

INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE ECONOMÍA

SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

**GLOBALIZACIÓN Y DESIGUALDAD, ENFOQUE MULTIDIMENSIONAL
CON RE-MUESTREO DIRIGIDO A REDES NEURONALES ARTIFICIALES Y
ÁRBOLES CLASIFICADORES.**

T E S I S

**QUE PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS ECONÓMICAS
(DESARROLLO ECONÓMICO)**

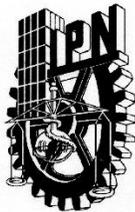
PRESENTA:

ANA CECILIA PARADA ROJAS



MÉXICO, CIUDAD DE MÉXICO

NOVIEMBRE DE 2019



INSTITUTO POLITÉCNICO NACIONAL SECRETARIA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REGISTRO DE TEMA DE TESIS Y DESIGNACIÓN DE DIRECTORES DE TESIS

Ciudad de México, 6 de septiembre de 2019

El Colegio de Profesores de Estudios de Posgrado e Investigación de la SEPI ESE-IPN en su sesión Ordinaria No. CCCXII celebrada el día 6 del mes de septiembre del año 2019 conoció la solicitud presentada por el(la) alumno(a):

Parada

Apellido paterno

Rojas

Apellido materno

Ana Cecilia

Nombre (s)

Con registro:

A	1	6	0	6	3	4
---	---	---	---	---	---	---

Aspirante de: Maestría en Ciencias Económicas

1.- Se designa a la aspirante el tema de tesis titulado:

Globalización y desigualdad, enfoque multidimensional con re-muestreo dirigido a redes neuronales artificiales y árboles clasificadores

De manera general el tema abarcará los siguientes aspectos:

Desigualdad y globalización. Descripción, análisis y manejo de la información. Proceso de modelados bootstrapping. Análisis de resultados

2.- Se designan como Directores de Tesis a los Profesores:

Dr. Humberto Ríos Bolívar

3.- El trabajo de investigación base para el desarrollo de la tesis será elaborado por la alumna en: SEPI ESE-IPN

que cuenta con los recursos e infraestructura necesarios.

4.- El interesado deberá asistir a los seminarios desarrollados en el área de adscripción del trabajo desde la fecha en que se suscribe la presente hasta la aceptación de la tesis por la Comisión Revisora correspondiente:

Directores de Tesis

Dr. Humberto Ríos Bolívar

Aspirante

M. en C. Ana Cecilia Parada Rojas

Presidente del Colegio

M. en C. Godfrey Orozco



S.E.P.
INSTITUTO POLITÉCNICO NACIONAL
E.S.E.
SECCION DE ESTUDIOS DE
POSGRADO E INVESTIGACION



COMITÉ TUTORIAL Y DE TESIS:

NOMBRE DEL PROFESOR	FUNCIÓN
Dr. Humberto Ríos Bolívar	Director de Tesis
Dr. Genaro Aguilar Gutiérrez	Asesor
Dra. Claudia Icela Martínez García	Asesor
Dr. Gerardo Ángeles Castro	Asesor
Dr. Guillermo Velázquez Valadez	Asesor
Dr. Salvador Cruz Aké	Asesor suplente

Ciudad de México, 6 de septiembre de 2019.

"LA TÉCNICA AL SERVICIO DE LA PATRIA"



M. EN C. GODFREY OROZCO LIRA
JEFE DE LA SECCIÓN DE ESTUDIO DE
POSGRADO E INVESTIGACIÓN
ESE-IPN

S.E.P.
INSTITUTO POLITÉCNICO NACIONAL
E.S.E.
SECCION DE ESTUDIOS DE
POSGRADO E INVESTIGACION

*GOL*HAC*tve

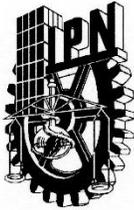


Dr. Humberto Ríos Bolívar

M. en C. Ana Cecilia Farías Rojas

S.E.P.
INSTITUTO POLITÉCNICO NACIONAL
E.S.E.
SECCION DE ESTUDIOS DE
POSGRADO E INVESTIGACION

M. en C. Godfrey Orozco Lira



INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de siendo las horas del día del mes de de se reunieron los miembros de la Comisión Revisora de la Tesis,

designada por el Colegio de Profesores de Posgrado de:

para examinar la tesis intitulada: por el (la) alumno (a):

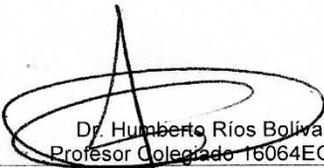
Apellido Paterno:	Parada	Apellido Materno:	Rojas	Nombre (s):	Ana Cecilia
-------------------	--------	-------------------	-------	-------------	-------------

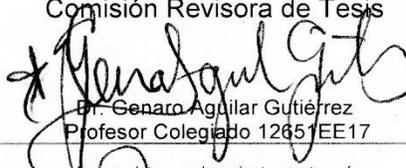
Número de registro:

Aspirante del Programa Académico de Posgrado:

Después de la lectura y revisión individual, así como el análisis e intercambio de opiniones, los miembros de la Comisión manifestaron **APROBAR** **NO APROBAR** la tesis, en virtud de los motivos siguientes:

Se concretaron satisfactoriamente las expectativas previstas con la investigación, tanto en los aspectos de fondo como de forma.


 Dr. Humberto Ríos Bolívar
 Profesor Colegiado 16064EG19
 Director de Tesis
 clave registro nombramiento y categoría

Comisión Revisora de Tesis

 Dr. Genaro Aguilar Gutiérrez
 Profesor Colegiado 12651EE17
 clave registro nombramiento y categoría


 Dra. Claudia Icela Martínez García
 Profesor de Asignatura 13602EC18
 clave registro nombramiento y categoría


 Dr. Gerardo Angeles Castro
 Profesor Colegiado 12652ED17
 clave registro nombramiento y categoría (en su caso)


 Dr. Guillermo Velázquez Valadez
 Profesor Colegiado 14065EE19
 clave registro nombramiento y categoría


 Presidente del Colegio de Profesores
 M. en C. Godfrey Orozco Lira
 Nombre



S.E.P.
 INSTITUTO POLITÉCNICO NACIONAL
 E.S.E.
 SECCION DE ESTUDIOS DE
 POSGRADO E INVESTIGACION



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México, siendo las 10:00 horas el día 25 del mes de noviembre del año 2019, el que suscribe **Ana Cecilia Parada Rojas** alumna del Programa de Doctorado en Ciencias Económicas, con número de registro **A160634**, adscrita a la **SEPI ESE-IPN**, manifiesta que es la autora intelectual del presente trabajo de Tesis bajo la dirección del **Dr. Humberto Ríos Bolívar** y cede los derechos del trabajo intitulado **GLOBALIZACIÓN Y DESIGUALDAD, ENFOQUE MULTIDIMENSIONAL CON RE-MUESTREO DIRIGIDO A REDES NEURONALES ARTIFICIALES Y ÁRBOLES CLASIFICADORES**, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director(es) del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección electrónica acecilia.parada@gmail.com Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

D. EN C. (C) ANA CECILIA PARADA ROJAS
Nombre y firma del alumno(a)

DEDICATORIA

Esta tesis es el resultado de cuatro años de trabajo, esfuerzo, desveladas y sacrificios, pero también de satisfacciones, recompensas y momentos que marcaron mi vida tanto personal como académicamente. Sin embargo, no he recorrido este trayecto sola, por lo que, dedico mi tesis de doctorado a todas aquellas personas que me han apoyado de diferentes maneras y acompañado a través de esta etapa.

Muy especialmente a mi esposo Jorge Omar Razo De Anda, quien estuvo siempre a mi lado comprendiendo, escuchando y apoyándome en los momentos más difíciles, siendo al mismo tiempo mi compañero, coautor y amigo.

Agradezco profundamente el apoyo incondicional de mi familia, de mis padres Leticia Rojas Ávila y Fernando Parada, quienes a través de sus consejos me exhortaron e impulsaron a seguir mis sueños y a no decaer inculcando en mí el coraje para perseverar y alcanzar mis metas, así como, a mis hermanos Alma, Aracely y Fernando quienes me han demostrado su cariño, fidelidad y amistad siempre que lo he necesitado.

Con especial cariño le agradezco su compromiso y orientación al Dr. Humberto Ríos Bolívar, director de la tesis quien a través de su experiencia y aportaciones ha moldeado este trabajo, así mismo, le agradezco su consejo invaluable para mi formación profesional, moral y académica

Agradezco ampliamente a la Escuela Superior de Economía del Instituto Politécnico Nacional que me ha acogido durante los últimos seis años además de fomentar e impulsar mi desarrollo académico, por lo que, buscare siempre poner en alto el nombre de esta noble institución. Así mismo, llevo con cariño y orgullo haber sido parte de esta comunidad en la que encontré grandes amistades.

Por último, agradezco al Consejo Nacional de Ciencia y Tecnología por la beca otorgada, durante mi estancia en el programa de doctorado, sin la cual no hubiera sido posible dedicar el tiempo necesario para terminar este trabajo de investigación.

ÍNDICE

ÍNDICE DE CUADROS Y FIGURAS	III
RESUMEN	V
ABSTRACT.....	VI
INTRODUCCIÓN.....	VII
CAPÍTULO I. DESIGUALDAD Y GLOBALIZACIÓN	1
1.1. Desigualdad	1
1.1.1. Determinantes de la desigualdad	2
1.1.2 Desigualdad del ingreso.....	4
1.1.3. Medidas de desigualdad.....	5
1.1.4. Coeficiente de Gini, fuentes de información y medición	8
1.2. Evolución de la desigualdad	10
1.2.1. Crecimiento generalizado de desigualdad	10
1.2.2. Redistribución del ingreso	12
1.3. Globalización como determinante de la desigualdad	13
1.3.1. Argumento clásico	14
1.3.2. Efecto de la globalización vía salarios (debate empírico)	15
1.3.3. Enfoque institucional	17
1.3.4. Papel del progreso tecnológico	18
1.3.5. Mercados financieros.....	20
1.4. Análisis agregado con un modelo de datos de panel: GINI y KOF.....	21
CAPÍTULO II. DIMENSIONES DE LA GLOBALIZACIÓN, DESCRIPCIÓN Y ANÁLISIS DE LA INFORMACIÓN.	25
2.1. El papel de la información en la investigación de fenómenos multidimensionales	25
2.1.1 Por qué divergen los resultados empíricos	26
2.3. Dimensiones de la globalización	28
2.3. Análisis estadístico y descriptivo.....	33
2.3.1 Análisis de valores nulos	36
2.4. Algoritmo de re-muestreo de factores con valores nulos (RM-CSR-VN).....	39
CAPÍTULO III. PROCESO DE MODELADO BOOTSTRAPPING.....	43

3.1. Redes neuronales artificiales	44
3.1.1. Perceptrón	46
3.1.2. Modelo perceptrón multicapa	47
3.1.3. Análisis de sensibilidad	48
3.2. Árboles de decisión clasificadores.....	49
3.2.1. Algoritmo de entrenamiento	51
3.2.2. Medidas de ajuste.....	52
3.2.3. Ventajas y desventajas	55
3.4. Proceso de modelado	56
3.4.1. Estimación y evaluación del modelo (subproceso B).....	58
3.4.2. Estabilidad	61
3.4.3. Mejor modelo.....	63
3.5. Implementación	64
3.5.1. Redes neuronales artificiales (RNAs).....	64
3.5.2. Árboles clasificadores (CART)	65
CAPÍTULO IV. ANÁLISIS DE RESULTADOS	68
4.1. Modelo de perceptrón multicapa de RNA	68
4.2. Análisis de sensibilidad de la RNA	69
4.3. Modelo árbol de decisión clasificador	79
4.4. Extracción y análisis de las reglas de conocimiento del CART	80
CONCLUSIONES	87
REFERENCIAS.....	90
APÉNDICE A. Niveles y cambios de desigualdad	97
APÉNDICE B. Índice de globalización.....	98

ÍNDICE DE CUADROS Y FIGURAS

Cuadro 1.1: Análisis estadístico del ingreso en México	6
Cuadro 2.1: Descripción de las variables macroeconómicas.....	29
Cuadro 2.2: Descripción de las variables de globalización	31
Cuadro 2.3: Niveles de ingreso.....	35
Cuadro 2.4: Número de países y observaciones por región.....	36
Cuadro 3.1: Calculo de las diferentes medidas de ajuste.....	55
Cuadro 4.1: Reglas de conocimiento que llevan a cambios de desigualdad cuando aumenta la globalización comercial	81
Figura 1.1: Curva de Lorenz	7
Figura 1.2: Desigualdad global	11
Figura 1.3: Desigualdad por regiones	12
Figura 1.4: Correlación entre el coeficiente de GINI y el índice KOF.....	22
Figura 1.5: Resultado de los cuatro modelos de panel de datos	24
Figura 2.1: Heterogeneidad de los datos en el tiempo y entre países	34
Figura 2.2: Diagrama entidad-relación de la base de datos	35
Figura 2.3: Visualización de valores nulos en la base de datos	37
Figura 2.4: Número de observaciones por número de países	38
Figura 2.5: Diagrama de flujo del algoritmo RM-CSR-VN	40
Figura 3.1: Arquitectura del perceptrón.....	46
Figura 3.2: Arquitectura de un modelo perceptrón multicapa.....	47
Figura 3.3: Ejemplo de un árbol de decisión	50
Figura 3.4: Matriz de confusión.....	54
Figura 3.5: Diagrama de flujo del proceso de modelado bootstrapping.....	57
Figura 3.6: Importancia por variables para 50 modelos RNA.....	65
Figura 3.7: Precisión vs Complejidad de los modelos CART	66
Figura 3.8: Precisión vs Complejidad de los modelos CART filtrados.....	66
Figura 4.1: Ajuste de la predicción y dispersión de errores.....	69
Figura 4.2: Efecto de variables macroeconómicas sobre el coeficiente de Gini	71
Figura 4.3: Efecto del empleo sobre la desigualdad	71
Figura 4.4: Efecto de los índices de gobernabilidad sobre el coeficiente de Gini.....	72

Figura 4.5: Efecto de otras variables macroeconómicas sobre el coeficiente de Gini.....	73
Figura 4.6: Efecto de los factores comerciales sobre la desigualdad.....	73
Figura 4.7: Efecto de las cuentas generales de la balanza de pagos	74
Figura 4.8: Efecto del mercado accionario	74
Figura 4.9: Efecto de la distribución del capital e inversión sobre la desigualdad.....	75
Figura 4.10: Efecto de las cuentas generales de la IIP sobre la desigualdad	75
Figura 4.11: Efecto del mercado accionario sobre la desigualdad	76
Figura 4.12: Análisis de sensibilidad.....	77
Figura 4.13: Importancia de las variables para las dimensiones de la globalización	78
Figura 4.14: Representación gráfica del árbol clasificador	79

RESUMEN

En este trabajo se analiza la relación entre las diferentes dimensiones de la globalización y la desigualdad en el ingreso. La cual ha aumentado en la mayoría de los países, algunos autores lo atribuyen a los procesos de globalización, sin embargo, no parece haber un consenso en la literatura. Los resultados empíricos de las diversas investigaciones dependen del enfoque teórico y de las características de los datos utilizados en el análisis, principalmente del periodo de estudio, de las variables y los países incluidos. Con el objetivo de aportar a este debate, se analiza dicho fenómeno con un enfoque de ciencia de datos, bajo la hipótesis de que los efectos dependen de las características macroeconómicas de cada país en combinación con los distintos factores de la globalización.

Para tal propósito, se utiliza el análisis de sensibilidad de las redes neuronales artificiales (RNA) el cual permite identificar los factores más importantes en la dinámica de la desigualdad; también se utiliza el análisis de las reglas de conocimiento dadas por un árbol de clasificación (CART) para analizar situaciones en las cuales aumenta o disminuye la desigualdad ante aumentos de la globalización. Estas herramientas también consideradas de minería de datos, captan relaciones no lineales desde un enfoque no paramétrico, sin embargo, suelen tener problemas de sobreajuste e inestabilidad derivados de la aleatoriedad en los parámetros iniciales. Por lo que se propone un algoritmo (RM-CSR-VN) para seleccionar el mejor modelo, a través del re-muestreo dirigido con y sin reemplazo (Bootstrapping). El re-muestreo dirigido depende de la importancia de las variables en modelos anteriores, con base en lo cual va descartando la variable que menos aporta, dada la importante pérdida de información en la base de datos analizada.

ABSTRACT

This paper analyzes the links between the different dimensions of globalization and income inequality. The inequality has increased in most countries, and some authors have attributed it to globalization processes, however, there is no accord in literature. The empirical results of the investigations depend on the theoretical approach and the data characteristics used in the analysis, mainly the study period, the variables and the countries included. In order to contribute to this debate, this phenomenon is analyzed with a data science approach, under the hypothesis that the effects depend on the macroeconomic characteristics of each country combined with the different factors of globalization.

For this purpose, the sensitivity analysis of artificial neural networks (RNA) is used, which allows identifying the most important factors in the inequality dynamics. The analysis of the knowledge rules given by the classification tree (CART) is also used to analyze situations in which inequality increases or decreases in the face of increases in globalization. These tools, also considered part of data mining, capture non-linear relationships from a non-parametric approach, however, they often have problems of over-fitting and instability, driven by randomness in the initial parameters. Therefore, an algorithm (RM-CSR-VN) is proposed to select the best model, through re-sampling directed with and without replacement (Bootstrapping). The directed re-sampling depends on the importance of the variables in previous models, based on which it discards the variable that contributes less, given the significant loss of information in the analyzed database.

INTRODUCCIÓN

El estudio de la desigualdad en el ingreso es un tema latente en la investigación económica, más allá del debate y de la discusión académica se ha vuelto una prioridad para la mayoría de los países del mundo. Aun cuando es un problema ampliamente abordado no parece haber un consenso, por el contrario, diferentes autores indican efectos adversos en distintos países. Se trata de un fenómeno complejo que surge y se transmite a partir de múltiples causas económicas y sociales, y se manifiesta mediante una connotación negativa asociada a un sesgo muy marcado en la concentración de los ingresos de las personas. En la actualidad existen estudios que advierten de un palpable aumento en la desigualdad atribuido al proceso de globalización (Harrison y Hanson, 1999) y que se potencializa con el cambio tecnológico (Bourguignon, 2017) principalmente en lo concerniente a las tecnologías de la información y comunicación (Dollar, 2004).

Los determinantes de la desigualdad han cambiado a través del tiempo debido a choques que impactan de manera importante la forma de vida de la población, como las revoluciones industriales, las guerras, el progreso tecnológico y la globalización. Los resultados de los diferentes canales dependen de sus efectos individuales combinados. La estructura de estos mecanismos sufrió un cambio fundamental entre los años ochenta y noventa (Corina, 2004), de tal manera que los aumentos de desigualdad no pueden ser explicados por causas tradicionales; a partir de lo cual un gran número de trabajos se han enfocado en explicar la relación entre la distribución del ingreso y la globalización (Nissake & Thorbecke, 2005; Mills, 2008). La tendencia hacia los tratados del libre comercio de la época derivaron en la percepción de incrementos de desigualdad en la mayoría de los países, especialmente de los más desarrollados (International Monetary Fund [IMF], 2007). Entre los trabajos empíricos que sustentan estos resultados, es decir, efectos diferentes o contrarios, se encuentran Atkinson et al. (2011), IMF (2007), Balan et al. (2015), Asteriou et al. (2014).

La relación entre la globalización y la creciente desigualdad de un país, no es clara; no parece haber un consenso, ni en el sentido de efecto ni en el de los factores involucrados. La globalización tiene distintos efectos sobre la desigualdad dependiendo de las circunstancias de cada país, como: la estructura de mercado laboral (Kuznets, 1955), la

educación de la fuerza de trabajo, el tipo de exportaciones (Dollar, 2004), además, de factores institucionales y políticas sociales (Atkinson et al., 2001; Bourguignon, 2017).

La hipótesis de este trabajo indica que, en el contexto de la economía mundial actual, parte de los cambios de desigualdad, padecidos en los últimos años en la mayoría de los países, se deben a la globalización potencializada por el progreso tecnológico, pero la dirección del efecto depende de las características macroeconómicas e institucionales de cada país en combinación con ciertos factores de la globalización, específicamente de su componente financiero, tecnológico y comercial.

Por otro lado, los distintos resultados empíricos presentados en diversas investigaciones dependen del periodo de estudio, países incluidos y de las características de los datos (Deininger & Squire, 1996; Milanovic, 2007). A pesar de que la cantidad de información ha crecido en las últimas décadas, la temporalidad, estandarización y calidad de la misma, sigue siendo un problema común en la investigación empírica, especialmente para fenómenos multidimensionales como la globalización.

Parte del debate empírico entre investigaciones sobre el efecto de un fenómeno multidimensional sobre la desigualdad, como la globalización, parte de los datos con los que se realiza el análisis. A pesar de la creciente cantidad de información que se genera actualmente, las cifras oficiales publicadas por instituciones internacionales no son suficientes para sustentar generalizaciones globales; la temporalidad, estandarización y calidad de la misma, sigue siendo un problema común en la investigación empírica, especialmente para fenómenos multidimensionales como es la globalización y la desigualdad.

El objetivo general de la tesis es analizar la relación entre globalización y desigualdad del ingreso, desde una perspectiva multidimensional, considerando tanto factores macroeconómicos como las diferentes dimensiones de la globalización, específicamente comerciales, financieras y tecnológicas, a través de modelos no paramétricos de minería de datos, definidos por un proceso de modelado para reducir la variabilidad de las relaciones funcionales, además de minimizar la pérdida de información por valores nulos.

El análisis incluye una descripción y conceptualización de los fenómenos involucrados, así como una revisión de la literatura sobre el tema. Asimismo la estimación de un modelo de panel de datos entre los índices Gini y Kof que miden la desigualdad y globalización respectivamente, incluyendo más de 140 países con 3800 datos, sin embargo, los resultados del modelo sugieren un análisis desagregado multivalente. Considerando este enfoque multidimensional, se propone una serie de variables a incluir en la investigación con base en algún argumento teórico que relaciona ambos fenómenos. Después del análisis estadístico y descriptivo de la información, se observa una pérdida potencial de datos si se eliminan los registros con valores nulos, a partir de lo cual, se propone un algoritmo iterativo para discriminar factores poco importantes o que no tengan suficiente información para aportar a la investigación.

