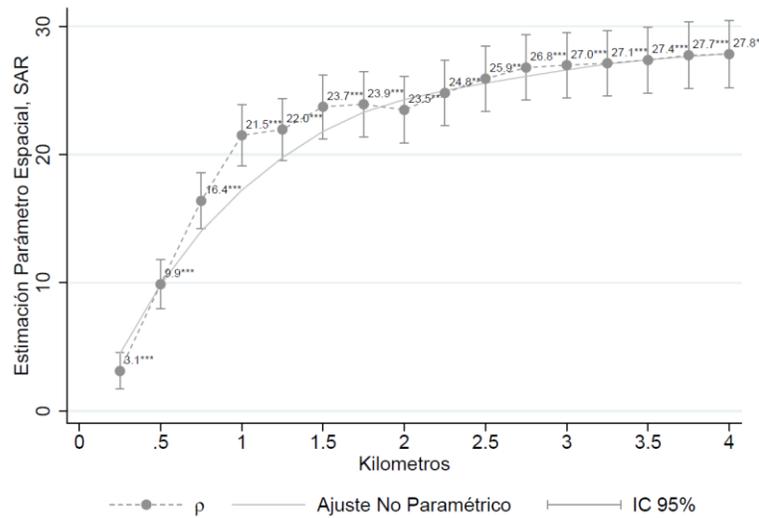


Fuente: Estimaciones de clusters jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clusters se computaron como distancias geodésicas.

## 5.2 Región Caribe: Atlántico, Bolívar, Córdoba, Magdalena y Sucre

Atlántico tiene un área aproximada de  $3,232 \text{ km}^2$  sin contar el área de Barranquilla, Bolívar tiene un área aproximada de  $24,885 \text{ km}^2$ , Córdoba tiene  $23,980 \text{ km}^2$ , Magdalena tiene  $23,188 \text{ km}^2$  y Sucre tiene  $10,670 \text{ km}^2$ . Estos departamentos cuentan respectivamente con 59, 117, 160, 106 y 94 estaciones de servicio. Luego, en la Región Caribe, se tiene una densidad aproximada de 1.2 estaciones de servicio cada  $200 \text{ km}^2$ . La Figura 37 presenta los resultados de estimar el modelo autorregresivo a partir del panel de precios de gasolina para la región considerada. Se puede observar, que a partir de 250 metros, existe evidencia de dependencia espacial en los precios, es decir, hay evidencia de competencia. Similar a los resultados presentados anteriormente, se encuentra que la importancia del coeficiente de correlación espacial decrece marginalmente.

La Figura 37. Estimaciones del Modelo SAR, Región Caribe



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Las Tablas 30 y 31 presentan los resultados del procedimiento para determinar la distancia crítica óptima. Dicho algoritmo señala que, al considerar un rango de 4 kilómetros, la distancia óptima es de 3,75 kilómetros. Usando esta distancia, el método de conglomerados indica que existen 198 mercados relevantes en los departamentos considerados para la Región Caribe. Los resultados del método de conglomerados se presentan en las Figuras 38 a 42. En Santa Marta se identifican 4 mercados relevantes, en Soledad se identifican 3 mercados, en Cartagena se encuentran 6 mercados y en Montería se identificaron estaciones pertenecientes a 9 mercados.

Tabla 30. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	9,25								
2,25	2620,23	81,82							
2,50	473,58	200,87	152,55						
2,75	18538,81	504,34	27227,08	93,80					
3,00	3437,97	660,70	1563,91	100,28	11,63				
3,25	2308,35	792,49	1134,08	113,04	23,77	12,87			
3,50	1026,17	1576,04	553,53	105,48	28,72	17,69	8,54		
3,75	593,02	5271,72	346,45	100,99	36,93	27,10	19,14	10,75	
4,00	456,98	1546,54	271,37	87,47	32,46	23,91	17,29	9,72	0,84

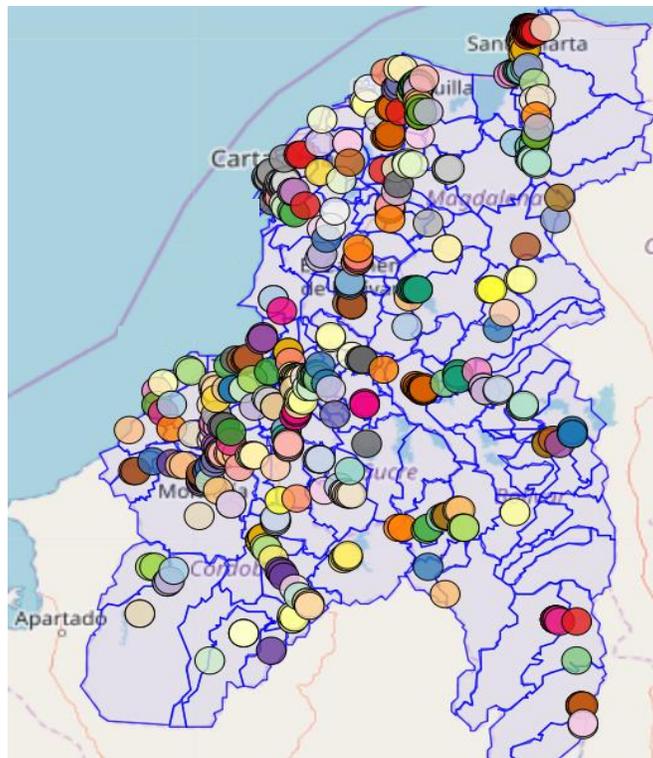
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 31. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,002								
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,001				
3,25	0,000	0,000	0,000	0,000	0,000	0,000			
3,50	0,000	0,000	0,000	0,000	0,000	0,000	0,003		
3,75	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	
4,00	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002	0,358

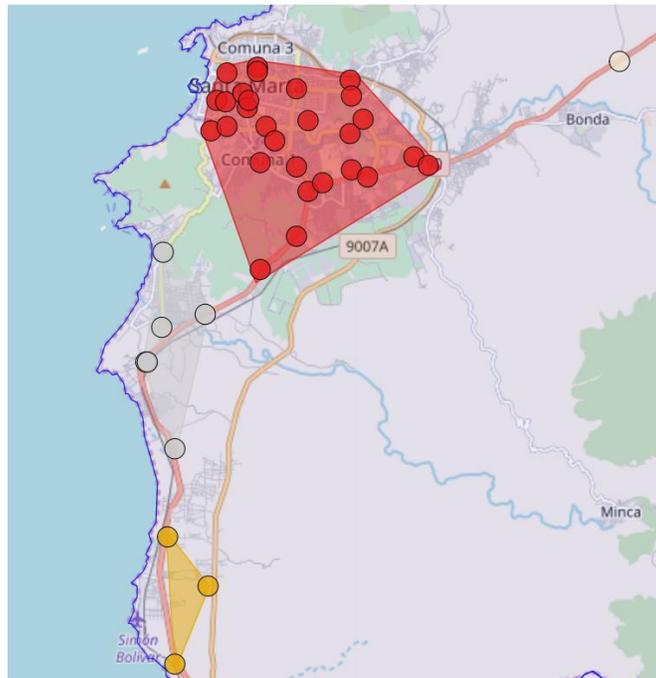
Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 38. Estimación de Conglomerados, Región Caribe.



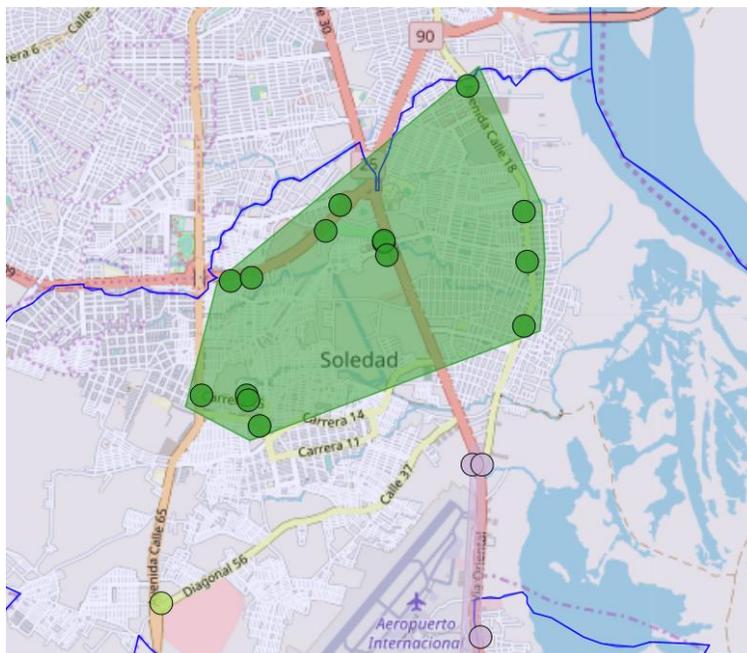
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 39. Estimación de Conglomerados, Santa Marta.



