

# Documentos de Trabajo

Análisis de elasticidades en el  
mercado automotor colombiano  
(2009 – 2011) mediante un modelo  
*logit* anidado

*Dennis Sánchez Navarro*

**No. 8**

**2013**

Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivadas 2.5 Colombia](#).

**Usted es libre de:**

Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra

**1.1.1 Bajo las condiciones siguientes:**

- **Atribución** – Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante. Si utiliza parte o la totalidad de esta investigación tiene que especificar la fuente.
- **No Comercial** – No puede utilizar esta obra para fines comerciales.
- **Sin Obras Derivadas** – No se puede alterar, transformar o generar una obra derivada a partir de esta obra.

Los derechos derivados de usos legítimos u otras limitaciones reconocidas por la ley no se ven afectados por lo anterior.



La serie Documentos de Trabajo es una publicación de la Superintendencia de Industria y Comercio. Los documentos son elaborados por los miembros del Grupo de Estudios Económicos o funcionarios de la entidad, y son de carácter provisional. Los análisis, opiniones y posibles errores son de responsabilidad exclusiva de los autores y no representa la posición de la Superintendencia de Industria y Comercio en la materia.

Para cualquier duda, sugerencia, corrección o comentario, escribir a: [estudioeconomicos@sic.gov.co](mailto:estudioeconomicos@sic.gov.co)

# **Análisis de elasticidades en el mercado automotor colombiano (2009-2011) mediante un modelo *logit anidado***

Dennis Sánchez Navarro<sup>1</sup>

## **Resumen**

El presente documento busca analizar el comportamiento de los hogares colombianos en la compra de vehículo nuevo para uso particular, y modelar el funcionamiento de la demanda de vehículos particulares en Colombia, entre 2009 y 2011. Esto a partir de la estimación de un modelo *logit anidado* que permite tomar como variables explicativas las características o atributos de los vehículos como determinantes de la decisión de compra de vehículo. Finalmente, se estiman las elasticidades con el fin de analizar el grado de sustituibilidad o contestabilidad del mercado automotor en Colombia.

Palabras clave: *logit anidado*, demanda, elasticidad, sustituibilidad, Colombia.

JEL: L62, C39, D12.

## **Abstract**

This paper analyzes the behavior of Colombian households in purchasing a new vehicle for private use, and model the performance of the demand for private vehicles in Colombia, between 2009 and 2011. This paper estimates a Nested Logit model that takes the characteristics or attributes of vehicles as explicative variables as determinants of the vehicle purchase decision. Finally, the elasticities are estimated in order to analyze the substitutability or contestability automotive market in Colombia.

Key words: Nested Logit, demand, elasticity, substitutability, Colombia.

JEL: L62, C39, D12.

---

<sup>1</sup> Magister en Ciencias Económicas. Economista del Grupo de Estudios Económicos de la Superintendencia de Industria y Comercio. E-mail: [dsanchez@sic.gov.co](mailto:dsanchez@sic.gov.co). Dirección de correspondencia: Carrera 13 No. 27-00, Piso 10 (Bogotá, Colombia).

La autora agradece especialmente a Miguel de Quinto, Juan Pablo Herrera y Jenny Paola Lis, compañeros de la entidad, por sus valiosos comentarios y a Natalia Cantor por sus aportes al documento.

## INTRODUCCIÓN

En los últimos años, los modelos de elección discreta han sido ampliamente utilizados en los trabajos empíricos de organización industrial, en particular los relacionados con análisis de estimación de demanda para productos o bienes claramente diferenciados. Estos modelos como lo señalan Akerberg & Rysman (2005) han permitido analizar las características no observables de los productos, mediante la inclusión de lo que denominan un componente simétrico no observado de diferenciación de producto (*symmetric unobserved product differentiation* –SUPD en inglés).

Este elemento resulta de gran utilidad para tratar de modelar los patrones de comportamiento de los consumidores en su proceso de elección y capturar patrones de sustitución complejos. Particularmente, los modelos de elección discreta como el modelo *logit* condicional propuesto por McFadden (1973) permite representar las decisiones individuales de elección a partir de:

- i) El conjunto de alternativas disponibles en el mercado.
- ii) Los atributos observables de dichos bienes.
- iii) El modelo de elección individual y distribución de los patrones de comportamiento en la población.

En línea con lo anterior, el modelo *logit* ha sido empleado, frecuentemente, por las facilidades computacionales que ofrece para su estimación y porque su especificación proviene de un problema de maximización de utilidad, en el cual las variables explicativas corresponden, principalmente, a las características o atributos de los productos. Asimismo, este tipo de modelos permite superar el problema de *dimensionalidad* intrínseco de la estimación de un modelo con bienes diferenciados. Sin embargo, la propuesta original de McFadden (1973) asume que la sustitución entre productos está dada entre las participaciones de mercado, más que por el grado de similitud de los mismos como lo señala Nevo (2000) e impone restricciones sobre las estructuras de sustitución y distribución del error como lo indican Wen & Koppelman (2001).

Estas limitaciones, resultan bastante restrictivas para modelar el mercado automotor colombiano puesto que en este último se observan diferentes segmentos y productos claramente diferenciados, sobre los cuales se definen características disímiles que son determinantes en la decisión de compra de un vehículo. De esa forma, no sólo influirán las características físicas u observables de los vehículos, sino de las diferentes alternativas e incluso elementos que no son fácilmente observables o cuantificables en términos econométricos, pero que para un consumidor pueden resultar definitivos, por ejemplo, los servicios post-venta.

Por esta razón, se decidió trabajar con el modelo *logit anidado* (Nested Logit NL en inglés), que se deriva del modelo de valores extremos generalizados (GEV en inglés) propuesto por McFadden (1978). Este modelo, provee como lo indican Wen & Koppelman (2001) mayor

flexibilidad en la estimación de patrones de sustitución y elasticidad cruzada entre pares de alternativas.

De esta forma, el presente trabajo emplea el modelo *logit anidado* con dos propósitos: (i) analizar el comportamiento de los hogares colombianos en la compra de vehículos nuevos para uso particular; (ii) calcular las elasticidades precio de la demanda, con el fin de identificar el grado de sustituibilidad o contestabilidad del mercado automotor colombiano. Para esto se utilizó: (i) la información de ventas mensuales de vehículos (automóviles) nuevos, suministrada por los principales concesionarios a nivel nacional, para el periodo enero de 2009 y diciembre de 2011; (ii) las principales características o atributos de los vehículos como variables explicativas; y (iii) las participaciones de mercado siguiendo a Berry (1994).

Asimismo, este documento hace una descripción detallada del proceso de estimación aplicado y de las pruebas empleadas, con el fin de fortalecer el carácter ilustrativo del texto, y apoyar futuras investigaciones relacionadas con la metodología empleada.

El presente documento cuenta con 5 secciones. En la primera parte se presentará una breve revisión de literatura de los principales modelos de elección discreta, empleados para mercados con bienes diferenciados y su aplicación empírica. Posteriormente, se describe la metodología a emplear, seguido por los resultados de la estimación. Por último, se presentan las conclusiones y consideraciones finales.

## 1. REVISIÓN DE LITERATURA

En la presente sección se hace una breve revisión de literatura en torno a los principales modelos de elección discreta empleados para la estimación de ecuaciones de demanda y cálculo de elasticidades, en particular, para mercados con bienes diferenciados.

Los modelos de elección discreta o de respuesta cualitativa resultan de gran utilidad para efectos del análisis de los productos diferenciados puesto que, como lo señalan Anderson, De Palma, & Thisse (1988) han sido desarrollados para describir el comportamiento de los consumidores cuando se enfrentan a un conjunto de posibilidades mutuamente excluyentes. Asimismo, los autores señalan que estos modelos asumen que la percepción individual de las características es heterogénea y no completamente observable.

Entre los modelos de elección discreta, el modelo *logit multinomial* condicional propuesto por McFadden (1973), ha sido ampliamente utilizado para estudiar el comportamiento de la elección de la población, partiendo de las reglas de decisión individuales, las alternativas disponibles en el mercado y los atributos de los bienes. Este modelo, en su versión original, asume que la relación entre cada par de alternativas resulta independiente de todas las demás. Este supuesto descrito por McFadden (1978, 1980) como la independencia de alternativas no relevantes (Independence of Irrelevant Alternatives –IIA en inglés) se refiere a que las probabilidades relativas de dos alternativas son independientes de los atributos de las mismas o incluso a la presencia de una tercera alternativa.

