



APERTURA COMERCIAL, POBREZA Y DESIGUALDAD: UN ANÁLISIS MEDIANTE ÁRBOLES DE REGRESIÓN

Fecha de presentación: 7 de agosto de 2020

Fecha de aceptación: 18 de noviembre de 2020

Ernesto R. Gantman¹

Resumen

El presente trabajo presenta un análisis empírico preliminar de las relaciones entre apertura comercial, pobreza y desigualdad en el ingreso mediante técnicas de análisis cuantitativo frecuentemente utilizadas en aprendizaje automático (*machine learning*), los árboles de regresión. En primer lugar, se observa que la desigualdad es menor en países con menor apertura comercial cuando esta última variable se operacionaliza con el coeficiente apertura económica (volumen de comercio exterior sobre el PBI), relación que no se verifica si utilizamos un indicador alternativo de la variable, el nivel medio de tarifas a las importaciones. En segundo lugar, ocurre algo similar con el nivel de pobreza pero en forma inversa: el coeficiente de apertura no parece tener influencia sobre la misma, pero un menor nivel de tarifas sí aparece asociado a un menor nivel de pobreza. Los hallazgos apoyan la hipótesis de que la apertura comercial mejora la distribución del ingreso y el nivel de pobreza de un país, pero esta asociación depende del tipo de indicador elegido para operacionalizar la apertura comercial.

Palabras clave

Pobreza, desigualdad del ingreso, apertura comercial, árboles de regresión

Códigos JEL: F14, F62, I30

¹ Universidad de Buenos Aires y Universidad de Belgrano
Email: ernesto.gantman@comunidad.ub.edu.ar, egantman@economicas.uba.ar



Abstract

This work presents a preliminary empirical analysis on the relationship between trade openness, poverty and income inequality using regression trees, a quantitative method that is mostly used in machine learning. First, it can be observed that income inequality tends to be lower in countries with low trade openness when this latter variable is operationalized with the openness coefficient (trade over GDP ratio). In contrast, this relationship is not verified if average tariffs is used as an alternative indicator of trade openness. Second, something similar happens with the poverty level of poverty, but in an inverse way. The openness coefficient does not seem to have an influence on poverty, but a lower level of tariffs appears to be associated with lower poverty. The findings support the hypothesis that trade openness improves the distribution of income and the poverty level of a country, but this association is contingent upon the indicator utilized to operationalize trade openness.

Keywords

Poverty, income inequality, trade openness, regression trees

JEL codes: F14, F62, I30



Introducción

El problema del “efecto derrame” del crecimiento económico tiene especial relevancia en nuestros días y ciertamente no es nuevo. En América Latina durante los años noventa, se argumentaba que los cambios en procura de un menor control del Estado de la economía, incluyendo menor presencia directa como productor de bienes y servicios, redundarían a la larga en menor pobreza y desigualdad en los países donde se implementasen. En muchos casos, las promesas implícitas en tales diagnósticos no estuvieron a la altura de las circunstancias. La reacción frente a las políticas pro-mercado de los noventa no tardó en hacerse esperar y el surgimiento de distintas variantes de regímenes de izquierda en la región (desde izquierdas más moderadas hasta regímenes más radicalizados como el bolivarianismo venezolano) es prueba de ello. Sin embargo, la performance de los regímenes “nacionales y populares” no se demostró claramente superior. El caso venezolano proporciona una evidencia contundente sobre el fracaso de los llamados populismos en la región. Es difícil pensar en una nación con un gobierno más antitético a las políticas de libre mercado y un grado más alarmante de pobreza. Además, hay ejemplos inversos como el caso chino, que constituye un caso claro al respecto. El abandono de la doctrina maoísta y la apertura económica convirtieron a China en un país con características marcadamente capitalistas, y este giro impensado tuvo a la larga un correlato en un incremento del nivel de vida de amplios sectores de la población.