Para identificar los factores más importantes, se realiza un análisis de sensibilidad sobre un modelo de redes neuronales artificiales (Perceptrón multicapa) con alto grado de eficiencia. Sin embargo, los pesos estimados de una red dependen de los valores asignados como condiciones iniciales los cuales se toman aleatoriamente, esto genera ciertas diferencias en los resultados de análisis de sensibilidad (De Oña & Garrido, 2014), considerando este problema de inestabilidad que también se presenta en los modelos CART (ya que, los criterios de decisión pueden cambiar con base en los parámetros y conjunto de datos de entrada), se propone un método de remuestreo dirigido que incluye un proceso de Bootstrapping o muestreo aleatorio del 75 % de los datos para construir una cantidad suficiente de modelos, de entre los que se selecciona el más confiable considerando su eficiencia y estabilidad.

La red que presenta una mejor estimación es un MLP con un capa oculta de 6 neuronas., para este modelo incluye efectos no lineales considerando más de 30 factores con diferentes grados de importancia. Con base en el cual se extrae información relevante sobre la magnitud y dirección del efecto que los factores más importantes en la dinámica entre la globalización y el contexto macroeconómico y su impacto en el coeficiente de Gini. Dado los resultados del modelo perceptrón, que enfatizan una importante proporción del efecto sobre la desigualdad derivado de factores macroeconómicos internos, se analizan las

relaciones como una combinación de factores a través del análisis de las reglas y criterios específicos en un modelo CART.

Con el objetivo de enfatizar los efectos derivados de la integración económica, se incluyen en este análisis solo las observaciones en las que se registra un aumento de la globalización, medida a través de un índice de globalización comercial construido con base en la distancia de Mahalanobis.

La construcción y discriminación de árboles clasificadores -técnica de minería de datos, flexible y no paramétrica- consiste en encontrar el mejor modelo para analizar las relaciones entre las variables y como se combinan para que una situación se presente, en este caso, cambios en la distribución del ingreso. Esta herramienta de clasificación permite entender el fenómeno de estudio, ya que, aporta información a través de la extracción y análisis de reglas de conocimiento (Faraway, 2016).

Este trabajo contribuye a la comprensión del fenómeno de la desigualdad en el ingreso al proporcionar primero una delimitación de los efectos derivados de una mayor interacción comercial. Por una parte identifica los factores más importantes en la determinación de la desigualdad, por otro lado se enfoca en circunstancias (combinación de factores) en momentos de mayor globalización.

Con base en el análisis de sensibilidad de la red neuronal, se recomienda el fortalecimiento del sector industrial, los índices de gobernanza y el impulso del progreso tecnológico a través de créditos o apoyo a la investigación y desarrollo. Se enfatiza el papel de las características individuales de los países en el efecto de la globalización sobre la desigualdad.

En el análisis de las reglas proporcionadas por el modelo CART, se identifica a la estructura del mercado de trabajo como característica determinante. Discriminando países agrícolas pero con potencial de urbanización como los beneficiarios del comercio, mientras que los países con empleo en el sector servicios aumentan su desigualdad en periodos de estancamiento o cuando no exportan alta tecnología.

En el primer capítulo se definen ambos fenómenos de estudio, sus características y como se relacionan entre ellos, también se presenta una revisión de la literatura que aborda

el debate entre las teorías que relacionan a la globalización con la desigualdad, además de un análisis econométrico entre el coeficiente de GINI y el índice KOF.

En el segundo capítulo, se considera un enfoque desagregado de la globalización en sus componentes comercial, tecnológico y financiero, a partir de lo cual se propone una lista de variables como factores de las diferentes dimensiones. Entonces, se presenta el problema de la estandarización, temporalidad y ausencia de datos, para finalizar con la descripción y especificación del algoritmo propuesto, con el objetivo de reducir la pérdida de información (RM-CSR-VN).

En el tercer capítulo, se describen los modelos propuestos: Redes neuronales artificiales y árboles clasificadores, incluyendo las metodologías que permiten analizar sus relaciones funcionales (análisis de sensibilidad, perturbación y extracción de reglas de conocimiento), así como, sus ventajas y desventajas. Se describe e implementa el proceso de modelado para cada técnica, enfatizando los criterios de selección o discriminación de modelos según sus medidas de ajuste.

Por último, en el cuarto capítulo se presenta la red neuronal perceptrón multicapa resultante del proceso de modelado, y se analizan las relaciones funcionales a través de un análisis de sensibilidad y perturbación, para identificar los factores más importantes en la dinámica de la desigualdad. También se presenta el árbol clasificador más estable, a partir del cual se extraen y analizan las reglas de conocimiento que relacionan las variables cuando se perciben aumentos de globalización. Para terminar con un análisis de los resultados y conclusiones generales de la tesis.

CAPÍTULO I. DESIGUALDAD Y GLOBALIZACIÓN

En este capítulo se presenta una revisión de la literatura que aborda el debate entre las teorías que relacionan ambos fenómenos. Dentro del contexto de la globalización se analiza el fenómeno de la desigualdad en el ingreso, para lo cual, es importante definir ambos conceptos, sus características y como se relacionan entre ellos. En este capítulo se responden las siguientes preguntas ¿Desigualdad con respecto a qué y entre quienes?, ¿Cómo la medimos?, ¿Qué fuente de datos utilizar para obtener información confiable y comparable? ¿La desigualdad ha aumentado en la mayoría de los países? De ser así, ¿cuáles son sus determinantes? y ¿Cuál es su relación con la globalización? considerando un enfoque desagregado de la globalización en sus componentes comercial, tecnológico y financiero, además de un análisis econométrico agregado.

1.1. Desigualdad

En las últimas décadas, el problema de la desigualdad ha venido tomando importancia en la investigación económica, ya que viene acompañada de una serie de problemas sociales, culturales y económicos. Desde el punto de vista económico podemos encontrar desigualdad de oportunidades, de salarios, de riqueza, e incluso de habilidades, entre grupos u agentes económicos, y cada una de ellas acompañada de sus respectivos problemas. Si bien, la igualdad de oportunidades de estudio, por poner un ejemplo, es importantes para una mejor sociedad, la distribución del ingreso puede reducir otros tipos de problemas sociales, además de que es considerada un punto de referencia aceptable para medir el bienestar, donde el ingreso puede verse como la cantidad de bienes que puede consumir una persona u hogar en cierto periodo.

Los constantes incrementos de éste tipo de desigualdad conducen a problemas económicos, sociales y políticos, así como, a “procesos de toma de decisiones disfuncionales” (Stiglitz, 2015). Este fenómeno es claramente un problema moral e incluso es una preocupación social para los más ricos cuando las diferencias se vuelven insostenibles y extremas, situación donde la mitad de la riqueza mundial pertenece solo al 1% de la población. También genera fragilidad financiera; cuanto más ricas son las personas menos adversos al riesgo se vuelven, por lo que, tienden a invertir en activos más riesgosos. Lo anterior es una de las razones que llevaron a la economía estadounidense a la crisis

inmobiliaria del 2008 (Stockhammer, 2015), más aún su lenta recuperación se debe a sus altos niveles (Cynamon & Fazzari, 2015). Por otra parte, y de acuerdo con el argumento poskeynesiano, el consumo depende del ingreso, de tal manera que la desigualdad reduce la demanda agregada debido a que la propensión marginal de consumir de las personas con menos ingresos es mayor que la propensión de los más ricos.

1.1.1. Determinantes de la desigualdad

Desde la época de Ricardo y Marx, el problema de la distribución del ingreso era tema latente, se preguntaban: ¿por qué unos se volvían cada vez más ricos mientras que los demás se volvían cada vez más pobres? En 1817 Ricardo argumenta que los terratenientes se hacían cada vez más ricos debido a la escasez de la tierra; Marx por su parte puntualiza en 1848 que eran los capitalistas los que terminarían por adueñarse de una porción cada vez más grande del capital debido a las plusvalías generadas por la productividad del capital, lo cual llevaría inevitablemente a descontentos sociales. Situación que sucedió en ciertos lugares, pero que no se generalizó gracias al progreso tecnológico (Piketty, 2014).

En 1950 Kuznets identifica una relación entre el progreso y la desigualdad, explicada por la existencia de al menos dos fuerzas que incrementan la brecha de distribución del ingreso (cambios a largo plazo entre clases) debido al desarrollo de los países (producto per cápita). (1) La concentración de ahorro de los más ricos, ya que, permite recibir rendimientos y acumular capital para sus siguientes generaciones, y (2) la proporción de población rural e industrial, debido a que la población rural obtiene menores salarios que las personas que viven en zonas urbanas -conocido como fenómeno de urbanización. La concentración de ahorro es afectada principalmente por políticas fiscales, fenómenos demográficos, emprendimiento de nuevas industrias (libertad de oportunidades) y cambios en la porción de remuneración del trabajo por servicios. Por su parte, la movilidad del sector agrícola hacia el sector industrial fue reconocida como proceso de crecimiento económico. Cuando las ciudades se industrializan el ingreso de las personas que trabajaban en éste sector es mayor que en el sector agrícola, por lo que la desigualdad aumenta, pero conforme se fue dando la movilidad esta disminuyó. Entonces el cambio estructural se caracterizó por un crecimiento económico debido al proceso de industrialización pero seguido de un proceso de desarrollo que conjunta mecanismos de reducción de la población, políticas fiscales y sociales. A partir

de lo cual, se establece la hipótesis de Kuznets que indica que la desigualdad es inherente al crecimiento en la primera etapa pero una transferencia de bienestar en la segunda etapa, de tal manera que la relación entre la distribución del ingreso y crecimiento económico tiene forma de U invertida (Kuznets, 1955). A partir de entonces, esta relación se generalizó y la investigación y políticas de gobierno se enfocaron en el crecimiento económico, confiando en que después de un proceso de expansión económica vendría un proceso de redistribución de la riqueza.

En este contexto surge la hipótesis de Kaldor (1957) que establece que la propensión marginal a ahorrar de los ricos es mayor que la de los pobres, lo que implica mayor desigualdad en el ingreso al principio pero después se verá reflejado como mayor ahorro agregado, mayor inversión en capital y por consiguiente mayor crecimiento económico. Éste efecto potenciador del crecimiento está basado en los incentivos y el hecho de que la inversión es indivisible (Nissanke & Thorbecke, 2006: 1342p), propiciando el crecimiento.

La teoría convencional del bienestar social sostiene que incrementar el producto per cápita de un país mejorará el ingreso de toda la población incluyendo a los más pobres. Estos argumentos clásicos exponen a la desigualdad como un determinante del crecimiento (Deininger & Squire, 1996). Investigadores aseguran que “mientras la proporción de ingreso que reciben los pobres no cambie los pobres se benefician del crecimiento. Por lo que el comercio internacional es bueno para el crecimiento y el crecimiento es bueno para la pobreza” (Dollar y Kraay, 2001), entonces al menos a largo plazo, la apertura comercial promueve el crecimiento económico y capital humano (Shahbaz, 2012). Estos argumentos implican que la distribución del ingreso no debe sufrir ningún cambio ante un aumento del ingreso nacional, de tal manera, que si el producto aumenta en un 10% el ingreso de todos y cada uno de los individuos debe aumentar en exactamente la misma proporción. Lo que dependerá de la naturaleza del progreso que haya propiciado el crecimiento, sin embargo, es una situación poco probable. Aun suponiendo que tal condición se cumpla, solo sería aceptable socialmente si la brecha fuera pequeña.

Desde otro punto de vista, una teoría de economía política que sostiene, por el contrario, que la desigualdad contrae el crecimiento a través de mecanismos más sociales. Por ejemplo, cuando los ricos buscan actividades no productivas con mayores rendimientos

y que atentan contra los derechos de propiedad se generan monopolios; o cuando la inestabilidad política y social genera incertidumbre se desincentiva la inversión; y cuando la mayoría de los votantes son relativamente pobres y buscan políticas de redistribución del ingreso a través de impuestos se provocan distorsiones económicas (Thorbecke and Charumilind, 2002: 1480).

1.1.2 Desigualdad del ingreso

El concepto de desigualdad es un adjetivo que compara dos objetos, e indica que dichos elementos no son iguales en cuanto a una característica. Desde el punto de vista económico, su estudio puede referirse a la comparación de diferentes propiedades, por ejemplo de: oportunidades, capacidades, educación, salud, riqueza, salarios, etc., ya sea, entre personas, hogares, clases sociales o países.

En este trabajo, se analiza la desigualdad en el ingreso definida por el valor monetario de lo que recibe una persona periódicamente; ya sea, por su trabajo a través de sueldos, salarios, primas y honorarios, o por su capital que incluye rentas, dividendos, intereses, regalías, todo importe derivado de riqueza acumulada como edificios, maquinaria y equipo. Hay que mencionar que no es lo mismo que riqueza, esta es todo lo que posee el individuo, es decir, el patrimonio (heredado y acumulado), el capital inmobiliario (vivienda) y capital financiero o profesional utilizado por las empresas (edificios, patentes), excluyendo el capital humano. Mientras que el ingreso es la cantidad de dinero extra que recibe un individuo, flujos salariales o rendimientos generados por inversiones en un periodo de tiempo. (Piketty, 2014)

Es fácil comparar el ingreso de dos personas y determinar si son iguales o si hay desigualdad; estrictamente hablando, si el individuo A tiene un ingreso de 5000 pesos y el B uno ingreso de 5500, podemos decir que hay desigualdad en el ingreso. Sin embargo, es más interesante comparar el ingreso de todos los individuos de un país que solo de dos individuos, y más útil es conocer que tan desiguales son estos ingresos.

Según Cowell (2011), existen tres importantes aspectos a considerar para definir un tipo de desigualdad, él los llama ingredientes del principio de medición de la desigualdad, y consisten en especificar: la unidad social (el individuo u hogar), el atributo particular de interés y el método de representación o agregación (población). Con base en estos criterios,

nuestro objeto de estudio es *la desigualdad en el ingreso entre los hogares a nivel nacional*. La unidad económica más pequeña suele ser un hogar, ya que, es común que los hogares reciban ingresos de una sola persona. Así mismo, se estudia la desigualdad entre hogares dentro de cada país, debido a que el objetivo es encontrar generalizaciones, es decir, factores clave en común o similares que determinantes de este fenómeno, para lo que es necesario analizar una gran cantidad de países.

1.1.3. Medidas de desigualdad

Uno de los principales problemas a los que se enfrenta un investigador que busca responder preguntas como: ¿Qué tan desigual es un país?, ¿Hasta qué punto es un problema? ¿Cuánto ha amentado o disminuido?, es la medición de la desigualdad del ingreso.

Para saber que tan pobre o rica es una persona u hogar en su país, tenemos que conocer qué parte del ingreso nacional del país representa su ingreso individual en comparación con lo que representa el de los demás, es decir, que parte del pastel le toca a cada quien. Si graficáramos el pastel correspondiente al producto nacional con las proporciones de todos los habitantes de un país, sería difícil distinguir cada una de las partes para realizar una comparación entre individuos. Lo más sano es agrupar los valores individuales, con base en su nivel de ingreso ordenado, por ejemplo, dividir la población entre los individuos con montos altos, medios y bajos. Estas agrupaciones entre tres, cinco, diez o cien son representaciones de los ingresos de las personas en un diagrama a partir de la función acumulada de distribución conocido como la curva de Lorenz, el cual muestra las proporciones acumuladas del ingreso nacional que gana cierto porcentaje de la población.

Por ejemplo, en México¹ la curva de Lorenz indica que más del 10% de los hogares reciben el 33% del ingreso nacional (ver figura 1.1), es decir, si el ingreso total de 10 personas fuera de 1,000 pesos esta distribución indicaría que una persona se queda con 330 pesos mientras que las otras 9 se reparten los 670 pesos restantes.

¹ Los cálculos realizados en esta sección corresponden a un análisis del ingreso promedio mensual calculado a partir de la Encuesta Nacional de Ingresos y Gasto de los Hogares (ENIGH 2016).

Con el objetivo de captar este tipo de desigualdad de manera general y comparable utilizamos el coeficiente de Gini, desarrollado por Corrado en 1921 y que desde entonces ha sido ampliamente utilizado para analizar la distribución del ingreso vista como una medida escalar. En México el coeficiente de Gini es de 0.4481 para el año 2016, lo que indica que es un país bastante desigual, esto no es de sorprender sabiendo que uno de los hombres más ricos del mundo vive en un país con un alto grado de pobreza. Este es un valor de desigualdad muy alto debido a que la diferencia entre su valor mínimo y su valor máximo es mayor a 370 mil. Analizando las estadísticas del ingreso en México, se aprecia que menos del 0.01% de la población gana más de 200 mil pesos mensuales, sin embargo, las tres quintas partes de la población reciben ingresos por debajo de los 16 mil pesos.

Cuadro 1.1: Análisis estadístico del ingreso en México

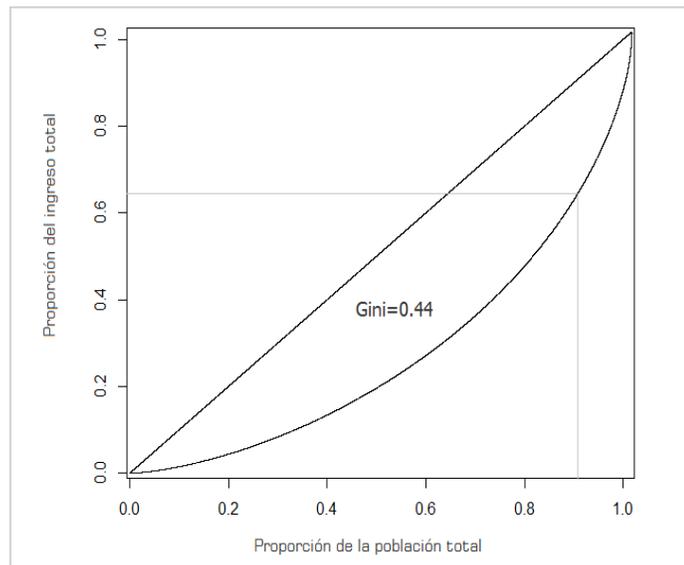
Mínimo	2do Quintil	Mediana	Media	3er Quintil	Máximo
\$ 11.41	\$ 5,852.50	\$ 9,795.10	\$ 13,818.20	\$ 16,470.00	\$ 371,842.17

Fuente: Elaboración propia con datos de la ENIGH 2016

El coeficiente de Gini mide la desigualdad en el ingreso de los hogares del país; su valor se obtiene del el área entre a la curva de Lorenz y la línea de igualdad perfecta que se muestra el figura 1. En México se determina con la siguiente fórmula $Gini = \frac{2 \sum z_i p_i l_i}{\sum p_i \sum p_i l_i} - 1$, donde p es un vector de pesos o factores de expansión (variable utilizado para estimar valores poblacionales a partir de datos muestrales), l es el vector de ingresos de una muestra de n individuos $i = \{1, 2, \dots, n\}$ (Aguilar, 2000)

Específicamente en la figura 1.1, la línea de 45° representa la situación en la cual el ingreso nacional se distribuye en cantidades iguales entre todos los hogares de la población, con una relación uno a uno representa una igualdad de ingresos perfecta. Si la curva de Lorenz fuera igual a esta línea, el área entre ellas sería cero, y por lo tanto, el coeficiente de Gini también. Por el contrario, la única forma en la que el coeficiente fuera igual a uno, sería una situación de desigualdad perfecta, donde una persona recibiera todo el ingreso (Cowell, 2011).

Figura 1.1: Curva de Lorenz



Fuente: Elaboración propia con datos de la ENIGH 2016

Existen otras medidas de desigualdad en el ingreso, sin embargo, todas estas medidas son aún más criticadas que el coeficiente de Gini. El rango R se enfoca en la diferencia entre el ingreso más alto y el más bajo, por lo que, se ocupa de los valores extremos, en nuestro ejemplo sería de 370 mil. Una medida menos escandalosa es la relación con respecto a su media y varianza, pero si se duplica el ingreso de todos los individuos esta medida se cuadruplica, cuando la desigualdad en realidad permanece igual, aunque la varianza logarítmica soluciona este problema.

El coeficiente Dalton, por otra parte, se basa en el principio de transferencia donde la distribución del ingreso mejora si se transfiere una cantidad z de ingreso del individuo A al B cumpliendo la siguiente condición $I_B \leq I_A - z$. Dalton analiza los cambios en la distribución con base en una función de bienestar agregada, donde la sociedad prefiere distribuciones más igualitarias, de tal manera que la sociedad está dispuesta a renunciar a una proporción de la producción a cambio de reducir la desigualdad. Atkinson incluye el coeficiente de aversión a la desigualdad α , en esta la función de bienestar social $U = \frac{I^{1-\alpha}}{1-\alpha}$ (Atkinson, 1970). Este método se utiliza principalmente para medir los cambios en la distribución ante políticas sociales (Stiglitz, 2015:144p)

Una relación más intuitiva son las tablas sociales, que permiten un análisis enfocado en el último decil de la distribución del ingreso, correspondiente a la población más rica conocido como el “Top Income”. Se trata de dividir la distribución acumulada del ingreso en diez partes, de tal manera que el primer nivel de ingreso -el más bajo- representa la parte del ingreso nacional (pastel) que pertenece al 10 por ciento de población. Estas medidas son utilizadas para analizar un tipo de clase económica en específico (Atkinson et. al., 2011)

Piketty analiza la desigualdad incluyendo la riqueza de los individuos, a través de un parámetro β que representa el ratio capital-ingreso, pues considera que existe mayor desigualdad en el capital que en el ingreso (Piketty, 2014: 67p), es importante mencionar que este trabajo se basa en datos fiscales, no disponibles en la mayoría de los países no desarrollados incluyendo a México.

La variable del ingreso que se utilice para su cálculo debe ser medible y comparable con el de otras personas u hogares. Sin embargo, estas dos características no son tan fáciles de asegurar y menos a nivel agregado. Aunque podemos decir que, tomamos el ingreso en la medida monetaria de cada país y lo hacemos comparable a través de su conversión a una moneda de referencia, debemos estar conscientes de la forma en la cual cada institución estadística recauda y presenta los datos originales.

Siendo el objeto de estudio la desigualdad entre los ingresos de los hogares de la población de un país, y su finalidad encontrar generalidades a partir de los datos de la mayoría de los países, se utiliza el coeficiente de Gini, por ser una medida más sensible ante cambios en los ingresos extremos (clase alta y baja) que en los medios (clase media) y tomar valores entre cero y uno, donde 0 es la igualdad perfecta y 1 es la desigualdad total. Aun cuando la distribución del ingreso es considerada como un fenómeno multidimensional, para efectos de esta investigación, en donde parte del objetivo es saber si la desigualdad ha aumentado o disminuido, es indispensable medirla con un solo valor numérico; un índice que pueda compararse a través del tiempo y, hasta cierto punto, entre países.