Este supuesto, como lo señalan Wen & Koppelman (2001) conduce a que las elasticidades cruzadas entre pares de alternativas resulten idénticas, lo cual no siempre puede ajustarse a la realidad y además puede generar sesgos en la estimación. Ante esta limitación, McFadden (1978) postuló un modelo que admite patrones de comportamiento de no-independencia entre las alternativas, conocido como el modelo de valores extremos generalizado (Generalized Extreme Value, GEV en inglés), que considera la no independencia entre alternativas extremas de la distribución.

El modelo *logit anidado* (Nested Logit en inglés), como lo señalan Koppelman & Wen (1998) es un caso especial del modelo GEV<sup>2</sup> de McFadden (1978) que es consistente con la teoría de maximización de la utilidad. Este modelo admite la interrelación entre alternativas de los distintos grupos o segmentos, lo que permite que el cambio de un atributo en alguna de ellas incida sobre la probabilidad de elección de otra opción para el consumidor, y que se asocie directamente a la pertenencia o no del mismo segmento o nido. Esta característica conduce a resultados más consistentes, puesto que las elasticidades estimadas se diferencian entre las alternativas, dependiendo de si cada par de productos está en el mismo nido o no. De esa

---

<sup>2</sup> Otros modelos como el logit combinatorio pareado (Paired Combinatorial Logit, PCL en inglés) y el logit anidado de sección cruzada (Cross-nested Logit, CNL en inglés), también corresponden a posibles casos del modelo GEV, de acuerdo con Wen & Koppelman (2001).

forma, el modelo en mención captura patrones de sustituibilidad complejos y diferenciados entre nidos o grupos de productos.

A pesar de sus múltiples ventajas, el modelo *logit anidado*, como lo señalan Wen & Koppelman (2001), mantiene restricciones sobre la igualdad de las elasticidades cruzadas entre pares de alternativas, dependiendo si estas hacen parte o no de un nido o segmento común, lo que podría resultar no ajustado a la realidad en algunos casos.

Ante esta limitación, se han propuesto métodos alternos para admitir elasticidades cruzadas diferentes entre pares de alternativas. Entre estos se encuentra el modelo *logit* combinatorio pareado desarrollado por Koppelman & Wen (2000) que permite la estimación de relaciones diferenciales entre cada par de opciones y el modelo de diferenciación de producto propuesto por Bresnahan, Stern & Trajtenberg (1997) que presenta una definición alternativa de similitud de producto, permitiendo así que aquellos productos pertenecientes al mismo segmento sean considerados sustitutos más cercanos entre sí, que aquellos productos que se encuentran en diferentes grupos.

El modelo *logit* mixto constituye otra alternativa que permite incorporar la heterogeneidad en las preferencias y relaja el supuesto IIA, permitiendo que la varianza del error no sea constante entre las diferentes alternativas según Bliemer & Rose (2010). Estos autores trabajan así con el modelo *logit* mixto con datos panel que incorpora explícitamente la posibilidad de que se correlacionen las diferentes alternativas de elección y la heterogeneidad de las preferencias. Hensher, Beck & Rose (2011) aplican este modelo al sector automotor, para investigar si hay diferencias significativas entre las preferencias de cada individuo cuando es aislado de las preferencias de su hogar.

El modelo *logit* y *logit anidado*, sin embargo, ha sido empleado por diversos autores para analizar los patrones de elección en múltiples aplicaciones para sectores como vivienda, transporte, inversión extranjera directa, televisión y sector automotor en trabajos como Chattopadhyay (2000), Yang, Zheng, & Zhu (2013), Rashidi, Auld, & Mohammadian (2012), Ravibabu (2013), Rasciute & Pentencost (2010), Danaher & Dagger (2012), Siriwardena, Hunt, Teisl & Noblet (2012), entre otros.

La aplicación extendida de estos modelos se explica sobre todo por su correspondencia con el problema neoclásico de maximización de utilidad del consumidor. En particular, Anderson, De Palma & Thisse (1988) confirman la relación explícita entre el modelo *logit* y la función de entropía, que puede ser considerada como una función de utilidad de un consumidor representativo, con un comportamiento de búsqueda de la variedad. Este autor concluye que el sistema de la demanda generado por el modelo *logit* se puede asociar a un consumidor que resuelve el problema de maximización de utilidad en un enfoque neoclásico.

Asimismo, el trabajo de Koppelman & Wen (1998) confirma la consistencia entre el modelo *logit anidado* y el problema de maximización de utilidad del consumidor representativo, al compararlo con el modelo *logit* anidado no normalizado (NNLM, por sus siglas en inglés)

adelantado por Daly (1987). Estos autores concluyen que este último no guarda relación directa con el problema de maximización de utilidad y recomiendan la utilización del modelo *logit anidado* al reconocer, entre otras cosas, que este obtiene elasticidades razonables y provee una clara interpretación de los parámetros entre las distintas alternativas.

En particular, el presente documento emplea el modelo *logit anidado* debido a: las facilidades interpretativas de sus parámetros, su consistencia con el problema de maximización de utilidad del consumidor representativo y el hecho de que permite reconocer la existencia de subgrupos de productos o nidos, que resultan clave para el sector automotor colombiano. Lo anterior, debido a que este sector se caracteriza por un alto grado de diferenciación, explicado por la multiplicidad de productos disponibles en el mercado, la diversidad de proveedores, marcas y servicios asociados que se ofrecen.

Entre los trabajos que han aplicado el *logit anidado* como metodología para la estimación de demanda de vehículos, se encuentran los adelantados por Berry (1994), Berry, Levinsohn & Pakes (1995), McCarthy & Tay (1998), Cecchini (2001), Verboven (2002), Requena-Silvente & Walker (2007), Restrepo (2010), Thomassen (2010), quienes realizan el análisis, considerando las características y atributos de los vehículos, así como los gustos idiosincráticos.

En particular, el presente documento seguirá de cerca la metodología propuesta por Berry (1994) quien parte del modelo *logit anidado* tradicional, y propone que la demanda sea descrita por ese modelo de elección discreta y que los precios sean endógenamente determinados por las firmas. Esta propuesta también trabajada por Nevo (2000) resulta de gran utilidad para el proceso de estimación, puesto que se toman las participaciones de mercado como instrumentos para encontrar los niveles de utilidad de cada bien, como se expone detalladamente en la siguiente sección.



## 2. METODOLOGÍA

En esta sección se presenta la metodología empleada para la estimación de la demanda de vehículos, que se adelantó con el propósito de analizar el comportamiento de los hogares colombianos en la compra de vehículo nuevo para uso particular. En primer lugar, vale la pena señalar que el sector automotor en Colombia se caracteriza por un alto grado de diferenciación debido a la multiplicidad de productos disponibles en el mercado y la gran diversidad de proveedores, marcas y servicios asociados que se ofrecen (Grupo de Estudios Económicos, Superintendencia de Industria y Comercio, 2012). Estos elementos hacen que la modelación de la decisión de compra de un vehículo nuevo resulte compleja al momento de capturar los esquemas de sustitución que pudieran derivarse en dicho mercado.

Específicamente, en este sector se reconoce que si bien dos bienes (en este caso automóviles), pueden tener características físicas similares, la percepción de un individuo puede diferir en términos de calidad, durabilidad, estatus o servicios, en el punto de venta, lo que dificulta la estimación de los efectos separados de cada característica, como señalan Berry (1994), McCarthy & Tay (1998).

De esta forma, un modelo de elección discreta que permite recoger las consideraciones hechas anteriormente puede partir, siguiendo a Berry (1994) y Berry *et al.* (1995) de la descripción de un modelo de elección que expresa la función de utilidad de un consumidor representativo en términos de las características observadas de los productos, así como el precio y otros aspectos no observados. De esa forma, una posible especificación de la utilidad del consumidor  $i$  frente al producto  $j$  sería:

$$u_{ij} = x_j \tilde{\beta} - \alpha p_j + \xi_j + \epsilon_{ij} \quad [1]$$

Donde,  $x_j$  corresponde al vector de características observadas,  $p_j$  al precio del producto  $j$ ,  $\xi_j$  a las características no observadas,  $\epsilon_{ij}$  es el término de error, finalmente  $\tilde{\beta}$  y  $\alpha$  son los parámetros a estimar.

