



Combinación del método CART y de econometría de corte transversal para la identificación de determinantes: supervivencia de exportaciones argentinas

GONZÁLEZ, GERMÁN

Instituto de Investigaciones Económicas y Sociales del Sur y Departamento de Economía
Universidad Nacional del Sur (Argentina)
Correo electrónico: ghgonza@uns.edu.ar

DELBIANCO, FERNANDO

Instituto de Matemáticas de Bahía Blanca y Departamento de Economía
Universidad Nacional del Sur (Argentina)
Correo electrónico: fernando.delbianco@uns.edu.ar

MICHALCZEWSKY, KATHIA

Instituto para la Integración de América Latina y Departamento de Economía
Universidad Nacional del Sur (Argentina)
Correo electrónico: kmichalczewsky@gmail.com

RAMIREZ, MICAELA

Departamento de Economía, Universidad Nacional del Sur (Argentina)
Correo electrónico: micaela.ramirez@uns.edu.ar

RESUMEN

Este trabajo identifica factores explicativos de la supervivencia de flujos de comercio argentinos utilizando como método de identificación la combinación del método de Análisis de Árbol de Clasificación y Regresión (CART, con sus siglas en inglés) y econometría de corte transversal. Mientras que el primero determina los mejores predictores, el segundo ofrece una función global que vincula a la variable dependiente con aquellos. La combinación de estos métodos permite reducir el número de variables potencialmente explicativas en problemas en los que no existe consenso teórico, o salvar situaciones en las que existen relaciones no lineales, datos faltantes u observaciones anómalas. La aplicación al caso de estudio permite observar que existe una diferencia significativa entre flujos de exportaciones preexistentes y nuevos flujos, a pesar de que las variables explicativas de su duración resulten ser las mismas.

Palabras clave: flujos de comercio, exportaciones, supervivencia, análisis de árbol de clasificación y decisión, econometría de corte transversal.

Clasificación JEL: C41; F10.

MSC2010: 62H30; 62J05.

Artículo recibido el 9 de mayo de 2018 y aceptado el 5 de marzo de 2019

Combination of the CART method and cross-section econometrics for the identification of determinants: Survival of Argentine exports

ABSTRACT

This paper identifies factors that explain the survival of Argentine trade flows using the combination of the Classification And Regression Tree analysis (CART) method and cross-section econometrics as the identification method. While the first determines the best predictors, the second offers a global function that links the dependent variable with those. The combination of these methods allows to reduce the number of potentially explanatory variables in problems in which there is no theoretical consensus, or to resolve situations in which there are non-linear relationships, missing data or anomalous observations. The application to the case study allows us to observe that there is a significant difference between pre-existing export flows and new flows, even though the explanatory variables of their duration turn out to be the same.

Keywords: trade flows, exports, survival, CART, cross section econometrics.

JEL classification: C41; F10.

MSC2010: 62H30; 62J05.



1. Introducción.

Sucesivos trabajos científicos desde comienzos de los años 2000 han fortalecido la hipótesis de que el crecimiento de las exportaciones en las economías en desarrollo ha estado impulsado, predominantemente, por el margen intensivo más que el margen extensivo -esto es, el crecimiento de los flujos existentes más que a partir de nuevos flujos de comercio (Besedes & Prusa, 2007; Brenton, Saborowski & von Uexkull, 2009). De todas formas, este fenómeno no sería exclusivo de las economías en vías de desarrollo, sino un fenómeno de carácter general.

En cambio, lo notable es que las economías en desarrollo experimentan una muy baja tasa de supervivencia en los nuevos productos exportados y esto contrarresta sus intentos por mejorar sus resultados comerciales. Argentina no es la excepción. Claramente, el estudio de este fenómeno es sumamente importante desde el punto de vista de la política práctica. No solo requiere atención la consideración de los factores conducentes a generar nuevos flujos de exportación, sino también de aquellos que proporcionen sustentabilidad a esos flujos; y, por consiguiente, garanticen el crecimiento del flujo de exportaciones agregadas, uno de los principales objetivos de estas economías.

Por lo general, se define la tasa de riesgo como la probabilidad de que finalice el flujo comercial en el momento t dado que no ocurrió antes y se estima su complemento, es decir, la tasa de supervivencia, mediante el estimador de Kaplan-Meier, o se regresan modelos de duración del tipo Cox, Weibull, o Prentice-Gloeckler.

A diferencia de estos ejercicios, se propone aquí un abordaje que parte de una medida alternativa de supervivencia y utiliza el método de análisis de árbol de clasificación y regresión (CART en sus siglas en inglés) para identificar los mejores predictores de la supervivencia de los flujos de exportaciones, combinado con econometría de corte transversal para fortalecer las conclusiones. Los justificativos de esta propuesta metodológica se presentan en la siguiente sección. En la sección 3 se muestran los datos, fuentes y variables. La sección 4 expone los resultados y en la sección 5 se proponen las consideraciones finales.

2. Antecedentes.

La propuesta metodológica que se presenta es la combinación de dos técnicas de análisis estadístico, una de ellas utilizadas desde hace tiempo en las ciencias médicas, aunque recientemente incorporada a la ciencia económica, y otra con el camino inverso.

El análisis de árbol de clasificación y regresión (CART por sus siglas en inglés) es un análisis de partición recursiva, capaz de manejar eficientemente relaciones no lineales, datos faltantes y robustos ante observaciones anómalas. El método involucra la discriminación de diferentes valores de variables de clasificación utilizando un árbol de decisión compuesto de divisiones binarias (nodos) que se van agregando progresivamente. Todos los valores de cada predictor son considerados un potencial divisor, siendo elegido el divisor óptimo a partir del criterio de impureza. Necesariamente, la división de un nodo debe aumentar la homogeneidad de los subconjuntos resultantes.

González-Fernández y González-Velasco (2016) son un claro ejemplo de la utilidad de aplicar CART en la ciencia económica. Este artículo ofrece un estudio sobre los determinantes de que algunos países paguen más por su deuda de largo plazo. Muchas de las conclusiones a las que los autores llegan hubieran sido difíciles de alcanzar con métodos tradicionales de regresión, por las no linealidades no paramétricas que se desprenden del análisis de árbol de clasificación.

Otros artículos que utilizan CART en las ciencias económicas, ya sea como metodología principal o suplementaria al análisis econométrico, son Gepp y Kumar (2015), Sohn y Moon (2004), y Tso y Yau (2007). El primero de ellos utiliza CART para analizar la posibilidad de

crisis financieras, comparando los resultados que arroja con los que arroja un análisis de supervivencia de Cox, y encuentran que los resultados son similares. Sohn y Moon utilizan árboles de clasificación para identificar los determinantes que hacen que la inversión en investigación y desarrollo (I+D) de las empresas logre ser efectivamente comercializada por ellas. Con el resultado de la clasificación obtenida, los autores proveen un enfoque para decidir qué variables ambientales deberían jugar un rol (y cómo) en la decisión de financiación de futuros proyectos de I+D. Finalmente, Tso y Yau (2007) incorporan CART junto con redes neuronales en un contexto de predicción de consumo de energía, para concluir que dichas metodologías tienen propiedades deseables a la hora de predecir, y deberían ser consideradas para complementar el análisis clásico de regresión paramétrica.

La desventaja que presenta el método CART es que carece de una función global y, por consiguiente, requiere de un gran número de datos para asegurarse que las relaciones detectadas sean significativas. Blanco, Irima y Vázquez (2016) enfrentan este problema combinando el método CART con estimaciones paramétricas. En particular, pre-seleccionan variables mediante el análisis del área bajo la curva (ROC) luego de hacer una regresión logística, y, con el subconjunto de variables seleccionadas, pasan a trabajar con el algoritmo de clasificación. Sin embargo, los autores pierden de vista que variables eliminadas en la etapa de preselección, podrían haber sido útiles en la clasificación de alguna manera no paramétrica (i.e. existen no linealidades difíciles de captar por el *logit*)

A diferencia de aquel trabajo, aquí se propone combinar el análisis de árbol de clasificación con una serie de estimaciones de mínimos cuadrados óptimos (MCO) que ofrezca dicha función global sin reasignar atributos. La naturaleza del problema estudiado y la definición de la variable dependiente, permiten alcanzar grados de libertad más que suficientes como para elegir los nodos y las divisiones del árbol de clasificación. Dicho de otro modo, si las observaciones son suficientes, es preferible que el algoritmo sea capaz de utilizar todos los atributos presentes en los datos. La complementación con MCO rinde, eficientemente, como análisis de robustez.

3. Fuentes de datos y variables.

La variable primaria ha sido el flujo anual de exportaciones argentinas de bienes desagregados a seis dígitos del Sistema Armonizado en su versión 1992, por socio y para el período 2000-2015. La base incluye la totalidad de flujos de exportaciones de cada año para ese período, que abarca un total de 4959 productos destinados a 201 países, alcanzando un total de 123.589 flujos. Si se considera como referencia el total de productos y el total de destinos, la base de datos contiene el 12.4% de los flujos de exportaciones potenciales argentinas.

Con el objeto de reducir una dimensión, se definió una variable secundaria, exportaciones por producto-destino para cada año de la ventana temporal. Finalmente, el estudio de supervivencia se realizó sobre la variable *maxduration*, definida como la cantidad de años en que la variable exportaciones desagregada por producto-destino fue positiva en forma continua. Es decir que, si el valor x_{ijt} representa el volumen de exportaciones del producto i al destino j en el año t , y si x_{ijt} fue positiva durante 5 años y luego el flujo se detuvo, *maxduration* toma valor 5 y representa la cantidad de años que sobrevivió ese flujo de exportaciones. Considerando que un flujo puede reiniciarse varias veces durante un período suficientemente largo de tiempo, se ha tomado por defecto el período de máxima duración.