La lógica subyacente a los planteos acerca de la relación entre pobreza, desigualdad y apertura comercial, en forma restrictiva, o liberalización económica, como marco más amplio o general del cual la reforma del comercio exterior es sólo un elemento, se refiere al mecanismo que vincula variables de diseño del sistema económico con variables referidas al bienestar social de la población. Justamente es en este sentido que el “efecto derrame” entra en escena. El razonamiento plantea que la reformas pro-mercado en diversos sectores de la economía generan mayor crecimiento económico. En el corto o mediano plazo, se presume que este crecimiento va a afectar en forma positiva también a las clases más desaventajadas de la sociedad. No obstante, dejando de lado ejemplos extremos como algunos de los mencionados precedentemente, la evidencia empírica no es conclusiva en torno a la existencia del efecto derrame. Si bien creemos que la mejor forma de estudiar este problema es a partir de un análisis longitudinal, en este trabajo introducimos a efectos exploratorios preliminares una técnica metodológica que no es utilizada con frecuencia en ciencias sociales. Nos referimos a CART (*classification and regression trees*) y al método de bosques aleatorios (*random forest*). El resto de la ponencia se estructura del siguiente modo. La próxima sección discute en forma muy breve algunos antecedentes sobre la problemática. Luego presentamos los métodos a utilizar, los datos y algunos resultados obtenidos. Por último, cerramos el trabajo destacando las principales conclusiones del estudio.

Desigualdad, pobreza y apertura económica

Como señalamos anteriormente, se presume que el mecanismo de transmisión del efecto de las políticas de liberalización económica, de la cual la apertura comercial es un elemento importante, sobre la pobreza y la desigualdad en el ingreso es el crecimiento económico. Pero dentro de los enfoques de la ortodoxia económica, lo central es lograr



una asignación más efectiva de los recursos y se presume que las barreras al comercio exterior constituyen un impedimento a tal efecto. Formulaciones como las de Heckscher-Ohlin (Heckscher, 1950) y Stolper-Samuelson (1941) sugieren que una mayor apertura económica tiende a incrementar la producción de los sectores más competitivos de la economía y a desincentivar la producción del resto de los sectores.

En este sentido, se ha afirmado que niveles elevados de crecimiento económico no son fácilmente alcanzables a partir de estrategias orientadas al mercado interno, ya sea porque el tamaño del mercado interno es reducido, porque cerrar la economía involucra una menor eficiencia en el sistema económico, o ambos (Sharma y Morrissey, 2006). Así, puede considerarse que la apertura comercial reduce la ineficiencia productiva al morigerar distorsiones de precios o introducir incentivos para que los productores aumenten su productividad. Sin embargo, la complejidad de los procesos económicos y la influencia de numerosas variables intervinientes indican que en la práctica la relación entre apertura comercial, pobreza y desigualdad es incierta. Tras un análisis de once casos en Africa, Asia y América Latina, Sharma y Morrissey (2006) concluyen que sólo en unos pocos de ellos la liberalización comercial ha producido una reducción de la pobreza, aunque tampoco hay casos en que se haya generado un sustancial incremento de la misma. Respecto a la desigualdad, el efecto fue ambiguo y en varios casos la apertura comercial produjo un aumento de la misma, posiblemente porque los sectores más dinámicos asociados a la exportación, que es donde se verifican los beneficios reales de tales políticas, hayan tenido una utilización más intensiva de mano de obra calificada.

Por su parte, Vos et al. (2006), analizando los casos de varios países latinoamericanos que atravesaron algunas experiencias de liberalización del comercio exterior y que luego vieron una disminución en su tasa de crecimiento (posiblemente debido al efecto de otras variables económicas), creen que lejos de ser causantes de incrementos en pobreza y desigualdad social, tales reformas pueden haber tenido efectos levemente positivos tanto para el crecimiento como para la reducción de la pobreza. No obstante, al igual que Sharma y Morrissey, advierten que mientras algunos sectores de la fuerza laboral pueden beneficiarse con la liberalización comercial, particularmente los más calificados, aquellos que cuentan con menor nivel de calificaciones pueden verse perjudicados, lo que explica el incremento de la desigualdad observado en algunos casos.