No obstante, la existencia de características no observadas puede generar un problema de endogeneidad, puesto que podrían estar relacionadas con el nivel de precios y estar recogidas por el término de error, violando así el supuesto de no relación entre el término de error  $\epsilon$  y las variables explicativas  $E(\epsilon|x) = 0$ . Además, es importante señalar que no necesariamente se tiene una relación lineal entre las características no observadas y el término de error o las otras variables explicativas.

Por lo anterior, la propuesta de Berry (1994) parte de descomponer el parámetro de gustos ( $\beta$ ) del consumidor  $i$  para la característica  $k$ , y realizar una transformación que conduce a la siguiente expresión del nivel medio de utilidad del producto  $j$ , descrita en términos lineales<sup>3</sup>:

$$\delta_j(s) = x_j \beta - \alpha p_j + \xi_j \quad [2]$$

Como se observa, ahora esta expresión se presenta en términos de las participaciones o cuotas de mercado ( $s$ ), lo que permitirá estimar la elasticidad precio-propia de la demanda y la elasticidad precio-cruzada de la demanda.

Por otro lado, cabe señalar que en la ecuación (2), aun se admite una posible correlación entre las características no observadas  $\xi_j$  y el nivel de precios, por lo cual se tendrán que buscar instrumentos para la estimación de esta ecuación y evitar así el problema de endogeneidad.

Adicionalmente, reconociendo que este mercado particular corresponde a un modelo de productos diferenciados, podría tenerse una ecuación como (2) para cada gama de productos como mercados particulares. En ese caso, cada ecuación podría ser estimada si se tuviera una muestra lo suficientemente amplia. No obstante, si se asume que  $\xi_j$  es independiente entre las firmas, como lo propone Berry (1994), puede estimarse el modelo descrito en la ecuación (2) manteniendo información de tipo de corte transversal de un amplio número de firmas dentro de un único mercado.

Así las cosas, este autor señala dos mecanismos de estimación que permiten resolver el problema del consumidor en términos de la función de utilidad, expresada como una función de las participaciones de mercado observadas: i) el modelo de elección discreta *logit* y el ii) modelo de diferenciación vertical.

Para efectos del presente trabajo, se decidió trabajar con el modelo *logit* que permite expresar las participaciones de mercado del producto  $j$  de la forma:

$$s_j(\delta) = \frac{e^{\delta_j}}{\sum_{k=0}^N e^{\delta_j}} \quad [3]$$

Para la estimación de este modelo, se incluye lo que se denomina un bien externo denominado en la literatura como “*outside good*”, que permite modelar la posibilidad de que los consumidores tomen la decisión de no comprar vehículo, en lugar de condicionar el modelo a que todos los agentes tengan que comprar 1 vehículo en alguno de los productos considerados.

De esa forma, al incluir un “*outside good*” y asumir que la utilidad media de ese bien se puede normalizar a cero, el modelo descrito en (2) resultaría así:

$$\ln (s_j/s_o) = \delta_j \equiv x_j \beta - \alpha p_j + \xi_j \quad [4]$$

---

<sup>3</sup> Para mayor información sobre los problemas de no linealidad de las características no observadas, y la transformación aplicada para poder estimar la demanda por medio de la técnica de variables instrumentales véase Berry (1994).

Siendo  $s_j$  la participación o cuota de mercado del producto  $j$ , y  $s_0$  la participación del “*outside good*”, es decir, la fracción del mercado potencial que no compra vehículo en el periodo analizado y cuya construcción se expondrá más adelante.

De otra parte, si se busca que el modelo admita la interacción entre las diferentes características de los vehículos y las preferencias de los consumidores, se puede utilizar el modelo *logit anidado* propuesto por McFadden (1978), que admite que los gustos de los consumidores estén correlacionados entre diferentes productos, lo que permite representar importantes patrones de sustituibilidad que no recogería el modelo *logit* tradicional.

En este caso, se toma un conjunto de productos en un segmento o grupo  $g$  y se asume que el “*outside good*” es el único bien de un grupo denominado 0 (opción de no comprar). En este caso, la utilidad del consumidor  $i$  puede ser expresada como:

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\epsilon_{ij}^4 \quad [5]$$

Retomando (2) y asumiendo que el parámetro  $\zeta_{ig}$  sería idéntico para el consumidor  $i$ , para todos los productos del grupo  $g$ , cuya distribución está asociada al parámetro  $\sigma \in [0,1)$ , el modelo *logit anidado* permitiría capturar el grado de correlación entre grupos de productos similares y podría describirse como:

$$\ln(s_j / s_0) = x_j \beta - \alpha p_j + \sigma \ln(s_{j/g}) + \xi_j \quad [6]$$

Ecuación en la cual,  $s_j$  es la participación o cuota del mercado del producto  $j$ ,  $s_0$  es la participación del “*outside good*”,  $x_j$  son las características observadas del producto  $j$ , mientras que  $s_{j/g}$  corresponde a la participación del producto  $j$  en el segmento  $g$ . Así, se tendría un nuevo parámetro en la estimación  $\sigma$ , entendida como la medida del grado de correlación de las preferencias entre los consumidores al interior de un segmento  $g$ . Este parámetro podrá indicar entonces algunos casos de interés:

- i. Si  $\sigma \rightarrow 0$ , se tendría el modelo *logit* expuesto antes sin que los segmentos o grupos de productos resulten relevantes.
- ii. Si  $\sigma \rightarrow 1$ , se tendría un caso donde las preferencias están fuertemente correlacionadas al interior del segmento o grupo de productos, mostrando que los consumidores diferenciarían entre segmentos pero al interior de un grupo de bienes similares podrían tomarse estos últimos como posibles sustitutos.
- iii. Si  $\sigma = 1$ , sería el caso en el cual los productos o bienes al interior de un mismo grupo o segmento podrían considerarse perfectamente sustituibles.

---

<sup>4</sup> Se asume que el término de error  $\epsilon_{ij}$  es idénticamente e independientemente distribuido.

### Construcción de elasticidades

A partir de la estimación de los modelos descritos en las ecuaciones (4) y (6), se puede obtener la elasticidad propia y elasticidad cruzada de la demanda, al tomar las derivadas de  $s_j$  con respecto al precio del bien  $P_j$ , como lo han señalado autores como Berry (1994) y Restrepo (2010).

De esta forma, para el modelo *logit* tradicional se tendría que la elasticidad-precio del bien  $j$  sería:

$$\eta_j = \frac{\partial s_j}{\partial p_j} * \frac{P_j}{s_j} = -\alpha s_j (1 - s_j) * \frac{P_j}{s_j} \quad [7]$$

Así, la elasticidad precio-cruzada de la demanda se puede obtener al hallar la derivada de  $s_j$  con respecto al precio de otros productos  $P_k$ . En este caso, se tiene:

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} * \frac{P_k}{s_j} = \alpha s_j s_k * \frac{P_k}{s_j} \quad [8]$$

El modelo *logit anidado* por su parte, al considerar la existencia de grupos o segmentos de productos, expresaría la elasticidad precio de la demanda del bien  $j$  como:

$$\eta_j = \frac{\partial s_j}{\partial p_j} * \frac{P_j}{s_j} = -\frac{\alpha}{1-\sigma_j} s_j (1 - \sigma_{j/g} - (1 - \sigma_j) s_j) * \frac{P_j}{s_j} \quad [9]$$

Para obtener la elasticidad precio-cruzada de la demanda, derivada del modelo *logit anidado* se tendría lo siguiente:

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} * \frac{P_k}{s_j} = \frac{\alpha}{1-\sigma_j} s_j (\sigma_{j/g} + (1 - \sigma_j) s_k) * \frac{P_k}{s_j} \quad [10]$$

Con la especificación anterior, se procedió a realizar la estimación de los modelo de elección discreta expuestos en la ecuación (4) y ecuación (6), para lo cual se expondrán a continuación los datos empleados en el análisis, los resultados de la estimación y los principales resultados.

### 3. ESTIMACIÓN Y RESULTADOS

#### *Datos empleados*

Para la estimación de los modelos descritos en (4) y (6), se empleó una muestra de 715 vehículos para el periodo 2009-2011, empleando datos de ventas anuales al detal de vehículos (automóviles) nuevos, reportados por las ensambladoras y principales distribuidores cabeza de red, así como el precio promedio de venta al detal de cada vehículo, reportado por la Revista Motor.