1.1.4. Coeficiente de Gini, fuentes de información y medición

La confiabilidad de las investigaciones sobre la evolución de la desigualdad, así como la comparación de resultados entre dichos trabajos, dependen de los datos utilizados para la

estimación del coeficiente de Gini, es importante especificar el periodo, la cantidad y fuente de la información, así como la población que se considere en el cálculo (Fisher, Johnson y Smeeding, 2013: 184). En la literatura existen una gran cantidad de trabajos que utiliza datos de panel transversal de varios países, sin embargo, no hay seguridad de que los datos sean comparables, debido a las diferentes metodologías utilizadas para medir o calcular la variable. Estas diferencias dependen de las circunstancias de cada país, por ejemplo, las instituciones de estadísticas oficiales son más transparentes en ciertas economías avanzadas. Por lo que, “la ausencia de datos en la distribución de la riqueza para una cantidad suficiente de países hace que los investigadores realicen estudios empíricos a proxis.” (Aghion et al., 1999). Atkinson et al. (2001) clasifican las bases de datos en primarias (recolectada por organizaciones con procesos estandarizados como las instituciones de los miembros de la OECD), secundarias (construidas a partir de varias fuentes pero con cierto grado de compatibilidad), y de los usuarios (datos que el investigador recolecta). Advierte sobre el uso de bases de datos secundarias, indicando que no hay seguridad de que los datos sean comparables, debido a las diferentes metodologías utilizadas para medir o calcular la variable; entre las dificultades que conlleva medir los ingresos de una persona, se encuentran la consideración de pensiones, flujos por trabajos informales y el llamado ingreso en especie (autoconsumo de alimentos cosechados); además no es lo mismo comparar el bienestar a través del ingreso de una familia con tres integrantes con el de un hogar de seis integrantes (Cowell, 2011).

El coeficiente de Gini es un ejemplo típico de este problema, a pesar de que es una medida agregada comparable y relativamente aceptable muy común para medir la desigualdad, no hay un consenso con respecto a su definición. Por ejemplo, el coeficiente de un país podría ser calculado con base en el ingreso antes de impuesto o considerando el segmento de la población desempleada (con ingresos de 0), afectando el valor de esta medida.

Una buena opción sería trabajar con datos originales, obtenidos directamente de las encuestas o bases de datos fiscales, el problema de utilizar datos originales, aparte de que también son datos sesgados, es el acceso y manejo de una cantidad tan grande de información para obtener un solo coeficiente y aun cuando podría solucionarse con técnicas de minería

de datos, las observaciones sería insuficientes para el análisis, por lo que optaríamos por interpolar y estandarizar una cantidad considerable de datos. Por ejemplo, en México la ENIHG 2016 proporciona más de 65,500 registros de ingresos de los cuales solo obtenemos un dato: el Gini para un país en un año, estos datos no se publican anualmente de hecho solo se han realizado encuestas para 15 años desde 1982.

Conscientes de este problema, se opta por el uso de una base de datos secundaria SWIID por sus siglas en inglés “The Standardized World Income Inequality Database” propuesto por Frederick Solt, en el cual se estandarizan los coeficientes GINI estimados por las principales instituciones globales², a través de algoritmos de imputación múltiple de datos perdidos lo cual permite la comparación de los valores considerando la incertidumbre con cien imputaciones (Solt, 2019), la versión utilizada SWIID 6.0 cuenta con más de 5 mil observaciones entre 1960 y 2016 para poco más de 180 países, además de ser un recurso utilizado en diversas investigaciones sobre el tema (Palma & Stiglitz, 2016; Heathcote et. al., 2017; Jeon & Kabukcuoglu, 2018; Rhee & Kim, 2018).

1.2. Evolución de la desigualdad

1.2.1. Crecimiento generalizado de desigualdad

La desigualdad promedio entre los ingresos de los hogares de un país, creció a partir de 1980. El reporte del Fondo Monetario Internacional (FMI) del 2007 y 2009, así como, Atkinson, Piketty, & Saez, (2011), entre otros, mencionan que en las últimas décadas aumentó en la mayoría de los países incluyendo los desarrollados (Harrison & Hanson, 1999). Estos incrementos constantes durante los 80s en varios países se atribuyen a los procesos de globalización que se profundizaron en dicho periodo. En la figura 1.2, se muestra el coeficiente de Gini³ promedio de 30 países desde 1980 hasta el año 2012. Lo que indica que la globalización no solo dejó ganadores y perdedores entre países sino también entre los

² Datos estandarizados y obtenidos de “United Nations University's World Income Inequality Database version, OECD Income Distribution Database, the Socio-Economic Database for Latin America and the Caribbean generated by CEDLAS and the World Bank, Eurostat, the World Bank's PovcalNet, the UN Economic Commission for Latin America and the Caribbean, the World Top Incomes Database, the University of Texas Inequality Project, National statistical around the world, and many other sources. Luxembourg Income Study data serves as the standard” (Solt, 2019)

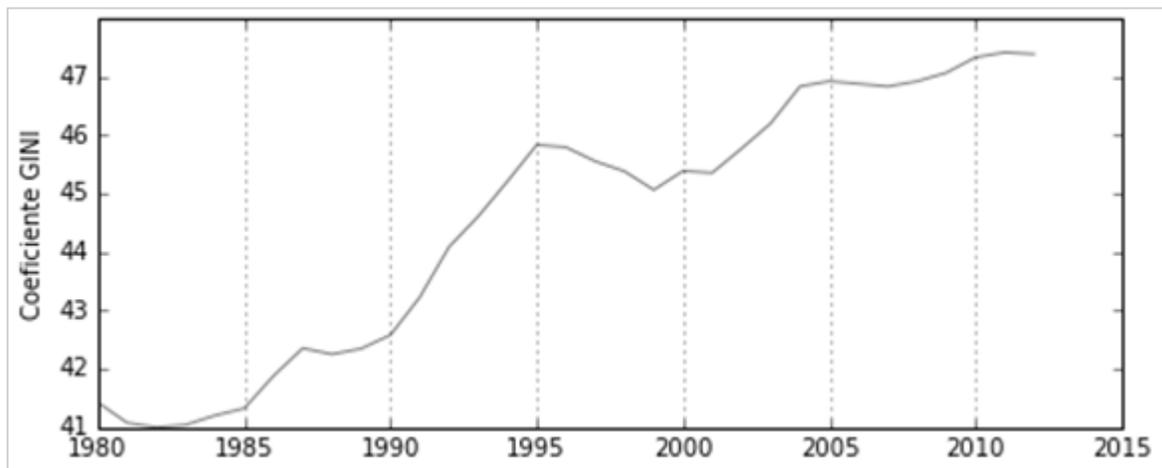
³ Este coeficiente se obtiene del valor calculado en promedio de todos países con datos obtenidos de la base de datos completos en la SWIID, es decir con GINIs para todos los años entre 1980-2012.

hogares de estas naciones, en palabras de Noam Choms “La desigualdad aumenta en periodos de globalización, entre y dentro de los países” (Sala-i-Martin, 1992).

El efusivo interés, que se ha dado desde los años noventa, por estudiar los cambios en la distribución del ingreso tiene justificación en los datos que muestran drásticos cambios en desigualdad global a partir de los años ochenta, como se muestra en la figura 1.3; también se puede apreciar que a partir del 2012 estamos en el punto más alto en promedio, aunque los aumentos en los últimos años son más pequeños.

Tampoco es posible generalizar este fenómeno clasificando a las economías por su nivel de desarrollo o región, debido a la heterogeneidad de las mismas. Aun, a nivel regional, se distingue como los cambios en el coeficiente de Gini varían, por ejemplo, en América latina nunca es menor a 0.40 siendo la región con mayor desigualdad pero también ha mostrado decrementos -a partir de 2005- que las demás regiones no.

Figura 1.2: Desigualdad global



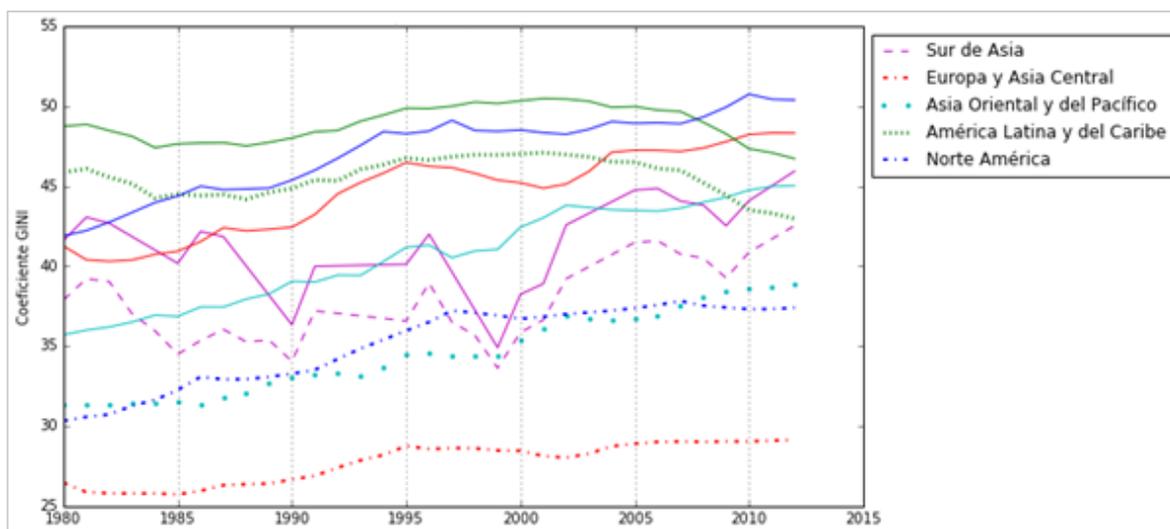
Fuente: Elaboración propia con datos de la SWIID

Desde 1970 el coeficiente de Gini ha aumentado considerablemente en los Estados Unidos e Inglaterra mientras que en Alemania y Francia no, a pesar de tener niveles de desarrollo y crecimiento similares. De la misma manera, Filipinas y Corea del Sur tenían similares características en 1960 (PIB per cápita, inversión, tasa de ahorro) pero diferente coeficiente de Gini, sin embargo, el país más desigual creció más lentamente. De lo anterior, surge el debate sobre la relación entre la desigualdad y el crecimiento económico, globalización, progreso o desindustrialización.

1.2.2. Redistribución del ingreso

Un factor importante que afecta la desigualdad es el sistema fiscal, ya que tiene como objetivo gravar los ingresos o utilidades a través de los impuestos para después redistribuir lo recaudado, buscando así mejorar el bienestar general de la población. Sin embargo, cuando el sistema permite que los más ricos no paguen el mismo porcentaje de su ingreso que los demás, la política de fiscalización podría tener un efecto contrario (Atkinson, 2015).

Figura 1.3: Desigualdad por regiones



Las líneas continuas indican desigualdad a partir de ingreso antes de impuestos

Fuente: Elaboración propia con datos de SWIID

El comportamiento del Gini por regiones⁴ se muestra en la figura 1.3 donde las líneas discontinuas representan la desigualdad en el ingreso disponible, es decir, después de impuestos. La diferencia entre el coeficiente neto y bruto refleja la eficiencia de las políticas fiscales; los países europeos tiene una brecha de aproximadamente 20 puntos, Alemania por ejemplo implementa una tasa progresiva que varía de 0 a 45% sobre ingresos, EU también tiene una diferencia considerable, en cambio los países latinoamericanos y asiáticos apenas

⁴ Los países con datos completos en la base de datos SWIID utilizados en las gráficas 1 y 2 pertenecientes a América latina y de caribe son: Colombia, México, Panamá, Uruguay y Venezuela. Para Asia Oriental y del Pacífico: Australia, China, Indonesia, Republica de Korea, Nueva Zelanda y Singapur. En Europa y Asia Central son: Bélgica, Bulgaria, Dinamarca, Finlandia, Francia, Alemania, Grecia, Hungría, Irlanda, Italia, Luxemburgo, Países Bajos, Noruega, Polonia, Portugal, Suiza y El Reino Unido. De Norte América solo se considera a Estados Unidos, y del Sur de Asia a Sri Lanka

logran reducir su desigualdad en un 5 o 7%. A pesar de que, muchos países también tienen cuotas progresivas la administración del gasto público, la corrupción, la informalidad e imperio de la ley, entre otras características, influyen en la eficiencia de las políticas fiscales, es por ello, que en este trabajo se incluyen los índices de gobernanza entre las características macroeconómicas.

Atkinson argumenta que son las fuerzas tanto políticas como económicas, a través de transferencias de gobierno, lo que determina la desigualdad. Al igual que Kuznets, Piketty indica que la brecha entre capital y ahorro es más grande que entre los ingresos.

1.3. Globalización como determinante de la desigualdad

La globalización es un proceso de integración económica entre los países del mundo, económicamente la podemos separar en dos etapas, la primera de 1870 a 1914 enfocada principalmente a la integración financiera y movilidad, y la segunda considerada la etapa contemporánea que se da a partir de 1914 (Milanovic, 2016). A pesar de que durante la segunda guerra mundial se dio un importante retroceso, el fin de la guerra restableció el rumbo de la integración económica, bajo el cobijo de los Acuerdos Generales de Aranceles y Comercio (GATT) en los que se fomenta el libre comercio. Sin embargo, Dollar (2004) identifica un hito detonado por la apertura comercial de China en 1978 que contribuyó al desarrollo de las crisis originadas por la deuda externa en los países latinoamericanos, a su vez, llevó a estas economías a cambiar sus estrategia de sustitución de importaciones para fortalecer la industria interna a una estrategia orientada hacia el exterior, es decir, al incremento de exportaciones principalmente hacia economías desarrolladas. Debido a lo cual este trabajo se enfoca en la globalización contemporánea a partir de 1980, periodo en el que este fenómeno se potencializó debido a los tratados de libre comercio entre países en desarrollo y desarrollados aunado a los avances tecnológicos y de comunicación.

A continuación se analizan los cambios en la distribución del ingreso; por un lado debidos a la oferta y demanda de los factores de producción en función de su productividad (enfoque clásico), y por otro lado, debido a cambios en el entorno económico externo del país como las condiciones tecnológicas y la regulación de los mercados, sin despreciar el papel de las instituciones e implementación de políticas sociales que incluyen las reformas fiscales (Bourguignon, 2017).

1.3.1. Argumento clásico

El efecto de un mundo cada vez más globalizado sobre la desigualdad, ya sea, entre países, entre las personas o entre los hogares de un país, sigue siendo un debate entre los investigadores. A pesar de ser un tema ampliamente estudiado, no hay un consenso sobre si la globalización aumenta la desigualdad o reduce incluso la pobreza (Dollar & Kraay, 2001).

El argumento clásico para justificar el intercambio de mercancías, recae en la ventaja comparativa asociada a la asignación óptima de los factores productivos; a partir de lo cual se equilibran los precios entre las economías involucradas. Un país con una industria más eficiente tiene una ventaja comparativa que le permite exportar su producción, según la teoría del bienestar esto empeora el consumo de las personas pero mejora el bienestar de los productores. Desde la perspectiva del país importador sus consumidores se ven beneficiados al pagar un precio menor por la mercancía, pero afecta la industria local al enfrentarse a una competencia internacional con una ventaja comparativa (Ricardo, 1817). De tal forma, que esta integración industrial permite una mayor competencia, equilibra el precio de los bienes y provoca la salida del mercado de las empresas menos productivas. Las pérdidas o ganancias sociales derivadas del proceso de apertura, dependen entonces de las ventajas comparativas de los países, aquellos con mayor oferta de capital producen bienes intensivos en capital y aquellos con mayor oferta de trabajo producirán bienes intensivos en mano de obra (Heckscher & Ohlin, 1991).

El efecto de la apertura comercial sobre la economía y el bienestar, no es nuevo, las creencias de que la protección de mercado interno es necesaria ante la inminente competencia de los mercados internacionales, ya era, un argumento común cuando Stolper & Samuelson presentan su famoso teorema (SS) en 1941, asegurando que era posible “inferir inequívocamente” los efectos del comercio en la remuneración del trabajo. En un contexto de competencia perfecta y cambios en los precios relativos, la apertura comercial disminuye la desigualdad entre los salarios de los trabajadores de un país en desarrollo dotado de fuerza de trabajo no calificada debido a una mayor demanda de este tipo de mano de obra, sin embargo, sucede lo contrario en países desarrollados, donde disminuye la demanda de trabajadores no calificados y aumenta la desigualdad (Stolper & Samuelson, 1941)

Los movimientos de la demanda de trabajo, afectan los precios de bienes y de los factores productivos, incluyendo el salario por ser considerado el precio del trabajo. Con el fin de captar el efecto del teorema de SS, se suele diferenciar entre el trabajo calificado o no calificado, con base en la educación, capacitación o habilidades de la mano de obra (FitzGerald, 1996). Por lo que, es importante considerar los mecanismos que incluyen las características del mercado de trabajo, como factores de la globalización.

Los argumentos en contra de la globalización se originan desde el punto de vista de los países desarrollados e industrializados⁵, países en los que la desigualdad aumentó tras firmar tratados de libre comercio (Alderson & Nielsen, 2002), y los datos que sustentan una relación positiva entre los incrementos de desigualdad con la globalización y progreso tecnológico corresponden a aquellas economías con bases de datos más completas.

Desde la perspectiva de los países en desarrollo, la globalización es una fuente de oportunidades para el desarrollo industrial, que permite incrementar los salarios de la mano de obra no calificada, lo que a su vez, reduce la desigualdad (Dollar, 2004). Aun cuando la cantidad de información no es abundante, existe evidencia de que países latinoamericanos han percibido decrementos de desigualdad (Bourguignon, 2017).

1.3.2. Efecto de la globalización vía salarios (debate empírico)

El origen en la divergencia de resultados empíricos parte de la calidad de la información, así como de las especificaciones de las variables con las que se mide tanto la desigualdad como la globalización (Ravallion, 2003). Este problema se presenta en mayor medida en los países en desarrollo que además entraron a la globalización más rápidamente (Dollar, 2004).

Conviene mencionar que los países con mayor cantidad de información son precisamente los más desarrollados, este sesgo de información reduce la confiabilidad de los resultados como la generalización de un efecto negativo.

⁵ Dollar (2004) indica que el efecto de la globalización sobre los cambios de desigualdad parecen estar correlacionada con la nacionalidad de los intelectuales y los datos de análisis, sugiriendo que el efecto positivo o negativo depende de dónde y cómo les tocó vivirla, si en países desarrollados o no.

A partir del teorema de Stolper & Samuelson, surgen una serie de investigaciones empíricas sobre el efecto de la globalización comercial en la desigualdad salarial. Indicando que tras procesos de liberación comercial la brecha salarial aumenta entre estratos sociales (Attanasio, Goldberg, & Pavcnik (2004), Menezes-Filho, Muendler, & Ramey (2008), y también dentro de estos grupos (Akerman et al., 2013), tanto en países desarrollados como en países en desarrollo Goldberg & Pavcnik (2007).

En México, por ejemplo, el argumento de Heckscher & Ohlin no se cumple en la década de los ochentas, cuando la desigualdad salarial aumentó a pesar de tener una ventaja comparativa por producción de bienes intensivos en mano de obra no calificada (Harrison & Hanson, 1999), sin embargo, disminuyó en la década de los años 90 consistente con la discriminación entre países desarrollados y en vías de desarrollo atribuido a su ventaja salarial en el sector industrial; hay que mencionar que la apertura comercial aumentó el nivel salarial pero redujo la prima salarial de la mano de obra calificada (Chiquiar, 2008).

Según Helpman et al. (2010) la apertura comercial incrementa la brecha salarial, pero podría tener distintos efectos sobre el desempleo, debido a la complementariedad de las habilidades de los trabajadores y la productividad de las empresas.

Entonces, las empresas tienen incentivos para seleccionar a los trabajadores más capaces, además, con la apertura comercial desaparecen las empresas menos productivas, aumentando la prima por capacidades, esto genera una reacción en el mercado de trabajo y un aumento de desigualdad en los salarios. Pero una vez abierta la economía, el efecto sobre la desigualdad salarial depende del nivel de globalización comercial, de tal manera que en las primeras etapas aumenta para después disminuir. En esta misma línea Han et al. (2012) tampoco encuentra evidencia contundente, analizan el efecto de las políticas de apertura de un país desarrollado y uno en desarrollo desde una perspectiva microeconómica, y encuentran un efecto pequeño sobre el empleo, argumentando que los ajustes en el mercado laboral no se deben a la apertura comercial.

Aunque existe evidencia, no es consistente con este enfoque en países latinoamericanos, donde la apertura comercial no redujo la brecha salarial (Rodrik, 1997). Ante esta crítica Wood & Ridao-Cano (1999) trata de explicar que esto podría ser debido que la abundante oferta de mano de obra no calificada en los países asiáticos, o porque las nuevas

tecnologías hacen que se demande más fuerza de trabajo calificada. Por otro lado, Mah encuentra una relación significativa entre apertura comercial y el coeficiente de Gini, al igual que Nielsen quien argumenta que la inversión multinacional incrementa la desigualdad (Chen & Tsai, 2012).

El ingreso por definición también considera los flujos por rendimientos de capital, y aun cuando parte de la desigualdad en el ingreso proviene de esta brecha salarial, aunado a que, la proporción del ingreso recibida por concepto de salario corresponde en su mayoría a las personas que no son dueños del capital, la relación no es tan simple en esta desigualdad, incluso podrían moverse en diferentes direcciones (Dollar, 2004), por lo que es necesario analizar distintos mecanismos que involucran otros factores.

Williamson (1997) estudia una combinación de factores en específico y el efecto sobre la distribución del ingreso como consecuencia de la globalización, países ricos en recursos con escasa fuerza de trabajo, y países pobres en recursos con abundante fuerza de trabajo. Aunque encuentra que las migraciones masivas tienen efectos más calados, la desigualdad en países ricos aumentó derivado del intercambio comercial en periodos entre guerras. Sin embargo, los países latinoamericanos y asiáticos redujeron su desigualdad durante la segunda mitad de siglo pasado.