En rigor, el efecto de la apertura comercial puede diferir incluso en un mismo país según se trate de pobreza o desigualdad en regiones urbanas o rurales, siendo negativo en las primeras y positivo en las segundas, como lo demuestran Castilho et al. (2012) para el caso de Brasil. Por otra parte, también hay experiencias que demuestran que la desigualdad salarial entre trabajadores calificados y no calificados puede disminuir tras la apertura al comercio exterior, como indican Kumar y Mishra (2008) para el caso de la India. Por su parte, Atolia (2007) sugiere que los efectos de la apertura comercial pueden ser negativos a corto plazo pero revertirse en el largo plazo. En definitiva, hay que remarcar que la complejidad del problema motiva que algunos autores destaquen la inexistencia de relaciones unívocas y recomienden analizar variables adicionales, no sólo en la estructura económica de los países considerados, sino en los aspectos institucionales que dan forma a las mismas (Bliss, 2007). Obviamente, no sólo la especificidad particular de la economía en los países analizados condiciona los resultados de los hallazgos empíricos, sino fundamentalmente las metodologías de



análisis. Sin pretensión de rigurosidad porque consideraremos muy pocos factores en nuestro análisis, introduciremos en nuestra aproximación preliminar a este tema una técnica metodológica que no ha sido empleada hasta el momento.

Aspectos metodológicos

El método CART (*classification and regresión trees*) se usa mayormente para realizar predicciones, tanto respecto a variables cuantitativas como cualitativas (Breiman et al., 1984). Es un método muy utilizado en inteligencia artificial y las predicciones que realiza deben ser validadas con un conjunto de datos distinto al empleado para generar el árbol de regresión. Su comprensión es intuitiva. Un algoritmo abre variantes o ramas, cada una de las cuales corresponde a una condición. En este último aspecto el parecido con los árboles de decisión es evidente, pero las ramas no indican un curso de acción a seguir (no es un modelo prescriptivo), sino que nos llevan a nuevas condiciones que se refieren a valores en las variables predictivas de interés. El árbol culmina en las ramas terminales y cada partición es seleccionada por un algoritmo que minimiza la suma de desvíos al cuadrado en cada rama (cuando la variable dependiente de interés es continua). A los efectos de su utilización en ciencias sociales lo más relevante son las conclusiones que pueden extraerse al ver cómo determinadas combinaciones sucesivas de variables se asocian a niveles menores o mayores de la variable dependiente que constituye el foco de la investigación. En este sentido, los árboles de regresión, a diferencia de las técnicas tradicionales de regresión, no dependen de ningún supuesto acerca de la distribución de los datos ni de la forma funcional de las relaciones entre las variables.

Se han propuestos distintos algoritmos para generar árboles de decisión, pero en términos generales, podemos describir el proceso iterativo de la siguiente forma. En primer término, se toman todas las unidades de análisis como perteneciendo a una misma rama o a un tronco único y se calcula el valor correspondiente a dicha rama o tronco a partir de una medida que es la que intentará minimizar mediante un proceso de división.

La fórmula para un árbol de regresión implica calcular la sumatoria de desvíos al cuadrado para dicho árbol. La fórmula que se utiliza es la siguiente:

$$S = \sum_{c_i} \sum_{z \in c_i} (y_z - y_c)^2$$

En la misma, c_i denota una rama i del árbol de regresión, z significa un elemento cualquiera perteneciente a la rama c , y_z es el valor de la variable dependiente para el elemento z , e y_c es la media de la variable dependiente para todos los valores de la rama c .

Si consideramos, el tronco o raíz del árbol, lo que se obtiene es la sumatoria de los desvíos al cuadrado de valores de la variable dependiente respecto a la media de todas las unidades.