En cuanto a las variables relacionadas con las características o cualidades de los automóviles, se utilizó la base de Fasecolda y las fichas técnicas de los vehículos correspondientes al modelo del año de referencia. La

Tabla 1 describe en detalle las variables consideradas con la respectiva fuente de información.

Para el modelo *logit anidado*, fue necesario clasificar los diferentes modelos considerados, en segmentos o subgrupos, para lo cual se utilizó la clasificación presentada por RENAULT – SOFASA presentada por Chagüi (2011) y disponible en el anexo 1 de este documento.

**Tabla 1. Variables empleadas en el análisis**

Variable	Nombre variable	Descripción	Fuente
Mercado potencial	<i>houshlds</i>	Corresponde al número de hogares colombianos reportado para el periodo 2008-2010 <sup>1</sup> que potencialmente comprarían carro durante cada año.	Encuesta de Calidad de Vida (ECV-DANE) 2008 Encuesta de Calidad de Vida (ECV-DANE) 2010
Precio vehículos nuevos	<i>princ</i>	Corresponde al promedio anual del precio al detal de los automóviles nuevos. Esta variable fue deflactada para ser expresada en términos de precios constantes de 2011. Asimismo, para representar las posibles mejoras en la capacidad adquisitiva de la población, se presenta el precio como fracción del ingreso per cápita real anual.	Revista motor mensual (Ene-2009 - Dic 2011) DANE (2011) Banco Mundial (2009-2011)
Ventas	<i>qu</i>	Cantidad vendida de vehículos nuevos al detal por ensambladoras y distribuidoras cabezas de red.	Ensambladoras y cabezas de red
"Outside good"	$s_0$	Se construyó como la diferencia entre el mercado potencial y las ventas totales de vehículos nuevos de cada año, con el fin de representar el número de hogares que toman la decisión de no comprar vehículo en ese año. Lo anterior se expresa en términos de participación del mercado, para lo cual se divide por el mercado potencial.	DANE Ensambladoras y cabezas de red
Participación de mercado	$s_j$	Corresponde al cociente entre las ventas anuales del vehículo (producto) $j$ y el mercado potencial de ese año.	Elaboración propia

Segmento	<i>seg</i>	Corresponde al segmento al cual pertenece el vehículo de referencia.	Segmentación Renault Sofasa
Aire Acondicionado	<i>aa</i>	Variable dicotómica que toma el valor de 1 si el vehículo cuenta con aire acondicionado, 0 en cualquier otro caso.	Fasecolda (Sistema informativo) Fichas técnicas de los vehículos
Ancho	<i>wi</i>	Corresponde al ancho del vehículo considerando los espejos retrovisores (en mm)	
Largo	<i>li</i>	Corresponde a la longitud del vehículo (en mm)	
Alto	<i>he</i>	Corresponde a la altura del vehículo (en mm)	
Caja	<i>caja</i>	Variable cualitativa que describe el tipo de caja de cambios con la que cuenta cada vehículo, (variables dummies) -Caja mecánica → 0 -Caja automática o triptonic → 1	
Cilindraje	<i>cy</i>	Cilindraje expresado en litros para cada vehículo en el modelo del año correspondiente.	
Combustible	<i>comb</i>	Variable cualitativa que describe el tipo de combustible con el que opera el vehículo (variable dummy): -Gasolina GSL → 1 -Diesel DSL → 0	
Forma	<i>forma</i>	Variable dummy cualitativa que describe la forma del vehículo: -Hatchback -Sedán -Coupé	
Gama	<i>gama</i>	Variable cualitativa que describe la gama a la cual pertenece el vehículo (variables dummies) -Gama alta -Gama media -Gama baja	
Importación	<i>import</i>	Variable cualitativa que describe si el vehículo fue importado o ensamblado en el país. Variable dummy: -Importado → 0 -Ensamblado → 1	
Pasajeros	<i>pas</i>	Número de pasajeros que admite el vehículo	
Potencia	<i>hp</i>	Caballos de fuerza reportados para cada vehículo en el modelo del año correspondiente.	

Fuente: elaboración propia.

<sup>1/</sup> El dato correspondiente al año 2009, se obtuvo por interpolación debido a la frecuencia con la que se realiza la ECV, mientras que para el año 2011, se hizo una extrapolación de la tasa de crecimiento registrada en los años anteriores.

Como se observa en la Tabla 1, se tienen diferentes variables con el fin de hacer la caracterización de los diferentes vehículos, que pueden explicar los factores diferenciadores

de la demanda en cada caso. Asimismo, se considera el número de hogares colombianos como el tamaño del mercado potencial y con esta variable se construye el “*outside good*”<sup>5</sup>, que posteriormente será empleado en la definición de la variable dependiente. Como se expuso anteriormente, la variable a explicar corresponde al cociente entre la participación del producto *j* y la participación del “*outside good*”.

Precisamente esta forma de definir la variable dependiente, pretende incluir la opción de no comprar vehículo por parte de los hogares colombianos. Considerando lo anterior, se exponen a continuación: i) el análisis preliminar de las variables, ii) los modelos estimados, y iii) los principales resultados.

### *Análisis preliminar*

Debido a la multiplicidad de variables cualitativas involucradas en el análisis, es válido preguntarse, previo a la formulación del modelo, qué tan correlacionadas podrían estar las variables consideradas, con el fin de evitar incluir variables que puedan resultar redundantes entre sí y que puedan generar un problema de multicolinealidad que afecte la estimación.

De esta forma, se presentan algunas estadísticas básicas de las variables, con el fin de identificar posibles relaciones importantes que valga considerar previo al proceso de estimación, en especial porque como se expuso antes, el modelo propuesto presenta un problema de endogeneidad que deberá ser tratado.

Adicionalmente, se analiza la matriz de correlaciones entre las variables explicativas del modelo con el fin de detectar potenciales problemas de multicolinealidad. De esta forma, se encuentra un alto grado de correlación entre las variables *hp* y *cy*, con un coeficiente superior a 0,8, así como una importante relación entre *hp* y *peso*, y las variables *princ* y *cy*. De esta forma, vale la pena tener en cuenta que pueden presentarse posibles problemas de multicolinealidad, al incluir las variables *peso*, *cy* y *hp* simultáneamente.

---

<sup>5</sup> Corresponde a la variable resultante de la diferencia entre el número total de hogares y el número de ventas totales de vehículos nuevos en ese año del número total de hogares.

**Tabla 2. Estadística descriptiva variables de interés**

Variable	Obs	Mean	Std. Dev	Min	Max
<i>li</i>	715	4.299,9	435,1	1.531,0	5,247
<i>wi</i>	715	1.736,9	107,9	1.240,0	2.220,0
<i>he</i>	715	1.474,3	67,3	1.250,0	2.020,0
<i>peso</i>	715	1.330,7	283,5	720,0	2.830,0
<i>cy</i>	715	1.909,3	616,2	995,0	6.162,0
<i>cla</i>	715	3,4	1,8	1,0	7,0
<i>caja</i>	715	1.088,0	1,0	-	1,0
<i>forma</i>	715	2,4	0,7	1,0	3,0
<i>gama</i>	715	0,8	0,7	-	2,0
<i>import</i>	715	0,9	0,3	-	1,0
<i>aa</i>	715	0,9	0,3	-	1,0
<i>hp</i>	715	141,5	60,0	52,0	420,0
<i>pas</i>	715	4.918,0	0,4	2,0	5,0
<i>comb</i>	715	1,0	0,1	-	1,0
<i>qu</i>	715	366,6	1.069,5	2,0	14.210,0
<i>pop</i>	715	1.519.183,0	19.623,6	1.495.171,0	1.543.878,0
<i>princ</i>	715	6.098,0	4.118,0	1.349,0	34.242,0
<i>houshlds</i>	715	288.383,0	6.654,4	280.242,8	296.754,5
<i>so</i>	715	0,7	0,1	0,6	0,8

Fuente: cálculos propios.