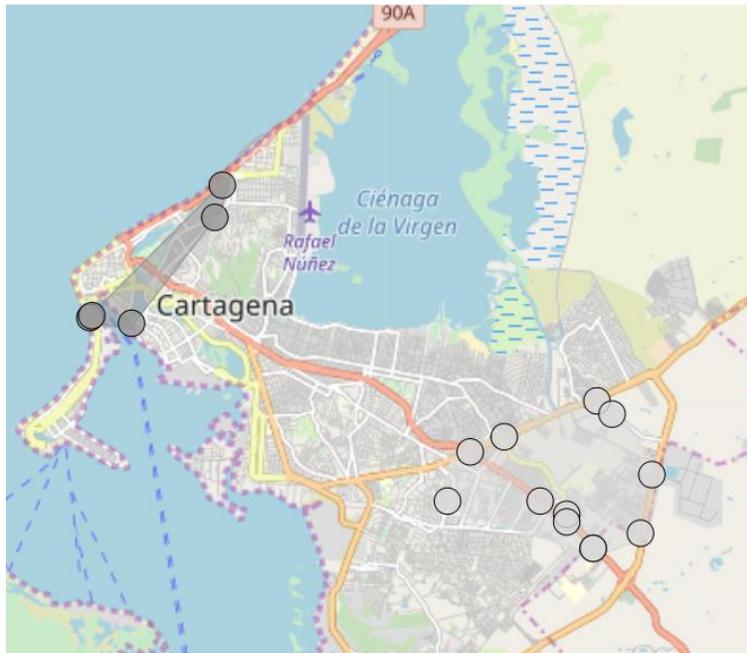
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 40. Estimación de Conglomerados, Soledad.



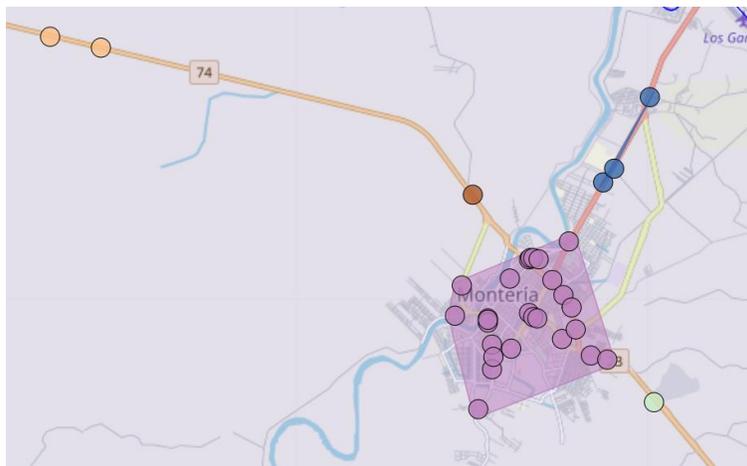
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 41. Estimación de Conglomerados, Cartagena.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

Figura 42. Estimación de Conglomerados, Montería.



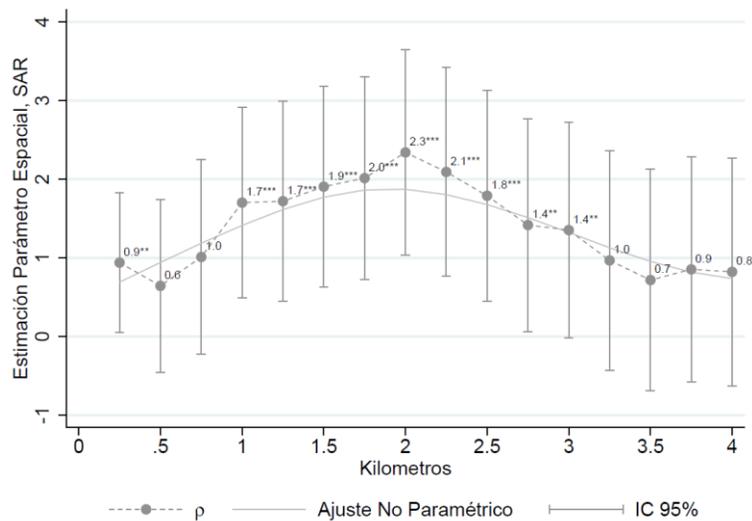
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

### 5.3 Región Centro Oriente: Cundinamarca y Santander

Cundinamarca tiene un área aproximada de 24,210 km<sup>2</sup> y Santander tiene un área aproximada de 30,537 km<sup>2</sup>. Estos departamentos cuentan con 275 y 215 estaciones de servicio, respectivamente. Luego, esta región tiene una densidad aproximada de 1.8 estaciones de servicio

por cada 200 km<sup>2</sup>. La Figura 43 presenta los resultados del modelo SAR para esta región. Se puede observar que a partir de 3 kilómetros no existe evidencia de competencia en precios, ya que los coeficientes de correlación espacial no son estadísticamente significativos. Note además que el comportamiento de la dimensión espacial para esta región empieza a decrecer a partir de 2 kilómetros, señalando un comportamiento algo diferente al de otras regiones analizadas en este documento. De esta manera, teniendo en cuenta los resultados presentados en la Figura 43 se considera que el coeficiente asociado a una distancia crítica de 2 kilómetros debe ser el óptimo.

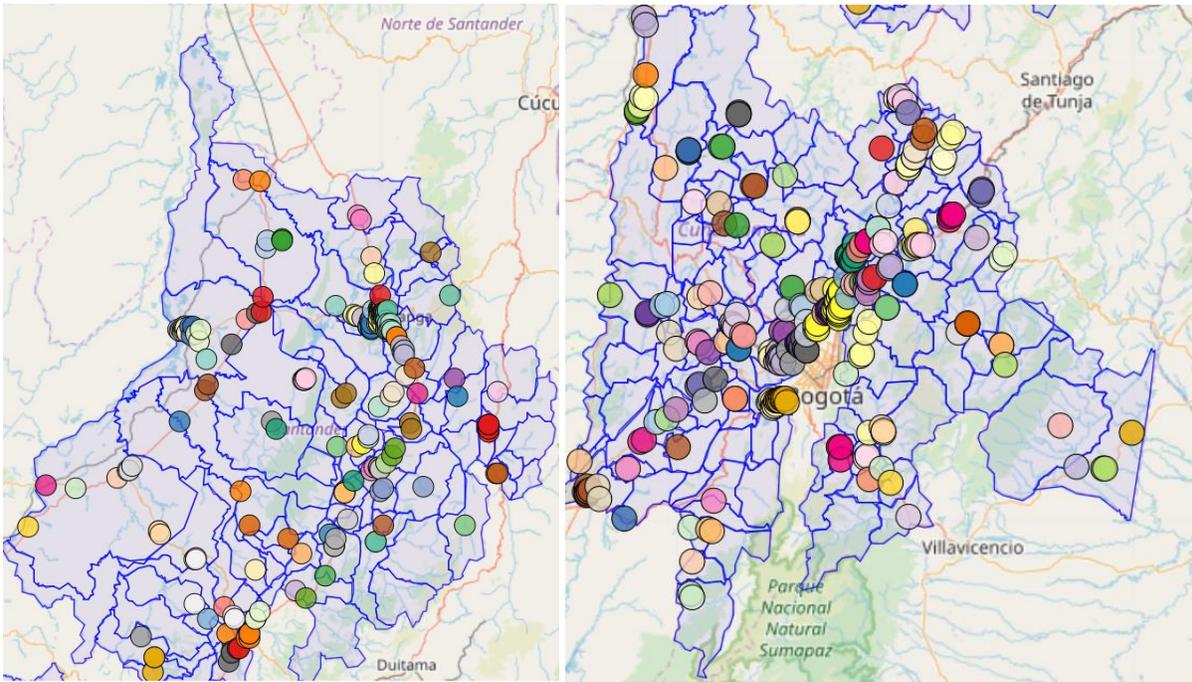
La Figura 43. Estimaciones del Modelo SAR, Región Centro Oriente