Luego, se procede a un proceso de subdivisión del árbol, computando el valor de S para distintas particiones binarias en base a valores de las variables independientes. La partición binaria cuyo valor minimiza S es la que se considera. Naturalmente, para que se forme una partición ese valor que minimiza S debe ser inferior al valor de S para el nodo anterior que contiene a ambas particiones. Consideremos una ilustración simple de este proceso de partición. Si tenemos una única variable dependiente con valores 1.3, 4.5, 5.6, 7.5 y 7.7, se calcularán cuatro particiones binarias tomando los puntos intermedios entre los cinco valores anteriores. Por ejemplo, usando la fórmula presentada previamente, esto implicaría dos ramas c_1 y c_2 para cada partición, asignadas según el valor de la variable dependiente sea menor a 2.9 (punto medio entre 1.3 y 4.5). Y también habría que calcular la fórmula para particiones fijadas en los valores 5.05, 6.55 y 7.6. Cada partición tendrá dos ramas y por cada rama obtendremos la media y la suma de los desvíos al cuadrado respecto a dicha media. Sumando ambas, se obtiene el valor para ese árbol. Así, para las cuatro particiones (cuatro árboles) calcularemos un valor S y aquel que corresponda al menor valor S será la apertura en ramas (nodos) que vamos a retener en nuestro árbol de regresión.

Los nodos o ramas obtenidas pueden a su vez dividirse en subramas, siguiendo el mismo procedimiento. El punto ahora consiste en qué regla adoptar para detener el proceso de división en ramas. Esto ocurre cuando (1) los valores de la variable independiente son los mismos para todas las unidades pertenecientes a la rama (es decir, la suma al cuadrado de los desvíos es cero), (2) se alcanza un número determinado de unidades para dicha rama que tomamos arbitrariamente como tope al proceso de división, o (3) la disminución obtenida en el valor de S es inferior a un valor arbitrariamente establecido como umbral.

El procedimiento precedente remite a una búsqueda exhaustiva y representa la lógica más simple para la construcción del árbol de regresión definitivo. Sin embargo, el proceso computacional puede tener demandas de memoria excesivas para ciertos conjuntos de datos, por lo cual se han formulado diversos algoritmos para su cálculo como M5 y GUIDE (Loh, 2011). En la práctica distintos programas para la obtención de árboles de regresión pueden dar lugar a soluciones levemente diferentes para un mismo conjunto de datos.

El método de bosques aleatorios (Breiman, 2001) se basa en árboles de regresión. Aquí el resultado surge como un promedio de los resultados de un número de árboles de decisión (típicamente se requieren 500 o 1000 de ellos) generados a partir de un proceso de *bootstrapping* sobre el conjunto de datos original. El *bootstrapping* es básicamente una técnica de remuestreo. Lo que propone esta técnica es tomar muestras de datos aleatoria a partir de una muestra original dada para realizar estimaciones e inferencias. De acuerdo a esta método, la muestra original hace las veces de población sobre la cual se extraen aleatoriamente nuevas muestras. Esta técnica estadística, que requiere gran poder computacional, ha sido utilizada en diversos campos incluyendo la econometría. (Horowitz, 2019). Obviamente, los diversos árboles generados mediante *bootstrapping* se computan sobre conjuntos de datos distintos (las muestras aleatoria obtenidas sobre la muestra original difieren entre sí). Consecuentemente, las variables que configuran las particiones en cada nodo pueden también ser diferentes. El resultado de interés para el científico social es un ranking de las variables que son más importantes en la generación de los árboles de decisión. Las mismas pueden entonces utilizarse en análisis exploratorios adicionales con técnicas paramétricas para evaluar su influencia sobre las



variables dependientes de interés. Con pocas variables, como es el caso de nuestro ejemplo, el empleo de este método puede parecer superfluo pero lo presentamos igualmente a efectos ilustrativos.

Las variables dependientes de nuestro análisis son pobreza y desigualdad en el ingreso. En el caso de la desigualdad utilizaremos el índice de Gini obtenido de los World Development Indicators del Banco Mundial (2018) para el año 2016. Para la variable pobreza también usaremos la misma fuente documental con la brecha de pobreza (*poverty gap*) para personas que viven con menos de 1.90 U\$A diarios (ajustados por PPP).