**Tabla 3. Matriz de correlación entre variables explicativas**

	<i>li</i>	<i>wi</i>	<i>he</i>	<i>peso</i>	<i>cy</i>	<i>D_caja1</i>	<i>D_caja2</i>	<i>D_forma1</i>	<i>D_forma2</i>	<i>D_forma3</i>	<i>D_gama1</i>	<i>D_gama2</i>	<i>D_gama3</i>	<i>import</i>	<i>aa</i>	<i>hp</i>	<i>pas</i>	<i>comb</i>	<i>princ</i>
<i>li</i>	<b>1,000</b>																		
<i>wi</i>	0,401	<b>1,000</b>																	
<i>he</i>	-0,085	-0,284	<b>1,000</b>																
<i>peso</i>	0,456	0,482	-0,111	<b>1,000</b>															
<i>cy</i>	0,513	0,485	-0,287	0,564	<b>1,000</b>														
<i>D_caja1</i>	0,326	0,329	-0,129	0,428	0,45	<b>1,000</b>													
<i>D_caja2</i>	-0,326	-0,329	0,129	-0,428	-0,455	-1,000	<b>1,000</b>												
<i>D_forma1</i>	-0,002	0,063	-0,255	0,123	0,25	0,0675	-0,0675	<b>1,000</b>											
<i>D_forma2</i>	-0,447	-0,279	0,216	-0,290	-0,39	-0,2356	0,2356	-0,2984	<b>1,000</b>										
<i>D_forma3</i>	0,435	0,228	-0,039	0,198	0,21	0,1831	-0,1831	-0,3805	-0,7691	<b>1,000</b>									
<i>D_gama1</i>	-0,192	-0,252	0,134	-0,254	-0,31	-0,2547	0,2547	-0,1503	0,1364	-0,0315	<b>1,000</b>								
<i>D_gama2</i>	0,446	0,505	-0,211	0,475	0,55	0,4612	-0,4612	0,1903	-0,2722	0,1363	-0,7489	<b>1,000</b>							
<i>D_gama3</i>	-0,383	-0,385	0,122	-0,339	-0,37	-0,3182	0,3182	-0,0694	0,2069	-0,1540	-0,2859	-0,4210	<b>1,000</b>						
<i>import</i>	-0,107	-0,038	0,091	-0,049	0,01	0,0011	-0,0011	0,0550	0,0625	-0,0973	-0,0391	0,0354	0,0023	<b>1,000</b>					
<i>aa</i>	0,260	0,299	-0,206	0,244	0,27	0,2136	-0,2136	-0,0017	-0,1862	0,1815	-0,0776	0,3337	-0,3764	-0,0936	<b>1,000</b>				
<i>hp</i>	0,510	0,537	-0,288	0,598	0,88	0,4926	-0,4926	0,2627	-0,3484	0,1617	-0,3438	0,5965	-0,3920	0,0797	0,2792	<b>1,000</b>			
<i>pas</i>	0,032	-0,085	0,287	-0,104	-0,23	-0,0976	0,0976	-0,2548	-0,0081	0,1785	0,1382	-0,1845	0,0777	-0,0523	-0,0633	-0,2602	<b>1,000</b>		
<i>comb</i>	0,069	-0,259	-0,021	-0,178	-0,13	-0,0990	0,0990	0,0502	0,0565	-0,0884	0,0932	-0,1244	0,0524	-0,0489	-0,0427	-0,1777	-0,0253	<b>1,000</b>	
<i>princ</i>	0,435	0,486	-0,236	0,542	0,688	0,4431	-0,4431	0,2090	-0,2671	0,1188	-0,3977	0,6301	-0,3668	0,1440	0,2636	0,8068	-0,1917	-0,2297	<b>1,000</b>

Fuente: cálculos propios.

Teniendo en cuenta lo anterior, se procede a describir los modelos que fueron estimados con el fin de analizar la demanda que realizan los hogares colombianos por vehículos nuevos de los diferentes segmentos y las diferentes gamas, teniendo como variables explicativas las características y atributos presentados previamente.



## Proceso de estimación y principales resultados

En primer lugar, vale la pena señalar que debido al alto grado de correlación identificado preliminarmente por las variables y para evitar caer en la “trampa de las variables ficticias”, no se incluyen todas las variables presentadas en la Tabla 1. Para las variables dummies se toman  $n - 1$  características para ser incluidas en la estimación; por ejemplo, aunque se tienen 3 gamas o 3 formas identificadas en la estimación sólo se incluyen 2 variables *dummy* asociadas a cada atributo.

Asimismo, para evitar redundancia entre las variables del modelo que pueden conllevar posibles problemas de multicolinealidad, no se incluyen de forma simultánea las variables *peso*, *cilindraje* (*cy*) y *potencia* (*hp*). Por último, vale la pena señalar que debido al problema de endogeneidad expuesto anteriormente, no se toma explícitamente la variable *princ* para la estimación, sino que se utilizan otras variables relacionadas con las características o atributos de los vehículos como instrumentos, siguiendo a Restrepo (2010) para la estimación de los modelos propuestos.

En particular, se toma la característica  $k$  del producto  $j$  para construir al menos dos instrumentos por cada atributo: i)  $k$  –ésima característica del vehículo  $j$ ; y la ii) sumatoria de la  $k$  –ésima característica de los competidores, y en el caso del modelo anidado se utiliza la sumatoria de la  $k$  –ésima característica de los competidores al interior de un mismo segmento.

Por lo anterior, se especificaron los siguientes modelos a ser estimados.

Modelo *logit*- estimado por MCO<sup>6</sup>

$$\ln\left(\frac{s_j}{s_o}\right) = \beta_0 + \beta_1 princ_j + \beta_2 d_{caja_{1j}} + \beta_3 d_{forma_{1j}} + \beta_4 d_{forma_{3j}} + \beta_5 d_{gama_{2j}} + \beta_6 d_{gama_{3j}} + \beta_7 pas_j + \beta_8 comb_j + \beta_9 aa_j + \beta_{10} import_j + \beta_{11} wi_j + \beta_{12} hp_j + \beta_{13} li_j + \beta_{14} he_j + \xi_j \quad [13]$$

Modelo *logit anidado* – estimado por MCO<sup>7</sup>

$$\ln\left(\frac{s_j}{s_o}\right) = \beta_0 + \beta_1 princ_j + \beta_2 d_{caja_{1j}} + \beta_3 d_{forma_{1j}} + \beta_4 d_{forma_{3j}} + \beta_5 d_{gama_{2j}} + \beta_6 d_{gama_{3j}} + \beta_7 pas_j + \beta_8 comb_j + \beta_9 aa_j + \beta_{10} import_j + \beta_{11} wi_j + \beta_{12} hp_j + \beta_{13} li_j + \beta_{14} he_j + \sigma_j(\ln s_{j/g}) + \xi_j \quad [14]$$

<sup>6</sup> Toma como variables explicativas (regresores) las diferentes características de los vehículos.

<sup>7</sup> Es similar al modelo anterior, pero incluye el parámetro de la participación en el mercado del producto  $j$  al interior del segmento  $g$ . Donde  $\sigma$  será el parámetro que mide el grado de correlación entre las preferencias de los consumidores en un segmento dado.

Modelo *logit* IV – variables instrumentales (IV)

$$\ln\left(\frac{s_j}{s_o}\right) = \beta_0 + \beta_1 \widehat{princ}_j + \beta_2 d_{caja1j} + \beta_3 d_{forma1j} + \beta_4 d_{forma3j} + \beta_5 d_{gama2j} + \beta_6 d_{gama3j} + \beta_7 pas_j + \beta_8 comb_j + \beta_9 aa_j + \beta_{10} import_j + \xi_j \quad [15]$$

Siendo  $\widehat{princ}$  una variable instrumento construida a partir de los atributos del vehículo que tienen un importante grado de correlación el precio. Asimismo, siguiendo a Restrepo (2010) no sólo se utilizan las características del producto  $j$  como instrumentos en la estimación, sino las características promedio de los demás competidores. De esa forma, por cada variable o atributo considerado al menos se tienen dos instrumentos para las variables de interés que se desee instrumentalizar.