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

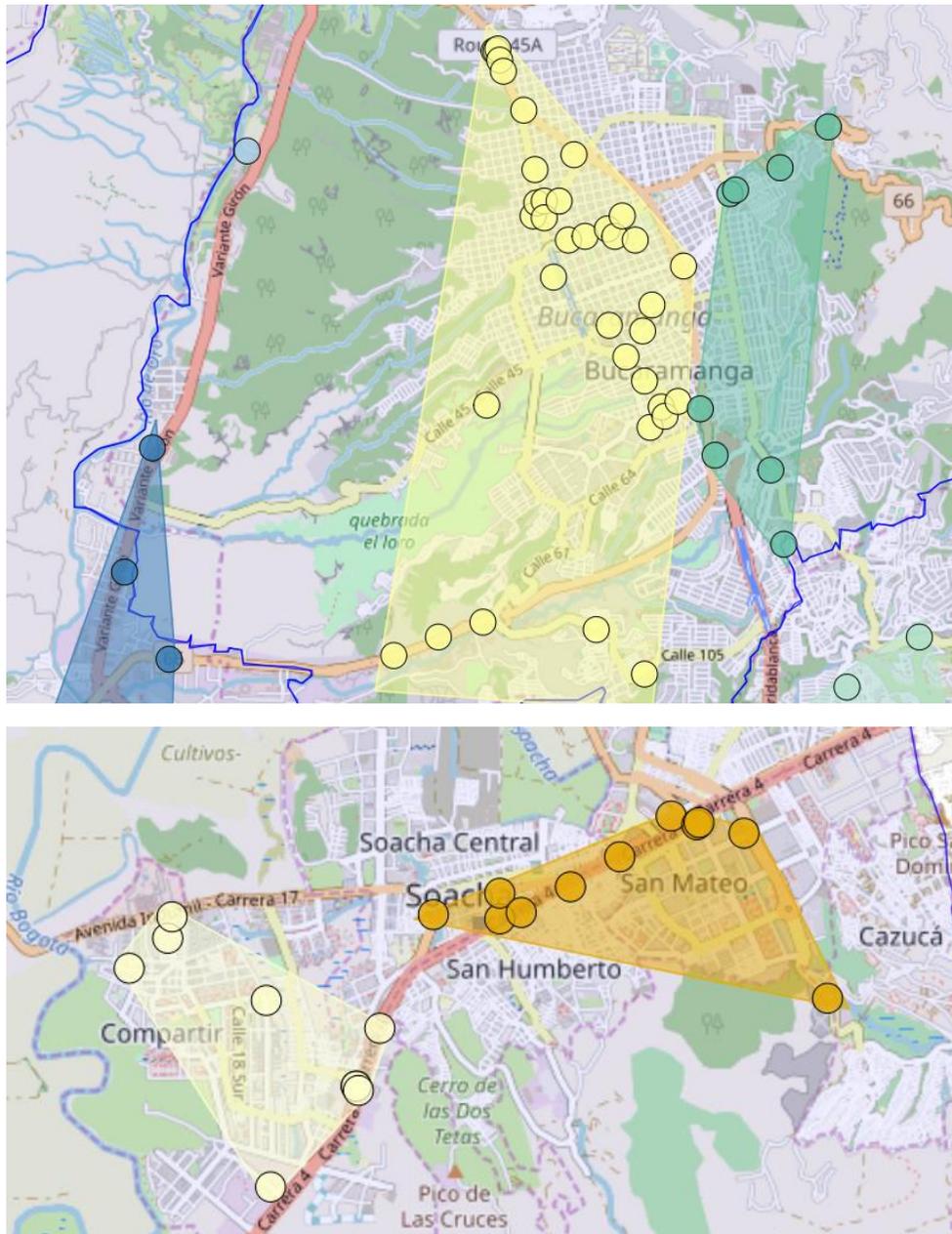
A partir de la distancia óptima señalada, el método de conglomerados identifica 204 mercados relevantes para los departamentos de Santander y Cundinamarca. Así, como se puede observar a partir de las Figuras 44 y 45, en Bucaramanga hay evidencia de que las estaciones de servicio pertenecen a al menos 5 mercados relevantes y, en Soacha, se encuentra evidencia de 2 mercados.

Figura 44. Estimación de Conglomerados, Santander y Cundinamarca.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 2 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 45. Estimación de Conglomerados, Bucaramanga y Soacha.

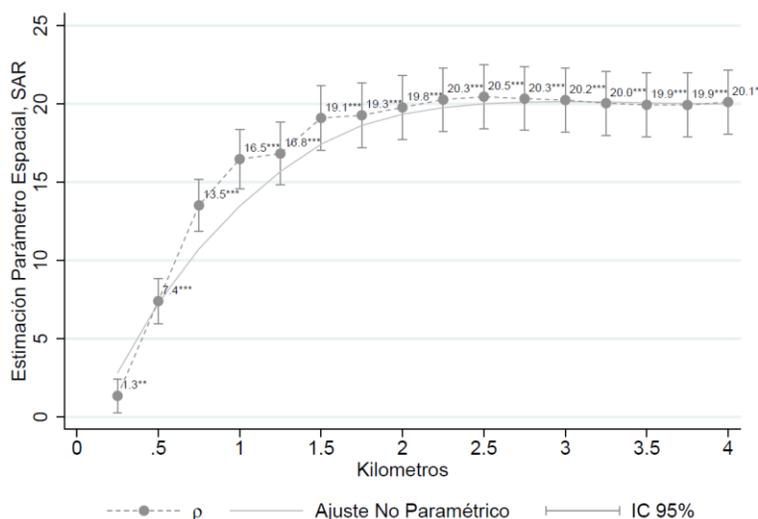


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 2 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

#### 5.4 Región Eje Cafetero: Antioquia, Caldas, Quindío y Risaralda

Antioquia tiene un área aproximada de 63,230 km<sup>2</sup> y 390 distribuidores, sin contar el área y las estaciones de Medellín. Los departamentos de Caldas, Quindío y Risaralda tienen áreas aproximadas de 7,888 km<sup>2</sup>, 1,845 km<sup>2</sup> y 4,410 km<sup>2</sup>. Estos últimos tres departamentos cuentan con 89, 65 y 82 estaciones de servicio, respectivamente. Luego, esta región tiene una densidad aproximada de 1.6 estaciones de servicio por cada 200 km<sup>2</sup>. La Figura 46 presenta los resultados del modelo SAR.

La Figura 46. Estimaciones del Modelo SAR, Eje Cafetero



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Las Tablas 34 y 35 presentan los resultados del procedimiento para determinar la distancia crítica óptima. Ya que para todas las distancias consideradas se rechazó la hipótesis nula de igualdad, al menos en uno de los casos, se toma la distancia máxima como distancia óptima. Es decir, la distancia crítica considerada en el modelo de conglomerados será 4 kilómetros. Ahora, con esta distancia, el método de conglomerados da como resultado 234 mercados relevantes para la región Cafetera. Un panorama general se presenta en la Figura 47, mientras que los mapas a nivel de ciudad se detallan en las Figuras 48 a 51 dan cuenta de los resultados para algunas ciudades de esta región. Entre Bello y Copacabana existen 3 mercados relevantes, en Manizales las estaciones de servicio pertenecen a al menos 4 mercados, en Pereira se tiene evidencia de 7 mercados relevantes y en Armenia se identificaron 4 mercados.

Tabla 34. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	34,20								
2,25	49,28	19,17							
2,50	67,96	35,12	68,72						
2,75	49,03	20,03	1,42	7,45					
3,00	48,09	18,01	1,19	42,77	2,36				
3,25	36,85	8,21	12,58	37,19	12,33	11,44			
3,50	24,60	2,56	52,25	103,42	32,14	60,65	4,80		
3,75	26,74	2,74	28,48	60,98	23,95	28,18	127,20	0,06	
4,00	50,93	17,69	3,78	17,43	5,19	2,87	2,95	7,66	16,96

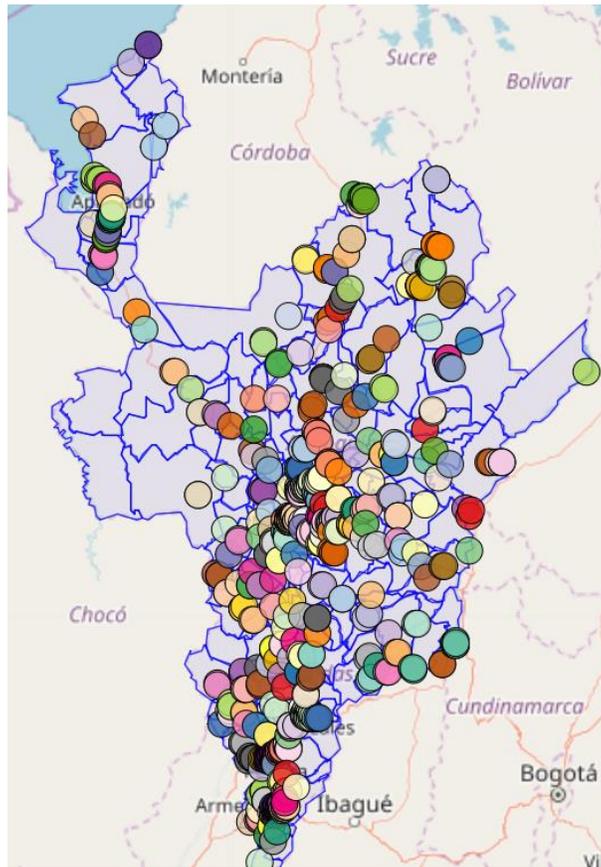
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 35. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,000								
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,234	0,006					
3,00	0,000	0,000	0,275	0,000	0,125				
3,25	0,000	0,004	0,000	0,000	0,000	0,001			
3,50	0,000	0,109	0,000	0,000	0,000	0,000	0,028		
3,75	0,000	0,098	0,000	0,000	0,000	0,000	0,000	0,802	
4,00	0,000	0,000	0,052	0,000	0,023	0,090	0,086	0,006	0,000