La operacionalización de la variable apertura comercial ha sido objeto de debate (Pritchett, 1991). En su mayor parte, el gran problema reside en la falta de datos comparativos confiables a lo largo de un período extenso de tiempo para un número grande de países (Gantman y Dabós, 2018). En esta oportunidad, y como trabajaremos con datos de sección cruzada, no tendremos ese problema. Sin embargo, usaremos dos indicadores alternativos para esta variable. En primer lugar, utilizaremos uno de los indicadores más comunes para estimar la variable apertura comercial: la suma de exportaciones e importaciones como porcentaje del PBI, usualmente denominada coeficiente de apertura económica. En segundo lugar, utilizaremos un indicador de la tasa media de las tarifas aplicadas a las importaciones, que asume que un país menos proteccionista aplica tasas menores a productos importados. Es posible, además, emplear otros indicadores como cuotas al ingreso de importaciones, números de convenios de comercio exterior suscriptos por el país o proporción de importaciones sobre PBI, pero en esta oportunidad dado el nivel preliminar del estudio limitaremos el análisis a los dos indicadores de apertura comercial precedentemente indicados.

Además, utilizaremos dos variables de control como el nivel de desarrollo del país, medido a partir de su PBI per cápita ajustado por PPP a valor constante en dólares 2011, y el tamaño de la economía, que es el indicador de PBI ajustado por PPP a valor constante en dólares 2011 (expresado en millones de dólares).

Debido a la falta de datos actualizados para 2016 para muchos países, la muestra de países es reducida (25). Se trata de los siguientes: Ucrania, Moldavia, Kyrgyzstan, Bielorrusia, Rumania, Mongolia, Bangladesh, Armenia, Georgia, Uruguay, Sri Lanka, El Salvador, Estados Unidos, Turquía, Argentina, México, Perú, Bolivia, Ecuador, Rep. Dominicana, Paraguay, Costa Rica, Honduras, Panamá y Colombia. Es problemático aseverar que se trata de una muestra representativa. Creemos que habría un sesgo hacia países en desarrollo, a pesar de que los países desarrollados son justamente los que presumiblemente tienen sistemas estadísticos nacionales de mayor calidad. En cualquier caso, el presente trabajo es básicamente exploratorio con el objetivo de indagar qué tipo de resultados se obtienen con esta nueva metodología. El procesamiento de datos se realizó con el programa R y los paquetes de software “tree” y “randomForest”.

Resultados

Si analizamos el efecto de la apertura comercial sobre la desigualdad con un árbol de regresión, añadiendo el desarrollo y el tamaño de la economía como variables adicionales, el resultado del árbol indica que existen un punto de corte en términos de la

apertura comercial operacionalizada en términos del coeficiente de apertura que determina varios niveles de desigualdad (Figura 1). Los valores que se reproducen en la misma son el número de nodo, la condición de decisión asociada al mismo, el número de observaciones del nodo, el desvío standard y la media de la variable dependiente para las observaciones del nodo. El primer nodo es el nodo raíz que contiene 25 observaciones con una media de 39.29 para la desigualdad y un desvío standard de 1641.

Mayores niveles de apertura comercial (superiores a 101,8% del PBI) se asocian a menores niveles de desigualdad en el ingreso. Obsérvese que la suma de los desvíos de los nodos 2 y 3 es 888.37, inferior al nodo 1 (raíz) que tiene 1641, tal como indicamos en nuestra descripción del proceso de generación del árbol. Idénticas precisiones se obtiene al examinar las sumatorias de los valores de otras ramas menores respecto a sus respectivas ramas madre. La suma de los desvíos standard de los nodos terminales (marcados con un asterisco) es obviamente bastante inferior a dicho valor, ascendiendo a 154.19.

Figura 1 – Arbol de regresión (Vd: desigualdad del ingreso – apertura = coef. de apertura)

1) raíz	25 (obs)	1641.00 (ds)	39.29 (mvd)
2) Coef. apert. < 101.839	20 (obs)	802.30 (ds)	42.03 (mvd)
4) PBI < 405000	13 (obs)	427.50 (ds)	43.08 (mvd)
6) PBI < 71038.5	6 (obs)	277.10 (ds)	40.40 (mvd) *
7) PBI > 71038.5	7 (obs)	70.61 (ds)	45.37 (mvd) *
5) PBI > 405000	7 (obs)	334.40 (ds)	40.10 (mvd) *
3) Coef. apert. > 101.839	5 (obs)	86.07 (ds)	28.32 (mvd) *

Nota: (*) asterisco indica nodo terminal

Con menores niveles de apertura comercial, los distintos valores de la desigualdad dependen del tamaño de la economía en una relación no lineal: de acuerdo a la partición en los nodos 4 y 5, a menor PBI (referencia: 405000), mayor desigualdad, pero en la partición del nodo 4 en los nodos 6 y 7, a mayor PBI (referencia: 71038.5), mayor desigualdad). Precisamente, la ventaja de los árboles de regresión es que permiten estas formas no lineales complejas, La variable desarrollo de la economía no tiene efecto en la construcción del árbol de regresión.