Modelo *logit anidado* IV – variables instrumentales (IV)

$$\ln\left(\frac{s_j}{s_o}\right) = \beta_0 + \beta_1 \widehat{princ}_j + \beta_2 d_{caja1j} + \beta_3 d_{forma1j} + \beta_4 d_{forma3j} + \beta_5 d_{gama2j} + \beta_6 d_{gama3j} + \beta_7 pas_j + \beta_8 comb_j + \beta_9 aa_j + \beta_{10} import_j + \sigma_j (\ln \widehat{s_{j/g}}) + \xi_j \quad [16]$$

Donde,  $\widehat{princ}_j, \ln(\widehat{s_{j/g}})$  son variables instrumentales, pero la última está construida con las características de los diferentes vehículos, diferenciando por el segmento al cual pertenece y las características promedio de los demás competidores al interior del mismo segmento. Reconociendo un posible problema de endogeneidad en los modelos expuestos anteriormente, se emplea la metodología de variables instrumentales. Esto con el fin de evitar la correlación entre variables no observadas que estarían recogidas en el término de error del modelo y que podrían sesgar los resultados del modelo al estar correlacionadas con otras variables explicativas consideradas en la estimación.

A continuación se presentan los resultados de la estimación de los modelos expuestos anteriormente.

**Tabla 4. Resultados de la estimación**

Variable	Logit MCO	Logit Anidado MCO	Logit IV	Logit Anidado IV
<i>princ</i>	-0,1266*	-0,0414*	-0,3426*	-0,235*
<i>s.e.</i>	0,027	0,009	0,044	0,024
<i>P-value</i>	0,000	0,000	0,000	0,000
<i>D_caja1</i>	-0,0723	0,0635	0,1000*	0,1566*
<i>s.e.</i>	0,161	0,053	0,169	0,078
<i>P-value</i>	0,654	0,233	0,554	0,044
<i>D_forma1</i>	0,6337*	0,1347**	0,7473*	0,3576*
<i>s.e.</i>	0,215	0,071	0,228	0,115
<i>P-value</i>	0,003	0,057	0,001	0,002

<i>D_forma3</i>	0,9585*	-0,0597	0,9238**	0,1705**
<i>s.e.</i>	0,152	0,056	0,158	0,101
<i>P-value</i>	0,000	0,289	0,068	0,091
<i>D_gama2</i>	-0,8613*	-0,7688*	-0,2697	-0,3953*
<i>s.e.</i>	0,191	0,072	0,221	0,108
<i>P-value</i>	0,000	0,000	0,223	0,000
<i>D_gama3</i>	-0,4293	0,0671	-0,5427**	-0,1211
<i>s.e.</i>	0,299	0,065	0,298	0,119
<i>P-value</i>	0,152	0,301	0,068	0,311
<i>pas</i>	0,0106	0,0332	-0,034	-0,0091
<i>s.e.</i>	0,113	0,036	0,138	0,075
<i>P-value</i>	0,925	0,352	0,806	0,903
<i>comb</i>	1,105*	0,1536	0,066	-0,3166
<i>s.e.</i>	0,319	0,197	0,471	0,36
<i>P-value</i>	0,001	0,436	0,888	0,38
<i>aa</i>	-0,5175**	0,059	-0,4516	-0,0907
<i>s.e.</i>	0,293	0,094	0,288	0,129
<i>P-value</i>	0,078	0,529	0,116	0,483
<i>import</i>	-0,5343*	-0,2721*	-0,22	-0,0073
<i>s.e.</i>	0,255	0,064	0,242	0,101
<i>P-value</i>	0,036	0,000	0,364	0,467
<i>wi</i>	0,0004	-0,0008*		
<i>s.e.</i>	0,001	0,003		
<i>P-value</i>	0,508	0,006		
<i>hp</i>	-0,0039*	-0,0019*		
<i>s.e.</i>	0,002	0,0005		
<i>P-value</i>	0,032	0,001		
<i>li</i>	-0,0002	-0,0003*		
<i>s.e.</i>	0,000	0,000		
<i>P-value</i>	0,169	0,000		
<i>he</i>	0,0008	0,001		
<i>s.e.</i>	0,001	0,000		
<i>P-value</i>	0,457	0,026		
<i>ln_cond_s/g</i>		0,9471		0,6651*
<i>s.e.</i>		0,011		0,072
<i>P-value</i>		0,000		0,000
<i>_const</i>	-8,1299*	-1,6961	-5,8787*	-3,0657*
<i>s.e.</i>	2,204	0,85	1,031	0,753
<i>P-value</i>	0,000	0,046	0,000	0,000

/\* Corresponde a los coeficientes que son significativos al 5% de significancia.

\*\* Corresponde a los coeficientes que resultan significativos al 10% de significancia.

Fuente: cálculos propios.

Adicionalmente, se presentan los resultados de la prueba de sobre-identificación J-Hansen que fue realizada para evaluar la correspondencia de los instrumentos propuestos, en los modelos (15) y (16).

**Tabla 5. Prueba de sobre-identificación J-Hansen**

Modelo	Estadístico J-Hansen	P-value
<i>Logit IV</i>	6,183	0,28885
<i>Logit anidado IV</i>	8,081	0,15185

Fuente: cálculos propios.

Esta prueba permite confirmar que los instrumentos considerados para la estimación, tanto en el modelo *logit* como en el modelo *Logit Anidado* estimados por el método de variables instrumentales son los adecuados, puesto que como se observa en la Tabla 5, el  $p - value > \alpha = 0,05$  (nivel de significancia). En otras palabras, no se encuentra evidencia para rechazar la hipótesis nula que plantea la exogeneidad de los instrumentos utilizados, por lo cual, los instrumentos considerados son válidos con un nivel de significancia del 5%<sup>8</sup>.

Teniendo en cuenta lo anterior, al confrontar los estimadores de los modelos *logit* y *logit anidado* obtenidos por MCO con los modelos estimados con variables instrumentales, se confirma el sesgo de los primeros ante los problemas de endogeneidad antes descritos. Así las cosas, se considera que la estimación por variables instrumentales ayuda a corregir los problemas de endogeneidad identificados. De esa forma, el siguiente paso para la selección del modelo definitivo consiste en evaluar la significancia del parámetro  $\sigma$ , que se expuso previamente como la medida del grado de correlación de las preferencias entre los consumidores al interior de un segmento.

Para lo anterior, se analiza la significancia del parámetro en el modelo *logit Anidado IV* y se encuentra que el parámetro  $\sigma_j$ , correspondiente al coeficiente de la variable *ln\_cond\_s/g* resulta significativo al presentar un  $p\text{-value} < 0,05$ . Lo anterior, señalaría la importancia de la existencia de segmentos o sub-grupos en el mercado automotor que son diferenciados para el consumidor y que presentan unos patrones de sustituibilidad complejos.

De esa forma, al encontrarse significativo el parámetro  $\sigma_j$ , se encuentra que resulta importante la segmentación al interior del mercado automotor, puesto que un individuo representativo percibe sub-grupos de productos claramente diferenciados. Esto indicaría que el modelo adecuado sería el *logit anidado* estimado por variables instrumentales.

Asimismo, el valor del coeficiente 0,665 sería indicativo de un nivel de sustituibilidad importante al interior de cada segmento, que es consistente con la estructura de un mercado

<sup>8</sup> Sin embargo, es importante señalar que el modelo presenta cierta sensibilidad a los instrumentos seleccionados, aunque mantiene en la mayoría de los casos la consistencia en la significancia de los principales estimadores.

que es claramente diferenciado, por lo cual, incluso al interior de un segmento, no se podrían caracterizar los diferentes productos como bienes sustitutos perfectos.

Reconociendo todo lo anterior, vale la pena señalar que se encuentra consistencia en los signos esperados de los estimadores. En primer lugar, se observa una incidencia negativa y significativa del precio sobre la participación de mercado, que es razonable en el caso de un bien normal, puesto que en la medida en que aumenta el precio se esperaría que, ante un mercado con sustitutos cercanos, la demanda disminuya. Asimismo, se identifican como características diferenciadoras el tipo de caja de cambios, la gama y la forma del vehículo, mientras que otras variables como el aire acondicionado, el tipo de combustible o el carácter de importado del vehículo no sería significativas<sup>9</sup> en la determinación de la participación de mercado de un producto en el mercado automotor para la muestra analizada.

El hecho de encontrar al tipo de caja de cambios, la gama y la forma de vehículo como atributos o características diferenciadoras podría ser indicio de una clara percepción que están tomando los consumidores de esas características como factores valiosos y diferenciadores de producto.