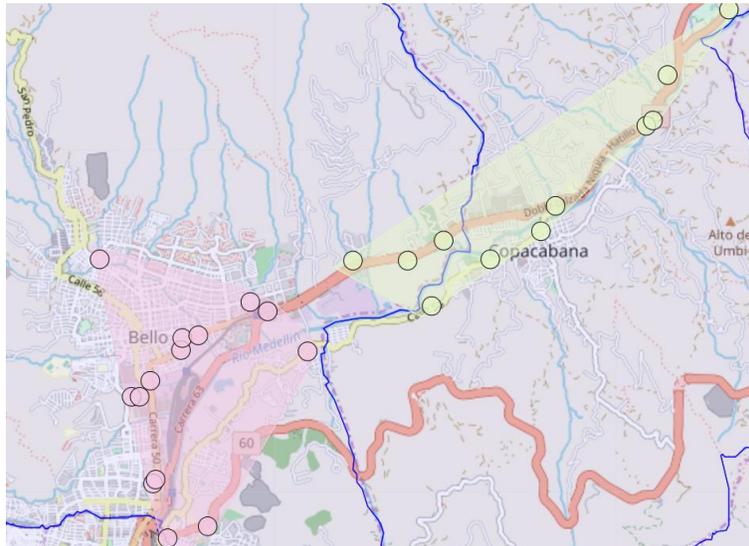
Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 47. Estimación de Conglomerados, Eje Cafetero



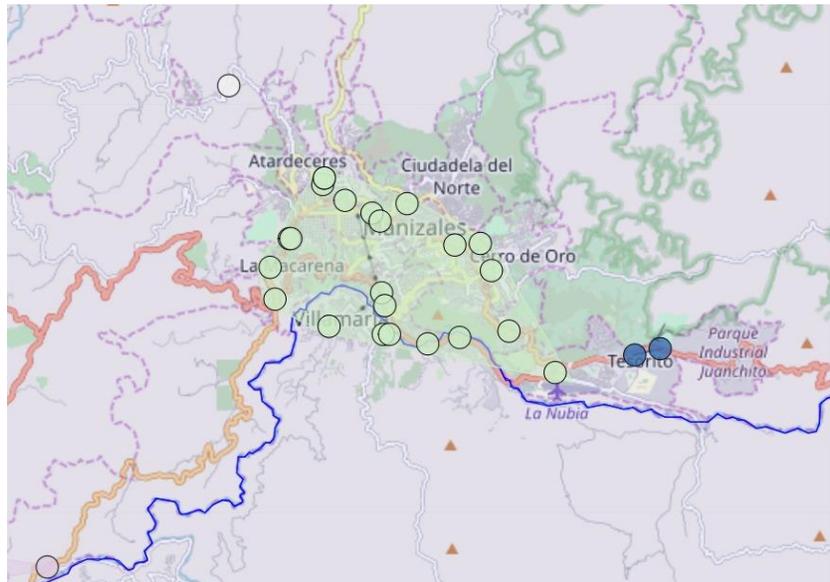
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 48. Estimación de Conglomerados, Bello y Copacabana.



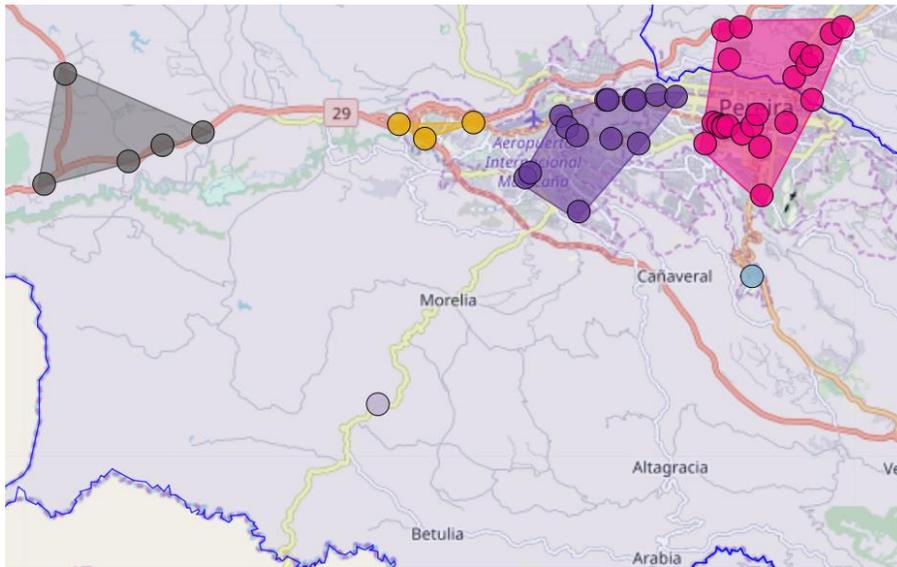
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 49. Estimación de Conglomerados, Manizales.



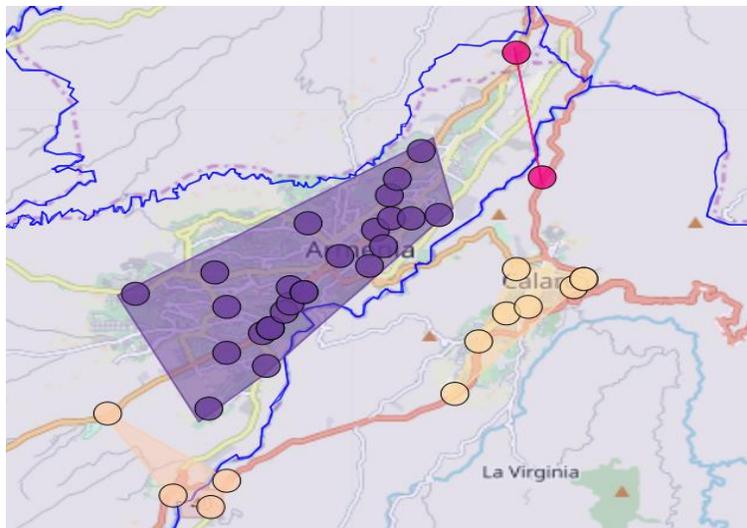
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 50. Estimación de Conglomerados, Pereira.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 51. Estimación de Conglomerados, Armenia.



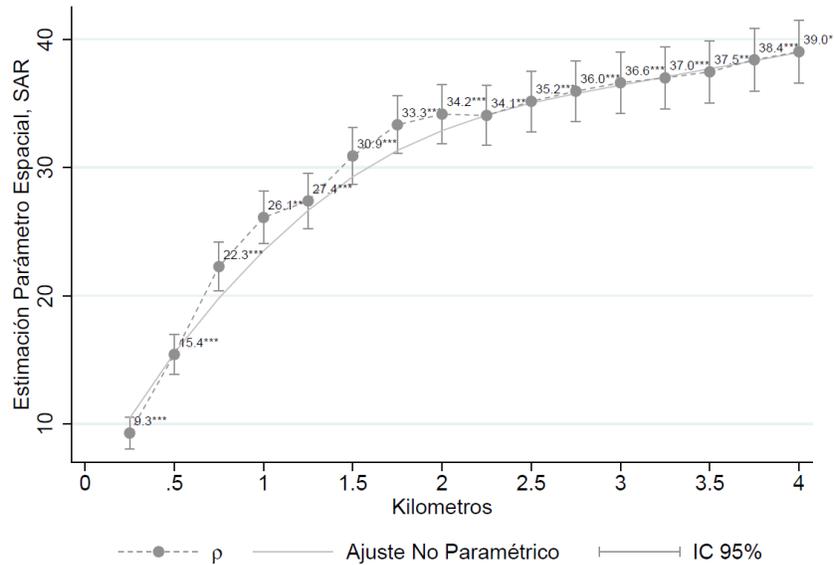
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

### 5.5 Región Pacífico: Chocó, Valle del Cauca y Cauca

Valle del Cauca tiene un área aproximada de 21,576 km<sup>2</sup> y 202 estaciones de servicio, sin contar el área y los distribuidores de Cali. Los departamentos de Chocó y Cauca tienen áreas

aproximadas de 46,530 km<sup>2</sup> y 29,308 km<sup>2</sup>. Estos últimos dos departamentos cuentan con 71 y 113 estaciones de servicio, respectivamente. Luego, esta región tiene una densidad aproximada de 0.8 estaciones de servicio por cada 200 km<sup>2</sup>.

La Figura 52. Estimaciones del Modelo SAR, Pacífico



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Las Tablas 36 y 37 presentan los resultados del procedimiento usado para determinar la distancia crítica óptima. Ya que para cada distancia se rechazó al menos una vez la hipótesis nula de igualdad, se asume que la distancia óptima es la máxima establecida, de 4 kilómetros. A partir de este resultado se implementa el análisis de conglomerados, que da como resultado 93 mercados relevantes en la región Pacífico. Un panorama general de estos resultados se presenta en la Figura 53. Por otro lado, las Figuras 54 a 56 dan cuenta de los resultados para algunas ciudades en los departamentos de esta región. Así, en Buenaventura y en Popayán se identifican estaciones de servicio pertenecientes a 2 mercados relevantes, respectivamente, y en Quibdó se identifica 1 mercado para toda la ciudad.