La base de datos se construyó a partir de varias fuentes estadísticas. En cuanto a los datos de exportaciones argentinas, éstos fueron obtenidos a partir de la Base pour l'Analyse du Commerce International del Centre d'Études Prospectives et d'Informations Internationales (CEPII) por ser una base con alto nivel de desagregación y armonizada en términos de clasificación de productos. Por lo tanto, permite hacer comparaciones de las canastas entre períodos de tiempo alejados.

Para identificar los predictores de la cantidad de años que se espera sobreviva un flujo de exportaciones de un producto determinado a un destino identificado, se consideraron una batería de variables. Partiendo de la premisa de que el comercio bilateral se ve afectado tanto por factores que facilitan, como de otros que lo entorpecen, es plausible que estos mismos factores tengan alguna responsabilidad en la permanencia de los flujos en el tiempo. Es así como fueron consideradas potencialmente explicativas las variables típicas de los modelos gravitacionales del comercio: el tamaño del mercado abastecido fue aproximado por el producto bruto doméstico (PPPDP), mientras que se utilizaron varias medidas alternativas de distancia (*dist*). A esta lista se agregaron variables dicotómicas para la existencia de lengua oficial común (*comlang_off*) y para la existencia de límites geográficos compartidos (*contig*).

Nitsch (2007) encuentra que el riesgo de que un flujo de exportaciones se detenga aumenta con la distancia y disminuye con el tamaño del mercado abastecido. Mientras que la existencia de límites comunes, lengua común e, incluso, lazos culturales o coloniales pueden resultar facilitadores del comercio cuando los productos son diferenciados¹ dado que reducen los costos de búsqueda (Rauch, 1999).

Tras el trabajo de Rauch (1996), otros estudios suministraron evidencia a favor de las diferencias en el comercio de bienes homogéneos frente a los diferenciados². En particular, Besedes y Prusa (2006a) encontraron que el tamaño de la transacción inicial (*maxvinitial*) tiene un efecto positivo en la tasa de supervivencia de los bienes diferenciados, y mayor es este efecto cuanto mayor es la duración del flujo, lo cual hace que el efecto sea persistente.

El hecho de que el flujo sea nuevo también puede implicar cierta fragilidad debido a que el exportador no cuenta con información perfecta sobre su productividad y capacidad para competir, ni cuenta con conocimiento pleno del mercado abastecido, hasta que comienza a exportar. Para capturar este fenómeno se construyó una variable dicotómica (*nuevo*), que toma valor 1 si las exportaciones en el momento *t* son positivas y tomaron valor nulo en *t-1*.

Además, los flujos comerciales de bienes homogéneos son más frágiles debido a las mayores posibilidades de sustitución que presentan estos productos, y este efecto interactúa con el tipo de mercado abastecido. Se podría esperar que, en economías más avanzadas, la competencia, las condiciones de acceso y la elasticidad de sustitución de la demanda fuesen mayores, lo que elevaría el estándar de las firmas exportadoras y también las posibilidades de entablar rápidamente nuevas relaciones comerciales que sostengan el flujo. Para considerar este efecto se incorpora una variable dicotómica (*incomegroup1*) que toma valor 1 cuando la economía importadora pertenece al estrato de ingreso per cápita alto, de acuerdo a la clasificación del Banco Mundial.

Para considerar posibles efectos asimétricos, en este trabajo se han realizado reagrupamientos alternativos de los bienes exportados: siguiendo la clasificación utilizada por Giordano (2016), por rubro (manufacturas de origen agropecuario, manufacturas de origen industrial, manufacturas de origen mineral, productos primarios agropecuarios, productos primarios minerales, combustibles y energía); y, siguiendo la clasificación de Lall (2000), por contenido tecnológico (productos primarios, manufacturas basadas en recursos naturales, manufacturas de bajo, de medio y de alto contenido tecnológico).

¹ Se entiende por productos homogéneos tanto a los bienes que son comercializados en mercados organizados, como aquellos que poseen precios de referencia establecidos por entidades especializadas. Mientras que los bienes diferenciados no tienen un precio de referencia por la multiplicidad de características y variedades, por lo que no es posible el arbitraje y los compradores y vendedores se involucran en un proceso de búsqueda que finaliza cuando se alcanza un precio de reserva (Rauch, 1996).

² Besedes y Prusa (2006b) enumeran los principales hechos estilizados.

Las características del mercado fueron consideradas a partir del nivel de ingreso per cápita (alto, medio y bajo) según el Banco Mundial, y la región (América Latina y El Caribe; América del Norte; África del Norte y Este Medio; África Subsahariana; Europa y Asia Central; Asia del Este y Pacífico; Asia del Sur). También se ha considerado la posibilidad de diferencias en la calidad de las instituciones mediante la variable IPD2016, calculada como el promedio de los ítems B, C y D del Institutional Profiles Database (CEPII).

Finalmente, la escala del flujo de exportaciones de un producto a un mercado específico puede ser un determinante significativo; por ello, se incluye la variable share calculada como el cociente entre el tamaño del flujo específico y la suma de todos los flujos al mismo destino en el momento t . Desde el punto de vista de los costos (logísticos, financieros, de información) puede resultar ventajoso que un mismo destino cumpla el rol de multimercado, siempre que existan economías de escala y vínculos intersectoriales que permitan alcanzarlas. Sin embargo, la permanencia de un share significativo responde a la preexistencia de un capital intangible intransferible, llámese historial de acuerdos comunes exitosos, red de contactos de confianza, etc. que elimina costos de búsqueda, facilita las negociaciones, y redundaría en una mayor supervivencia del flujo.

Tanto las condiciones macroeconómicas como los cambios en las políticas podrían afectar la supervivencia de los flujos. Este efecto puede ser controlado mediante un efecto fijo en las regresiones de datos de panel si la base de datos es balanceada. En este caso, el panel construido no cumple con esa característica debido a la definición de la variable de interés. De todas formas, las variables omitidas no representan un problema para el método de identificación utilizado.

4. Resultados.

4.1. Descripción estadística.

La Tabla 1 describe estadísticamente a las variables seleccionadas. La base cuenta con 804563 observaciones, cada una de ellas corresponde a la duración del flujo de exportaciones de un bien identificado a un destino determinado. El 25% de las observaciones se encuentra censurado a la izquierda, es decir, se observa comercio desde el inicio del período considerado, y el 26% está censurado a la derecha, es decir, el flujo alcanza el extremo mayor del período. En ambos casos, el corte podría introducir error debido a que, en el primer caso, puede suponerse un nacimiento cuando el flujo ya existía desde antes del año 2000; mientras que, en el segundo caso, puede suponerse una muerte cuando en realidad el flujo no terminó en 2015. Por tal motivo, se han tomado las definiciones para minimizar la probabilidad de error.

La duración promedio de los flujos de comercio medido en términos de producto-destino y para el período considerado (16 años entre 2000 y 2015) supera los 7 años. Comparado con otros estudios, la media resulta significativamente alta³. Esto se debe a que, por cuestiones metodológicas, la base de datos no incluye los valores nulos y, por definición, la variable toma el período máximo en aquellos casos en que existe intermitencia en los flujos durante los 16 años. Esta situación ocurre en el 31% (38539 flujos) de los casos, siendo 1.49 años la media de las observaciones de los flujos descartados.

Es preciso tener en cuenta, también, que el valor promedio de duración de los flujos disimula la dinámica interna, que podría caracterizarse en algún producto particular por altas tasas

³ González (2015) señala que la duración media de los flujos de exportaciones argentinas -al mismo nivel de desagregación que el presente trabajo- para el período 1993-2012 es 3 años, Besedes y Prusa (2007) y Besedes y Blyde (2010) encuentran valores menores a los 4 años para países con diferentes niveles de desarrollo y perfiles productivos. Del mismo modo se presenta en Fugazza y Molina (2011), con un promedio de 10 años de duración para países desarrollados y de 2 para países en desarrollo.

de entrada y salida de firmas exportadoras; y también hay que considerar que el desvío estándar de la duración media se encuentra cerca de los 6 años, un valor que deja en evidencia la gran disparidad existente al nivel de producto-destino.

Tabla 1. Estadística descriptiva.

Variable	Obs	Media	Desvío estándar	Min	Max
Maxduration	804563	7.154509	5.568756	1	16
Maxviniacial	804563	564.2249	10333.99	1	1157631
Share	792505	.0807566	.1869666	0	1
Dist	798180	7701.01	5076.709	215.0746	19447.35
Contig	798180	.2429527	.428867	0	1
comlang_off	798180	.45826	.498255	0	1
IPD2016Prom	778939	2.299403	.2636286	.8873239	2.842105
incomegroup1	799854	.4618755	.4985447	0	1
PPPGBP	787766	1462.749	3164.803	.037	19695.74

Fuente: Elaboración propia (ver Anexo).

Es interesante ver en la Tabla 2 que tanto la duración media como el valor inicial medio se reduce con el nivel de ingreso del mercado de destino: desde 7 hasta un valor cercano a 4 años, para la primera variable, mientras que 586.58 a 247.63 en la segunda.

También muestra una gran disparidad cuando las observaciones se reagrupan por regiones geográficas: por encima de los 8 años para los flujos destinados a América Latina y El Caribe, y un valor similar para América del Norte, mientras que se reduce a 4/5 años para las restantes regiones. Los valores iniciales medios también varían sustancialmente según la región abastecida, pero no se observa un patrón que sintetice la relación entre ambas variables.