En particular, el resultado de encontrar el tipo de caja como cualidad diferenciadora puede responder al relativo carácter de novedad del tipo de caja automática y triptonic, siendo esta última desde 2008 una característica que ha empezado a masificarse en múltiples segmentos en lugar de concentrarse en un conjunto particular de automóviles de gama alta. Asimismo, la diversidad de productos que se ha incorporado al mercado colombiano ha permitido mejorar variedad de formas y modelos en los diferentes segmentos, que están siendo valorados por los consumidores como características particulares.

Por el contrario, otras características que se tienen en el mercado colombiano desde tiempo atrás, como son la tenencia de aire acondicionado, el tipo de combustible, o el carácter de importado, no operan como patrones diferenciadores de productos, en tanto se han convertido en características o cualidades generalizadas en el mercado. En particular, el aire acondicionado o el tipo de combustible son elementos que pueden tenerse en un tipo de vehículo particular, pues al momento de tomar la decisión de compra es común encontrar la posibilidad de adquirir las distintas versiones de los vehículos.

De igual forma, la característica *import*, esto es, ser un vehículo importado no presenta un atributo que resulte marcadamente diferenciador contrastado con un vehículo ensamblado, puesto que el coeficiente *import* no resulta significativo en el modelo analizado. Esto podría explicarse por varios elementos en consideración:

- i) Carácter de un mercado globalizado, con mayor reconocimiento del producto por marca que por país de origen.

---

<sup>9</sup> Es importante señalar que las otras características presentadas previamente como cilindraje, potencia, altura, longitud, peso, se emplearon como instrumentos para la modelación del precio.

- ii) La existencia de los acuerdos comerciales han reducido relativamente los diferenciales en términos de precios que podrían originarse entre los vehículos nacionales y ensamblados.
- iii) La presencia de plantas de producción de una misma marca en diferentes locaciones, que operan como plataformas exportadoras, lo cual permite que una misma marca provea un mercado con vehículos que han sido ensamblados en distintos países y podría fisurar un poco la percepción del importado o la afiliación a un país-proveedor particular.

En cuanto a las variable asociada a la gama del vehículo se observa que sólo resulta significativo el coeficiente asociado a gama 2, que corresponde a la gama alta, es decir, que los vehículos de gama alta tienen una participación menor en el mercado, lo cual resulta razonable.

#### Estimación de elasticidades

Para efectos de analizar el grado de sustituibilidad o contestabilidad del mercado automotor en Colombia, se utiliza el modelo *logit anidado IV* para la estimación de las elasticidades del mercado. De esa forma, se remplazan los valores obtenidos en el modelo referido, en las ecuaciones 9 y 10, para la estimación de la elasticidad propia de la demanda y la elasticidad precio-cruzada de la demanda, manteniendo los segmentos identificados.

Los resultados de estos cálculos por segmento se observan en la Tabla 6, en la cual se señala la elasticidad precio propia de la demanda al interior de cada segmento y la elasticidad precio-cruzada de la demanda.

**Tabla 6. Elasticidades precio de la demanda – Modelo Logit Anidado VI**

Segmento		Elasticidad precio-propia	Elasticidad precio-cruzada
		2009-2011	2009-2011
i1	Vehículos compactos de entrada de gama	-1,362	3,08E-04
i2	Vehículos pequeños deportivos motor 1,6L	-2.658	1,01E-05
M0	Vehículos pequeños sedanes entre 1,4 y 1,6 L	-1.993	4,49E-04
M1A	Vehículos medianos entre 1,8 y 2,0L	-3.746	1,34E-04
M1BM	Vehículos medianos entre 1,6L y 1,8L	-2.751	1,29E-04
M2	Vehículos medianos grandes	-4.850	1,30E-04
S	Vehículos sofisticados, gama Premium	-7.393	9,43E-06

Fuente: cálculos propios.

Lo anterior, permite señalar que para la muestra analizada en el periodo 2009-2011, se encuentra un mercado elástico. En todos los segmentos se observa una elasticidad-precio de la demanda mayor a 1, lo cual sería indicativo de una amplia oferta en el mercado, incluso si se analizan como segmentos diferenciados.

Así mismo, se observa que el segmento gama Premium y los vehículos medianos grandes, considerados en la muestra, son los que presentan mayor sensibilidad a las variaciones del precio, lo cual a primera vista podría resultar contra-intuitivo. Se esperaría que los consumidores que hacen parte del segmento sofisticado fueran menos sensibles a pequeñas variaciones en los precios, e incluso existieran otros parámetros diferentes al precio que determinarían su compra. Sin embargo, esto podría explicarse por tres razones principalmente.

1. En primer lugar, este resultado corresponde al carácter de bien de lujo de los vehículos sofisticados o Premium, lo que ocasionaría que un consumidor de este tipo de vehículos de alta gama pudiese prescindir fácilmente de este bien, comparado con otros bienes.
2. En segundo lugar, podría pensarse que en los segmentos más bajos, en los cuales realmente la posibilidad del consumidor de desviarse a otro segmento es más restringida, y el carácter de bien necesario podría predominar, el mercado sería bastante elástico en estos segmentos.
3. En tercer lugar, la forma como se construyó la variable dependiente como la probabilidad de compra del producto  $j$  versus la posibilidad de no compra (“*outside good*”) puede incidir en la sensibilidad de los resultados.

Otra posible explicación a este resultado radica precisamente en la forma como se definió la variable dependiente, es decir la probabilidad de compra del producto  $j$  versus la posibilidad de no compra (“*outside good*”). De esa manera, se refuerza el carácter de bien de lujo de los vehículos más sofisticados, puesto que la opción de no compra en este segmento sería más marcada, aún más reconociendo que la variedad en el mismo es más limitada.

De otra parte, se observa que las elasticidades cruzadas de la demanda en los diferentes segmentos resultan significativamente bajas, y el valor del coeficiente del *logit Anidado IV*,  $\sigma_j = 0,665$  sería indicativo de un nivel de sustituibilidad importante al interior de cada segmento. Lo anterior, podría ser un indicio de la existencia de distintos mercados que no ejercen presiones competitivas importantes entre ellos, sino que el grado de sustituibilidad se da al interior de cada segmento particular.

Igualmente, vale la pena señalar como lo hacen Akerberg & Rysman (2005), que el modelo *logit* y *logit anidado*, y en general los modelos de elección discreta, que admiten la inclusión de la “diferenciación simétrica no observada de producto” (SUPD en inglés), podrían subestimar las elasticidades de los mercados con mayor número de productos por un posible efecto congestión, y así que podría generarse alguna distorsión sobre las elasticidades de los nidos con mayor variedad de productos. De esa manera, los autores consideran que el método de estimación *logit anidado* estándar, podría sobre-estimar en algunos casos la elasticidad propia de la demanda y que en presencia del efecto congestión, los métodos estándar de estimación son más propensos a sobre-estimar las elasticidades precio cruzadas al interior de cada grupo, y subestimar las elasticidades entre los grupos.

Autores como Huang & Rojas (2013) atribuyen el posible sesgo a la especificación del “*outside good*” que incide en la definición del mercado potencial de forma inadecuada, puesto que si se asume un mercado potencial demasiado grande podrían encontrarse elasticidades precio-cruzadas artificialmente bajas. No obstante, el modelo estimado permite reconocer que existe una segmentación importante al interior de ese mercado y es consistente con el comportamiento de un bien normal en la medida en que los vehículos considerados responden de forma inversa a la dinámica de precios. Asimismo, señala que si bien hay una amplia disponibilidad de los vehículos tanto con caja automática como con caja mecánica, se tiene una relativa preferencia por los vehículos con caja mecánica, así como por los vehículos *coupé* y sedán, en términos relativos a los vehículos con forma *hatchback*.

De otra parte, se observa que variables como el número de pasajeros, el tipo de combustible y el aire acondicionado no resultan ser determinantes en la decisión del consumidor, lo cual se puede explicar, principalmente, por la predominancia en el mercado de vehículos de 5 pasajeros, y con gasolina.

## **CONCLUSIONES Y CONSIDERACIONES FINALES**

El presente documento describió de forma detallada la estimación del modelo *logit* y el modelo *logit anidado* mediante la metodología de variables instrumentales, siguiendo la propuesta de Berry (1994), para analizar la demanda nacional de vehículos nuevos realizada por los hogares colombianos, utilizando una muestra de 715 vehículos (automóviles) para el periodo 2009-2011.