Tabla 36. Estadísticos  $\chi_{gl=1}^2$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	22,51								
2,25	12,98	0,94							
2,50	60,10	39,73	81,74						
2,75	111,60	103,37	173,41	107,81					
3,00	154,65	153,58	226,57	151,29	53,79				
3,25	176,30	176,14	244,88	165,39	74,78	22,60			
3,50	202,77	203,05	267,33	186,01	100,72	49,65	27,08		
3,75	277,11	288,63	362,98	283,05	191,83	138,14	118,10	102,17	
4,00	329,59	348,28	428,14	349,21	255,73	201,97	183,82	170,66	68,50

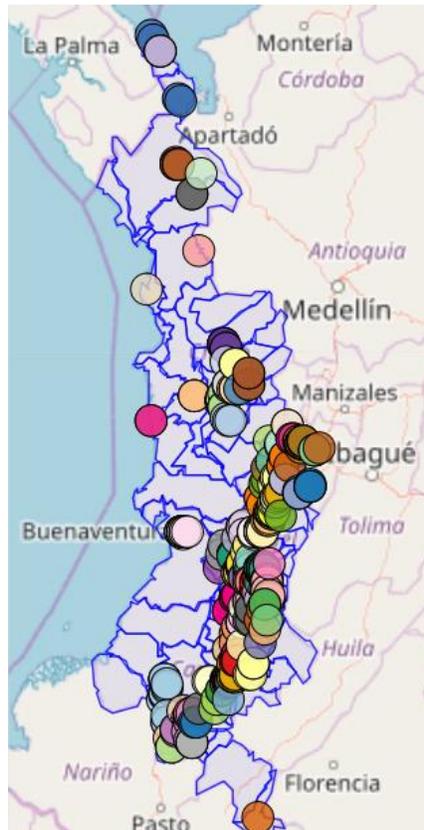
Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc'')^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

Tabla 37. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,000								
2,25	0,000	0,332							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,000				
3,25	0,000	0,000	0,000	0,000	0,000	0,000			
3,50	0,000	0,000	0,000	0,000	0,000	0,000	0,000		
3,75	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
4,00	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

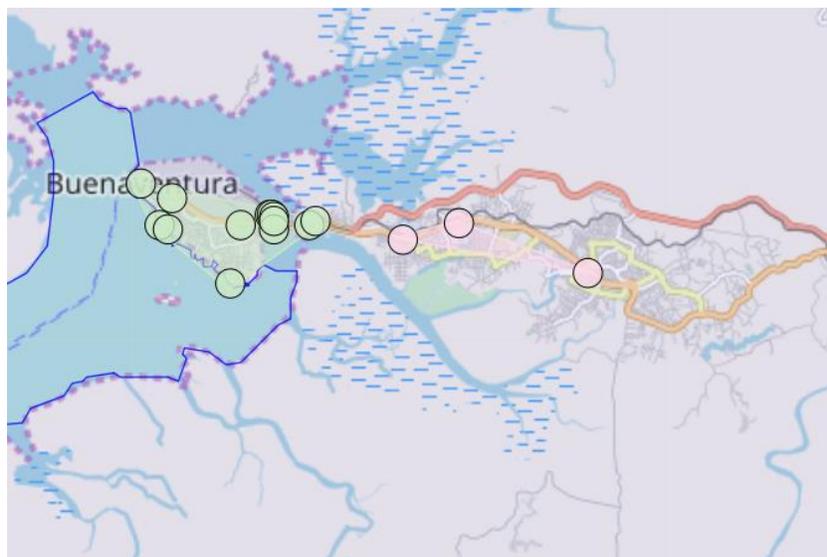
Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Figura 53. Estimación de Conglomerados, Pacífico.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 54. Estimación de Conglomerados, Buenaventura.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 55. Estimación de Conglomerados, Popayán.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

Figura 56. Estimación de Conglomerados, Quibdó.

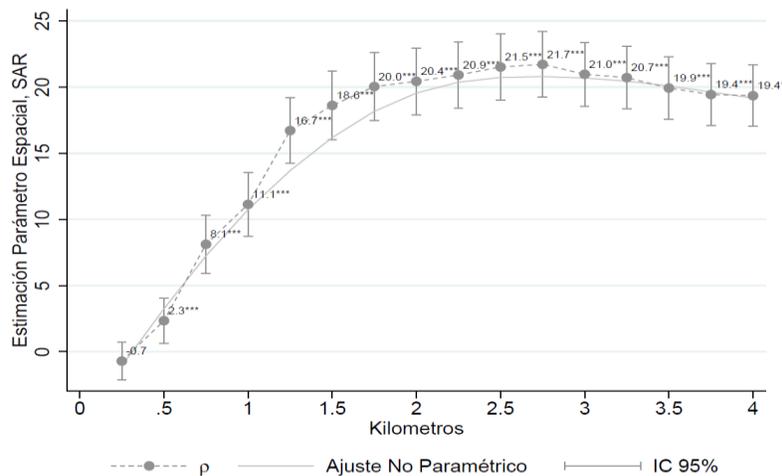


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

## 5.6 Región Llanos: Casanare, Guaviare y Meta

Casanare tiene un área aproximada de 44,640 km<sup>2</sup>, Guaviare tiene 53,460 km<sup>2</sup> y Meta tiene un área aproximada de 85,635 km<sup>2</sup>. Estos departamentos cuentan con 52, 31 y 113 estaciones de servicio, respectivamente. Luego, esta región tiene una densidad aproximada de 0.2 estaciones de servicio por cada 200 km<sup>2</sup>. La Figura 57 presenta los resultados del modelo SAR para la región de los Llanos. Se observa que a partir de 500 metros existe competencia en precios de la gasolina. Por otro lado, las Tablas 38 y 39 dan cuenta de los resultados del procedimiento para determinar la distancia crítica óptima. Como es posible observar, este procedimiento indica que la distancia óptima en la región es de 3,75 kilómetros.

La Figura 57. Estimaciones del Modelo SAR, Llanos.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Tabla 38. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	6,16								
2,25	25,40	46,17							
2,50	62,06	115,90	70,86						
2,75	52,25	57,35	27,16	2,11					
3,00	10,02	4,70	0,06	5,89	17,12				
3,25	4,58	1,07	0,56	9,96	21,70	4,75			
3,50	0,11	2,84	11,39	31,47	51,63	36,87	39,01		
3,75	2,85	9,66	22,31	46,76	70,08	56,47	58,32	19,84	
4,00	3,59	10,87	23,74	48,02	70,51	55,96	55,98	19,13	1,44

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |Var(\rho_{dc'}) - Var(\rho_{dc''})|$ . Donde  $Var(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

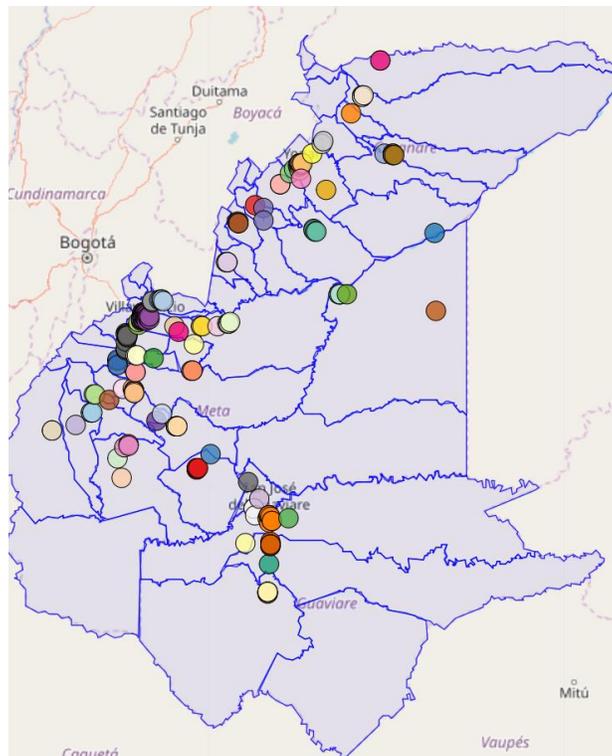
Tabla 39. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,013								
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,147					
3,00	0,002	0,030	0,813	0,015	0,000				
3,25	0,032	0,301	0,454	0,002	0,000	0,029			
3,50	0,736	0,092	0,001	0,000	0,000	0,000	0,000		
3,75	0,091	0,002	0,000	0,000	0,000	0,000	0,000	0,000	
4,00	0,058	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,230

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor < 0.05 entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