En cuanto a los reagrupamientos según rubro y según contenido tecnológico, las Tablas 3 y 4 muestran las duraciones medias más altas en productos primarios, principalmente agropecuarios, y manufacturas basadas en recursos naturales y de tecnología media, los cuales concuerdan con aquellos donde Argentina presenta ventajas comparativas⁴ (Corbella, 2016). Sin embargo, los valores rondan los 7 años en todos los casos y las diferencias no son significativas.

Tabla 2. Duración y valor inicial por grupos de mercados destino.

Variable	Obs	Media	Desvío estándar	Min	Max
Países de ingresos altos					
maxduration	369433	7.339718	5.687787	1	16
maxviniacial	369433	586.5832	10351.3	1	1157631
Países de ingresos bajos					
maxduration	15706	3.76805	3.758535	1	16
maxviniacial	15706	247.6323	1483.407	1	53581.02
América Latina y El Caribe					
maxduration	393147	8.646758	5.693496	1	16
maxviniacial	393147	542.6462	12419.39	1	1157631
América del Norte					
maxduration	44921	8.342913	5.733946	1	16
maxviniacial	44921	1212.218	14897.66	1	558359.7

⁴ En Nicita, Shirotori y Tumurchudur (2013) se estudia si la distancia entre la canasta de bienes exportados y aquellos en los que presentan ventaja comparativa afecta la duración de las exportaciones. Los autores encuentran que solo después de 10 años de duración de la relación comercial, el efecto es similar entre países menos adelantados y países del MERCOSUR. Sin embargo, en los primeros años, ese efecto es mayor en el primer grupo respecto del segundo.

Variable	Obs	Media	Desvío estándar	Min	Max
Africa del Norte					
maxduration	47183	4.379798	4.260487	1	16
maxviniacial	47183	595.289	5208.854	1	195222.3
Asia del Este y Pacífico					
maxduration	70224	5.302162	4.78129	1	16
maxviniacial	70224	590.0801	8906.881	1	531285.4
Asia del Sur					
maxduration	13076	4.94662	4.502039	1	16
maxviniacial	13076	987.8443	11136.07	1	204149.6
Europa y Asia Central					
maxduration	195913	5.96541	5.08905	1	16
maxviniacial	195913	463.541	5729.625	1	317231.8

Fuente: Elaboración propia (ver Anexo).

Tabla 3. Duración por rubro del bien exportado.

Rubro	Obs	Media
Combustibles y energía	3400	7.387941
Manufactura de origen agropecuario	103496	7.762329
Manufactura de origen industrial	558502	7.018021
Manufactura de origen minero	63914	6.677442
Producto primario agropecuario	70411	7.774666
Producto primario mineral	4840	7.020868

Fuente: Elaboración propia (ver Anexo).

Tabla 4. Duración por grupo según contenido tecnológico del bien exportado.

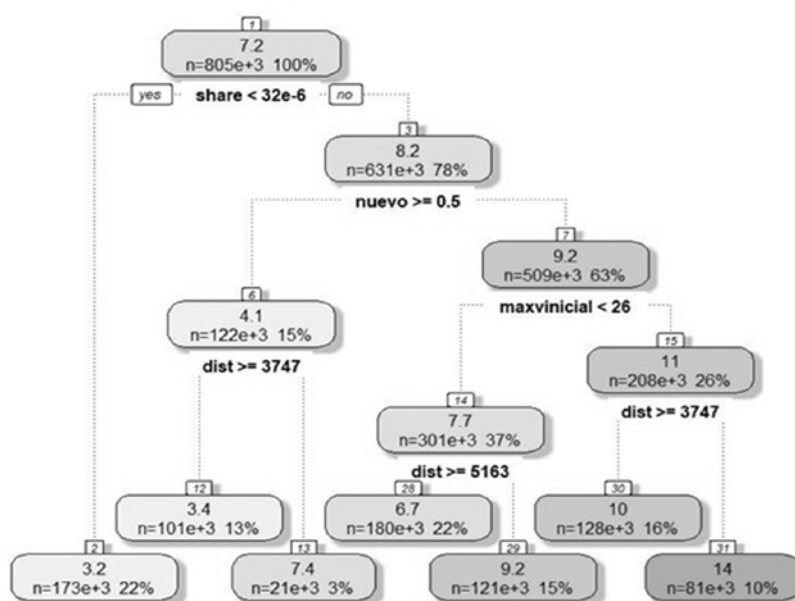
Grupo	Obs	Media
Producto primario	92120	7.756741
Manufactura basada en recursos naturales		
Agrícola/forestal	95237	7.437897
Otros	69268	7.096451
Manufacturas de baja tecnología		
Textil	90673	6.689819
Otras	120542	6.95828
Manufactura de tecnología media		
Productos automotores	17821	7.746591
Industrias de procesos de tecnología media	79568	7.59945
Ingeniería de tecnología media	142712	6.910764
Manufactura de alta tecnología		
Productos electrónicos y eléctricos	46039	6.372336
Otros	38731	7.140508
Otras transacciones	11852	8.228316

Fuente: Elaboración propia (ver Anexo).

4.2. Análisis multivariado.

La Figura 1 muestra el árbol de clasificación para la muestra completa. El número de observaciones, como se mencionó antes, está cerca de los 805000 registros, cada uno de los cuales se corresponde con la duración de flujos de exportaciones según producto-destino para el período 2000-2015. La división estadísticamente más significativa, ubicada en la base del árbol, separó a los flujos de acuerdo con share. Si bien el valor umbral para la clasificación representa un valor muy bajo comparado con su media, es claramente significativo en términos de duración del flujo: mientras que, si no lo supera, la duración media es algo más de 3 años (Grupo 2; 22%); si lo supera, la supervivencia aumenta en un año (Grupo 3; 78%).

Figura 1. Árbol de clasificación para la muestra completa



Fuente: Elaboración propia.

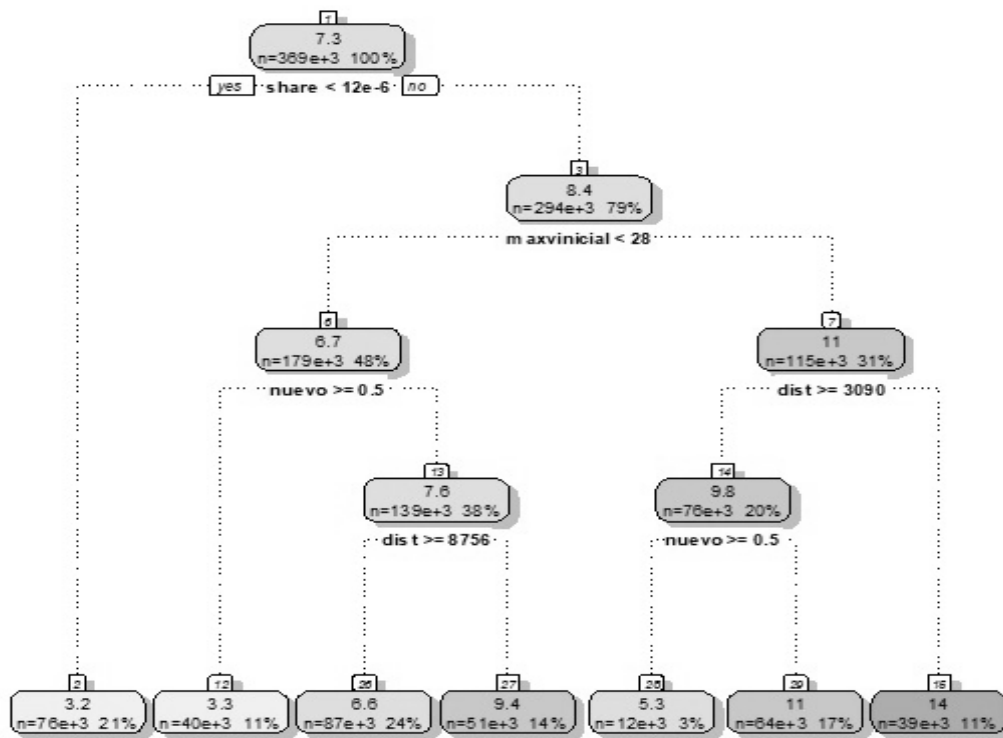
El grupo 3 es dividido según sea un flujo nuevo (Grupo 6; 15%) o no lo sea (Grupo 7; 63%). Si el flujo es nuevo y el mercado es una economía relativamente lejana, la duración esperada se reduce a 3 años (13%); mientras que, si la distancia es menor al umbral, la duración supera los 7 años (3%).

En cambio, si el flujo no es nuevo, gana relevancia el valor inicial del flujo correspondiente y se anticipa a la distancia como predictor, y las expectativas de supervivencia mejoran cuando el valor inicial supera al umbral computado por el algoritmo. Tras ello, tanto para el Grupo 6 (no nuevo, con valor inicial por debajo del umbral) como para el Grupo 7 (no nuevo, con valor inicial por encima del umbral), la duración del flujo se reduce con la distancia.

Finalmente, el mejor pronóstico de supervivencia lo tienen los flujos de exportaciones que son significativos en términos de escala (aunque con un umbral bajo), con cierta experiencia acumulada y que, preferentemente, correspondan a mercados cercanos. El 10% de las observaciones comparten estas características y alcanzan los 14 años promedio (Grupo 31), mientras que el 13% reúne las condiciones contrarias y su pronóstico se reduce a 3 años (Grupo 12).