1.3.3. Enfoque institucional

Desregular los mercados permiten que las compañías más fuertes compitan con las más débiles, de tal forma que las pequeñas empresas –por lo general originarias de los países en desarrollo o subdesarrollados- desaparezcan y las grandes empresas de los países desarrollados se vuelvan más poderosas, por ende la desigualdad entre países aumente, pero incluso podría aumentar la desigualdad en el país desarrollado. “Las grandes beneficiarias de la globalización han sido las empresas internacionales, y no el trabajador promedio quienes han visto cómo sus ingresos bajan año tras año” (Stiglitz, 2015:233)

Gereffi (2008) enfatiza que la forma en la que las dos economías en desarrollo China y México entraron a la globalización comercial, definieron su resultado sobre crecimiento económico; México abrió sus mercados reduciendo las barreras arancelarias afectando así la industria interna; China por el contrario ha protegido sus industrias y entrando gradualmente

a la globalización (Dollar, 2004) convirtiéndose en una de las economías con mayor crecimiento en las últimas décadas.

Stiglitz (2015) indica que la mayoría de las condiciones bajo las cuales el libre mercado funciona no se cumplen en la realidad, pues la existencia de asimetrías de información y los mercados incompletos distorsionan la eficiencia. Por otra parte, este enfoque se refiere a una asignación de recursos en el sentido de Pareto, donde no es posible mejorar el bienestar de un agente económico sin afectar a otro. Como consecuencia la desigualdad depende de las dotaciones iniciales las cuales no tienen que ver con equidad.

La eficiencia de las instituciones gubernamentales puede incluso revertir el efecto de las medidas de globalización, y explicar los resultados adversos de la apertura comercial en países en desarrollo. Atkinson (2015) enfatiza el papel de los índices de gobernanza y las medidas proteccionistas como factores clave en la determinación de la desigualdad. Debido a las políticas internas y estructuras sociales los beneficios de la globalización no llegan a los pobres, lo que aumenta la desigualdad. Akerman et al. (2013) encuentran aumentos de desigualdad salarial en países sin instituciones que protejan el mercado de trabajo, resaltando la variación entre sectores y ocupaciones.

Según Krugman, las políticas muchas veces no funcionan o tienen efectos adversos debido a la ineficacia del estado para implementarlas. Los países con altos niveles de corrupción difícilmente pueden redistribuir los recursos. Aunque la globalización también puede inducir cambios en el ambiente institucional, estos pueden ser muy lentos (North, 1990).

1.3.4. Papel del progreso tecnológico

El hito que cambió la globalización a partir de 1980 no solo se dio en el nivel de integración económica, sino también en la manera en la que se distribuyeron los mercados y el tipo de exportaciones de países en desarrollo hacia economías desarrolladas; “los reproductores de CD de China, los refrigeradores de México y los programas informáticos de Tailandia” (Dollar, 2004:150p). Este cambio se desarrolló a partir de los avances tecnológicos de producción y comunicación.

Las innovaciones tecnológicas generan nuevas industrias y aumentan la productividad del capital ya existente, a su vez la globalización potencializa su efecto abriendo los mercados (Jaumotte et al., 2013). La globalización acelera la privatización especialmente de las investigaciones que generan tecnología, incrementando la desigualdad entre quien puede comprar o acceder a estos progresos y quien no, esto es aplicable tanto a nivel personal como entre países e industrias.

Lawrence y Slaughter (1993) indican que existe poca evidencia de que la globalización beneficie a los trabajadores debido a la teoría clásica; encuentra que la desigualdad viene mayormente por el cambio tecnológico. Por otro lado, el auge de mercados tecnológicos afecta la demanda de trabajo, pues las empresas dedicadas a las innovaciones requieren empleados con un mayor grado de educación, e incluso ciertas habilidades específicas como la creatividad, la habilidad para solucionar problemas, otros lenguajes, etc. En consecuencia, la demanda de trabajo especializado aumenta los salarios de la fuerza laboral capacitada, pero también el desempleo en ciertas actividades (destrucción creativa), aumentando la brecha salarial.

De acuerdo con los resultados de Asteriou et al. (2014) la desigualdad disminuye en los países europeos que exportan alta tecnología, específicamente maquinaria para procesos industriales generada con alto grado de investigación y desarrollo como las innovaciones aeroespaciales, computacionales, maquinaria eléctrica e instrumentos científicos. Para que un país cree inversiones de este tipo es necesaria una importante cantidad de recursos destinados a la investigación y desarrollo, recordando que en todo programa social la asignación de recursos debe ser eficiente, es decir, que se generen resultados y que para este caso se traduce en alta tecnología.

Los cambios tecnológicos nacen en los países desarrollados o industrializados debido su inversión en investigación y desarrollo. Los avances tecnológicos están enfocados en ahorrar mano de obra y sustituir capital por fuera de trabajo (no capacitada), de tal manera que se demandan cada vez menos trabajadores y los que se requieren son trabajadores capacitados. Lo anterior provoca desigualdades en los salarios tanto en países en desarrollo como desarrollados, además de generar una dependencia tecnológica hacia los países avanzados. Jaumotte et al. (2013) encuentra que los progresos tecnológicos como la

integración financiera, benefician al 20 por ciento de la población más rica del país. Aunque argumenta que los países en desarrollo también podrían beneficiarse de los progresos tecnológicos a través de la globalización; China es un claro ejemplo de que es posible desarrollar tecnología con base en los avances tecnológicos de otros países.

1.3.5. Mercados financieros

Los capitales se mueven hacia países en desarrollo en periodos de crisis, lo cual se traduce en mayor desigualdad para los países en desarrollo y menor para las economías desarrolladas (Milanovic, 2002). Si esto es cierto, los países desarrollados captan mayor inversión de cartera en tiempo de crisis, y se vuelve más difícil la inversión en países en desarrollo. Entonces, una mayor integración económica hace a los países más dependientes entre ellos, aumentando la volatilidad y fragilidad financiera de las economías. Existe una relación sistemática entre inestabilidad económica y desigualdad económica, la primera afecta mayormente a los hogares pobres; los hace más vulnerables en situaciones de crisis, como sucedió con la burbuja inmobiliaria del 2008 en E.U., donde millones de personas perdieron sus hogares mientras que los responsables el sistema bancario era rescatados con billones de dólares de los contribuyentes (Stiglitz, 2015). Sin embargo, cuando el Fondo Monetario Internacional recomendó, en los años 80, a los países en desarrollo abrir sus mercados financieros, el objetivo era que los países con altas tasas de ahorro invirtieran en el desarrollo (principalmente en inversión física de capital) de los países emergentes con altas expectativas de crecimiento, por supuesto que a una tasa de interés lo suficientemente atractiva. Estos modelos, que sugerían aumentar la deuda pública, no lograron sus objetivos. La demanda de activos seguros aumenta cuando aumenta la volatilidad del ingreso no seguro en un ambiente de globalización financiera, fenómeno conocido como fuga de capitales Azzimonti et al. (2014). Según Easterly (2004) el problema es que estos flujos de capital no van precisamente a inversiones productivas sino a portafolios especulativos, y entonces generan inestabilidad financiera que se agrava en momentos de crisis con la fuga de capitales.

Por otro lado, uno de los aspectos claves que relacionan la desigualdad con la globalización financiera es la regularización de los mercados financieros. Desde el punto de vista neoclásico, la intervención en los mercados provoca distorsiones e ineficiencias, por lo que se recomienda la liberalización de los mercados. Según Obstfeld et al. (2005), son

precisamente los mercados financieros mundiales los menos regulados, razón por la cual sucesos como la crisis suprimió impulsó las regulaciones, así como, investigaciones que retoman las ideas de Minsky sobre las crisis financieras.

1.4. Análisis agregado con un modelo de datos de panel: GINI y KOF

En este apartado se presenta un análisis econométrico entre estos dos fenómenos económicos, la globalización y desigualdad medidos a través de índices agregados, el análisis a través de dichas variables agregadas (construidas a partir de estimaciones numéricas) permite realizar un análisis considerando una cantidad importante de países.

El análisis parte de lo general a lo específico, analizando primero la correlación, para después estimar un modelo de regresión de panel de datos. La desigualdad en el ingreso se mide de manera general y comparable utilizando el coeficiente de Gini, específicamente el valor promedio neto del Gini sintético obtenido de la base de datos SWIID. La globalización por su parte se mide utilizando el índice KOF desarrollado por Dreher et al. (2006, 2008), el cual representa el grado de interacción de un país al considerar las principales dimensiones de este fenómeno. Los tres componentes de éste índice son: la globalización económica (flujos de bienes, capital y servicios), la globalización política (difusión de las políticas gubernamentales) y la globalización social (difusión de ideas, información, imágenes y gente).

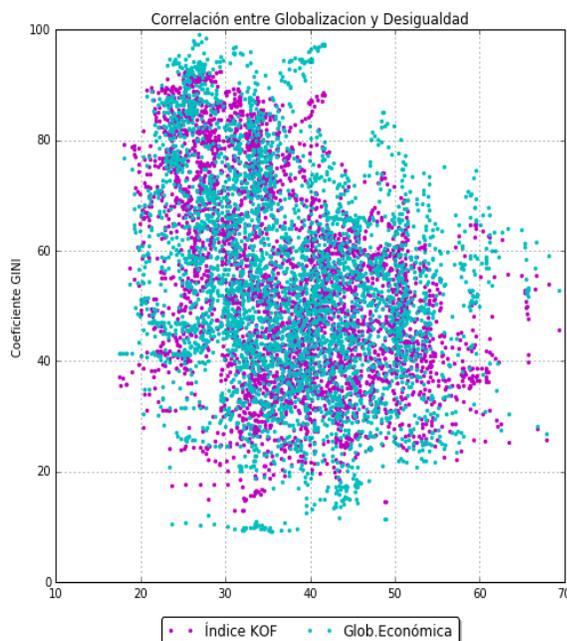
La globalización económica incluye dos aspectos, por un lado los flujos reales y por el otro las restricciones. Los flujos reales se miden con un coeficiente que capta efectos del comercio, la inversión extranjera directa y las inversiones de cartera, incluyendo variables como el pago por ingresos de capital. Las restricciones se consideran a través de un índice referente a las barreras de importación ocultas, tipos arancelarios medios, impuestos al comercio internacional y al control de capitales; entre más ingresos por aranceles, menor es el valor del índice. La globalización social incluye tres aspectos, contacto personal, flujos de información y cercanía cultural entre los individuos de diferentes países. La globalización política se representa con el número de embajadas, comisiones, organizaciones internacionales a las que el país es miembro, así como, las misiones de paz de la ONU y los tratados firmados entre dos o más naciones. Si bien la variable de interés para este trabajo es

la parte económica, no se descarta que la parte social y la política pueden tener efectos importantes.

La matriz de correlaciones entre las variables muestra una relación negativa entre el coeficiente de Gini y el índice de globalización KOF, igual que con sus tres dimensiones, lo que indica que entre más globalizado sea el mundo los países se vuelven más igualitarios. Sin embargo, esta la correlación no es muy clara, por lo que se analiza econométricamente.

Figura 1.4: Correlación entre el coeficiente de GINI y el índice KOF

	GINI	KOF	G. Económica	G. Social	G. Política
GINI	1.0000000				
KOF	-0.4327857	1.0000000			
G. Económica	-0.2941859	0.8758909	1.0000000		
G. Social	-0.4815939	0.9452250	0.8146637	1.0000000	
G. Política	-0.3031123	0.7173870	0.3802609	0.5393093	1.0000000



Fuente: Elaboración propia

Para examinar este posible efecto se estiman diferentes regresiones econométricas de panel de datos desbalanceados, con 3820 observaciones de 147 países que abarcan el periodo de 1970 a 2013. Se utiliza el logaritmo natural del coeficiente de Gini como variable dependiente, y el índice de KOF como independiente, e incluyendo algunas características categóricas de los países.

El modelo 1, representa un panel de datos de efectos fijos temporales, con cuatro variables dicotómicas para incluir el nivel de ingreso (N_{ING}) que puede ser alto, bajo, medio alto y medio bajo.

$$\ln(Gini) = \beta_t + \beta_{1t}KOF_i + \sum_{j=2}^4 (\beta_{jt}N_{ING})_{jt}$$

El modelo 2, también es un panel de datos de efectos fijos temporales, pero incluye las dimensiones de la globalización definidas por tipo: política (GP), social (GS) y económica (GE) para conocer los efectos individuales.

$$Gini = \beta_t + \beta_{1t}GS_i + \beta_{2t}GE_i + \beta_{3t}GP_i + \sum_{j=2}^4 \beta_{jt}N_{ING}_{jt}$$

Se incluyen las siete regiones clasificadas por el Banco Mundial (vea cuadro 2.3) en el modelo 3 que es un panel de datos efectos fijos entre países.

$$\ln(Gini) = \beta_i + \beta_i KOF_i + \sum_{j=2}^7 \beta_{ji} REG_{ji}$$

Se obtiene una mejor estimación con un panel de datos efectos fijos entre países modelo 4 desagregando el KOF en sus componentes.

$$\ln(Gini) = \beta_i + \beta_{1i}GS_i + \beta_{2i}GE_i + \beta_{3i}GP_i + \sum_{j=2}^7 \beta_{ji} REG_{ji}$$

En general, el aumento del índice KOF tiene una relación negativa y significativa con los aumentos la desigualdad, indicando que la globalización reduce la desigualdad. Sin embargo, el coeficiente de determinación es solo del 27%, además de ser significativo solo para países del Medio Oriente y África del Norte o en países con alto nivel de ingresos.

Cuando se analiza la desigualdad con respecto al índice KOF desagregado en sus tres dimensiones, notamos que la globalización económica tiene una relación positiva. Indicando que el componente económico aumenta la desigualdad; el efecto de la parte social tiene un coeficiente mayor y negativo, siendo este aspecto la razón por la cual en general el índice KOF reduce la desigualdad.

Figura 1.5: Resultado de los cuatro modelos de panel de datos

Variable dependiente Ln(GINI)	Modelo 1	Modelo 2		Modelo 3	Modelo 4
KOF	-0.225*** (0.008)		KOF	-0.233*** (0.008)	
G. Económica		0.004*** (0.000)	G. Económica		0.004*** (0.000)
G. Social		-0.008*** (0.000)	G. Social		-0.008*** (0.000)
G. Política		-0.000* (0.000)	G. Política		-0.000* (0.000)
África Sub-Sahara	0.055 (0.736)	-0.015 (0.019)	Ingresos Medio Bajos	-0.768 (0.514)	0.006 (0.013)
América Latina y del Caribe	-0.367 (0.735)	-0.010 (0.019)	Ingresos Medio Altos	-0.611 (0.504)	0.003 (0.013)
Norte América	1.830 (1.140)	0.045 (0.029)	Ingresos Altos	-2.037*** (0.528)	-0.041** (0.014)
Europa y Asia Central	1.491* (0.696)	0.020 (0.018)			
Medio Oriente y África del Norte	-2.878*** (0.842)	-0.073*** (0.022)			
Asia Oriental y del Pacífico	-1.786* (0.758)	-0.015 (0.020)			
R ²	0.212	0.283	R ²	0.196	0.280
Adj. R ²	0.209	0.279	Adj. R ²	0.194	0.276
Num. obs.	3820	3820	Num. obs.	3820	3820

***p < 0.001, **p < 0.01, *p < 0.05

Fuente: Elaboración propia con datos de Dreher & Solt.

De lo anterior, podemos enfatizar la importancia de la globalización de la información, la cultura y las relaciones interpersonales entre los ciudadanos de todos los países para reducir la desigualdad.

Los aumentos de desigualdad están relacionados con la globalización, sin embargo, el efecto no es claro ni contundente, con el fin de aportar claridad en este trabajo se analizan los efectos de la desigualdad considerando el entorno macroeconómico de los países. Otro punto importante, es que no podemos captar los efectos de la globalización de manera general, debemos desagregar este fenómeno en sus componentes y analizar detalladamente el caso de la globalización económica pues marca una relación positiva sobre el coeficiente de Gini pero con coeficientes de determinación muy pequeños.

CAPÍTULO II. DIMENSIONES DE LA GLOBALIZACIÓN, DESCRIPCIÓN Y ANÁLISIS DE LA INFORMACIÓN.

Existen varias y diferentes teorías económicas, e incluso estudios empíricos, que muestran una relación entre la desigualdad y la globalización, sin embargo no hay un consenso con respecto a la dirección del efecto o los mecanismos de transmisión. Por lo que, es necesario desagregar la globalización y analizarla desde tres perspectivas económicas: comercial, financiera y tecnológica (considerando a su vez que el comportamiento de cada dimensión depende de una serie de factores que las componen), además de incluir aspectos macroeconómicos de los países.

Parte del debate empírico entre investigaciones de fenómenos multidimensionales, como el efecto de la globalización sobre la desigualdad, se debe a las características de los datos con los que se realiza el análisis. A pesar de la creciente cantidad de información que se genera actualmente, las cifras oficiales publicadas por instituciones internacionales no son suficientes para sustentar generalizaciones globales una vez que los datos se han preparado. Con el objetivo de reducir esta pérdida de información, se presenta un algoritmo de remuestreo con y sin reemplazo de variables a incluir, el cual maximiza el número de observaciones sin valores nulos (RM-CSR-VN); su bondad recae en una selección de muestras a través de un enfoque dirigido al seleccionar la muestra con base en la importancia de los factores resultado de estimaciones anteriores.

2.1. El papel de la información en la investigación de fenómenos multidimensionales

El método científico tradicional consiste en proponer una hipótesis sustentada teóricamente y entonces utilizar una herramienta experimental apropiada sobre el objeto de estudio para rechazar la hipótesis, de tal forma, que al no poder rechazarla se considera una afirmación válida científicamente.

En las ciencias sociales el objeto de estudio es complejo y dinámico, aunado a lo difícil que resulta experimentar con teorías generales, ¿cómo corroborar, por ejemplo, que la globalización disminuye la pobreza? ¿Incentivando el libre comercio con unos países y con otros no?, y es que, no podemos controlar un experimento social de este tipo, debido a que el objeto de estudio es la población o la economía misma. Lo que si podemos hacer, es probar

las hipótesis analizando el pasado, las situaciones y circunstancias en las que se generaron ciertos fenómenos, a partir de lo cual predecir, generalizar o construir teorías.

La ciencia económica se ha soportado en gran medida por la econometría, una serie de herramientas de análisis de datos que combina elementos de estadística, matemáticas, física e incluso computación. Sin embargo, es común encontrar distintas teorías bien fundamentadas pero contrarias, sobre todo cuando se trata de las teorías más generales. El problema puede venir de dos fuentes: el planteamiento y las características de los datos, aunque al final ambas recaen en el conjunto de información con base en el cual se construyen los modelos.

2.1.1 Por qué divergen los resultados empíricos

Uno de los temas más investigados en economía es la relación entre la globalización y la desigualdad en el ingreso, sin embargo, los resultados de investigaciones empíricas difieren entre sí, esta falta de consenso se debe en parte a la complejidad del fenómeno y en parte a las características de los datos utilizados. Entonces por un lado, la relación no es simple porque el efecto depende de muchos factores y el mecanismo de transmisión cambia dependiendo de las circunstancias de cada país, lo que implica que algunos trabajos consideran factores que otros no; por otro lado, los países incluidos, la cantidad de observaciones, temporalidad, fuente y acceso a la información con la que se estiman y evalúan los modelos, influyen en el resultado de las investigaciones.

Los efectos económicos tienen múltiples determinantes; dependen de las circunstancias en las que se desarrollen, dadas por la combinación de ciertos factores en un momento dado. El problema que se aborda en este trabajo es identificar cual es la combinación de estos factores que permiten cambios de la desigualdad derivados de factores de la globalización. Si bien la revisión de la literatura presentada nos ofrece un conjunto de factores que podrían influir en el fenómeno de estudio, no podemos asegurar que todos sean importantes, debido a que se proponen a partir de distintos enfoques, teorías o mecanismos. Para obtener conclusiones más generales, es necesario considerar que no todas las variables resultan importantes en todos los casos o que un factor que por sí solo no parece importante resulta ser indispensable en combinación con otro. Para abordar este problema se presenta un proceso de modelado Bootstrapping que discrimina factores.

En toda investigación empírica (basada en datos) es importante especificar las características de los datos utilizados para el análisis, ya que los resultados dependen de ello. Generalizaciones teóricas sustentadas a partir de casos específicos, pueden afectar el rumbo de la investigación científica.

El trabajo de Kuznets fue uno de los primeros en utilizar datos de ingresos para apreciar los cambios en la distribución, lo que dio mayor fuerza a su argumento, y aunque solo tenía información para Estados Unidos de 1913 a 1948 (periodo de posguerra), su hipótesis se generalizó para todos los países, a partir de lo cual la investigación y políticas de gobierno se enfocaron en el crecimiento económico confiando en que después de un proceso de expansión económica vendría un proceso de redistribución. Aunque parece que se cumple en 1970 para la mayoría de los países de la OECD, no es así para países en desarrollo o sub desarrollados. Investigaciones posteriores han presentado una gran variedad de casos en los que esta hipótesis no se cumple, de tal forma que a partir de 1980 la curva de Kuznets ya no es tan clara (Piketty, 2014), dando paso a nuevas teorías (Aghion et al., 1999).

Desde el punto de vista empírico, existen una gran cantidad de trabajos que buscan dar luz al debate sobre la relación entre desigualdad y globalización, fenómenos que en sí mismos son complejos y multidimensionales.

La calidad y consistencia de los datos son consideradas entre los principales obstáculos de la investigación económica (Deininger & Squire, 1996; Atkinson et. al. 2001, Milanovic, 2007). Con respecto a la calidad de la información con la que se realizan las investigaciones, Atkinson (2003) muestra lo sesgado que pueden estar los resultados con base en la forma en la que recaban los datos. Estos problemas se reflejan en la construcción de casi cualquier modelo tanto econométricos de series de tiempo, datos panel y logísticos, como en modelos no paramétricos incluyendo las redes neuronales artificiales y los árboles de decisión.

Sin embargo, es posible obtener ciertas conclusiones; aunque difícilmente encontraremos una teoría o regla general e inequívoca para relacionar estos fenómenos si suponemos que los efectos se dan según una combinación de factores o características. Si bien los mecanismos son complejos, es posible identificar los factores que influyen en el efecto, como el nivel de desarrollo, la estructura del mercado del empleo, los niveles de

corrupción, tecnología, etc. Tanto las características de cada país como la forma en la que los países deciden implementar medidas de globalización son importantes en el efecto que estas tengan sobre la desigualdad del país. Por lo que se proponen una serie de factores que podrían influir en el mecanismo u origen por el cual esta impacta el bienestar de los hogares en los países con la mayor cantidad de datos disponibles.