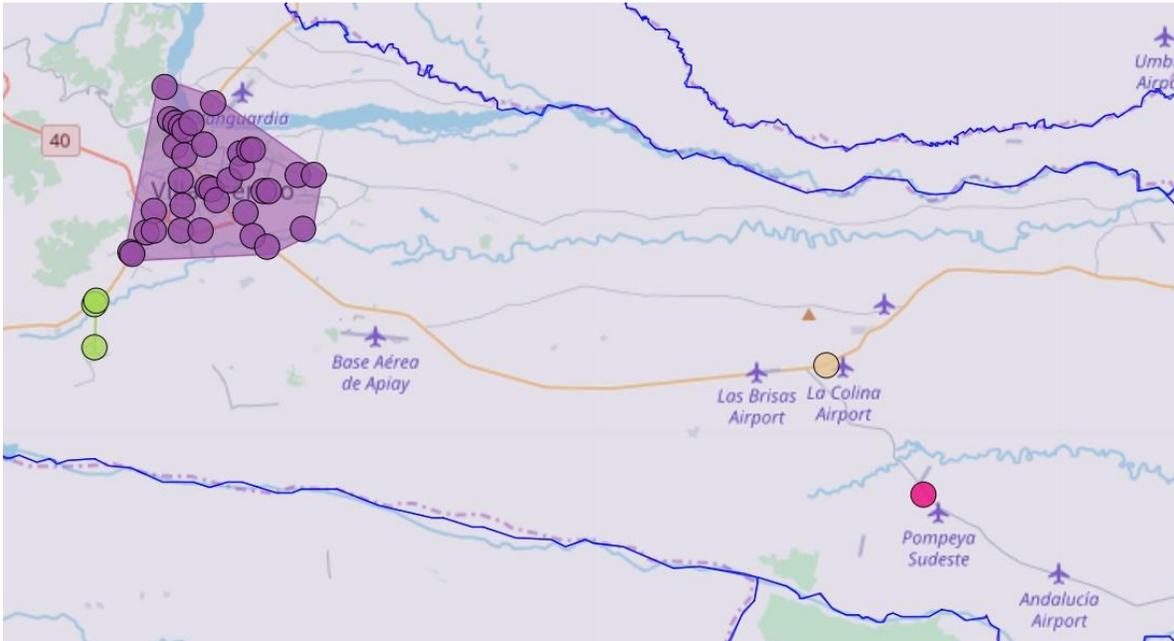
Las Figuras 58 a 61 presentan los resultados del modelo de conglomerados usando la distancia óptima, 3.75 kilómetros. Dicho modelo genera 61 mercados relevantes para la región de los Llanos. En Villavicencio y en San José del Guaviare se identificaron estaciones de servicio pertenecientes a 4 mercados, respectivamente, y en Yopal se encuentran distribuidores minoristas pertenecientes a 6 mercados.

La Figura 58. Estimaciones del Modelo SAR, Llanos.



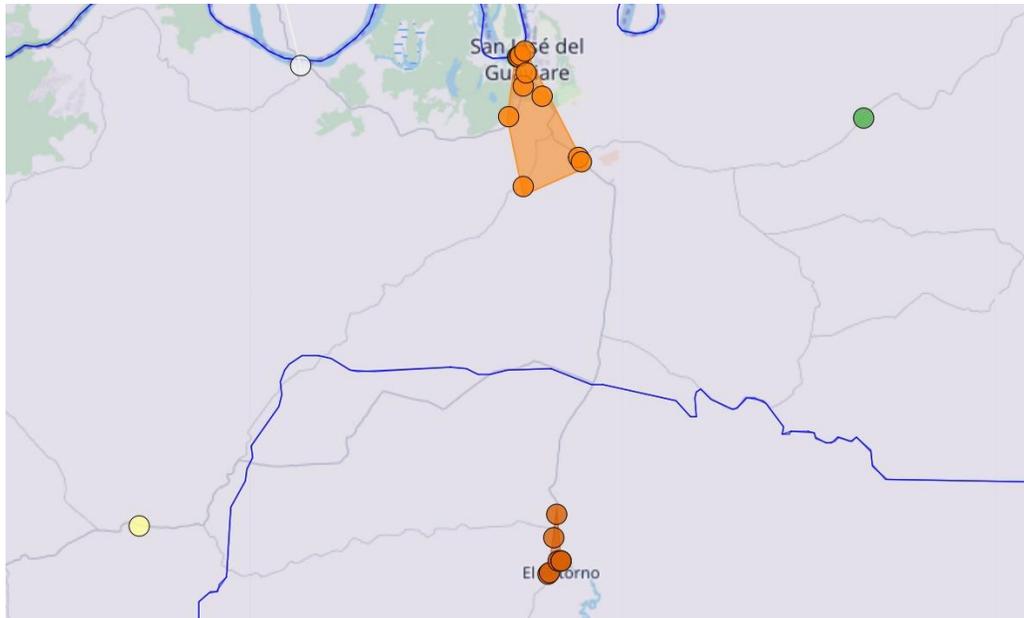
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

La Figura 59. Estimaciones del Modelo SAR, Villavicencio.



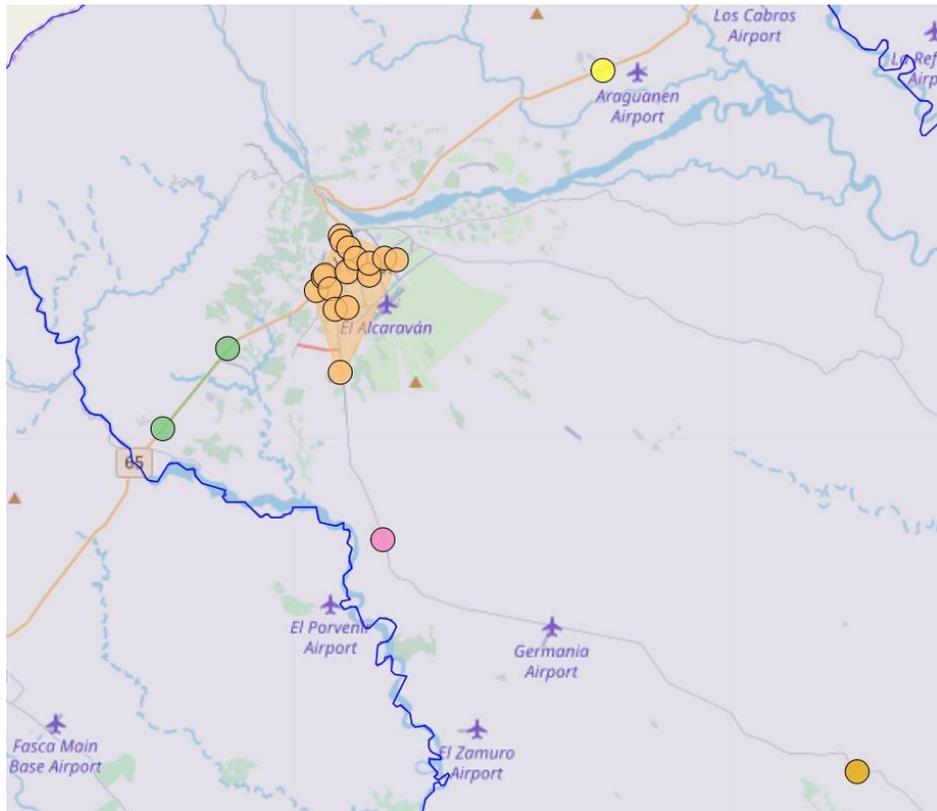
Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

La Figura 60. Estimaciones del Modelo SAR, San José del Guaviare.



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

La Figura 61. Estimaciones del Modelo SAR, Yopal.

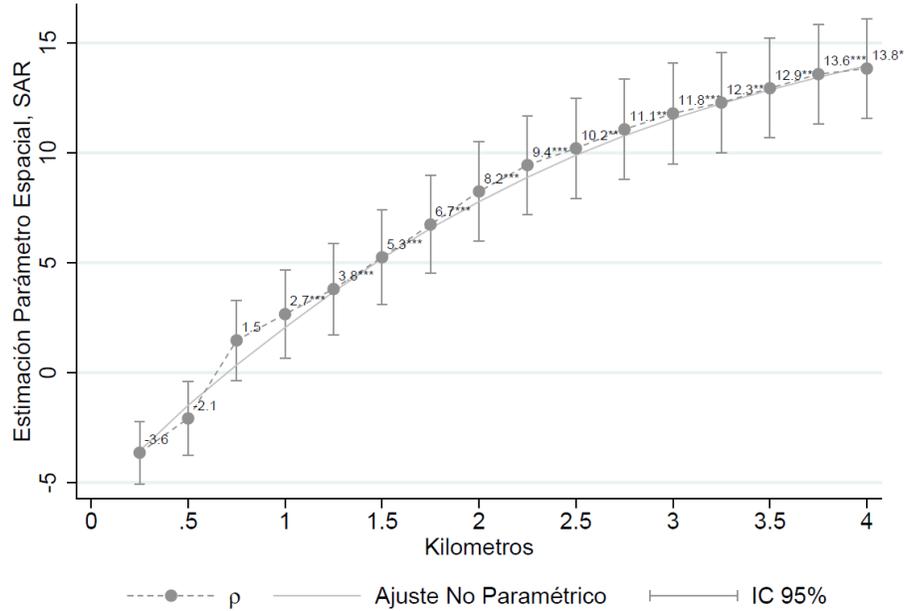


Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 3.75 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias euclídeas.

### 5.7 Región Centro Sur: Amazonas, Caquetá, Huila y Tolima

Amazonas tiene un área aproximada de  $109,665 \text{ km}^2$ , Caquetá tiene  $88,965 \text{ km}^2$ , Tolima  $23,562 \text{ km}^2$  y Huila  $19,890 \text{ km}^2$ . Estos departamentos tienen respectivamente 7, 92, 160 y 140 estaciones de servicio. Así, la región tiene una densidad de 0.33 estaciones de servicio por cada  $200 \text{ km}^2$ . La Figura 62 presenta los resultados del modelo autorregresivo espacial para la región Centro Sur. Se puede observar que a partir de 1 kilómetro se tiene evidencia de competencia en precios del combustible. Similar a los resultados de otras regiones se observa una importancia de la dimensión espacial que es marginal decreciente. Las Tablas 40 y 41 permiten observar que el procedimiento de determinación de distancia crítica indica que para cada distancia considerada se rechazó la hipótesis nula de igualdad, y por lo tanto se asume que la distancia crítica óptima es de 4 kilómetros.