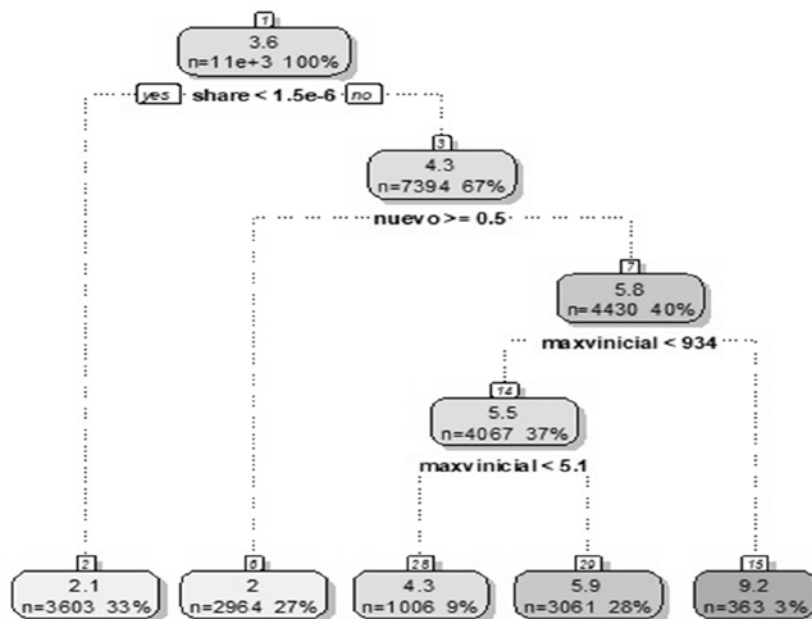
Cuando se agruparon las observaciones según las características del mercado abastecido, surgieron resultados interesantes para remarcar. Si el nivel de desarrollo es el criterio elegido, los mejores pronósticos (en términos de supervivencia) los muestra el árbol que corresponde a las exportaciones dirigidas a las economías con mayores ingresos per cápita (Figuras 2a y 2b). Tanto en su mayor como en su menor pronóstico, los años esperados resultan superiores que en los casos similares pero destinados a economías de ingresos bajos. La diferencia mayor, en términos porcentuales, se da en los flujos con menor escala, en cuyos casos el pronóstico para destinos ricos es 50% superior al pronóstico de los destinos relativamente pobres: 3.2 frente a 2.1 años. Sin embargo, la diferencia en años de supervivencia es significativa en el extremo superior: 14 años para los mercados ricos frente a 9.2 en los países relativamente pobres. La explicación parece estar vinculada con la mayor escala en los flujos, pero también podría deberse al tipo de producto comercializado y a las condiciones de acceso a esos mercados, que elevan el estándar de competitividad de las firmas exportadoras.

Figura 2a. Árbol de clasificación para destinos con ingreso per cápita alto



Fuente: Elaboración propia.

Figura 2b. Árbol de clasificación para destinos con ingreso per cápita bajo



Fuente: Elaboración propia.

Son la escala de los flujos, el valor inicial, la experiencia y la distancia, los predictores más relevantes para clasificar a los flujos de exportaciones dirigidas a países de ingreso per cápita alto. La escala inicial, maxvicial, divide al 79% de los flujos en dos grupos con medias de 7 y 11 años de supervivencia respectivamente. A partir de este punto, los flujos nuevos son los que peor

pronóstico presentan, y es aún más bajo cuanto más lejos sea el destino. De todas formas, estas características se disputan la supremacía como predictor, marcando una posible no-linealidad en la relación entre estas variables y maxduration. Esta observación se ve reforzada al notar que la distancia no es relevante como predictor para los flujos destinados a mercados incluidos en el grupo de países de bajos ingreso per cápita ni tampoco en gran parte de las regiones geográficas.

Observando los árboles por regiones geográficas de destino⁵ se identifica un patrón en el que la distancia no resulta relevante como predictor y en donde aparecen, manteniendo el mismo orden, share, nuevo y maxviniacial. Este patrón lo presentan Europa, Asia y Norte de África. Por lo que, la existencia de vínculos consolidados y la experiencia previa elevan notablemente el pronóstico de supervivencia. La carencia de alguna de ellas reduce la expectativa a 2 o 3 años, mientras que su existencia simultánea se relaciona con un mínimo de 5 años y a un máximo de 13 años como medias, dependiendo positivamente del tamaño inicial del flujo.

Los flujos dirigidos a América del Norte tienen un pronóstico similar (4 años para el peor de los casos y 13 años en el otro extremo). A diferencia de las regiones anteriores, en el segundo nivel del árbol, la escala inicial del flujo resulta mejor predictor que la experiencia. Algo similar ocurre con los mercados de América Latina y El Caribe. Considerando a las variables share, maxviniacial y nuevo (en ese orden) se alcanza una predicción de supervivencia para el 67% de los flujos con destino latinoamericano, que alcanza un máximo de 10 años para los casos en los que se cuenta con experiencia previa y no se supera un determinado umbral de escala en el primer año en que se registraron exportaciones. El mejor pronóstico para esta región es de 13 años, y corresponde a los casos en que se supera los umbrales definidos para share y maxviniacial, y los mercados son relativamente cercanos.

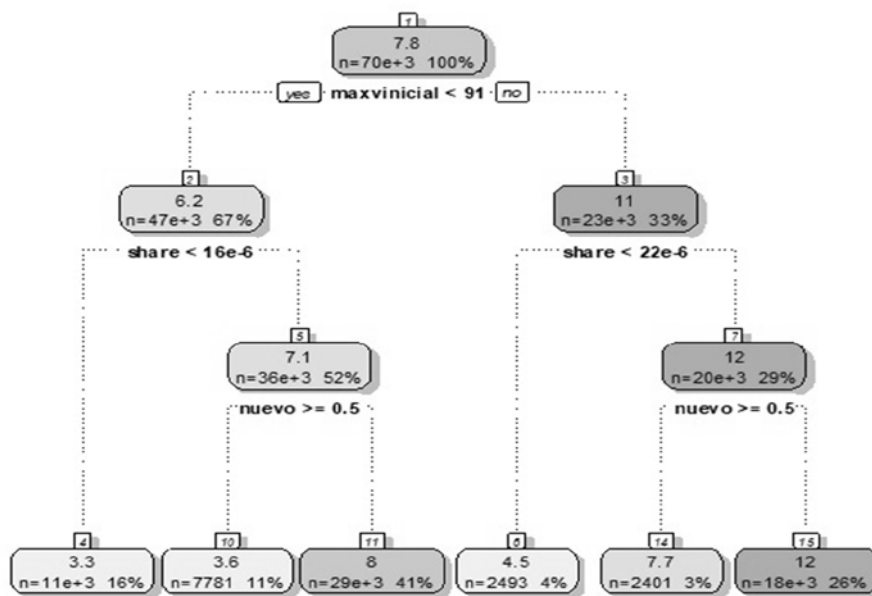
Nuevamente, la relevancia como predictor de la variable dist es discutida cuando se reagrupan los flujos de acuerdo con el tipo de producto. Para los productos primarios agropecuarios (PPA), la distancia no es relevante y el primer nodo separador corresponde a maxviniacial. En la Figura 3 se observa que este predictor identifica a un grupo (67% de los flujos) con media de 6 años, pero con un pronóstico máximo de 8 cuando se supera el umbral de share y cuando existe experiencia previa. El restante grupo identificado (33%) presenta una media de 11 años, pero alcanzaría los 12 años si el flujo no es nuevo.

Las manufacturas de origen agropecuario (MOA) y las manufacturas de origen industrial (MOI) presentan como primer separador a share. Los grupos conformados por los flujos que superan el umbral definido para esta variable contemplan el 74% y el 78%, respectivamente. Luego, las variables maxviniacial y nuevo se disputan el lugar como segundo separador, y dist aparece como último predictor relevante.

Aquellos flujos de MOA que presentan una escala significativa para el mercado abastecido, cuyo valor inicial superó el umbral de maxviniacial y no es un flujo nuevo, tienen un pronóstico de 11 o 14 años, dependiendo de si el mercado es lejano o cercano. Por su parte, los flujos de MOI que superan el umbral de share, no son flujos nuevos y su valor inicial fue suficientemente grande, presentan un pronóstico cercano a 7 en el peor de los casos (destinos muy lejanos) y de 13 en el mejor de ellos (destinos relativamente cercanos). Las figuras 4a y 4b muestran los árboles correspondientes.