Si hacemos el mismo análisis, pero operacionalizando la apertura comercial a partir del nivel promedio de tarifas, sólo la variable desarrollo aparece como relevante en el árbol resultante (Figura 2). Ni el tamaño de la economía ni el nivel promedio de tarifas aparecen en nodo alguno del árbol de decisión.

Figura 2 – Arbol de regresión (vd: desigualdad del ingreso - apertura = tarifa promedio)



- 1) raíz 19 (obs) 1310.0 (ds) 38.21 (mvd)
- 2) Desarrollo < 8530.81 6 (obs) 264.5 (ds) 31.27 (mvd) *
- 3) Desarrollo > 8530.81 13 (obs) 623.2 (ds) 41.42 (mvd)
- 4) Desarrollo < 16076.8 6 (obs) 125.1 (ds) 45.70 (mvd) *
- 5) Desarrollo > 16076.8 7 (obs) 293.5 (ds) 37.74 (mvd) *

Nota: (*) asterisco indica nodo terminal

Repetimos los análisis precedentes pero utilizando la pobreza como variable dependiente. Cuando la apertura comercial es operacionalizada mediante el coeficiente de apertura, el árbol de decisión sólo toma como relevantes los valores de la variable desarrollo (Figura 3). Ni la apertura comercial, ni tamaño de la economía contribuyen al mismo, como se observa más abajo. El árbol brinda tres particiones para distintos intervalos de valores del PBI per cápita. Como es esperable, a mayor nivel de riqueza del país, menor pobreza.

Figura 3 – Arbol de regresión (Vd: brecha de pobreza – apertura = coef. de apertura)

- 1) raíz 25 (obs) 411.800 (ds) 2.8840 (mvd)
- 2) Desarrollo < 7188.01 5 (obs) 215.400 (ds) 7.9000 (mvd) *
- 3) Desarrollo > 7188.01 20 (obs) 39.120 (ds) 1.6300 (mvd)
- 4) Desarrollo < 14750.2 11 (obs) 23.260 (ds) 2.2180 (mvd) *
- 5) Desarrollo > 14750.2 9 (obs) 7.409 (ds) 0.9111 (mvd) *

Nota: (*) asterisco indica nodo terminal

Interesantemente, la operacionalización de la apertura comercial con el valor medio de las tarifas sí resulta relevante en el árbol de regresión resultante. Menor pobreza se asocia a mayor apertura económica (menor nivel de tarifas). Dentro de una de las ramas del árbol, el valor de la pobreza depende a su vez del nivel de desarrollo (países con PBI per cápita superior a 14750 U\$A tienen menor pobreza que los que tienen un valor inferior a dicha cifra).

Figura 4 – Arbol de regresión (Vd: brecha de pobreza – apertura = tarifa promedio)

- 1) raíz 19 (obs) 224.800 (ds) 2.4740 (mvd)
- 2) Tarifas < 8.38 14 (obs) 27.360 (ds) 1.4860 (mvd)
- 4) Desarrollo < 14750.2 8 (obs) 18.560 (ds) 1.9380 (mvd) *



5) Desarrollo > 14750.2 6 (obs) 4.988 (ds) 0.8833 (mvd) *
 3) Tarifas > 8.38 5 (obs) 145.500 (ds) 5.2400 (mvd) *

Nota: (*) asterisco indica nodo terminal

Dado que el valor de los árboles de regresión depende fuertemente de los datos sobre los que ellos se calculan (lo que se denomina el *training set*) es necesario un extenso proceso de validación con datos adicionales. Por lo tanto es relevante añadir al análisis precedente el resultado de correr una simulación de 500 árboles de regresión con el método de bosques aleatorios. Uno de los principales resultados que se obtienen a partir de este método, es una clasificación de la importancia de las variables según su relevancia en los resultados de los árboles que se generan por *bootstrapping* (en este caso, tomamos 500 árboles). Este ranking de variables se puede hacer a partir de las disminuciones en impurezas de los nodos al hacer particiones en las variables, lo cual se promedia a lo largo de los distintos árboles aleatoriamente generados. En el caso de las regresiones se toma como medida de las impurezas el cuadrado de la suma de los residuos.