La Figura 62. Estimaciones del Modelo SAR, Centro Sur.



Fuente: Estimaciones por Máxima Verosimilitud con datos del Fondo Soldicom. Cada punto representa la estimación puntual del parámetro de correlación espacial, a partir de regresiones independientes con distintas matrices de ponderación. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabla 40. Estadísticos  $\chi^2_{gl=1}$  para comparar entre pares de coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	158,20								
2,25	542,87	1722,55							
2,50	486,39	370,11	51,76						
2,75	557,13	412,62	131,05	83,38					
3,00	839,01	780,98	325,22	440,09	159,86				
3,25	1106,47	1211,06	563,64	1405,30	249,94	91,87			
3,50	1899,82	3716,43	1806,01	1672,21	260,32	129,68	57,17		
3,75	3486,46	32081,16	331052,72	1007,23	310,65	187,91	116,77	59,75	
4,00	3397,33	61266,04	14323,87	1329,73	404,92	267,28	185,27	147,10	46,48

Fuente: Cada entrada se computa de la siguiente manera  $(\rho_{dc'} - \rho_{dc''})^2 / |\text{Var}(\rho_{dc'}) - \text{Var}(\rho_{dc''})|$ . Donde  $\text{Var}(\rho_{dc})$  es la varianza estimada para cada coeficiente  $\rho_{dc}$ .

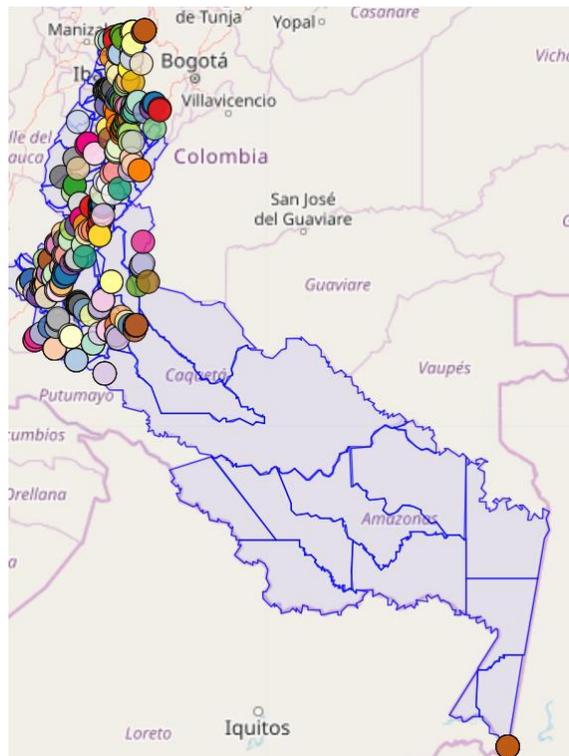
Tabla 41. P-valor asociado a la comparación entre coeficientes  $\rho_{dc}$ .

Distancia Crítica (Km)	1,75	2,00	2,25	2,50	2,75	3,00	3,25	3,50	3,75
2,00	0,000								
2,25	0,000	0,000							
2,50	0,000	0,000	0,000						
2,75	0,000	0,000	0,000	0,000					
3,00	0,000	0,000	0,000	0,000	0,000				
3,25	0,000	0,000	0,000	0,000	0,000	0,000			
3,50	0,000	0,000	0,000	0,000	0,000	0,000	0,000		
3,75	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
4,00	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Fuente: Cada entrada se computa de la siguiente manera:  $1 - F^{-1}(\chi_{gl=1}^2)$ . Si el P-valor  $< 0.05$  entonces se puede rechazar la hipótesis nula de que  $\rho_{dc'} = \rho_{dc''}$ .

Los resultados del modelo de conglomerados para la región Centro Sur, usando la distancia óptima, encuentra 124 mercados relevantes. La Figura 58 presenta los resultados de este modelo, de manera general. Las Figuras 63 a 67 dan cuenta de los resultados para algunas ciudades de esta región. En Ibagué se identificaron 5 mercados relevantes, en Florencia se encuentran 2 mercados y en Neiva y Leticia el modelo sugiere que existe 1 mercado respectivamente.

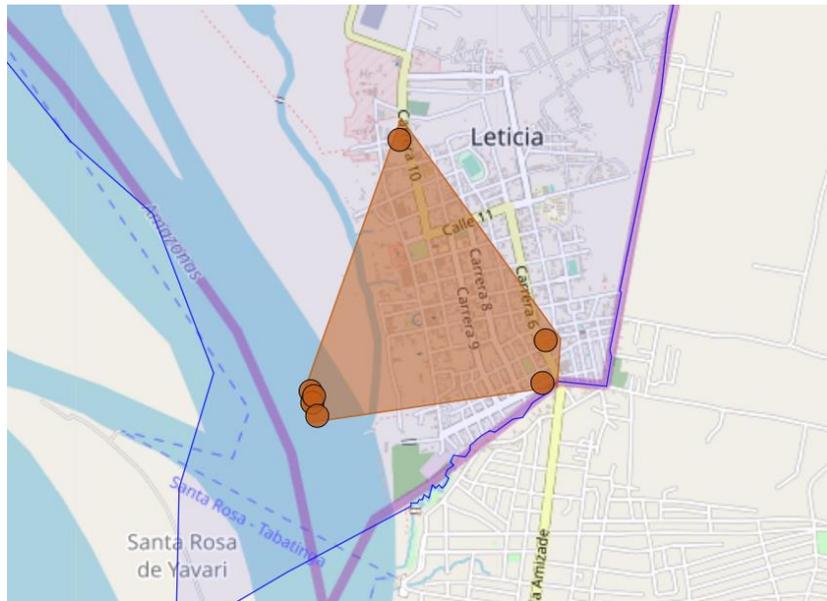
La Figura 63. Estimaciones del Modelo SAR, Centro Sur



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

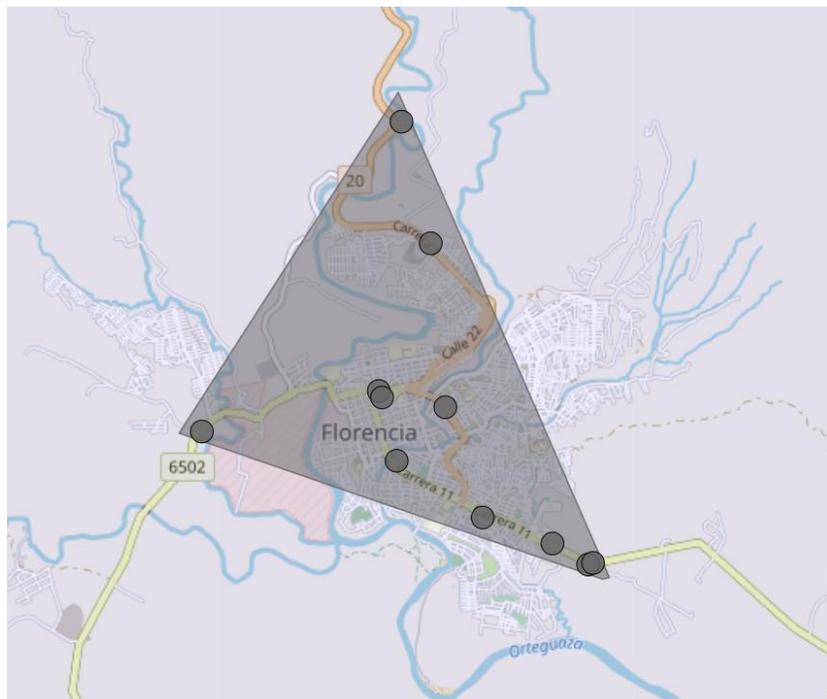


La Figura 66. Estimaciones del Modelo SAR, Leticia



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

La Figura 67. Estimaciones del Modelo SAR, Florencia



Fuente: Estimaciones de clústeres jerárquicos usando la metodología promedio y una distancia crítica máxima de 4 kilómetros. Las distancias usadas para generar los clústeres se computaron como distancias geodésicas.

## 6. Conclusiones

En el presente documento se muestran los resultados de aplicar la metodología de identificación de mercados relevantes en distribución de combustible para la totalidad del territorio colombiano esto es, para determinar el menor espacio físico en que existe evidencia de competencia en precios. Un estudio previo de la CREG (2016) planteaba una metodología para estimar la distancia óptima de definición de los mercados, pero su aplicación al contexto colombiano no arrojó resultados relevantes para la discusión de la regulación y la política pública sobre distribución minorista de combustibles. La metodología usada se basa en la estimación de un modelo Autorregresivo Espacial (SAR), el cual analiza los patrones de variación en el precio de los combustibles para determinar la distancia óptima de definición de un mercado relevante en cada contexto geográfico (Lee, 2007).

Se aprovechó la riqueza de datos que ha sido desarrollada por el fondo SOLDICOM a nivel de estación de servicio, para aterrizar la metodología planteada de manera general por la CREG (2016) a diferentes contextos de regulación de las estaciones de servicio minoristas. Para lograrlo, se propusieron e implementaron algunas modificaciones a la metodología mencionada, sugiriendo estimar un modelo SAR con datos panel y la inclusión de efectos fijos de tiempo y efectos fijos de estación de servicio, con lo que se deja de suponer que las estaciones de servicio son idénticas y se consideran los choques en el tiempo que afectan los precios del combustible. Utilizando estos radios de competencia o distancias óptimas, se implementó un modelo de conglomerados que permite agrupar estaciones de servicio, identificando finalmente los mercados relevantes en cada departamento.