2.3. Dimensiones de la globalización

La globalización es un fenómeno complejo, por lo cual, se analiza desde tres perspectivas: comercial, financiera y tecnológica, incorporando una lista de factores que con base en algún argumento se relaciona con el objeto de estudio (ver cuadro 2.2); considerando que su efecto sobre la desigualdad depende a su vez de las circunstancias y características de cada país, es decir, de factores macroeconómicos (ver cuadro 2.1).

2.3.1. Factores macroeconómicos

Entre los factores macroeconómicos se considera la densidad poblacional del país, habitantes por kilómetro cuadrado. La esperanza de vida al nacer (años de vida promedio) por ser un indicador de bienestar relacionado con la salud y la mortalidad de la población. El crecimiento económico que según la hipótesis de Kuznets viene acompañado de desigualdad, el cual se mide con el cambio porcentual anual del producto interno bruto. Debido a la relación entre desarrollo y desigualdad, explicada por la existencia de al menos dos fuerzas que influyen en la distribución del ingreso: la concentración de ahorro de los más ricos, que permite recibir rendimientos y acumular capital para sus siguientes generaciones, y la proporción de población rural e industrial, fenómeno de urbanización donde la población rural obtiene menor ingreso que la población que viven en las ciudades debido al tipo de industria en la que trabajan (Kuznets, 1955); se incluyen el crecimiento porcentual en zonas urbanas (indicador a proxi de industrialización), así como, de empleo en los tres sectores principales agrícola, industrial y de servicios.

La proporción de empleo en el sector agrícola está relacionada con aumentos de desigualdad, al igual que la disminución de trabajadores en el sector industrial, aunque aunado al deterioro de los derechos sindicales (Alderson & Nielsen, 2002). Además, Bluestone y Harrison (2001), muestran una relación entre los cambios en la estructura del empleo por sectores, coincidentes con una baja productividad y la polarización del salario en

E.U. Por otra parte, debido a la dificultad de medir las capacidades de los trabajadores a nivel agregado, como variables proxy se considera la proporción de los empleados en los tres principales sectores. Estas variables de empleo también están estrechamente relacionadas con los argumentos sobre premio a las habilidades o “Skill Premium” y el teorema de Samuelson-Stolper (SS). Con respecto a la concentración del ahorro se consideran las proporciones de inversión pública y privada, así como, el stock de capital objetivo de inversiones extranjeras (Azzimonti et al., 2014).

Cuadro 2.1: Descripción de las variables macroeconómicas

Código	Descripción	Fuente
GDP	Tasa de crecimiento del Producto Interno Bruto	WB
Población y Empleo		
POB_URB	Población Urbana (Tasa de crecimiento anual)	WB
POB_KM	Densidad de la población (habitantes por km ²)	WB
TFERA	Tasa de fertilidad en mujeres entre 15-19 años (nac. p/c 1000)	WB
ESP_VIDA	Esperanza de vida al nacer (años)	WB
EM_IND	Personas empleadas en el sector industrial (% empleados)	WB
EM_SERV	Personas empleadas en el sector servicios (% empleados)	WB
EM_AGR	Personas empleadas en el sector agrícola (% empleados)	WB
Gasto e inversión		
IGOV	Inversión pública (% PIB)	IMF
IPRIV	Inversión privada (% PIB)	IMF
KGOV	Stock de capital público (% PIB)	IMF
KPRIV	Stock de capital privado (% PIB)	IMF
GAS_EDU	Gasto en educación (USD)	
DEB_GDP	Deuda pública histórica (% PIB)	IMF
GMILITAR	Gasto militar (% PIB)	WB
Índices de gobernabilidad		
COR	Control de corrupción	WGI
GEF	Eficiencia de los gobernantes	WGI
ESP	Estabilidad política y ausencia de violencia	WGI
CREG	Calidad de las regulaciones	WGI
RL	Imperio de la ley	WGI
VA	Voz y responsabilidad	WGI

WB-World Bank Data, World Development Indicators.

WGI-Worldwide Governance Indicators

IMF-International Monetary Found

Los países con altos niveles de corrupción difícilmente pueden redistribuir los recursos, y las políticas muchas veces no funcionan o tienen efectos adversos, debido a la

ineficacia del estado para implementarlas. Un ejemplo claro es la corrupción en los puertos, con efectos directos en los costos de exportación que se incrementan en un 8%, lo que reduce la posibilidad de aumentar el intercambio de mercancías para países con corrupción en sus puertos (Dollar, 2004). Desde el punto de vista político el papel del gobierno también es importante, pues la desigualdad es directamente afectada por políticas sociales que reducen la polarización del ingreso a través de los impuestos (Ha, 2012).

Para tomar en cuenta el papel del gobierno, se incluyen los índices de gobernabilidad en los cuales se capta la percepción de la ciudadanía con respecto a la eficiencia del gobierno, entre los que se encuentran: la rendición de cuentas (1), la estabilidad política y ausencia de violencia (2), la efectividad gubernamental en cuanto a servicios públicos(3), la calidad de las regulaciones (4), el estado de derecho –confianza en el sistema judicial-(5) y el control de la corrupción (6) tanto en el sector público como en el privado e individual (Kaufmann et al., 2009).

También se considera la proporción de deuda pública con respecto al ingreso nacional que suele considerarse como un sustituto de la inversión extranjera directa con el objetivo de reducir la entrada de empresas multinacionales, aunque puede tener diferentes efectos dependiendo de la finalidad del crédito, que podría ser gasto militar, de educación, de inversión en tecnologías o gato corriente.

2.3.2. Globalización comercial

La globalización comercial es el grado de apertura de los mercados para el libre flujo de bienes y servicios entre las economías a nivel internacional. Para medirla se considera el volumen de bienes y servicios que se comercializan en los mercados internacionales, separado en exportaciones e importaciones. Para captar la importancia del sector industrial se incluyen las exportaciones de mercancías, es decir, sin considerar servicios o bienes intermedios. Así mismo, si se deterioran los términos de intercambio se deben aumentar las exportaciones para mantener el mismo nivel de importaciones, por ello esta es una característica inherente de la globalización.

Con base en el teorema de Samuelson-Stolper, el salario parte fundamental para la determinación del ingreso de las personas, y este depende de la función de la oferta y

demanda de trabajo. Su impacto sobre la desigualdad proviene de la distinción entre la remuneración del trabajo calificado y no calificado afectada por la demanda (FitzGerald, 1996). Para abordar esta diferenciación entre la demanda de trabajo partimos de que en el sector servicios requiere de personal más calificado que el sector industrial y que en el sector agrícola trabajan las personas menos calificadas, es decir, con menor nivel de educación (Ángeles-Castro, 2007).

Cuadro 2.2: Descripción de las variables de globalización

Código	Descripción	Fuente
Comerciales		
MT	Exportaciones más Importaciones de Mercancías (% PIB)	WB
TERM_TRADE	Índice de términos de intercambio	WB
IM_GS_P	Importación de Bienes y Servicios (% PIB)	WB
EXP_GS_P	Exportaciones de Bienes y Servicios (% PIB)	WB
PPP	Paridad del poder adquisitivo (factor de conversión PIB)	WB
Financieras		
SM_T	Valor de las acciones negociadas (USD)	WB
SM_D	Valor de las acciones nacionales con respecto al VMA (Tasa)	WB
FDI_IN	Inversión extranjera directa entrada de flujos netos(% PIB)	WB
FDI_OUT	Inversión extranjera directa salida de flujos netos(% PIB)	WB
LDI	Total de pasivos de inversión directa (USD)	IMF
ADI	Total de activos de inversión directa (USD)	IMF
RERV	Reservas totales excluyendo el oro SDRs	IMF
REMESAS	Remesas personales (USD)	IMF
AYUDA	Asistencia para el desarrollo y ayuda oficial recibida (USD)	WB
KAOPEN	Grado de apertura de la cuenta corriente	Chinn*
TX_ING	Ingresos Fiscales	IMF
Tecnológicas		
RD_P	Gasto en Investigación y Desarrollo (% PIB)	WB
HTE_T	Exportaciones de alta tecnología (USD)	WB
HTE_PM	Exportaciones de alta tecnología (% de mano facturas)	WB
CRED_TICS	Crédito a Tecnologías de información y computacionales	IMF
US_INT	Usuarios de internet (por cada 100 personas)	WB

WB-World Bank Data, World Development Indicators.

IMF- Fondo Monetario Internacional

*(Chinn & Hiro, 2006)

2.3.3. Globalización financiera

La globalización financiera se define como “la interpenetración de los mercados monetarios y financieros y su integración con los mercados mundiales”, su efecto en la economía depende del grado de apertura financiera y la política monetaria que cada país aplica.

Uno de los fenómenos más notables e importantes de la globalización es la internacionalización de los mercados financieros. Este concepto extiende las oportunidades de compra y venta de activos financieros, pues permite que el precio de una acción o bono gubernamental no solo dependa de su demanda interna sino de la demanda proveniente de todo el mundo, de tal forma que un inversionista puede decidir invertir su dinero en el país que menor riesgo tenga o mejor rendimiento ofrezca, el cual depende de la oferta y demanda de dichos activos.

Por otra parte, la posibilidad de mover los capitales para invertir a cualquier parte del mundo, permite que las industrias no solo exportan bienes finales, sino establecer sus procesos productivos en los países con mayor oferta de mano de obra para reducir sus costos vía salarios. Este tipo de acciones genera empleo, cadenas de valor, y a largo plazo, la clase trabajadora gana, debido a los aumentos de productividad derivados de una asignación más eficiente de los factores de producción. Sin embargo, impulsa la desindustrialización de las economías con una mano de obra mayormente calificada, porque los inversionistas nacionales prefieren invertir su capital (ahorro doméstico) en proyectos más rentables. Entonces la inversión extranjera directa se convirtió en un sustituto de la inversión doméstica para las economías con bajos niveles de ahorro nacional (Letto- Gillies 1992).

Siendo los flujos de inversión directa clave para una asignación eficiente de recursos a nivel mundial, sin embargo, algunos autores la consideran una táctica de globalización Alderson, & Nielsen (2002) o meramente otra forma de inversión (Bhagwati 1994), lo que si es cierto es que estos flujos se incrementaron con la apertura de los mercados financieros internacionales. Por lo que, incluimos el flujo de capitales medido por la inversión extranjera directa, el tamaño y capitalización del mercado bursátil, así como, el valor del mercado accionario (número de acciones por su precio), y el número de acciones nacionales con respecto al valor del mercado.

Las remesas, asistencia para el desarrollo y ayuda oficial recibida representan un ingreso importante para países subdesarrollados y en desarrollo, y aunque ayudan a reducir la desigualdad, también podrían afectar la oferta y demanda de trabajo.

2.3.4. Globalización tecnológica

En cuanto a la globalización tecnológica, las innovaciones generan nuevas industrias y aumentan la productividad del capital ya existente, así mismo, su efecto se potencializa con la apertura de los mercados (Jaumotte et al., 2013). La globalización acelera la privatización especialmente de las investigaciones que generan tecnología, incrementando la desigualdad entre quien puede comprar o acceder a estos progresos y quien no, esto es aplicable tanto a nivel personal como entre industrias y países. Para medir el acceso a la información global se toma el número de usuarios de internet, lo cual reduce problemas de asimetría de información. Por otro lado, el auge de los mercados tecnológicos afecta la demanda de trabajo, pues las empresas dedicadas a las innovaciones requieren empleados con un mayor grado de especialización.

Asteriou et al. (2014) encuentran que, los países exportadores de alta tecnología tienen menores niveles de desigualdad, para lo cual, el gasto en investigación y desarrollo (I&D) y los créditos enfocados a tal rubro juegan un papel fundamental, pues, permiten exportar tecnología en lugar de importarla. Se incluyen las exportaciones de alta tecnología del sector manufacturero, debido a que la eficiencia de este tipo de industria a través de la automatización reduciría el nivel de empleo sobre todo de la mano de obra no calificada.

En total se consideran 11 factores financieros, 5 comerciales, 5 tecnológicos, los 6 índices de gobernanza y 15 macroeconómicos entre empleo e inversión y poblacionales, descritos en los cuadros 2.3 y 2.4.

2.3. Análisis estadístico y descriptivo

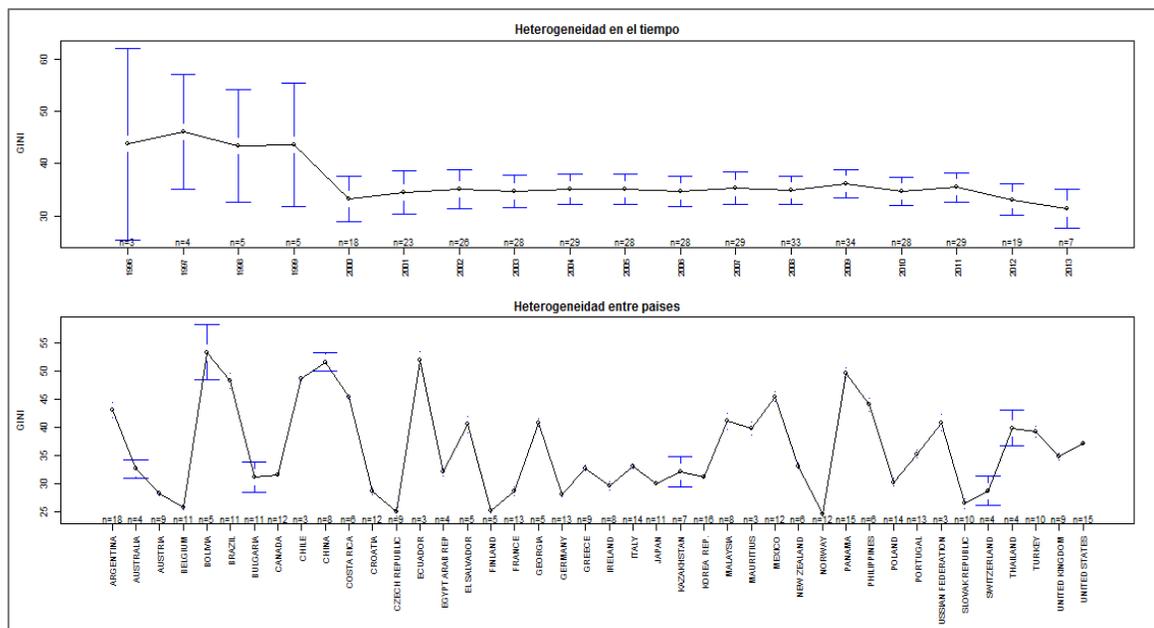
La desigualdad se mide con el coeficiente de Gini de la base de datos SWIID, como se describe anteriormente, la cantidad de observaciones es mayor a los cinco mil, su intersección con las 42 variables reduce la muestra a poco más de cuatro mil, ya que no todos los países tienen bases de datos estadísticos con todas las variables propuestas para todos los años. El coeficiente de Gini mínimo alcanza un valor de 18, asociado a los países escandinavos

específicamente República Checa y Eslovaquia, mientras que el valor máximo es de 60 y se relaciona con países que se encuentran en el sur de África como Suazilandia y Botsuana. Sin embargo, estos valores son poco probables de suceder, ya que la media corresponde aproximadamente a un coeficiente del 38.

En la figura 2.1 se presenta la heterogeneidad de una sub-muestra compuesta por 42 países con la mayor cantidad de observaciones entre el año 1996 y 2013; la variabilidad del Gini en cada país es relativamente pequeña, por ejemplo: en México está entre el 40 y el 50. Sin embargo, el Gini promedio entre los países es muy marcado; entre dos países el coeficiente puede variar hasta 30 puntos, por el contrario el valor promedio de la desigualdad varía poco en el tiempo a partir del 2000.

Los datos correspondientes a las variables que miden los diferentes factores de la globalización, se obtienen principalmente de las dos instituciones internacionales más confiables: el Banco Mundial y el Fondo Monetario Internacional.

Figura 2.1: Heterogeneidad de los datos en el tiempo y entre países



Fuente: Elaboración propia con datos de la SWIID 6.0

La base de datos con que se trabaja tiene el siguiente esquema relacional. En la tabla “Países” se guarda un identificador numérico por país “idPais” y los datos generales de cada nación, como nivel de ingreso per cápita y la región geográfica a la que pertenecen, así como

algunas variables categóricas binarias que indican si el país es miembro de la OCDE según el Banco Mundial. Las descripciones de las regiones y niveles de ingreso se almacenan en las tablas “Regiones” y “NivelesIng”, y se detallan en el siguiente cuadro.

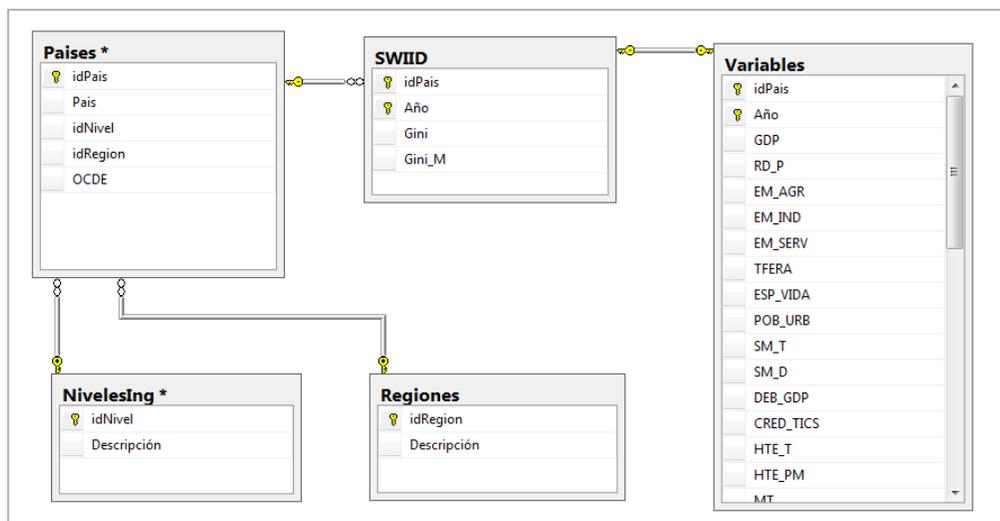
Cuadro 2.3: Niveles de ingreso

Nivel Ing.	Descripción
A	Bajo
B	Medio bajo
C	Medio alto
D	Alto
E	Otro

Fuente: Banco Mundial

Guardamos en la tabla “SWIID” los datos estandarizados del coeficiente de Gini antes y después de impuestos, se utiliza un promedio de las trayectorias estimadas (Solt, 2016).

Figura 2.2: Diagrama entidad-relación de la base de datos



Fuente: Elaboración propia en SQL Server 2016⁶

En la tabla “Variables” se concentran los datos recabados de distintas bases de datos institucionales, utilizando como llave primaria el “idPais” y el “Año” al que pertenece el

⁶ Las llaves señalan el campo que identifica una tupla, registro o conjunto de atributos, de tal manera que no puede haber dos registros con la misma llave primaria, por ejemplo, no puede haber dos valores del coeficiente Gini para un país en el mismo año, esta propiedad permite relacionar los datos de una variable con los del Gini correspondiente a un país en un año.

conjunto de atributos representados por las variables; la descripción, tipo (macroeconómica, comercial, financiera o tecnológica) y fuente se detalla en los cuadros 2.1 y 2.2.

Considerando que el objetivo es extraer generalizaciones globales dadas las características de cierto tipo de países, lo ideal sería tener datos de la mayoría de los factores propuestos de un grupo de países representativo de cada región. Sin embargo, los datos internacionales no están estandarizados, además de que no todos los países concentran las mismas estadísticas o no tienen datos para ciertos periodos y en algunos otros simplemente no tienen información suficiente.

Cuadro 2.4: Número de países y observaciones por región

Región	Global		Muestra				Ratio Global de la Muestra	
	Países	Obs.	Países	%	Obs.	%	% Países por región	% Obs. por región
1 Sur de Asia	8	288	8	4.62	180	4.25	100	63
2 Europa & Asia Central	59	2,124	49	28.32	1,415	33.43	83	67
3 Medio East & África del No	21	756	14	8.09	325	7.68	67	43
4 Este de Asia & del Pacífico	38	1,368	25	14.45	588	13.89	66	43
5 Sub-Saharan Africana	48	1,728	43	24.86	875	20.67	90	51
6 Latín América & El Caribe	41	1,476	32	18.50	780	18.43	78	53
7 Norte-América	3	108	2	1.16	70	1.65	67	65
	218	7,848	173	100	4,233	100	79	54

Fuente: Elaboración propia con datos obtenidos de diferentes fuentes según los cuadros 2.1 y 2.2.

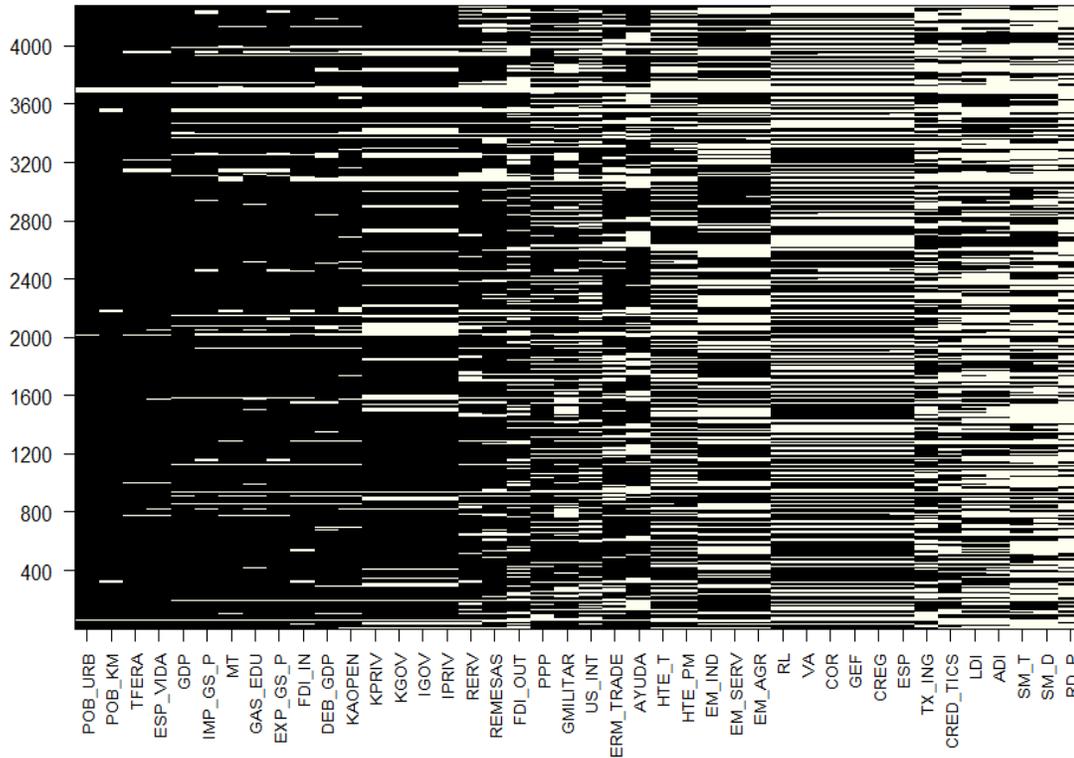
En el cuadro 2.4 se muestra el porcentaje de países y el número de observaciones por región; las dos primeras columnas corresponden al total de países registrados por el Banco Mundial, para un periodo de estudio de 1980 al 2015, es decir, considerando 36 observaciones por país, se debería tener un total de 7,848 registros; en las siguientes cuatro columnas se aprecia la cantidad de información obtenida y correspondiente a 173 países, conviene destacar que los que mayor cantidad aportan al análisis son los europeos; y en las últimas dos columnas podemos ver la representatividad de la muestra con respecto a las cantidades globales, de tal manera que se tienen registros del 79% de los países, aunque no para todos los años pues las observaciones totales representan el 54%.