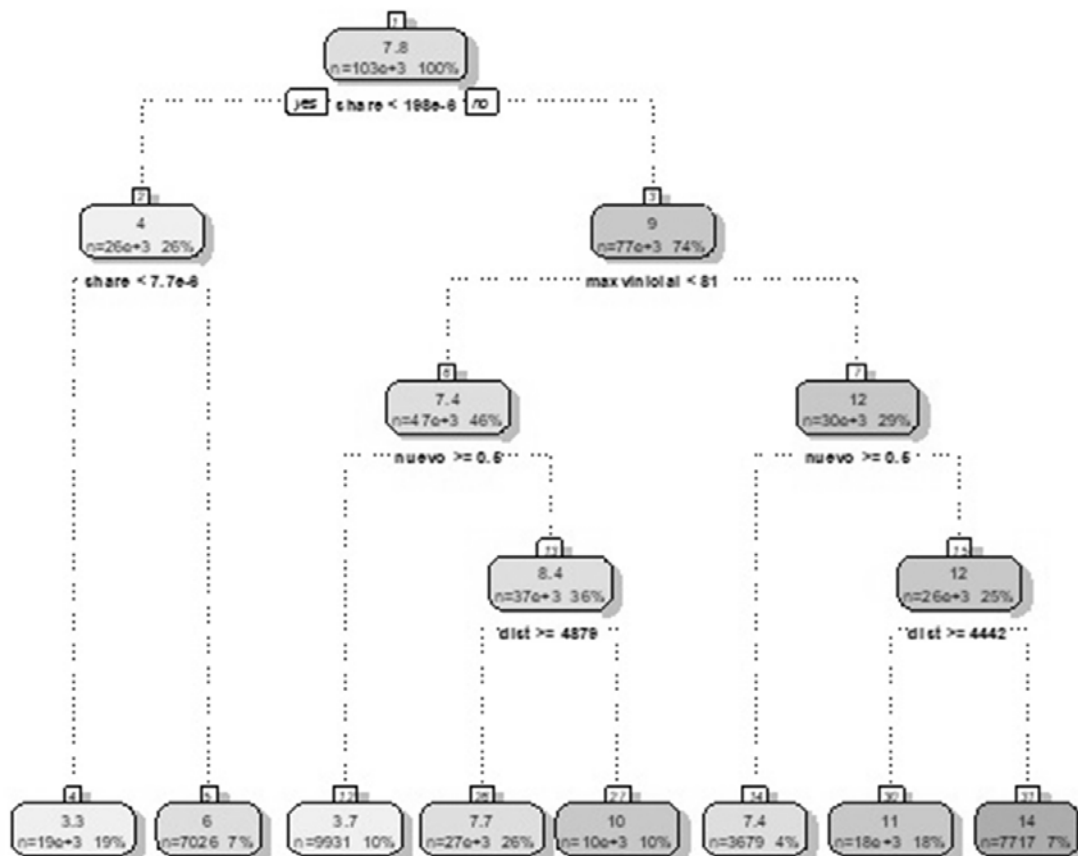
⁵ Los árboles según destino por regiones geográficas se encuentran en el Anexo.

Figura 3. Árbol de clasificación para tipo de producto: productos primarios agropecuarios.



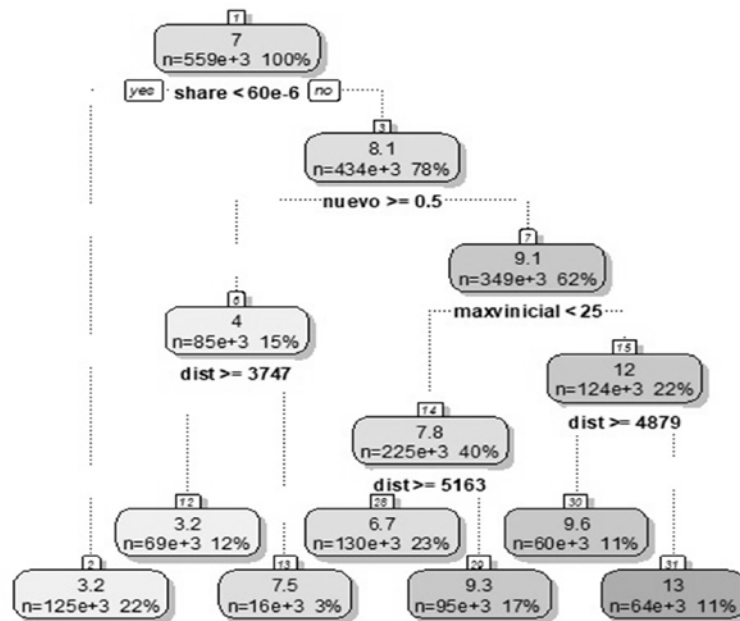
Fuente: Elaboración propia.

Figura 4a. Árbol de clasificación para tipo de producto: manufacturas de origen agropecuario.



Fuente: Elaboración propia.

Figura 4b. Árbol de clasificación para tipo de producto: manufacturas de origen industrial.



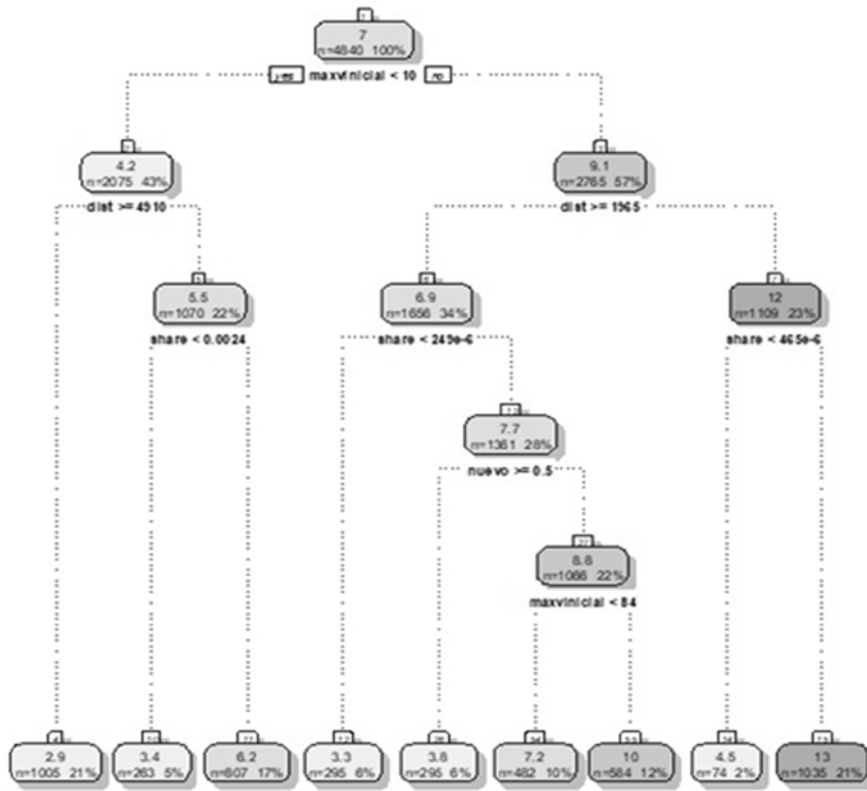
Fuente: Elaboración propia.

A diferencia de los rubros anteriores, los árboles correspondientes a flujos de exportaciones de productos primarios minerales (PPM), las manufacturas de origen minero (MOM) y los combustibles y energía (CYE) presentan a la variable dist como primer o segundo nodo separador. En los flujos de PPM, como puede observarse en la Figura 5a, y MOM, Figura 5b, el umbral de share es muy relevante a pesar de aparecer como tercer nodo. Al no superar el volumen indicado, el pronóstico de supervivencia es aún menor que para los casos en que los destinos son lejanos (es decir, máximos en las ramas izquierdas del árbol).

En los flujos de PPM es maxvicial el primer nodo separador, mientras que dist aparece en segundo lugar. Como puede observarse en la Figura 5a, las siguen las variables share y nuevo, según el caso. El 57% de los flujos superó el umbral definido en el primer nodo, mientras que la distribución por distancia fue más pareja, aunque superior para destinos relativamente lejanos. El pronóstico para los flujos de PPM que cumplieron esa primera característica mejora con la relevancia del producto en el mercado abastecido, con la experiencia y el valor inicial, alcanzando un máximo de 13 y un mínimo de 3 años.

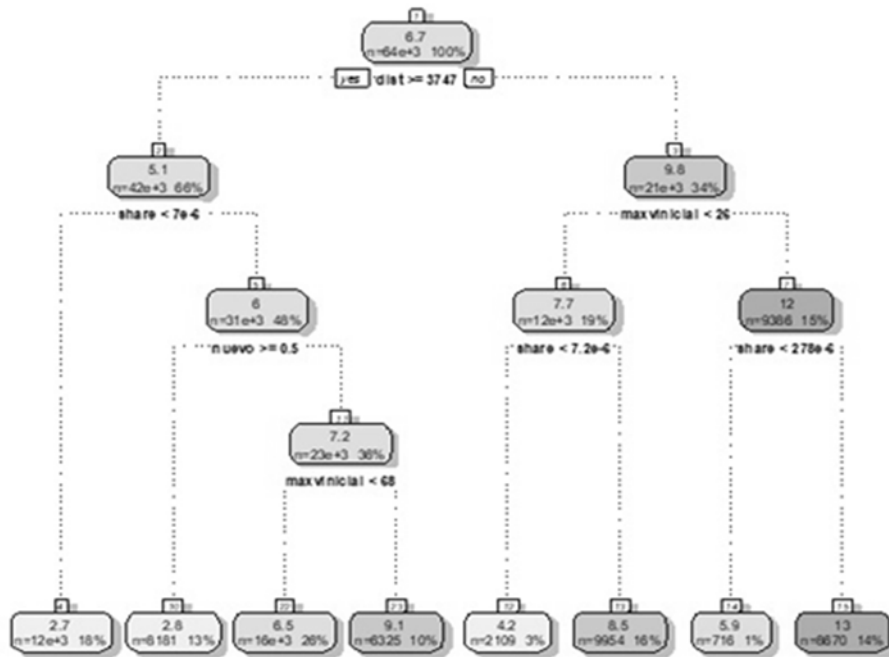
Los flujos de CYE son los de menor número. La Figura 5c muestra el árbol de clasificación. A pesar de esa primera característica, el árbol resultante presenta algunas diferencias interesantes respecto a los anteriores. En primer lugar, hay una mayor cantidad de nodos separadores, lo que puede asociarse a una mayor variabilidad interna. En segundo lugar, el nodo terminal con mejor pronóstico surge en el brazo correspondiente a los flujos enviados a mercados más distantes (16 años), aunque corresponde solo al 1% de los flujos. En tercer lugar, surge otra variable como predictor, la calidad institucional (IPD2016), aunque como último nodo separador, y con el sentido contrario al esperado, pero solo significativo para el 2% de los flujos. Finalmente, aquellos flujos que se destinan a mercados cercanos, cuyo valor inicial superó el umbral definido y corresponden a flujos comerciales significativos en términos de volumen, alcanzan los 14 años de supervivencia, valor máximo solo alcanzado por los flujos de MOA.