En particular, se tomaron tres grandes divisiones de territorios: departamentos de frontera, ciudades capitales y el resto del país. La metodología en cada caso es ligeramente diferente: para los departamentos de frontera y las ciudades capitales, las matrices de distancias se llevaron a cabo con distancias de ruta, obtenidas a través de los servicios de Google; adicionalmente, para las ciudades principales se incluyó una estimación previa de los mercados relevantes asociados a las vías principales de entrada y salida de cada ciudad, bajo el supuesto de que estas EDS tienen un comportamiento especial y diferenciado de las EDS del interior de la ciudad. En contraste, las estimaciones para el resto del país, dividido en las regiones del Sistema General de Regalías (más una región especial con los departamentos de la frontera sur, Nariño y Putumayo) se hicieron con distancias geodésicas, lo cual permite ser más eficiente en términos de carga computacional.

En los casos en los que los datos lo permitían, se comparó el modelo sencillo, con efectos fijos por EDS, con un modelo más complejo en donde se introdujo la información adicional de cada EDS -tales como la bandera, el número de accesos viales a la estación, el número de surtidores, la presencia de mini mercado, entre otras-, como controles independientes en la regresión. Esto permitió ofrecer un panorama más detallado de los diferentes determinantes del precio del combustible, así como identificar el mayor o menor nivel de precios al consumidor asociado a cada bandera en cada zona geográfica.

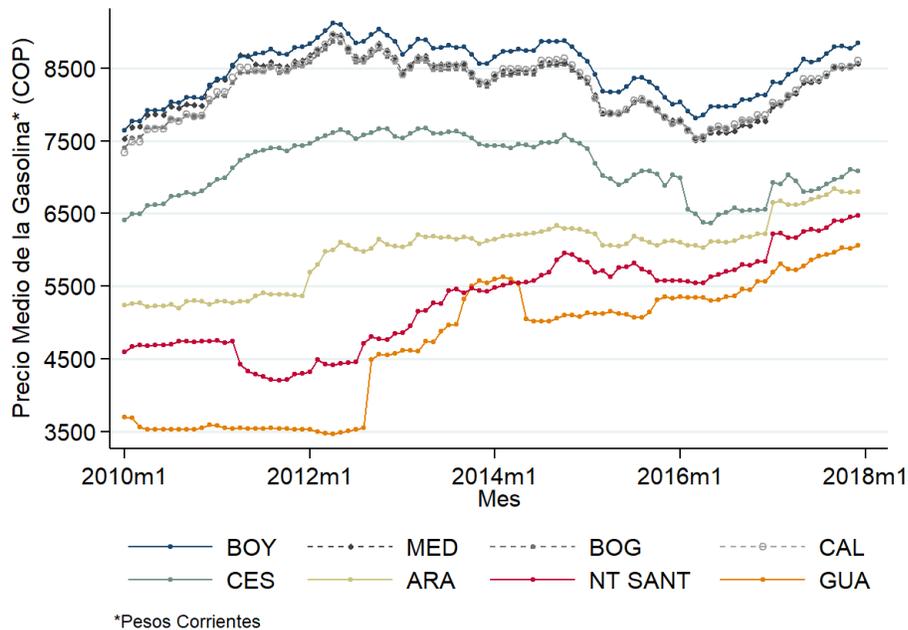
La presente investigación pretende hacer una contribución a la caracterización del mercado de distribución minorista de combustible. Se demostró que hay una correlación espacial de los

precios, la cual es inversamente proporcional a la distancia. Aunque cada área geográfica analizada arroja resultados diferentes, se puede resumir el ejercicio en unas cifras agregadas para todo el país. De las 4.792 EDS que cuentan con georreferenciación actualmente, se analizaron 4.295, dado que se encontró que 191 EDS estaban georreferenciadas por fuera de los límites del municipio al que pertenecían, y 287 EDS cuyos datos no aparecen en el panel de precios suministrado. A partir del análisis de las 4.295 EDS, se encontraron un total de 1.363 mercados relevantes en el país, lo que arroja un tamaño promedio de 3,2 EDS por cada mercado relevante. Por último, se encontró que la distancia óptima promedio que define los mercados relevantes de distribución de combustible en Colombia es de 2,8 km, no muy lejana a la cifra obtenida en los estudios previos, de 1 a 1,5 millas (Lee, 2007 y Hastings y Gilbert, 2005).

## Anexo A. Estadísticas Descriptivas

A continuación, se presenta los precios medios mensuales para las ciudades principales (Bogotá, Cali y Medellín) y las zonas de frontera (La Guajira, Cesar, Norte de Santander, Boyacá y Arauca) de la base de datos usada para las estimaciones del modelo SAR.

Figura 1A. Precios Promedio de la Gasolina por Ciudad o Departamento



Fuente: Estimaciones propias usando datos del Fondo Soldicom. Cada punto representa el precio promedio de la gasolina en las estaciones de servicio para un mes, entre los años 2010 y 2018.

## Anexo B. Manejo de la Información

Las fuentes principales de información que se usaron en este documento son: 1.) Censo de Caracterización de Estaciones de Servicio del Fondo Soldicom (2016), 2.) Panel Mensual de Precios de los Combustibles Líquidos del Fondo Soldicom (2010 - 2017) y 3.) Distancias de Ruta computadas usando Google Distance Matrix API y la georreferenciación de las estaciones de servicio a partir del Censo 2015.

Previo al uso de la información se realizaron algunas limpiezas al Censo 2015 y al Panel de Precios. A continuación se detallan en qué consistió la depuración de estas fuentes de datos:

- Se identificaron duplicados a nivel de código SICOM en la base del Censo. Algunos duplicados tienen incoherencias acerca del municipio de localización, observe que no es posible que una estación de servicio se encuentre en dos municipios diferentes. Debido

a que no es posible determinar el origen de las mencionadas EDS, éstas se excluyeron de la base de datos. 14 EDS presentaban la incoherencia señalada.

- Se excluyeron aquellas EDS para las cuales no se tiene información de longitud y latitud.
- Los nombres de los municipios señalados en el Censo Soldicom se unieron a los códigos de división administrativa (Divipola) y, usando los límites municipales a partir de la georreferenciación del DANE, se determinó si las coordenadas de cada EDS se encuentran al interior de sus respectivos municipios. Se encontraron 191 EDS cuyas coordenadas no están al interior de sus municipios.
- El panel de precios se encuentra desbalanceado, con estaciones de servicio que no registran precio para algunos meses. Para una estación de servicio, que en un mes determinado no registra precio, se imputó el precio del mes anterior más cercano para que sí se observa precio. Sin embargo, si el precio está por debajo del percentil 1 de los precios del municipio o por encima del percentil 99 se imputa respectivamente el percentil 1 o el percentil 99. Finalmente, se crea una variable categórica que indica si el precio para un mes dado fue imputado. Dicha variable fue incluida en todas las regresiones para controlar por el manejo que se le dio a los precios.

## Referencias

Belotti F., Hughes G. y Mortari, A. P. (2017). Spatial panel data models using Stata (March 25, 2016). CEIS Working Paper No. 373. In: <https://ssrn.com/abstract=2754703>

Comisión de Regulación de Energía y Gas (2016). Metodología para la definición de Mercados Relevantes en las Actividades de Distribución Mayorista y de Distribución Minorista de Combustibles Líquidos en Colombia. Circular 30 de 2017 mediante la cual se publica el estudio de Econometría para comentarios de los agentes dentro de la agenda de regulación.

Hastings, J. S. (2004). Vertical relationships and competition in retail gasoline markets: empirical evidence from contract changes in Southern California. *American Economic Review* 94, 317-328.

Hastings, J. S. y Gilbert R. J. (2005). Market power, vertical integration and the wholesale price of gasoline. *Journal of Industrial Economics* 53, 469-492.

Fondo Soldicom (2016) Censo de Caracterización de las Estaciones de Servicio.

Kao, Y. H. y Bera, A. K. (2016). Spatial regression: the curious case of negative spatial dependence. In: <https://drive.google.com/file/d/0B4iwIYuy9hBbDlmaWtkek4zMUU/view>

Lee, S. (2007) Spatial competition in the retail gasoline market: an equilibrium approach using SAR models. In: Three Essays on Spatial Econometrics and Empirical IO [https://etd.ohiolink.edu/!etd.send\\_file?accession=osu1211907715&disposition=inline](https://etd.ohiolink.edu/!etd.send_file?accession=osu1211907715&disposition=inline)

Neilson, H. y Bruce, D. (2007). The geographic extent of gasoline markets. In: <http://web.utk.edu/~dbruce/neilson.bruce.gasoline.pdf>

Ord, J. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

Teodorovicz, T., Leandro, T., y Esteves, L. A., (2015). The evolution of a methodology for relevant market definition: an analysis of CADE's jurisprudence on private higher educational markets.