2.3.1 Análisis de valores nulos

A pesar de la gran cantidad de información que existe en la actualidad, el análisis de fenómenos multidimensionales (independientemente de la técnica econométrica o herramienta estadística) se dificulta debido a la inconsistencia de la información. Siendo la

falta de datos temporales y entre países (valores nulos) uno de los principales problemas que conllevan las investigaciones referentes a la desigualdad (Harrison & Hanson, 1999).

Figura 2.3: Visualización de valores nulos en la base de datos

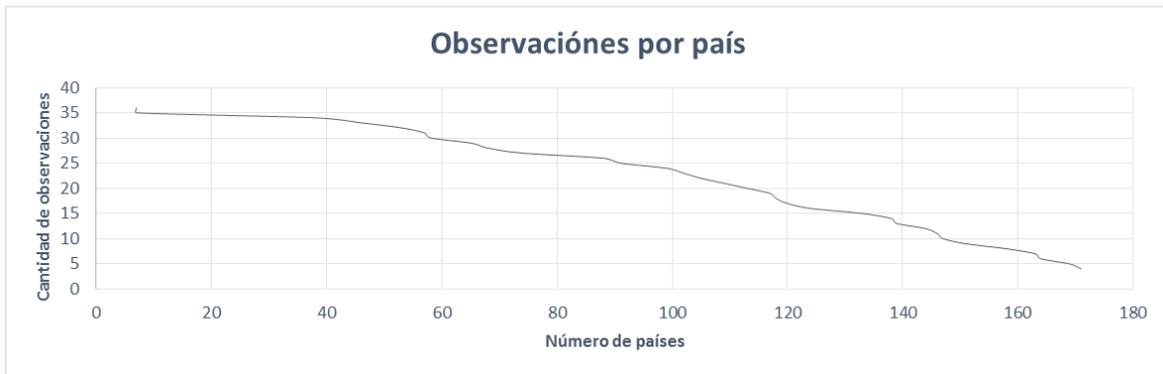


Fuente: Elaboración propia

El primer paso para preparar los datos de entrada de casi cualquier modelo suele ser eliminar las observaciones con al menos un valor nulo. Sin embargo, al limpiar la base de datos se pierde más del 95% de la información, al reducir la cantidad de registros de 4233 a solo 150. Por otro lado, en un análisis por dato se tienen 177,786 valores de los cuales 47,700 no contienen información, de tal forma que los valores nulos representan el 36.67%.

Una opción para abordar este problema, es descartar los países con menor cantidad de información disponible, analizando solo aquellos con suficiente información para todas las variables, esto implica estudiar prácticamente los países desarrollados, que tienen por lo menos 35 años de datos.

Figura 2.4: Número de observaciones por número de países



Fuente: Elaboración propia

En la figura 2.4, se muestra el número de países por cantidad de observaciones, con un promedio de 20 datos por país, se muestra que por lo menos 21 países tienen menos de 10 observaciones. Por otro lado, casi 40 tienen 35 registros, la mayoría de los cuales corresponden a países desarrollados, por lo que trabajar solo con estos países proporcionará resultados sesgados hacia estos países, además de tener una muestra menor a los 1400 registros.

Otra opción es eliminar los factores con menor cantidad de valores nulos, en este caso descartaríamos el porcentaje en investigación y desarrollo (RD_P), sin embargo, quitar esta variable apenas aumenta el número de observaciones sin nulos de 150 a 207, a pesar de que cuenta con 1389 registros para distintos países. Sí por el contrario este factor es en realidad importante, un análisis considerando solo esta variable podría generar buenos resultados.

También se consideró eliminar variables que podrían medir en cierto grado el mismo factor, por ejemplo, los flujos de inversión extranjera directa (FDI) y los pasivos o activos en la balanza de pagos (LDI, ADI). Eliminar las variables de uno de los grupos permitiría trabajar con mayor cantidad de datos, pero ¿cuál debemos eliminar?

Por otro lado, no sabemos si todos los factores propuestos son determinantes para el fenómeno de estudios, por lo que, podríamos estar incluyendo algunos aspectos no significativos y con poca información; una variable con muchos nulos podría no contener información suficiente para sustentar una relación o podría simplemente no tener una relación importante.

Entonces, hasta qué punto podemos justificar eliminar un factor por falta de información, considerando que se tienen variables con mayor información pero que igualmente tiene valores nulos; hasta qué punto podemos eliminar países con poca información sin sesgar los resultados hacia ciertas economías. Eliminar las observaciones que generan mayor pérdida de datos no asegura encontrar el mejor modelo, por lo que incluimos un criterio de discriminación que ordena las variables por nivel de importancia tras estimar un modelo.

2.4. Algoritmo de re-muestreo de factores con valores nulos (RM-CSR-VN)

Con el objetivo de reducir la pérdida de información pero manteniendo los factores más significativos dados los datos disponibles, se propone un proceso sistemático para seleccionar las variables que proporcionan mayor cantidad de observaciones sin datos nulos, considerando un criterio de discriminación por orden de importancia que depende del modelo utilizado para el análisis.

La idea consiste en generar conjuntos de información a partir de distintas combinaciones de factores. Note que el tamaño de la base de datos sin valores nulos cambia según el grupo de variables a incluir, de tal forma que la cantidad de información es mayor cuando se incluyen por ejemplo solo variables de globalización comercial que al incluir solo variables de globalización tecnológica.

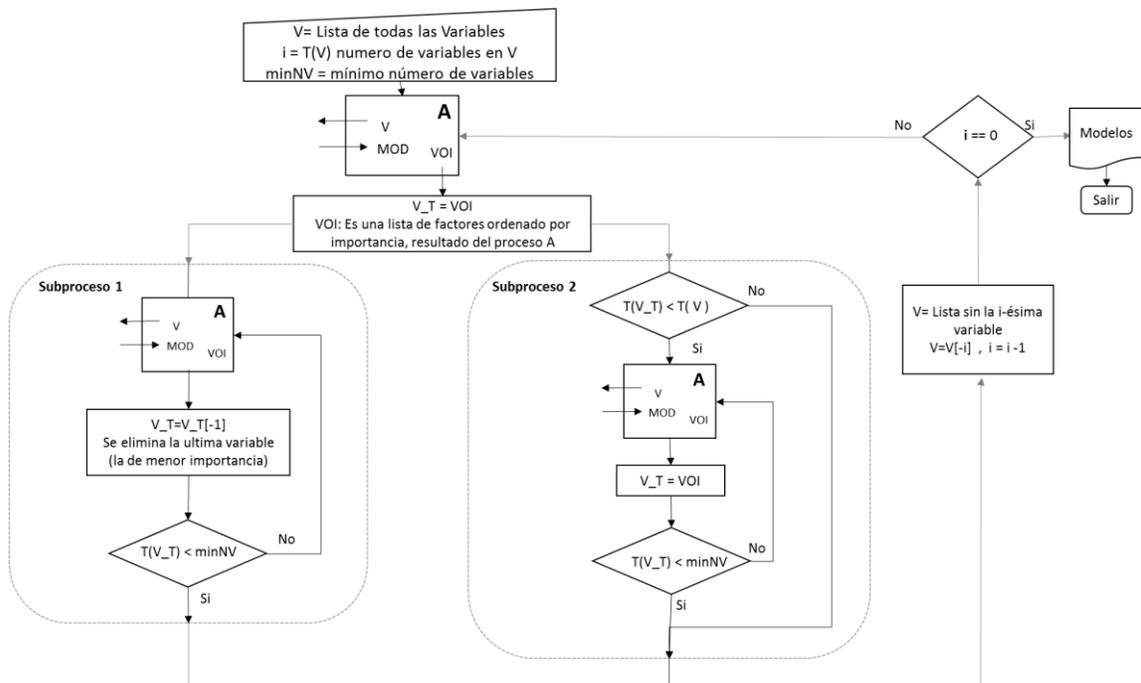
Si se generan todas las combinaciones posibles, el número de conjuntos aumenta exponencialmente con base en el número de factores. Con 10 variables son más de mil y con apenas dos más, es decir, con 12 aumenta más de 4 veces hasta 4095, entonces discriminar variables de una lista de 42 resulta una opción computacionalmente no manejable, por lo que se utilizan técnicas de remuestreo con y sin reemplazo, pero no de las observaciones sino de la lista con los $n = 42$ factores propuestos.

El remuestreo con reemplazo consiste en eliminar iterativamente de la lista original (V) una variable diferente cada vez, de esta primera parte del proceso se obtienen n listas diferentes de factores. Entonces para cada uno de estos conjuntos resultantes se elimina, sin reemplazo, la variable con peor desempeño hasta un mínimo de variables definido como $minV$ (subproceso 1, figura 9), aumentando las combinaciones a $n * (n - minV)$; el criterio de

desempeño depende del modelo utilizado para la investigación, si bien, podemos enfocarnos en maximizar la cantidad de observaciones de la base de datos suponiendo un orden de importancia en el análisis con base en la cantidad de valores no nulos, ordenar las variables acorde con los resultados de un modelo mejora considerablemente los resultados (como un algoritmo evolutivo). El modelo puede ser un panel de datos, una regresión logística, una red neuronal artificial o un árbol de decisión, o cualquier otro, siempre y cuando sea posible ordenar las variables que utiliza por un criterio de desempeño o importancia.

Si el modelo es capaz de discriminar factores por sí mismo, como es el caso de los árboles, se ejecuta paralelamente un proceso iterativo para cada lista resultante del primer remuestreo (con remplazo), en cada iteración se incluyen solo las variables utilizadas en el modelo anterior hasta un mínimo de variables definido (subproceso 2, figura 2.5). Con este método se obtienen un máximo de $2n * (n - \text{min}V)$ conjuntos de información, o modelos. Aunque podrían ser considerablemente menos, pues para evitar conjuntos repetidos, antes de estimar cualquier modelo se verifica que la combinación encontrada no se halla considerando anteriormente, además de que el algoritmo de aprendizaje del árbol discrimina varias variables al mismo tiempo.

Figura 2.5: Diagrama de flujo del algoritmo RM-CSR-VN



Fuente: Elaboración propia

En la figura 2.5 se muestra el diagrama de flujo del proceso de generación de conjuntos de información sin valores nulos. Donde V es un vector de strings que contiene los nombres de las variables propuestas, la función $T()$ regresa el tamaño del vector, i es un contador que controla la iteración del proceso de re-muestreo con reposición.

El subproceso A, recibe una lista de variables con base en la cual se filtra la base de datos y eliminan los valores nulos, dependiendo del tipo de análisis encuentra el mejor modelo dado un conjunto de datos sin nulos, regresa las medidas de ajuste y características del modelo incluyendo los nombres de los factores utilizados ordenados por importancia. Este subproceso representa una función conectora para implementación de distintos tipos de técnicas y modelos.

En este trabajo se utilizan los modelos: Perceptrón multicapa y un árbol clasificador, la forma en que se estiman, evalúan y selecciona el mejor modelo dado un conjunto de información, se detalla en el siguiente capítulo. A continuación se presenta un subproceso A general de discriminación de variables, independientemente del tipo de modelo, tomando como ejemplo el criterio para maximizar el número de observaciones.

1. Filtrar las columnas correspondientes a la lista de variables recibidas y eliminar valores nulos.
2. Estimar el modelo correspondiente. Se calcula el número de valores sin dato por cada variable
3. Ordenar las variables utilizadas con base en un criterio específico de desempeño. Las variables se ordenan empezando por aquella que tenga menor cantidad de datos nulos
4. Guardar los parámetros que identifican el modelo. En primer lugar, el número de registros sin nulos y número de variables y en segundo lugar la lista de variables recibidas ordenadas alfabéticamente.
5. Guardar las medidas de ajuste del modelo, incluyendo la lista de variables ordenadas, donde la lista de las variables están ordenadas por importancia.

Dado que el ejemplo anterior no discrimina variables por sí solo, la cantidad máxima alcanzan las 1554 combinaciones, sin embargo, se obtienen 884. El número mínimo de observaciones se genera incluyendo el total de las variables (150 registros), si aplicamos el mismo criterio de maximizar la cantidad de registros para seleccionar el mejor conjunto de

información, éste sería la combinación de POB_URB, POB_KM, TFERA, ESP_VIDA, GDP, con 3,933 observaciones sin nulos, lo anterior considerando 5 variables como mínimo $minVN = 5$. El conjunto de variables que máxima el número de observaciones sin valores nulos, corresponde a las primeras variables de derecha a izquierda de la figura 2.3, es decir, a variables macroeconómicas características de la población (POB_URB, POB_KM, TFERA, ESP_VID, GAS_EDU), además de Producto Interno Bruto (GDP) y varias comerciales (IMP_GS_P, EXP_GS_P, MT, KAOPEN) y la inversión extranjera directa (FDI_IN). Sin embargo, no todas estas variables resultan tan importantes cuando el criterio depende de un modelo con efectos sobre la desigualdad como se muestra en los siguientes capítulos.

Los criterios de selección pueden variar según el objetivo de la investigación, se puede penalizar los modelos que no incluyan algún factor de globalización comercial, tecnológica o financiera, encontrar un balance entre número de registros y número de factores, así como, agrupar combinaciones que generen cantidades cercanas de registros, por ejemplo entre 200 y 210, o 210 y 220, este criterio reduce los conjuntos de información a 236. En este caso es posible aplicar el algoritmo RM-CSR-VN, que permite generar conjuntos de información maximizando el número de observaciones y factores, ya que ordenar las variables por el número de datos nulos descendientemente, permite descartar las variables con menos datos de tal forma que sobreviven aquellas que aportan mayor información.

CAPÍTULO III. PROCESO DE MODELADO BOOTSTRAPPING

Dada la importancia de los datos y el actual crecimiento de la información, aunado a las cada vez más poderosas computadoras, un método científico inverso ha tomado fuerza, este consiste en analizar un conjunto considerablemente grande de datos relacionado al tema de estudio, y analizarlos con técnicas de minería de datos para extraer conocimiento, generalidades, comportamientos o relaciones, este cambio de paradigma se utiliza en la ciencia de datos. Esta disciplina, encargada de obtener conocimiento a partir de la información, no es nueva, sin embargo ha tomado fuerza debido al incremento y acceso a la información, dando lugar a un conjunto de técnicas y procesos que extraen conocimiento a través de la búsqueda de patrones dentro de grandes bases de datos (Olson & Delen, 2008).

El desarrollo de las tecnologías computacionales y de técnicas de minería de datos o inteligencia artificial, ha abierto un campo potencial de investigación aplicada en distintas áreas, como la medicina, la química, marketing, finanzas, entre otras. Actualmente la información se ha convertido en un activo valioso, considerada la materia prima para extraer conocimiento, siendo, las técnicas de minería de datos la maquinaria o herramientas de extracción sin las cuales el ser humano no podría, por sí solo, entender u obtener conocimiento de la gran cantidad de información actualmente disponible.

La mayoría de estas técnicas son no paramétricas y se basan en procesos iterativos de entrenamiento inductivo, no precisamente deductivo como suele hacerse ciencia tradicionalmente, es decir, presuponiendo que existe una relación (hipótesis) y luego comprobándola bajo ciertos supuestos, por el contrario este enfoque parte del análisis de una cantidad de información suficientemente grande para hacer inferencia o generalizaciones. Estos procesos de entrenamiento inductivos se basan en algoritmos de aprendizaje comúnmente utilizados en el campo de la inteligencia artificial.

La elección de la técnica a utilizar, depende principalmente del objetivo de la investigación, que puede incluir clasificar, inferir o predecir y agrupar. Los modelos con mayor grado de precisión son las redes neuronales artificiales, sin embargo, también son consideradas cajas negras, debido que es complicado interpretar las relaciones no lineales que generan tan acertados resultados, sin embargo, se han desarrollado técnicas de

sensibilidad que permiten conocer la importancia de cada variable en la estimación del modelo. Por otro lado, los árboles de decisión son quizá los modelos con mayor grado de interpretación, ya que indican la forma y condición de las relaciones entre los factores involucrados, pero con mayor margen de error (Rokach & Oded, 2008). En esta tesis se utilizan estas dos técnicas, para analizar la relación entre la globalización y la desigualdad.

3.1. Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son consideradas como modelos no paramétricos adaptativos de minería de datos, que buscan generalizaciones lineales y no lineales a partir de un conjunto de datos de entrada. También se considera parte de la inteligencia artificial, puesto que se basa en el proceso neurológico del aprendizaje biológico.

Las redes neuronales son consideradas herramientas de minería de datos, ya que permiten encontrar relaciones en una gran cantidad de datos, además de ser no paramétricas, lo que implica que no es necesario presuponer que los datos tienen un comportamiento determinado.

Pero también, son precursoras de la inteligencia artificial (IA), debido al origen de su diseño enfocado en emular el proceso de aprendizaje de la red neuronal biológica, especialmente del comportamiento entre sus conexiones sinápticas. Su diseño se basa en la red neuronal del cerebro humano, debido a su capacidad de aprender a través de las conexiones entre sus neuronas. La forma en la que aprenden se basa en un proceso de entrenamiento, que consiste en obtener generalizaciones a partir de un conjunto de datos de entrada (ejemplos), por lo que el algoritmo de aprendizaje es un elemento clave del modelo. Haykin (2005) define una red neuronal como:

"Un procesamiento masivamente paralelo de distribución compuesto por unidades simples de procesamiento, que tiene una propensión natural para el almacenamiento de conocimiento experimental de tal manera que esté disponible para su uso. Se asemeja al cerebro en dos aspectos: (1) El conocimiento se adquiere por la red de su entorno a través de un proceso de aprendizaje y (2) las fuerzas de conexión entre neuronas, conocidos como los pesos sinápticos, se utilizan para almacenar los conocimientos adquiridos "(Haykin, 2005: 24).

Como parte de la inteligencia artificial (IA) sus orígenes vienen desde 1943 cuando McCulloch & Pitts (1943) publicaron su trabajo en el que mostraron que cualquier función podía ser estimada mediante un modelo de neuronas, y propusieron la ley "todos para uno" que especifica una red con unidades simples -neuronas- y conexiones sinápticas operando simultáneamente entre sí. En 1958 Rosenblatt presentó el famoso modelo perceptrón con un método de aprendizaje supervisado; Gabor propone un modelo adaptativo no lineal, y plantea el modelo Adeline; en 1967 surge el libro de Minsky "Cálculo: Máquina finita e infinita" (Haykin, 2005). En ese período, el progreso de esta técnica explotó; en palabras de Haykin "durante el período clásico de la década de 1960 parecía que no había nada que el Perceptrón no pudiera hacer " (Haykin, 2005: 62p). Hasta que Minsky & Papert muestran las limitaciones fundamentales de una red de una sola capa (perceptrón) con un modelo matemático, diseñando un modelo para el problema de asignación de crédito conocida como red de Hopfield (que está enfocada en la optimización). Después de esto, hubo un período de olvido hasta 1980, cuando la potencia de los ordenadores alcanza a la teoría de las redes. Entonces Grossberg establece la teoría de la resonancia adaptativa principio que incluye una parte dinámica, en la cual los patrones se ajustan a través del aprendizaje de retroalimentación, resolviendo así el problema de Minsky y dando paso a la red de Hopfield, un gran progreso en esta área.

En los años 90 se desarrolló el algoritmo de propagación hacia atrás (backpropagation) que tiene aplicaciones interesantes en el control óptimo; hubo mucha atención en mejorar el proceso de aprendizaje, la base de lo que ahora se conoce como aprendizaje automático. En esa misma época aparece la red de aprendizaje supervisado (máquinas de vectores soporte), utilizados para la estimación, regresión y reconocimiento de patrones. En 1988 Broomhead y Lowe diseñan una red de alimentación directa utilizando funciones de base radial (Haykin, 2005).

Hasta 1988, una RNA se consideraba una caja negra aunque con excelentes resultados, pero sin una interpretación de valor teórico. Las redes neuronales habían sido vistas como modelos con muy buenos resultados pero de caja negra, ya que resulta difícil interpretar los pesos de los modelos no lineales. Sin embargo, en los últimos años varios trabajos han desarrollado técnicas que permiten obtener más información a partir del

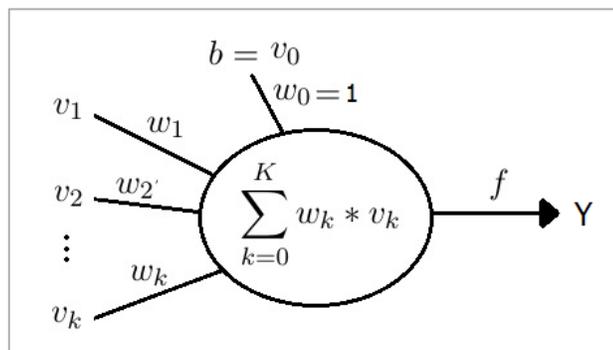
proceso de ajuste de las RNA, entre las que se encuentran la extracción de reglas de Ultsch (1993) y los métodos de sensibilidad que desarrollan Montañó & Palmer (2002), Gevrey et al. (2003).