Cuando utilizamos la metodología de bosques aleatorios (Tabla 1) con desigualdad como variable dependiente junto al nivel de desarrollo, el PBI y el coeficiente de apertura, este último indicador es que resulta más importante, seguido por PBI y desarrollo. Los números entre paréntesis indican un ranking de la importancia de cada variable independiente para explicar cambios en la variable dependiente. En cambio, si la apertura comercial se operacionaliza mediante el valor medio de las tarifas, el nivel de desarrollo surge como la variable más importante y la apertura comercial aparece en segundo término. Al tomar como variable dependiente la pobreza, junto al nivel de tarifas, el PBI y el indicador de desarrollo, el nivel de tarifas es lo que tiene mayor de importancia. Cuando la apertura comercial se operacionaliza mediante el coeficiente de apertura, la variable que surge como más relevante de acuerdo a este método es el nivel de desarrollo, seguida por el tamaño de la economía.

Tabla 1. Análisis de bosques aleatorios

Variables independientes	Variable dependiente: desigualdad		Variable dependiente: pobreza	
	Incremento pureza de nodo	Incremento pureza de nodo	Incremento pureza de nodo	Incremento pureza de nodo
Desarrollo	372.03 (3)	393.98 (1)	158.08 (1)	66.75 (2)
PBI	430.97 (2)	297.22 (3)	67.36 (2)	35.42 (3)
Coef. de apert.	575.48 (1)		62.50 (3)	



Tarifas	322.60 (2)	78.77 (1)
---------	------------	-----------

Nota: ranking de importancia relativa de las variables entre paréntesis

En definitiva, estos resultados son completamente compatibles con los hallazgos obtenidos con los árboles de regresión, los cuales en general son de más fácil interpretación en la medida en que no se trate de un árbol con una multiplicidad de nodos.



Conclusión

En este trabajo, hemos realizado un análisis empírico de las relaciones entre apertura económica, pobreza y desigualdad en el ingreso a partir de técnicas de aprendizaje automático (*machine learning*). Es importante subrayar el carácter preliminar de la presente investigación. Si bien los resultados tienen limitaciones propias por tratarse de un análisis de sección cruzada y de basarse en una muestra reducida de países, algunos hallazgos obtenidos que resumiremos a continuación aparecen como interesantes. Sin embargo, estamos conscientes respecto a que hay variables omitidas como las características de las políticas fiscales y sociales propias de cada Estado nacional. Sabemos que el combate de la desigualdad en el ingreso es una de las tareas centrales de la agenda pública del Estado contemporáneo (Oszlak, 2006; Oszlak y Gantman, 2007), y variables como el tipo de régimen político o incluso algunas dimensiones propias de la cultura de los países impactan en la forma en que se ataca la pobreza y la desigualdad desde el sector público. Por tal motivo, debemos admitir el carácter provisorio de los resultados y la necesidad de profundizar el análisis incorporando variables adicionales.

En términos generales y con las salvedades apuntadas precedentemente, la hipótesis de que la apertura económica mejora la distribución del ingreso y contribuye a la disminución de la pobreza recibe cierto apoyo empírico, pero el mismo no es robusto y depende del indicador elegido para operacionalizar la variable apertura económica. En primer lugar, se observa que la desigualdad es menor en países con menor apertura comercial cuando esta última variable se operacionaliza mediante el coeficiente de apertura económica. Al emplear un indicador alternativo, el nivel medio de tarifas a las importaciones, la relación ya no se verifica. En segundo lugar, ocurre algo similar con el nivel de pobreza pero en forma inversa: el coeficiente de apertura no parece tener influencia sobre la misma, pero un menor nivel de tarifas sí se presenta como algo relacionado con un menor nivel de pobreza. Estos resultados muestran que la elección de indicadores para operacionalizar la apertura comercial puede dar lugar a conclusiones disímiles. Pero, más importante aún, ponen de manifiesto la pertinencia de utilizar nuevas técnicas de análisis para indagar acerca de los determinantes de la pobreza y la desigualdad. Naturalmente, el carácter parcial y exploratorio de nuestros hallazgos no da lugar a extrapolación más allá de la muestra y período de tiempo considerados, aunque los mismos sugieren la necesidad de profundizar el análisis con más datos y variables adicionales.