Figura 5a. Árbol de clasificación para tipo de producto: productos primarios de origen mineral.



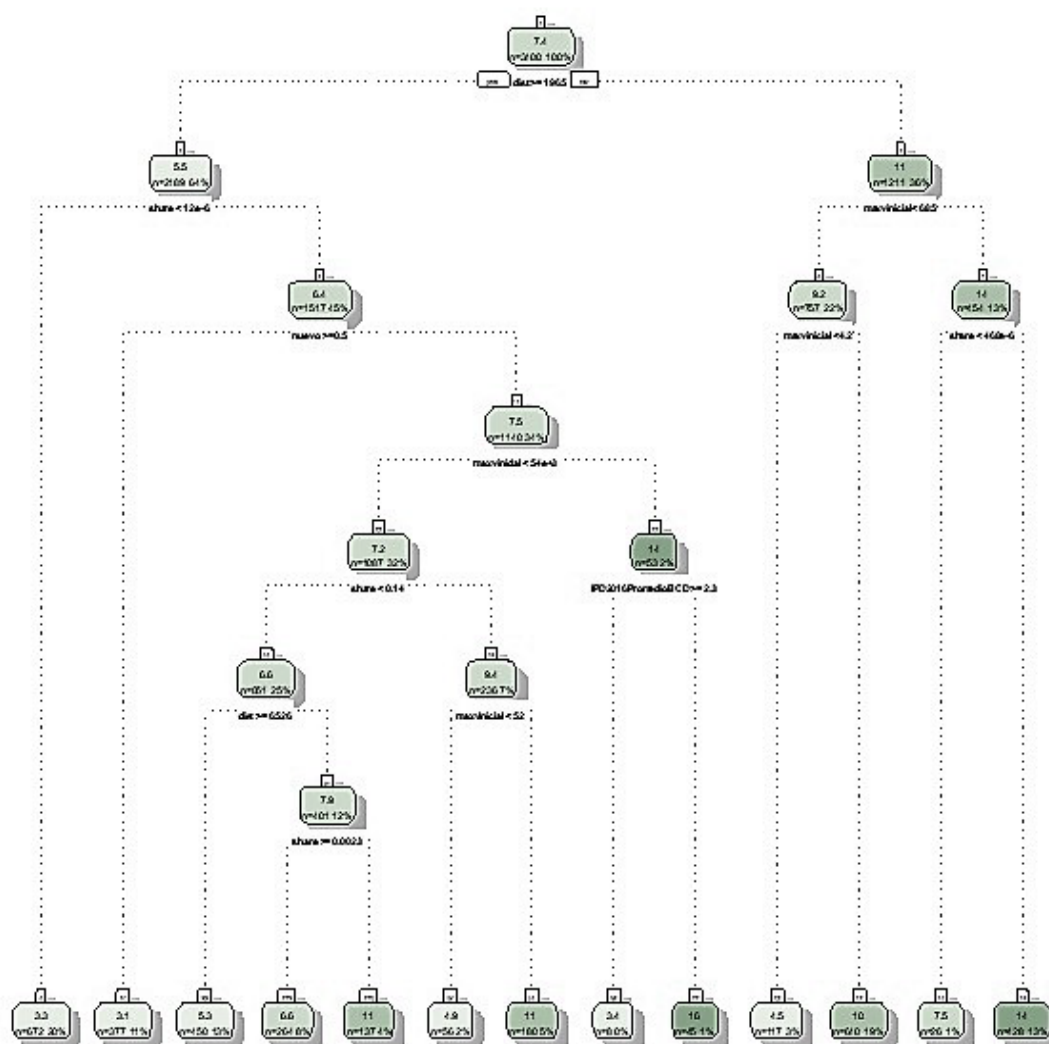
Fuente: Elaboración propia.

Figura 5b. Árbol de clasificación para tipo de producto: manufacturas de origen mineral.



Fuente: Elaboración propia.

Figura 5c. Árbol de clasificación para tipo de producto: combustible y energía.



Fuente: Elaboración propia.

Con el objetivo de convalidar la significatividad y el signo de los predictores sobre la duración de los flujos de exportaciones se realizó una serie de estimaciones por MCO de un modelo de corte transversal, suponiendo una relación lineal entre la variable dependiente y las restantes mencionadas. En líneas generales, se espera una relación positiva entre la dependiente, maxduration, y las variables maxviniat, share*, incomegroup1 e IPD2016; mientras que se espera signo negativo entre la dependiente y las variables dist. y nuevo*. Además, se agregan como variables de control a comlang_off, contig y PPPGDP.

Las variables share* y nuevo* son transformaciones de las variables share y nuevo descritas en la sección precedente. Dado que las variables originales tienen variabilidad en el tiempo y el modelo es de corte transversal, se ha optado por tomar los valores promedio de cada variable durante el período de supervivencia. Mientras que share* presenta la misma interpretación, es preciso hacer una aclaración con respecto a nuevo*.

La variable original, nuevo, es una dicotómica que toma como valor a la unidad si el flujo es positivo en el momento t cuando en el t-1 era nulo, y esta variable puede mostrarse no-nula en todo el período o en varias oportunidades durante el período considerado en el estudio, o mostrar un valor no nulo por única vez y luego desaparecer. Sumado a ello, para el promedio se toman en

consideración todos los subperíodos en que el flujo estuvo presente, por lo que los promedios no presentan un denominador común. La complejidad de la construcción de la variable nuevo* hace que dificulte su interpretación, aunque no le quita utilidad. La variable nuevo* es una variable continua con valores definidos entre 0 y 1, caracterizando al flujo de exportaciones como más inestable conforme la variable aumenta su valor y alcanza el 0.5, mientras que a partir de entonces aumenta su estabilidad, aunque reduce su supervivencia, debido posiblemente a la extraordinariedad de la relación comercial⁶.

La Tabla 5 muestra los resultados obtenidos. Las variables típicas de los modelos de gravedad muestran los signos y nivel de significatividad esperados. Cuanto mayor es la distancia menor resulta el período máximo de vida del flujo de exportaciones. En cambio, la existencia de una lengua oficial común y de vecindad, tienen el efecto contrario. El tamaño de la economía de destino, medido por PPPGDP resulta ser positivo, efecto que se ve reforzado con incomegroup1. En todos los casos, el resultado es robusto al modelo estimado.

Tabla 5. Regresiones de maxduration.

maxduration	(1)	(2)	(3)	(4)	(5)
maxviniacial	.0000292* (1.49e-06)	.0000408* (1.83e-06)	.0000291* (1.49e-06)	.0000389* (1.84e-06)	.0000403* (1.83e-06)
Share*	-.1211581** (.0584412)	-2.821541* (.070114)			-2.767507* (.0701669)
nuevo*	-13.48226* (.0564095)		-13.50487* (.0553461)		
Dist	-.000054* (3.92e-06)	-.0001254* (4.78e-06)	-.0000535* (3.91e-06)	-.0001149* (4.81e-06)	-.0001133* (4.75e-06)
contig	1.680714* (.0407069)	2.827541* (.0494296)	1.666351* (.0401135)	2.52658* (.0492021)	2.960119* (.0491448)
comlang_off	.3595107* (.3595107)	1.349918* (.0415846)	.3560142* (.034218)	1.305509* (.0418606)	1.311833* (.0416157)
IPD2016Prom	-.1475012* (.0468)	.6548864* (.0570799)	-.1537368* (.046704)	.5365859* (.0574027)	1.063163* (.0534982)
PPPGDP	.0000707* (3.74e-06)	.0001944* (4.53e06)	.0000693* (3.68e-06)	.0001665* (4.51e-06)	.0002013* (4.52e-06)
incomegroup1	.2904057* (.023804)	.5888492* (.0290676)	.2877254* (.0237692)	.5361144* (.029241)	
_cons	8.070512*	2.575238*	8.081367*	2.614916*	1.776122*
Obs	115351	115351	115351	115351	115351
R-squared	0.4332	0.1524	0.4331	0.1405	.1494

Nota: Errores estándar en paréntesis. *** p<0.01, ** p<0.05, * p<0.1

Fuente: Elaboración propia.

El resto de las variables también presentan alta significatividad. Mientras que el valor inicial del flujo resulta ser el esperado, la relevancia del flujo en el mercado de destino, share*, presenta el signo negativo. Una posible explicación de este último resultado se apoya en la distribución de los valores de dicha variable. Un conjunto de 2018 flujos de exportaciones posee, simultáneamente, share igual a 1 y maxduration con valor cercano a 1, es decir, fueron flujos de duración mínima y a destinos adonde no se exportó ningún otro producto. Estas observaciones podrían ser outliers que empujan la relación en el sentido contrario al esperado. La Tabla 6 muestra estadística descriptiva de este subconjunto de observaciones. Ni nuevo* ni incomegroup1

⁶ El valor máximo surge de considerar el caso límite de un flujo que nace y muere en el mismo año y no vuelve a nacer en los siguientes (1 nacimiento en 1 año de vida). En cambio, un flujo que nace y muere continuamente durante el período completo, es decir, toma valores 1 y 0 intercalados, alcanzará un valor igual a 0.5 (8 nacimientos en 16 años). En el extremo inferior, el flujo es positivo desde el momento inicial por lo que el valor de la variable nuevo* será 0.0625 (1 nacimiento y 16 años de vida).

parecen ser útiles para explicar la existencia de estos flujos. La eliminación de share de la regresión no cambia los signos ni la significatividad de las demás variables, por lo que no existen argumentos para retirarla.