En general los beneficios de las RNAs son: (1) que pueden ser redes distribuidas lineales y no lineales, (2) que tienen una capacidad incorporada de adaptación -diversificación en un entorno no estacionario, (3) que proporcionan información acerca de un patrón en particular, (4) que son tolerantes a fallos, (5) que tienen una integración a gran escala –capturando comportamientos complejos, (6) que tienen un diseño y análisis uniforme, así como un gran potencial de aplicaciones (Haykin , 2005: 25-28)

3.1.1. Perceptrón

El modelo de red con la arquitectura más simple es el perceptrón al representar una sola neurona; como se muestra en la figura 3.1, estima un valor de salida Y con base en un conjunto de datos de entrada $X = \{v_1, \dots, v_k\}$ y un sesgo b, siendo f la función de activación que define el umbral de energía de las neurona.

Figura 3.1: Arquitectura del perceptrón



Fuente: Elaboración propia con base en Haykin (2005)

Los parámetros del modelo (pesos w) se estiman a través del algoritmo BackPropagation (BP) que va ajustando los valores iterativamente hasta encontrar un mínimo en el error cuadrado (Rumelhart et al., 1995).

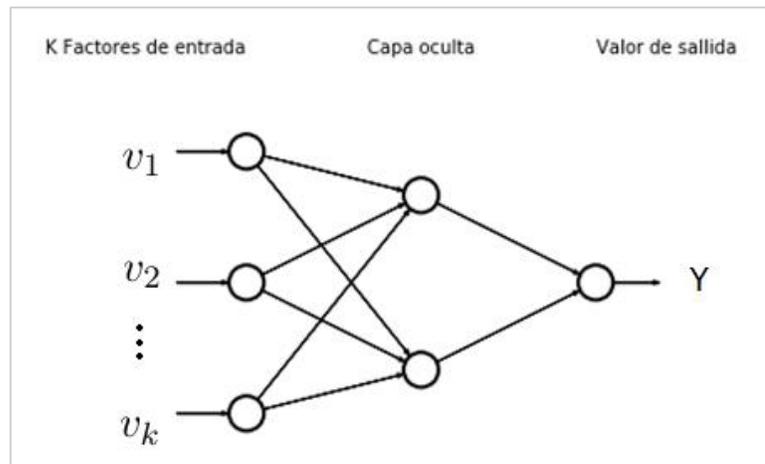
$$Y = f\left(b + \sum_{k=1}^K w_k v_k\right)$$

El modelo anterior es equivalente a una regresión de mínimos cuadrados ordinarios donde f es una función identidad, el sesgo b_M es la constante, los pesos w_k los coeficientes estimados y la función de activación es logística.

3.1.2. Modelo perceptrón multicapa

La arquitectura de un modelo perceptrón multicapa consiste en agregar capas ocultas con diferentes cantidades de neuronas, donde los valores de salida de las neuronas en cada capa oculta son los valores de entrada para las neuronas de la siguiente capa, como se muestra en la figura 3.2. Los factores son los valores de entrada y la capa oculta tiene dos neuronas que representan un perceptrón cada una.

Figura 3.2: Arquitectura de un modelo perceptrón multicapa



Fuente: Elaboración propia con base en Haykin (2005)

La arquitectura general del modelo perceptrón multicapa se representa con la siguiente función, donde H es el número de capas ocultas y N el número de neuronas por capa.

$$Y = f_M \left(\sum_{k=1}^K b_k + \sum_{j=1}^H w_{jk} * f_H \left(b_H + \sum_{i=1}^N w_{ij} X_{ki} \right) \right)$$

Se utiliza un modelo de entrenamiento supervisado por lo que su estimación se realiza en dos etapas, la de funcionamiento para obtener la arquitectura de la red y la de aprendizaje en la cual se ajustan los pesos de acuerdo al algoritmo de aprendizaje BP.

$$x_{ik} = \frac{x_{ik} - \bar{X}_k}{\sqrt{\sum \frac{(x_{ik} - \bar{X}_k)^2}{n-1}}} \quad \forall k$$

El modelo se estima con un conjunto de datos de entradas estandarizado representadas por X , para facilitar la convergencia cuando las k variables tienen diferentes escalas, lo cual mejorará el proceso de entrenamiento (Faraway, 2016); cada observación se estandariza con base en su media y desviación como se muestra en la ecuación anterior.

3.1.3. Análisis de sensibilidad

La naturaleza no lineal de las redes neuronales artificiales dificulta la interpretación de las relaciones entre las variables incluidas, sin embargo, existe una línea de investigación enfocada no en interpretar los pesos sino en analizar la sensibilidad del modelo ante cambios en los valores de entrada. Para extraer este tipo de información se han desarrollado una serie de métodos que analizan la contribución de cada variable en los modelos MPL, entre los que destacan el método de perturbación y de importancia relativa basado en los efectos marginales.

3.1.3.1. Método de perturbación

Para estimar la contribución de cada variable se utiliza el método de perturbación, que consiste en estimar el valor de salida, a través del modelo de redes neuronales MLP con los mismos datos de entrada excepto el valor de la variable en cuestión; este método analiza el efecto en el coeficiente GINI ante cambios de una variable manteniendo las demás fijas.

Dado un conjunto de información $X'(x_i, \dots, x_{-j}, \dots, x_k)$ los valores de x_j se estiman con base en el rango de los posibles valores que puede tomar el factor en cuestión, tomando como límites sus valores mínimo y máximo $\widehat{G}_k = f_M(\bar{x}_i, \dots, x_{-j}, \dots, \bar{x}_k)$. Para cada factor se estima y grafica su efecto con respecto a la variable objetivo, las gráficas se desarrollaron bajo el método de perturbación agregando ruido en cada variable y graficando las estimaciones bajo diferentes condiciones.

3.1.3.2. Importancia de los factores que afectan la desigualdad en el modelo.

La metodología de importancia relativa de las variables se basa en la interacción entre los pesos sinápticos. La importancia relativa de la k -ésima variable Q_k , se determina con base en la interacción entre los parámetros del modelo W y el porcentaje de influencia de una variable k sobre la de salida, como se muestra a continuación

$$Q_k = \frac{\sum_H R_{nk}}{\sum_N \sum_H R_{nk}}, \quad R_{nk} = \frac{w_{nj}}{\sum_N w_{kj}} v_{nk}$$

donde R_{nk} es el valor de importancia relativa de la variable k con respecto al valor total de entrada para la neurona n (Garson, 1991).

Los resultados del análisis de sensibilidad pueden ser inestables porque el algoritmo de aprendizaje BP es capaz de encontrar diferentes conjuntos de pesos (óptimos locales), es decir, diferentes redes -todas con un alto grado de ajuste-, esto depende de los pesos iniciales los cuales son aleatorios. Un método que reduce considerablemente este problema consiste en: entrenar la red con los mismos datos y la misma arquitectura, encontrar un conjunto de pesos óptimos, repetir este entrenamiento M veces con diferentes pesos iniciales y obtener las importancias relativas en cada iteración de tal manera que al final se tenga una matriz $M * K$; de las M importancias relativas se trabaja con el promedio (De Oña & Garrido, 2014). Sin embargo, en este trabajo se implementa el algoritmo (RM-CSR-VN).

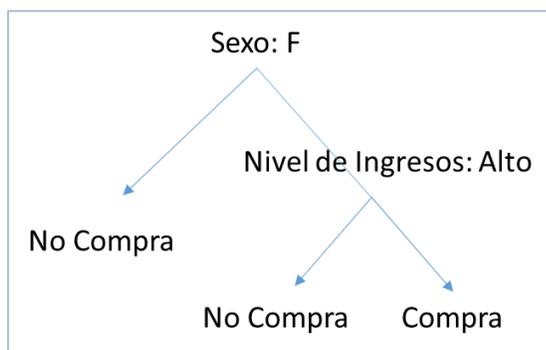
3.2. Árboles de decisión clasificadores

Los modelos econométricos y estadísticos presuponen el comportamiento de los datos o estiman ciertos parámetros basados en el análisis exploratorio de la información, sin embargo, esto se vuelve menos claro cuando los modelos son complejos, como cuando se incluyen una gran cantidad de variables con comportamientos no normales.

Los árboles de decisión son modelos flexibles aditivos, entre lineales y no paramétricos, que captan la interacción entre las variables al condicionar el comportamiento de las variables con el de otras. Los resultados dependen de las decisiones individuales que se van tomando en cada nodo del árbol, a esta combinaciones de pequeñas decisiones que llevan a un resultado final o combinación de condiciones se le conoce como regla de conocimiento (Faraway, 2016). Fueron desarrollados en 1970, y tienen raíces tanto

estadísticas como de la ciencia computacional; por una parte surgen con el fin de graficar las relaciones de un conjunto de datos a partir de características estadísticas, por parte de la ciencia computacional surge como una estructura de decisión que llega a una respuesta a través de un conjunto de reglas. Por ejemplo, el siguiente árbol indica si una persona es cliente potencial para comprar un auto rosa, considerando como atributos de entrada el sexo y el nivel de ingresos. El primer criterio de decisión es el sexo como se visualiza en la figura 3.3 donde las reglas se interpretan como sigue: 1) los hombres no compran el auto rosa, 2) las mujeres con bajos ingresos no compran el auto, 3) las mujeres con altos ingresos tienen alta probabilidad de comprar un auto rosa.

Figura 3.3: Ejemplo de un árbol de decisión



Fuente: Elaboración propia

Un árbol de decisión clasificadores y de regresión (CART) clasifica una instancia⁷ dentro de un conjunto de clases definidas con base en los valores de los atributos de entrada, cada punto de decisión es un nodo y por cada ruta, es decir, desde la raíz hasta que termina en una hoja, hay una regla que lleva a un resultado con una probabilidad asociada.

Formalmente el esquema un CART se define como una combinación de atributos $B(A \cup Y)$, donde A es el conjunto de n atributos (variables independientes) $A = \{ a_1, a_1, \dots, a_i, \dots, a_n \}$, Y la variable objetivo con un dominio $dom(Y) = \{ c_1, c_2, \dots \}$ que contiene las c clases que representan los posibles valores de Y (Rokach & Oded, 2008).

Los atributos pueden ser variables de dos tipos nominales y numéricas, el primero también se conoce como de tipo factor o categórica que representa valores no ordenados, los

⁷ Una instancia es la realización de un ejemplo, es decir, un conjunto o tupla, que contiene un valor de cada variable asociados al valor de la variable objetivo.

atributos de este tipo pueden ser etiquetas que identifican sexo, estado civil, etc., los segundos de tipo numérico contiene valores continuos en los números reales como ingresos, edad, índices, etc. Los arboles clasificadores suelen tener un atributo nominal como variable objetivo.

3.2.1. Algoritmo de entrenamiento

El algoritmo de entrenamiento es el proceso para encontrar los criterios de decisión que clasifican los ejemplos con el menor margen de error posible.

Si suponemos que el modelo solo tiene una parte determinista, es decir, sin ruido, será posible encontrar un árbol capaz de clasificar todos los ejemplos, sin embargo, se podrían estar estimando reglas muy específicas, y por ende, no muy útiles, entonces el árbol podría perder poder de interpretación. Por otro lado, este supuesto es poco plausible debido a lo cual se entrena un árbol solo con cierto porcentaje de los datos, muestra de entrenamiento.

Dado el espacio universal del modelo $U = X \times \text{dom}(y)$ donde X es el espacio de instancias (conjunto de todas las posibles combinaciones de los atributos de entrada) definido como el producto cartesiano entre el dominio de los atributos $X = \text{dom}(a_1) \times \text{dom}(a_2) \times \dots \times \text{dom}(a_n)$ estos pueden tener valores nulos. El conjunto de entrenamiento (DE) es subconjunto de U seleccionado aleatoriamente a través de una función de distribución. Es posible encontrar diferentes modelos estimados con base en distintos conjuntos de entrenamiento, es decir, diferentes observaciones, debido a la aleatoriedad. En este trabajo, cuando no se especifica se estiman 100 modelos y se elige el árbol de decisión con la menor tasa de clasificación incorrecta.

En general, el algoritmo para estimar un CART busca de entre los atributos el que mejor clasifica según sea el criterio de división ($a^* \forall a_i \in A$), entonces se define un nodo o criterio de decisión con el que se dividen los datos, entonces se vuelve a buscar el mejor atributo para cada subconjunto de datos, hasta que se cumpla el criterio de parada. Cada que se hace una división se evalúa el tamaño del cambio en la eficiencia (porcentaje de observaciones clasificadas correctamente), considerado el nivel de impureza como el criterio para dividir una rama del árbol que se mide a través de la desviación de clasificación (D) del

k-ésimo nodo final, donde C es el número de registros clasificados correctamente y NC su complemento.

$$D = \sum_k D_k = \sum_k -2n_k [n_k^C \ln(n_k^C) + n_k^{NC} \ln(n_k^{NC})]$$

Entre los criterios de parada en el proceso de construcción, se encuentran: profundidad (número máximo de nodos permitidos), número mínimo de observaciones clasificadas en el nodo hijo, la clasificación de todos los ejemplos en una clase (no recomendado ya que sobreestima el modelo y considera que los datos no tienen ruido)⁸.

Una forma de manejar el ruido es podando el árbol, es decir, eliminando los nodos finales que clasifican datos en condiciones muy específicas que podrían causar problemas de sobreajuste. El tamaño del árbol se define con base en la disyuntiva entre el tamaño de árbol y la desviación. Cada división depende de la decisión en el nodo anterior, entre más ramas tenga el árbol mayor es el orden de interacción y complejidad.

Para lo anterior, se eliminan los nodos que contribuyen en menor medida a reducir la desviación, es decir, los nodos más impuros. Determinar cuándo dejar de podar el árbol depende de la disyuntiva entre costo de complejidad y desviación; entre más grande el tamaño del árbol menor la desviación pero mayor el costo de complejidad.

$$C(T) = D(T) + \alpha \text{Size}(T)$$

El podado minimiza el costo de complejidad $C(T)$, donde $D(T)$ es la impureza del árbol de tamaño T, $\text{Size}(T)$ es el número de nodos finales y α es el término de penalización (Ledolter, 2013); a partir de esta función se tiene un costo para diferentes α cada una analiza la calidad de la predicción con datos con respecto al nivel de podado.

3.2.2. Medidas de ajuste

3.2.2.1. Tasa de clasificación incorrecta MC

La tasa de observaciones clasificadas incorrectamente (MC) por sus siglas en inglés (Miss Classification), es la proporción porcentual del número de ejemplos no clasificados

⁸ Los modelos CART presentados en este trabajo se estiman con la librería “tree” para R, y consideran una profundidad de 32.

correctamente (nd^{NC}) con respecto al total de datos en un conjunto de ejemplos específico $MC(U) = (nd^{NC}/nd) * 100$ donde nd es el número de registros en U . Mide la capacidad del modelo para generalizar un ejemplo con cierto grado de confianza (Rokach & Oded, 2008).

Se pueden calcular las MC en función de una muestra de observaciones, como la tasa de la muestra de entrenamiento $MC(DE)$, una tasa para los datos de prueba o fuera de la muestra $MC(DT)$ y una tasa para la muestra completa $MC(U)$.

3.2.2.2. Robustez de la tasa de validación cruzada CV

En cuanto a la robustez del modelo, se estima una tasa promedio de las tasas de clasificación incorrecta para distintas muestras de datos (tasa CV), técnica conocida como validación cruzada (Cross-Validation). Específicamente se generan cuatro sub-muestras de datos mutuamente excluyentes de forma aleatoriamente, se estima la MC en la primer sub-muestra para un modelo estimado con las tres sub-muestras restantes, este proceso se repite para cada sub-muestra y se promedian las cuatro tasas. Esto puede aumentar el error tipo I, sin embargo, métodos alternativos generan un error tipo II aún peor. El error tipo I se da cuando se rechaza una hipótesis nula que es verdadera (decir que algo que no es cierto), en este caso sería identificar una característica significativa de decisión cuando no existe, y el error tipo II se da cuando no se rechaza (se acepta) una hipótesis nula que es falsa (no decir algo que es cierto), en este caso sería no identificar una característica de decisión que si existe.

3.2.2.3. Precisión de la predicción

El error puede ser asociado a medidas de exactitud o precisión, la exactitud indica que tan cerca está el valor de una estimación con respecto a su valor real y la precisión indica que tan cerca está el valor de una estimación individual con respecto a los demás valores (Chapra, 2010). Este último es importante para problemas con una distribución de clases no balanceada.

Para problemas de clasificación binaria⁹, una medida de precisión indica “cuantos ejemplos clasificados como positivos son de hecho negativos”. Una medida de exactitud incluye tanto la sensibilidad (que tan bien el modelo puede reconocer ejemplos positivos) como la especificación (que tan bien puede reconocer ejemplos negativos), que podríamos ver como la regla y la excepción.

Para visualizar estos conceptos, utilizamos la matriz de confusión presentada en Rokach & Oded (2008), que muestra las propiedades discriminantes del modelo a través del número de ejemplos clasificados, esto suponiendo que las posibles calificaciones de clase son binaria (+: Positivos, -: Negativo, para efectos demostrativos).

Figura 3.4: Matriz de confusión

	Estimado (-)	Estimado (+)
Ejemplos (-)	A	B
Ejemplos (+)	C	D

Fuente: Rokach & Oded (2008)

El valor de A+B representa el número de ejemplos (observaciones) etiquetados con negativo en la muestra de datos, por lo que, aquellos C+D suman los ejemplos negativos. De tal forma que, si el modelo clasificara todos los ejemplos correctamente entonces C y B serían iguales a cero. Así mismo, los valores en la diagonal (A y D) son ejemplos clasificados por el árbol correctamente y aquellos que están fuera de la diagonal (C y B) ejemplos clasificados incorrectamente. Con base en la matriz de confusión, se estiman las siguientes medidas utilizadas para analizar bases de datos no balanceados o sesgados.

1. La tasa de exactitud (*ACR*)

$$ACR = \frac{(A + D)}{(A + B + C + D)}$$

2. La tasa de precisión (*P*)

$$P = \frac{D}{B + D}$$

⁹ Un problema de clasificación binaria, solo puede etiquetar un ejemplo como 1 o 0 (positivo o negativo)

3. La tasa de clasificación incorrecta (MC)

$$MC = \frac{(C + B)}{(A + B + C + D)}$$

En el cuadro 3.1 se presenta el porcentaje de clasificación según su tipo de error, con el objeto de identificar la regla o excepción y la precisión por clase {+: Positivos, -: Negativo}

Cuadro 3.1: Calculo de las diferentes medidas de ajuste

	Regla/ Excepción		Precisión	
	Estimado (-)	Estimado (+)	Estimado (-)	Estimado (+)
Ejemplos (-)	$A/(A+B)$	$B/(A+B)$	$A/(A+C)$	$B/(B+D)$
Ejemplos (+)	$C/(C+D)$	$D/(C+D)$	$C/(A+C)$	$D/(B+D)$

Fuente: Elaboración propia con base en Rokach & Oded (2008:37p)

Si la importancia recae en la clasificación de valores negativos, B sería el número de errores tipo I (número de veces que el árbol clasificó el ejemplo como negativo cuando no lo era) y C sería el número de errores tipo II (ejemplos que eran negativos y no se clasificaron como tal). Considerando que el error tipo 2 es mejor, si $C > B$ el modelo es mejor para clasificar valores negativos que positivos.

3.2.3. Ventajas y desventajas

Su mayor ventaja es su fácil interpretación, ya que no solo predice un valor o su tendencia sino que aporta información sobre la interacción entre las variables involucradas. Sin embargo su mayor desventaja es su inestabilidad la cual se desencadena también de utilizar modelos CART para analizar las relaciones entre variables. Ya que cambios en el conjunto de información, con base en el cual se estima el árbol, pueden tener diferentes efectos sobre los criterios de decisión.

Esto implica que al tomar muestras aleatorias de entrenamiento se pueden obtener modelos diferentes, algunos podrían ser más complejos y tomar una mayor cantidad de variables, considerar un factor en diferentes niveles o tener distintos valores como criterio de división. Este problema se potencializa en el análisis que involucran muchos factores o pérdidas de información por la eliminación de observaciones con valores nulos.

3.4. Proceso de modelado

Las técnicas presentadas tienen importantes ventajas sobre otros métodos econométricos multivalentes y estadísticos, entre las que se encuentran su capacidad de predicción e interpretación, sin que los datos de análisis tengan que cumplir con pruebas de normalidad o multicorrelación. Su principal problema, por otro lado, es el sobreajuste de los datos. En el caso de los árboles se suelen eliminar las últimas ramas para captar efectos de ruido, pero ligeros cambios en los datos de entrada pueden generar modelos con distintos resultados. De tal manera que podar el árbol resulta insuficiente para garantizar los resultados del modelo CART. Las redes neuronales por su parte, no aseguran haber encontrado el mejor modelo, debido a que el algoritmo de entrenamiento podría encontrar máximos locales sobre todo cuando se estima una gran cantidad de coeficientes.

Aun cuando los modelos estimen o clasifiquen con cierto nivel de precisión, el análisis de sensibilidad o de las reglas de conocimiento pueden variar significativamente. Dadas las características del conjunto de datos a analizar, descritas en el capítulo anterior, se pueden generar modelos con distintas estructuras, las cuales difieren con mayor grado conforme mayor sea el número de factores incluidos, aunado a una reducida cantidad de observaciones.

Con el objetivo de encontrar un modelo con un alto grado de generalización y estabilidad, se propone un proceso de modelado que conceptualmente es aplicable para ambos tipos. Consiste en estimar varios modelos con base en un re-muestreo con repetición aleatorio del 75% de los datos filtrados sin nulos, es decir, un Bootstrapping sobre un conjunto de datos dado, entonces, se filtran los modelos más estables a través de un análisis de frecuencia de variables penalizadas por el orden de importancia en el modelo.