Este modelo permite analizar los principales determinantes de la demanda de los hogares por vehículos nuevos, tomando como variables explicativas las características o atributos de los vehículos objeto de estudio. A partir del análisis se identificó que el mercado automotor en Colombia puede considerarse como un mercado elástico, lo que puede derivarse de la existencia de una variedad importante de vehículos, puesto que el modelo estimado reconoce una dinámica de sustituibilidad importante.

De igual forma, se encontró que para el mercado analizado existe una percepción de segmentos diferenciados por parte del consumidor, puesto que el parámetro de segmentación empleado en el modelo *logit anidado* resulta significativo en la estimación. En particular, el parámetro resultante es de 0,665 ( $P - value < 0,05$ ), lo cual es indicativo de un nivel de sustituibilidad importante al interior de cada segmento, que es consistente con la estructura de un mercado claramente diferenciado. Lo anterior indica que en el mercado analizado existen patrones de sustituibilidad complejos, pues los productos se diferencian entre segmentos y al interior de los mismos, los productos incluso al interior del nido no operan como bienes sustitutos perfectos.



Asimismo, la estimación reporta los signos esperados con relación al precio de los vehículos, y se identifican características diferenciadoras como el tipo de caja de cambios, la gama y la forma del vehículo, mientras que otras variables como el aire acondicionado, el número de pasajeros, el tipo de combustible o el carácter de importado del vehículo no resultan significativas.

Se encontró un resultado particularmente llamativo en torno a las altas elasticidades en los segmentos más sofisticados, lo cual podría resultar a simple vista contra-intuitivo. Sin embargo, este comportamiento puede explicarse por i) el carácter de bien de lujo de los vehículos sofisticados o Premium, ii) la posibilidad del consumidor en los segmentos más bajos para desviarse a otro segmento resulta más restringida, y iii) la definición de la variable dependiente como la probabilidad de compra del producto  $j$  versus la posibilidad de no compra ("*outside good*").

Ackerberg & Rysman (2005), también han encontrado este tipo de resultados en el modelo *logit* y *logit anidado*, y en general los modelos de elección discreta, señalando que estos modelos al incluir la "diferenciación simétrica no observada de producto" (SUPD en inglés), podrían subestimar las elasticidades de los mercados con mayor número de productos por un posible efecto congestión, así que podría generarse alguna distorsión sobre las elasticidades de los nidos con mayor variedad de productos. De esa manera, los autores consideran que el método de estimación *logit anidado* estándar, podría sobre-estimar en algunos casos la elasticidad propia de la demanda y que en presencia del efecto congestión, los métodos estándar de estimación son más propensos a sobre-estimar las elasticidades precio cruzadas al interior de cada grupo y subestimar las elasticidades entre los grupos.

En línea con lo anterior, Huang & Rojas (2013) atribuyen el posible sesgo de las elasticidades a la especificación del "*outside good*", que es determinante para la definición del mercado potencial y que podría conllevar elasticidades precio-cruzadas artificialmente bajas.

Ante estas limitaciones, resulta importante abordar y tratar estos posibles sesgos de elasticidades en futuras investigaciones. Específicamente, podrían trabajarse sobre medidas correctivas que permitan identificar de forma adecuada el mercado potencial y con esto mitigar desviaciones y sensibilidad de los resultados.

## Bibliografía

- Akerberg, D., & Rysman, M. (2005). Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects. *RAND Journal of Economics*, 36(4), 771-788.
- Anderson, S., De Palma, A., & Thisse, J. (1988). A Representative Consumer Theory of the Logit Model. *International Economic Review*, 29(3), 461-466.
- Berry, S. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, 25(2), 242-262.
- Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63, 4, 841-890.
- Bliemer, M., & Rose, J. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B*, 44, 720-734.
- Bresnahan, T., Stern, S., & Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: personal computers in the late 1980. *The RAND Journal of Economics*, 28, s17-s44.
- Cecchini, L. (2001). A logit analysis of the car market. *Applied economics letters*, 8, 199-201.
- Chagüi, Z. (2011). Competitividad por segmentos del sector automotor en Colombia. *Disertación tesis Colegio de Estudios Superiores de Administración*.
- Chattopadhyay, S. (2000). The effectiveness of McFadden's nested logit model in valuing amenity improvement. *Regional Science and Urban Economics*, 30, 23-43.
- Daly, A. (1987). Estimating "tree" logit models. *Transportation Research Part B: Methodological*, 21, 4, 251-267.
- Danaher, P., & Dagger, T. (2012). Using a nested logit model to forecast television ratings. *International Journal of Forecasting*, 28(3), 607-622.
- Hensher, D., Beck, M., & Rose, J. (2011). Accounting for Preference and Scale Heterogeneity. *Environ Resource Econ*, 49, 1-22.
- Huang, D., & Rojas, C. (2013). The Outside Good Bias in Logit Models of Demand with Aggregate Data. *Economics Bulletin*, 33(1), 198-206.
- FASECOLDA (2012). Sistema de Consulta Fasecolda-Colserauto. Consultada en <http://fasecolda.colserauto.com/fasecolda.web/publico/>
- Koppelman, F., & Wen, C. (1998). Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B*, 32(5), 289-298.
- Koppelman, F., & Wen, C. (2000). The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological*, 34(2), 75-89.
- McCarthy, P., & Tay, R. (1998). New vehicle consumption and fuel efficiency: a nested logit approach. *Transportation Research (Logistics and Transpn Rev.)*, 34(1), 39-51.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *University of California*, 105-142.
- McFadden, D. (1978). *Modelling the choice of residential location*. Amsterdam, Holland.
- McFadden, D. (1980). Econometric Models for Probabilistic Choice Among Products. *The Journal of Business*, 53, 3, S13-S29.

- Nevo, A. (2000). A Practitioner's Guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4), 513-548.
- Rasciute, E., & Pentencost, E. (2010). A Nested logit approach to modelling the location of foreign direct investment in the Central and Eastern European Countries. *Economic Modelling*, 27(1), 32-39.
- Rashidi, T., Auld, J., & Mohammadian, A. (2012). A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A: Policy and Practice*, 46(7), 1097-1107.
- Ravibabu, M. (2013). A nested logit model of mode choice for inland movement of export shipments: A case study of containerised export cargo from India. *Research in Transportation Economics*, 38(1), 91-100.
- Requena-Silvente, F., & Walker, J. (2007). The impact of exchange rate fluctuations of profit margins: the UK car market, 1971-2002. *Journal of Applied Economics*, X(1), 213-235.
- Restrepo, M. (2010). Los efectos de la política comercial: el caso del G-3 y la cuota de importación en el mercado automotriz colombiano. *Vniversitas*, 121, 233-262.
- Siriwardena, S., Hunt, G., Teisl, M., & Noblet, C. (2012). Effective environmental marketing of green cars: A nested-logit approach. *Effective environmental marketing of green cars: A nested-logit approach*, 17(3), 237-242.
- Thomassen, O. (2010). A generalised nested-logit model of the demand for automobile variants. *Center for Economic Studies - Discussion papers ces10.17*.
- Verboven, F. (2002). Quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars. *RAND Journal of Economics*, 33(2), 275-323.
- Wen, C., & Koppelman, F. (2001). The generalized nested logit model. *Transportation Research Part B*, 35, 627-641.
- Yang, L., Zheng, G., & Zhu, X. (2013, April ). Cross-nested logit model for the joint choice of residential location, travel mode, and departure time. *Habitat International*, 38, 157-166.

ANEXO 1. Segmentación mercado automotor colombiano

Segmentos		Sub-Segmento	Descripción
A	Automóviles	I1	Vehículos compactos de entrada de gama
		I2	Vehículos pequeños deportivos motor 1,6L
		M0	Vehículos pequeños sedanes entre 1,4 y 1,6 L
		M1BM	Vehículos medianos entre 1,6L y 1,8L
		M1A	Vehículos medianos entre 1,8 y 2,0L
		M2	Vehículos medianos grandes
		S	Vehículos sofisticados, gama Premium
SUV	Sport Utility Vehicle	SUV	Sport Utility Vehicle
P	Pick ups	PU1	< 1 Tonelada "Carga Util"
		PU2	= 1 Tonelada "Carga Util"
		PU3	> 1 Tonelada "Carga Util"

Fuente: Chagüi (2011).

Lo invitamos a visitar el micrositio del Grupo de Estudios de Estudios Económicos



La colección completa de la serie de documentos de trabajo se encuentra disponible en