Tabla 6. Observaciones con share = 1.

Variable	Obs	Media	Des. Est.	Min	Max
maxduration	2018	1.161546	.7094992	1	16
nuevo*	2018	.5034919	.1539845	.0625	1
Incomegroup1	2018	.5258193	.4994569	0	1

Fuente: Elaboración propia.

La variable nuevo* tiene un efecto negativo y significativo sobre la duración del flujo. Como se ha mencionado, el incremento en su valor implica inicialmente mayor inestabilidad y luego mayor extraordinariedad. En ambos casos, el efecto sobre la duración es menguante, por lo cual el resultado obtenido se condice con esta interpretación. El resultado, además, es robusto al modelo estimado, y el coeficiente de determinación de los modelos aumenta cuando se incluye nuevo* entre las explicativas, elevando así su poder predictivo.

Finalmente, un párrafo aparte merece el resultado de IPD2016prom. El signo positivo esperado, es decir, mayor pronóstico de supervivencia cuanto mejor fuese la calidad de las instituciones del mercado de destino, solo se da en aquellos modelos en los que no aparece la variable nuevo*. En los restantes, el signo es negativo, lo que resulta contraintuitivo. Lo mismo ocurre en el único árbol en el que aparece como nodo separador (esto es, en el subconjunto de flujos de exportaciones de combustible y energía). Si bien no resulta ser un separador eficiente en los restantes árboles, la falta de robustez mostrada en la regresión requeriría de un estudio minucioso.

5. Consideraciones finales.

La combinación del método CART y la econometría de corte transversal han permitido explotar dos vías de análisis distintas y hacerlo en forma complementaria. Por un lado, el árbol de clasificación ha permitido identificar los mejores predictores para el caso de estudio, entre una gran variedad de variables que diferentes enfoques teóricos señalan como relevantes, y ha podido identificar no-linealidades en las relaciones asociadas con la edad de los flujos de comercio y diferentes agrupamientos. Por otra parte, la econometría clásica de MCO ha permitido un análisis paramétrico y de significatividad suponiendo la existencia de una relación global entre las variables explicada y explicativas, tras una transformación de estas últimas acorde a los resultados obtenidos mediante el método CART.

El análisis CART y las regresiones coinciden en considerar relevantes al nivel de ingreso del país destino, la existencia de lengua común y vecindad, y al valor inicial del flujo. Mientras que el primero de los factores eleva el estándar de las firmas exportadora y asegura escala de colocación, la lengua común y la vecindad reducen el costo de búsqueda y facilitan las operaciones en los casos de firmas exportadoras sin experiencia. El valor inicial del flujo da cuenta del compromiso de las firmas involucradas, tanto exportadoras como importadoras, en la operación, y esto es más significativo en los casos de bienes diferenciados. Los resultados verifican la relación entre aquella variable y la duración del flujo.

Cuando el flujo es nuevo, ya sea debido al producto exportado o al destino, la expectativa de supervivencia es menor. Este resultado es observado en los árboles de clasificación, y es robusto a los reagrupamientos por destino y producto. Este resultado se contradice con la hipótesis de que el desconocimiento sobre las propias condiciones para afrontar la competencia afecta negativamente en los primeros años, luego la experiencia reduce el riesgo y amplía la expectativa

media de vida. También es observado en las regresiones y, en este caso, la variable se asocia a la inestabilidad y a la rareza de la relación comercial. La experiencia exportadora ofrece estabilidad, y ésta eleva la expectativa de vida del flujo comercial.

La escala del flujo en términos del mercado abastecido es relevante. Cuanto mayor es la participación del flujo específico en el total de compras del destino considerado, mayor es la supervivencia. La escala es resultado de la explotación de cierto capital intangible y específico, ventaja que se ve reflejada en la duración de los flujos. El resultado positivo entre escala y duración puede confundirse con la explotación de una ventaja comparativa significativa, aunque si éste fuera el caso debería sostenerse en todos los mercados para el mismo producto.

Por lo visto en los árboles de clasificación, cuanto mayor es la escala y la experiencia, y más cercano el mercado de destino, el pronóstico de vida se alarga hasta los 14 años de media, mientras que, si ocurre exactamente lo contrario, la expectativa se reduce a 3 años de media. Si los destinos son países desarrollados o pertenecen al continente americano, la duración esperada es mayor. Existen algunas diferencias en los resultados cuando se agrupan las observaciones por tipo de producto y destinos. Sin embargo, las conclusiones anteriores serían robustas. Finalmente, los resultados obtenidos hasta el momento dan cuenta de una diferencia significativa entre flujos preexistentes y nuevos flujos, a pesar de que las variables explicativas de su duración resulten ser las mismas.

Por último, el estudio resulta relevante por otro motivo. No se encuentran demasiados artículos que se concentren en los determinantes de la supervivencia de los flujos de exportaciones argentinas. Si bien existen antecedentes, la bibliografía presenta algunos resultados que merecen una mayor atención. En este sentido, la actualización del análisis hasta un período reciente, la utilización de una metodología alternativa y novedosa para el campo de estudio, y el análisis desagregado por tipo de producto y socio comercial ofrecen resultados que permiten una discusión política enriquecida por la información estadística.

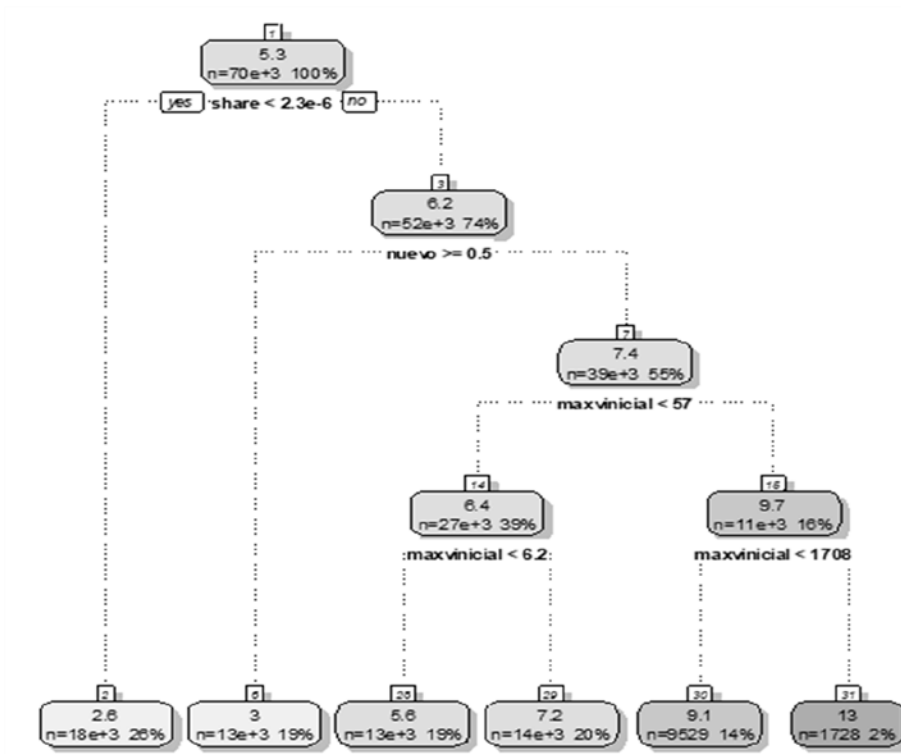
Referencias

- Besedes, T., & Blyde G. (2010). *What drives export survival? An analysis of export duration in Latin America*. Institute of Technology, Inter-American Development Bank.
- Besedes, T., & Prusa T. (2006a). Ins, outs, and the duration of trade. *Canadian Journal of Economics*, 39, 266-295.
- Besedes, T., & Prusa T. (2006b). Product differentiation and duration of US import trade. *Journal of International Economics*, 70, 339-358.
- Besedes, T., & Prusa, T. (2007). *The Role of Extensive and Intensive Margins and Export Growth*. Working Paper 13628, NBER.
- Blanco, A., Irimia, A., & Vázquez, M. (2016). Diseño de un modelo específico para la predicción de la quiebra de micro-entities. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 22, 3-18.
- Brenton, P., Saborowski, Ch., & von Uexkull, E. (2009). *What Explains the Low Survival Rate of Developing Country Export Flows?* The World Bank Poverty Reduction and Economic Management Network International Trade Department, June.
- Corbella, V. (2016). *Asimetrías Estructurales, Integración Comercial y Especialización Industrial: el caso de la UNASUR*. Tesis de Magister en Economía. Universidad Nacional del Sur.