En este sentido, dos estrategias son posibles. Por un lado, para comprender mejor las relaciones entre apertura comercial como variable independiente y pobreza y desigualdad del ingreso como variables dependientes sería recomendable un estudio longitudinal mediante técnicas econométricas como las series de tiempo en panel. Por otro lado, a los efectos de validar mejor la utilidad del enfoque de árboles de regresión, sería pertinente la realización de un estudio con validación cruzada, es decir utilizar un conjunto de países para la formulación de un modelo inicial y comprobar si el mismo produce buenas predicciones de los valores de la variable dependiente empleando otro conjunto de países. Además, sería deseable también comprobar si las predicciones generadas mediante la utilización de árboles de regresión son más acertadas que aquellas derivadas de modelos de regresión tradicionales.



Referencias

Atolia, M. (2007) Trade liberalization and rising wage inequality in Latin America: Reconciliation with HOS theory. *Journal of International Economics*, 71: 467–494

Banco Mundial (2018) World Development Indicators. Disponible en: <<http://datatopics.worldbank.org/world-development-indicators/>>.

Bliss, C. (2007) *Trade, Growth, and Inequality*. Oxford: Oxford University Press.

Breiman, L. (2001) Random Forests, *Machine Learning*, 45: 5 – 32.

Breiman, L., J.H. Friedman, R.A. Ohlsen y C.J. Stone (1984) *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Castilho, M., M. Menéndez y A. Sztulman (2012) Trade Liberalization, Inequality, and Poverty in Brazilian States. *World Development*, 40(4): 821–835.

Gantman, E. y Dabós, M.P. (2018) Does trade openness influence the real effective exchange rate? New evidence from panel time-series. *SERIEs*, 9(1): 91-113.

Heckscher, E. (1950) The Effect of Foreign Trade on the Distribution of Income. En H.S. Ellis y L.A. Metzler (eds.), *Readings in the Theory of International Trade* Homewood, Ill.: Richard D. Irwin.

Horowitz, J.L. (2019) Bootstrap methods in econometrics. *Annual Review of Economics*, 11, 193-224.

Kumar, U. y P. Mishra (2008) Trade Liberalization and Wage Inequality: Evidence from India. *Review of Development Economics*, 12(2): 291–311.

Loh, W.J. (2011) Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1): 14-23.

Oszlak, O. (2006) Oszlak, Burocracia estatal: política y políticas públicas. *POSTData: Revista de Reflexión y Análisis Político*, 11: 11-56.

Oszlak, O. y Gantman, E. (2007). La agenda estatal y sus tensiones: gobernabilidad, desarrollo y equidad. *Nordic Journal of Latin American and Caribbean Studies*, 37(1): 79-110.

Pritchett L (1991) Measuring outward orientation in developing countries: can it be done? World Bank, Working Paper WPS566.

Sharma, K y O. Morrissey (2006) Introduction: Trade, growth and inequality in the era of globalization. En K. Sharma y O. Morrissey (eds.) *Trade, growth and inequality in the era of globalization*. London: Routledge.



Stolper W., y P. Samuelson(1941) Protection and Real Wages. *The Review of Economic Studies*, 9(1), 58-73.

Vos R., E. Ganuza y S. Morley (2006) Introduction: Rising exports, slower growth and greater inequality: is trade liberalization to blame? En R. Vos, E. Ganuza, S. Morley y S. Robinson (eds.) *Who Gains from Free Trade? Export-led growth, inequality and poverty in Latin America*. London: Routledge.