- Fugazza, M., & Molina, A.C. (2011). On the determinants of exports survival. *Policy Issues in International Trade and Commodities Study Series*, 46. Geneva: United Nations.
- Gepp, A., & Kumar, K. (2015). Predicting financial distress: A comparison of survival analysis and decision tree techniques. *Procedia Computer Science*, 54, 396-404.
- Giordano, P. (2016). *Cambio de Marcha. América Latina y el Caribe en la nueva normalidad del comercio global*. Monitor de Comercio e Integración, BID, Washington DC. <https://publications.iadb.org/handle/11319/7942>
- González, B. (2015). *Supervivencia de los Flujos de Exportación. Análisis de la Duración a Nivel de Producto-Destino para Argentina 1993-2012*. Tesis de Maestría en Economía, Facultad de Ciencias Económicas Universidad Nacional de La Plata.
- González-Fernández, M., & González-Velasco, C. (2016). Which Countries Pay More or Less for Their Long Term Debt? A CART Approach. *Journal of Quantitative Methods for Economics and Business Administration*, 21, 103-116.
- Lall, S. (2000). *The Technological Structure and Performance of Developing Country Manufactured Exports, 1985-1998*. Working Paper 44. Queen Elizabeth House, Oxford University.
- Nicita, A., Shirotori, M., & Tumurchudur, B. (2013). Survival Analysis of the exports of least developed countries: the role of comparative advantage. *Policy Issues in International Trade and Commodities Study Series*, 54. Geneva: United Nations.
- Nitsch, V. (2007). *Die another day: duration in German Import trade*. CESifo Working paper 2085. Munich: CESifo.
- Rauch, J. (1996). *Networks versus markets in International Trade*. Working Paper 5617, National Bureau of Economic Research, June.
- Rauch, J.E. (1999). Networks versus markets in international trade. *Journal of International Economics*, 48, 7-35.
- Rauch, J. (2007). *Development through synergistic reform*. National Bureau of Economic Research, June.
- Sohn, S., & Moon, T. (2004). Decision tree based on data envelopment analysis for effective technology commercialization. *Expert Systems with Applications*, 26(2), 279-284.
- Tso, G., & Yau, K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761-1768.

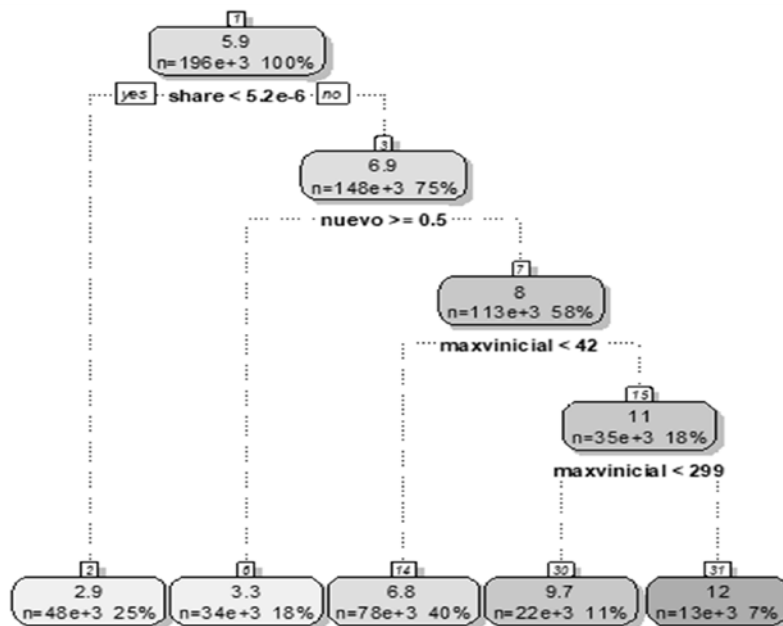
Anexo

Figura A1. Árbol de clasificación para destinos: Asia del Este y Pacífico.



Fuente: Elaboración propia.

Figura A2. Árbol de clasificación para destinos: Europa y Asia Central.



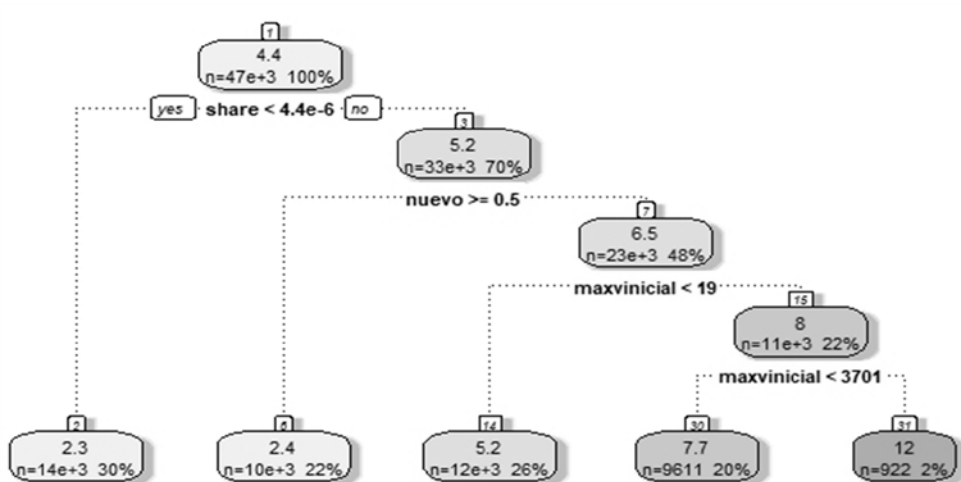
Fuente: Elaboración propia.

Figura A3. Árbol de clasificación para destinos: América Latina y El Caribe.



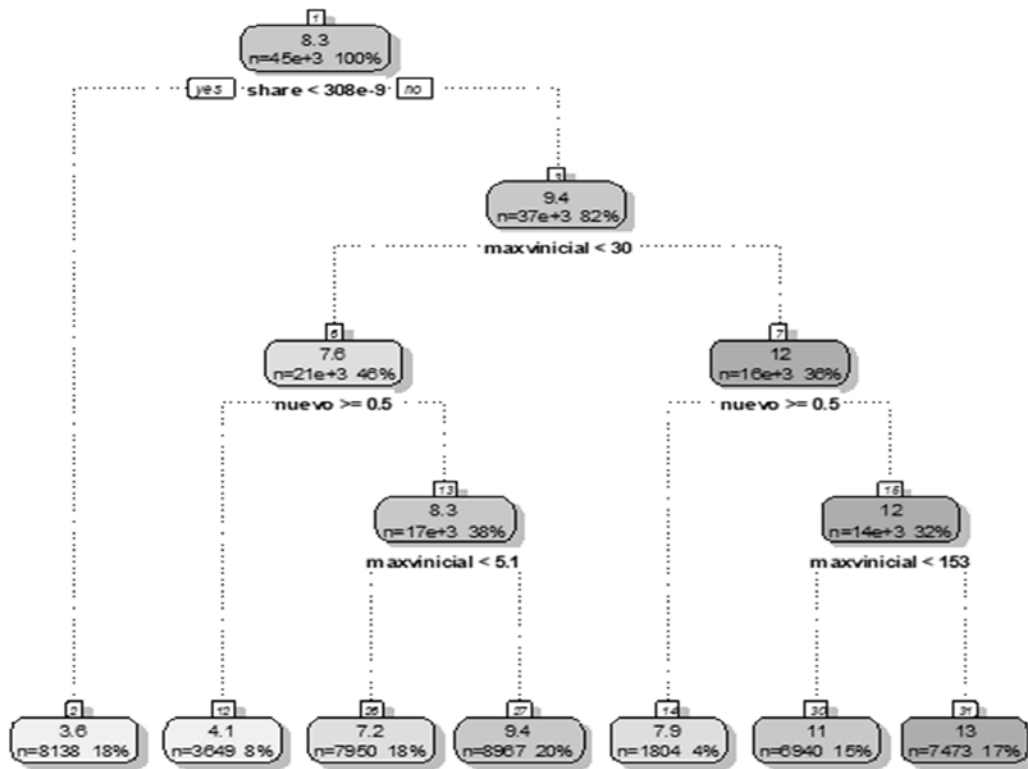
Fuente: Elaboración propia.

Figura A4. Árbol de clasificación para destinos: África del Norte y Este Medio.



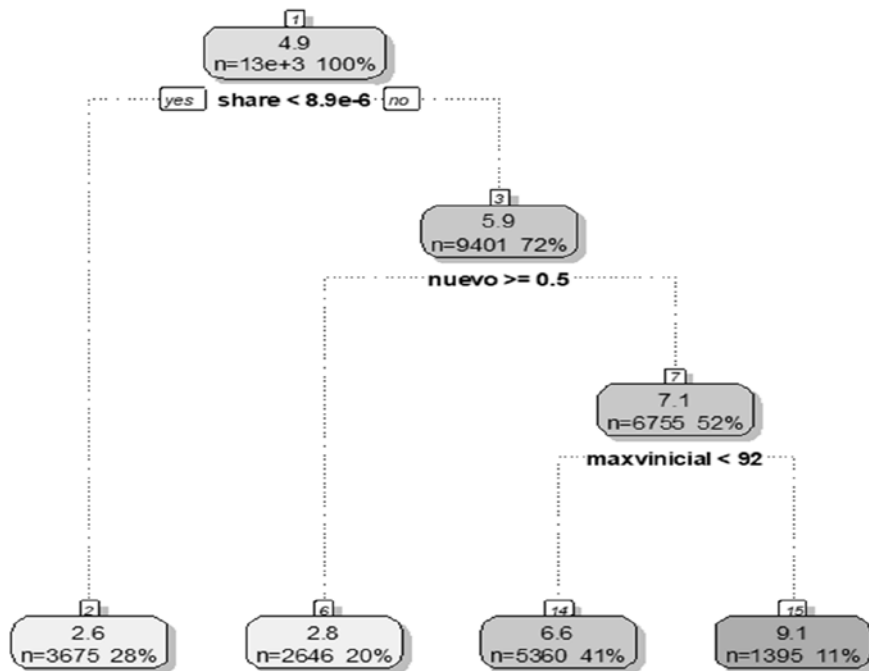
Fuente: Elaboración propia.

Figura A5. Árbol de clasificación para destinos: América del Norte.



Fuente: Elaboración propia.

Figura A6. Árbol de clasificación para destinos: Asia del Sur.



Fuente: Elaboración propia.