La cantidad de observaciones depende de las variables que se incluyan en el análisis, como se describe en el capítulo anterior, pero la aportación e importancia de cada factor dentro del fenómeno de estudio depende de su papel en el modelo. Entonces se busca el mejor modelo dado un conjunto de información definido dentro del algoritmo RM-CSR-VN, pero a su vez las combinaciones de variables generadas por dicho algoritmo dependen del orden de las variables dado por el mejor modelo de una combinación anterior. La interrelación de este argumento se aprecia en el diagrama de flujo de la figura 2.5, donde el proceso de

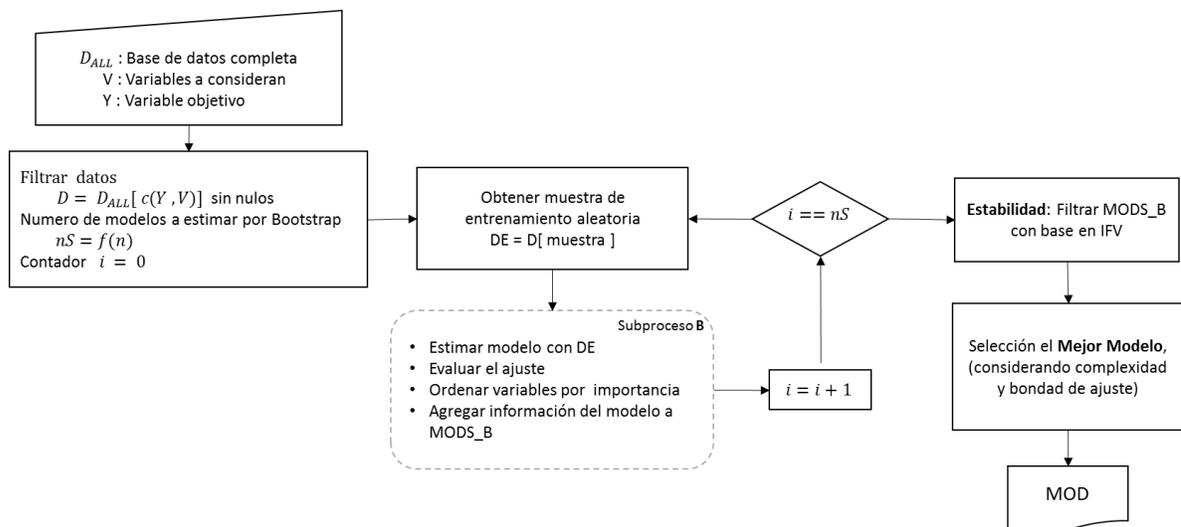
modelado se representa como el subproceso A, y en la figura 3.5 se muestra el flujo de este subproceso.

3.4.1 Bootstrapping

El proceso Bootstrapping selecciona muestras aleatorias para inferir el comportamiento promedio de una serie de modelos construidos a partir de un conjunto de datos.

Dada una combinación de factores específica, se obtiene un subconjunto $D \subset D_{ALL}$ filtrando las variables y eliminando los valores nulos, se construyen $nS = f(n)$ modelos dependiendo del número de observaciones n en D , cada estimación se realiza con un conjunto de entrenamiento DE correspondiente al 75% de los datos obtenidos aleatoriamente, para identificar cada modelo se fija una semilla inicial (aunque ésta bien podría ser aleatoria) la cual se incrementa en una cantidad definida para cada iteración. Por default, se toma $p=0.75$ correspondiente al 75% de los datos seleccionados aleatoriamente, n veces, controlando las semillas $s_n = s_{n-1} + ds$, donde ds es el aumento en el valor de la semilla, por default $ds = 5$ con una semilla inicial $s_1 = 5$. Se van guardando las características que permiten ver el modelo, además de las variables utilizadas ordenadas por importancia, y según sea el tipo los aspectos que indican que tan bueno es el modelo.

Figura 3.5: Diagrama de flujo del proceso de modelado bootstrapping



Fuente: Elaboración propia

La figura 3.5, muestra el proceso de modelado correspondiente al subproceso **A** del algoritmo RM-CSR-VN. Recibe una lista de variables a incluir, lo cual permite generar un subconjunto de información con respecto a la base de datos completa y el nombre de la variable objetivo, y regresa información del mejor modelo encontrado dada cierta combinación de factores, incluyendo las variables utilizadas por orden de importancia. A continuación se describen los procesos de filtrado mencionados en el diagrama anterior.

3.4.1. Estimación y evaluación del modelo (subproceso B)

La importancia de las variables depende de la técnica utilizada para el análisis, el subproceso B del diagrama de flujo en la figura 3.5 representa una función que ejecuta dicha técnica dependiendo del tipo de modelo predefinido. En este trabajo consideramos dos tipos RNAs y CARTs, aunque hay que mencionar que es posible aplicar la metodología a cualquier otro tipo de modelo multivalente siempre que las variables utilizadas tengan un ranking de importancia en los resultados del modelo.

3.4.1.1. Redes neuronales artificiales

Se estima un modelo perceptrón multicapa, considerando los siguientes parámetros generales:

1. Como variable objetivo el valor del coeficiente de GINI, entre 0 y 100 como se obtienen de la base de datos SWIID.
2. La función que define el número de modelos a estimar es $nS = \max\{50, n/10\}$
3. La cantidad de modelos a estimar puede incrementarse al considerar distintas estructuras, en este caso se estiman modelos con 4, 6, y 8 neuronas en la capa oculta.

Considerando las siguientes especificaciones en la estimación de cada modelo:

1. La muestra de entrenamiento es estandarizada, cada observación menos la media entre la desviación estándar, lo cual ayuda a encontrar el mejor conjunto de pesos en menor tiempo.
2. Los pesos del modelo se estiman con el algoritmo BackPropagation. Se utiliza la librería nnet de R, con una tasa de aprendizaje o decaimiento de 0.0001.

3. Un nivel de importancia mínimo aceptable es del 5% para mantener la variable en las siguientes iteraciones, esto permite que la red discrimine varias variables juntas.
4. Los pesos se ajustaron iterativamente hasta encontrar un mínimo de error $E = \sum_p \|G - \hat{G}\|^2$ para cada observación p; se alcanza la convergencia en menos de mil iteraciones.
5. Utiliza una función de activación f_h tipo logística para cada neurona, la capa de salida solo tiene una neurona $M = 1$ y una función identidad $f_m(x) = x$, la arquitectura del modelo se describe a continuación, donde w representa los pesos y b el sesgo.

$$G = f_m \left(b_m + w_h * f_h \left(b_h + \sum_{i=1}^H w_{ik} x_{ki} \right) \right)$$

Y la función de activación tipo logística es:

$$f_h = \varphi(v_p) = \frac{1}{1 + e^{-v_p}}$$

Las medidas de evaluación utilizadas para este tipo de modelos son:

1. Como las RNAs no están basadas en supuestos estadísticos, su evaluación consta de pruebas de ajuste: coeficientes de determinación $R^2 = 1 - \frac{E}{\sum_p \|y - \bar{y}\|^2}$ y su correspondiente ajustado, es decir, penalizado por el número de variables (nV)

$$R_A^2 = 1 - \left[\frac{n - 1}{n - nV - 1} \right] * (1 - R^2)$$

2. Se calcula la medida de ajuste para: la muestra de entrenamiento DE, la muestra de prueba DT y la muestra completa D.
3. También se calcula la importancia relativa de cada variable con base en sus pesos sinápticos y se genera una lista de los factores ordenados (VOI) según este criterio.

La información del modelo a guardar en MODS_B consiste en:

1. Parámetros para identificar la muestra de entrenamiento : semilla (s), lista de variables de entrada (VU), coeficiente de decaimiento (d)

2. Características de la estructura de la red: Número de Neuronas (N_N), Núm. de Variables Utilizadas (N_VU), variables ordenadas por importancia (VOI) descartando aquellas que aportan menos del 5% al modelo.
3. Los coeficientes de determinación ajustados para cada una de las muestras (R2A_DT), (R2A_DE) y para todos los datos (R2A_D).

3.4.1.2. Árboles de decisión clasificadores.

Se estiman los modelos, considerando los siguientes parámetros generales:

1. Como variable objetivo se utiliza una variable categórica CAMBIOG que representa aumentos o decrementos en el coeficiente de Gini de un año a otro. Los años donde no se generan cambios CAMBIOG = 'SC' son menos del 16%, lo cual dificulta encontrar una generalización que clasifique correctamente este tipo de cambio; considerando que el objetivo es analizar cuando si hay cambios de desigualdad, estas observaciones se descartan (para más detalle de su cálculo ver el apéndice A)
2. Con el objetivo de analizar los cambios de la desigualdad cuando aumenta la globalización comercial, se filtran las observaciones con cambios positivos en I_G de un año a otro y se descartaron aquellos registros donde no se presentaron cambios de desigualdad (ver apéndice B).

$$CI_G = 'C +' y CAMBIOG <> 'SC'$$

3. La función que define el número de modelos a estimar es $nS = \min\{500, n/4\}$

Considerando las siguientes especificaciones en la estimación de cada modelo:

1. Formalmente, el esquema del árbol $B(A \cup y)$ y el dominio de sus atributos se muestran a continuación.

$$B(A_{t-1} \cup C_{GINI_t}), \quad |dom(C_{GINI_t})| = \{C+, C-\}$$

$$A_{t-1} = \{ idRegion, idNivel, OCDE, V \}$$

$$|dom(idNivel)| = \{A, B, C, D, E\}^{10},$$

¹⁰ Correspondientes a los valores en la tabla NivelesIngreso del cuadro 4

$$| \text{dom}(idRegion) | = \{1,2,3,4,5,6,7\}^{11}$$

$$| \text{dom}(OCDE) | = \{0,1\}$$

2. Se utiliza la librería “tree” de R, para estimar los criterios de división
3. Todos los árboles se podan considerando un alfa igual a la mediana.

Las medidas de evaluación utilizadas para este tipo de modelos son:

1. La tasa de clasificación incorrecta MC para cada modelo en función de las observaciones fuera de la muestra $MC(DT)$, de entrenamiento $MC(DE)$, de la muestra completa $MC(D)$
2. La tasa de clasificación incorrecta considerando todos los datos $MC_{CN} = (D_{ALL})$.
3. La tasa de precisión TP y una tasa cross-validation (CV)

La información del modelo a guardar en MODS_B consiste en:

1. Parámetros para identificar la muestra de entrenamiento: semilla (s), lista de variables de entrada (VU), número de observaciones en el análisis (n) y número de variables (nV) de entrada.
2. Características del modelo CART: Número de nodos de decisión (N), Núm. de nodos finales (NF), el Núm. de variables utilizadas en árbol (NV) y la lista de variables ordenadas conforme el nivel en el que aparecen en el árbol (VOI).
3. Las medidas de ajuste descritas en el punto anterior (MC_DT, MC_D, MC_DE, MC_CN, CV, TP)

3.4.2. Estabilidad

El modelo con la mejor medida de ajuste podría ser también el modelo con menor probabilidad de generarse, es decir, contiene la combinación de factores más difícil y poco común de encontrar. Debido a que el objetivo de la tesis es entender cómo interactúan los factores en conjunto para generar cambios en la desigualdad, es importante abordar el problema de inestabilidad con respecto a los factores incluidos.

¹¹ Correspondientes a los valores en la tabla Regiones del cuadro 5

Consideramos que un modelo es estable cuando los resultados y los criterios de importancia entre variables utilizadas no varían notablemente utilizando los mismos datos de entrada, pero considerando diferentes muestras de entrenamiento con distintas semillas (Bootstrapping).

La estabilidad de los criterios depende de la claridad de las relaciones entre los factores, por lo que si las relaciones son suficientemente fuertes y estables, los modelos aunque diferentes son muy parecidos, por ejemplo, en los modelos CART rara vez el criterio raíz cambia, igual que los criterios en los niveles más altos, y en las redes los mismos factores se mantienen en los niveles de importancia más altos.

Par lo cual se propone filtrar los modelos más estables considerando la frecuencia de factores aunado a su nivel de importancia. Como medida de estabilidad utilizamos un índice *IFV* basado en la frecuencia con la que aparecen las variables en los modelos, penalizado por el nivel del criterio en el que aparezca (n), es decir, por cada vez que aparece la i -ésima variable en un modelo su índice se incrementa proporcionalmente a su nivel importancia; si es parte de la primera condición, el incremento es de uno; si forma parte de la segunda condición es de un medio, si la variable aparece más de una vez se suman sus proporciones.

$$\sum_{j=1}^M \sum_{i=1}^{v \in M_j} IFV_i = IFV_i + \frac{1}{n_j}$$

El factor con el índice más alto será aquel que aparezca mayor número de veces como el más importante, siendo este el criterio más estable el principal.

El filtrado de modelos (MODS_B) consiste en descartar aquellos que no consideren el factor con el índice más alto como el mejor, por ser los mayormente inestables, después se descartan los que no consideran como criterios importantes en los niveles 2 y 3 los factores con un índice entre los primeros cuatro mejores, y por último se suprimen aquellos que no contienen los factores con índices mayores a su media.

Entonces se selecciona el mejor modelo de entre los más estables, con base en su complejidad, presión, exactitud al clasificar o coeficiente de determinación, según sea el caso. A continuación se presenta el proceso de modelado detallado para ambos tipos de técnicas.

3.4.3. Mejor modelo

Una vez descartados los modelos inestables, con menor probabilidad de suceder, se puede seleccionar el mejor modelo. Muchos de los modelos con altos índices de predicción o clasificación, resultan demasiado grandes, es decir, con muchos factores, neuronas o criterios de decisión. Esto complica la interpretación de los resultados, razón por la cual se concentra un índice que capta el desempeño del modelo pero lo penaliza por su complejidad.

3.4.3.1. Complejidad

La interpretación de un árbol depende de su complejidad, comúnmente relacionada al número de nodos (N), en este trabajo se propone una medida que también considera el número de nodos finales (NF), el número de variables (NV) y penaliza los modelos que repiten variables en los criterios, como se muestra a continuación.

$$C_{CART} = \frac{NV + N}{NF} + (N - NF - NV)$$

En un modelo perceptrón multicapa, la interpretación de los resultados del análisis de sensibilidad depende del número de variables y más que de la estructura. Por lo que la complejidad de las redes se mide como se muestra a continuación.

$$C_{RNA} = \frac{1}{nV + NN}$$

3.4.3.2. Desempeño

Dado que la construcción de los árboles se realiza con base en los datos de entrenamiento, para cada modelo se cumple $MC(DT) > MC(DATA_{IN}) > MC(DE)$. Como medida de eficiencia, se considera un promedio ponderado de las tasas de clasificación incorrecta calculadas a partir de los datos de entrenamiento MC_{DE} , de prueba MC_{DT} y con toda la muestra MC_D , que representan la tasa de especificación incorrecta.

$$MC = (MC_{DT} + MC_{DE} + 2 * MC_D) / 4$$

Con el objetivo de encontrar el modelo más preciso, más exacto y menos complejo posible, primero se estandarizan los valores para realizar una comparación entre estas medidas: se utiliza el complemento de la tasa de precisión, es decir, una tasa de impresión $PI = 1 - TP$, y se utiliza una correspondencia proporcional del índice de complejidad C

con respecto a un rango de valores entre el mínimo de los mínimos y el máximo de los máximos de las PI y las MC ¹²

$$C_E = \left[\left(dif_R * \frac{(C - \min(C)) * 100}{dif_C} \right) / 100 \right] + \min R$$

El desempeño de una red neuronal, se obtiene con el promedio ponderado de los coeficientes de determinación ajustados, por muestra de datos. La complejidad se estandariza con base en los límites de R_A^2 de la misma manera que para los árboles.

$$R_A^2 = (RA2_{DT} + RA2_{DE} + 2 * RA2_D) / 4$$

Entonces el mejor modelo clasificador, es el que tenga menor índice de eficiencia calculado con el promedio de las tres medidas anteriores

$$EF_{CART} = (PI + MC + C_E) / 3 .$$

El mejor modelo perceptrón multicapa, es el que tenga el mayor índice de eficiencia

$$EF_{RNA} = (3 * R_A^2 + C_E) / 4 .$$

3.5. Implementación

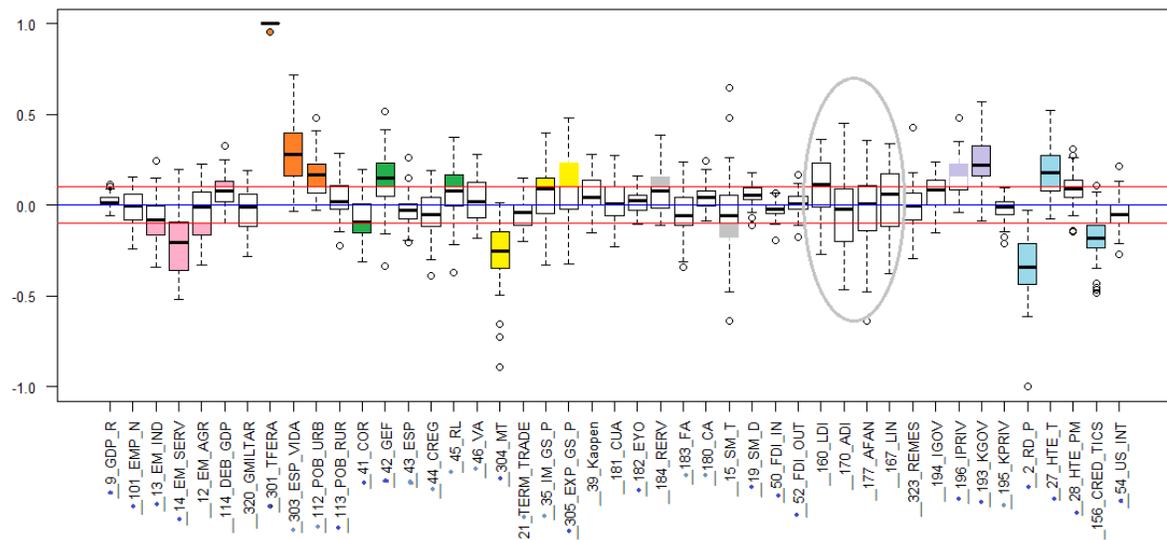
Implementación, descripción general de las estimaciones para ambas técnicas de minaría de datos generadas en el proceso de modelado descrito.

3.5.1. Redes neuronales artificiales (RNAs)

Al implementar el algoritmo de re-muestreo dirigido, se descartan pocos factores, por lo que los modelos de árboles se estiman a partir de una muestra relativamente pequeña. Aun así, los modelos muestran coeficientes de determinación superiores al 87%. En la figura 3.6 se muestra la distribución de la importancia de las variables por modelo estimado.

¹² $dif_R = (maxR - minR)$, donde $minR = \min(\min(PI), \min(MC))$, $maxR = \max(\max(PI), \max(MC))$
 $dif_C = (\max(C) - \min(C))$,

Figura 3.6: Importancia por variables para 50 modelos RNA



Fuente: Elaboración propia con base en el análisis de sensibilidad de las RNAs

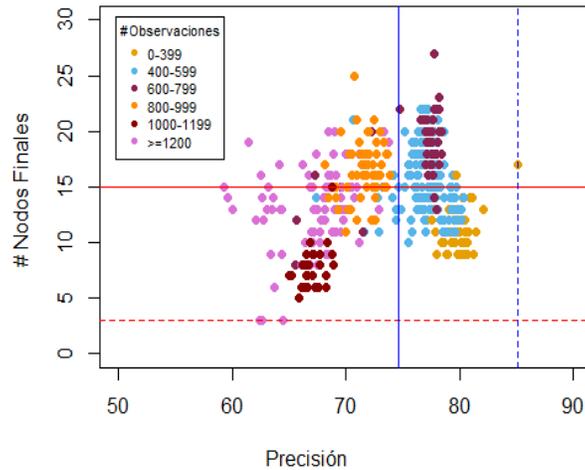
En la figura 3.6, resalta que la importancia de variables resultado de políticas sociales enfocadas en la salud y control demográfico tiene efectos en la desigualdad independientemente de las variables de globalización o factores externos, así mismo, es importante notar la alta variabilidad de los factores financieros. En el siguiente capítulo, se describe detalladamente el mejor modelo RNA encontrado, a partir del cual se identifican los factores más importantes en la determinación de este fenómeno.

3.5.2. Árboles clasificadores (CART)

Del proceso de modelado con Bootstrapping se encontraron 650 modelos estables para el conjunto de información a partir del cual se construyeron, cada conjunto correspondientes a distintas combinaciones de las 42 variables presentadas, que a su vez definen diferentes cantidades de observaciones.

De estos modelos, se selecciona el que mejor clasifica (más eficiente) con menor complejidad y a la vez mayor capacidad de generalización, al estimar con cierta eficiencia observaciones fuera de la muestra e incluyendo las 1241 observaciones con valores nulos. En un primer filtro, se descartan los modelos más complejos y menos precisos, con base en los valores medios de las medidas de ajuste correspondientes; quedando 202 modelos que cumplen la condición $TP \geq 74.6$ & $NF \leq 15$.

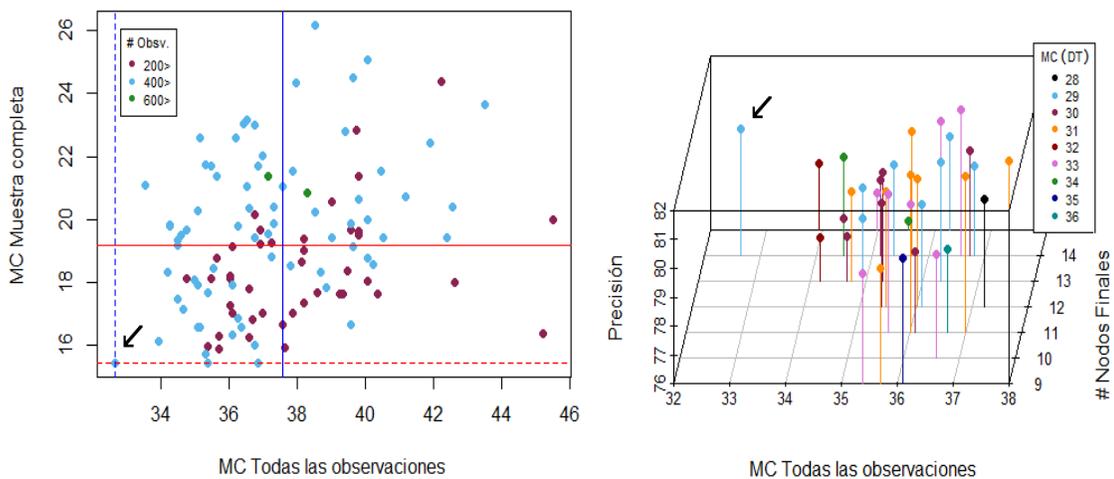
Figura 3.7: Precisión vs Complejidad de los modelos CART



Fuente: Elaboración propia con base en los CARTs estimados

La figura 3.7, muestra la tasa de precisión y nodos finales de los 650 modelos CART contruidos a partir de los diferentes conjuntos de observaciones según las variables incluidas, el cuadrante derecho inferior contiene los mejores 202 modelos filtrados. En un segundo filtro se considera la capacidad de los modelos para clasificar observaciones que incluyen valores nulos –medida por la MC_{CN} donde 63 modelos cumplen la siguiente condición $MC(DATA_{IN}) \leq 19.2 \ \& \ MC(DATA_{ALL}) \leq 37.6$, los límites dependen de los valores promedio de las medidas de ajuste.

Figura 3.8: Precisión vs Complejidad de los modelos CART filtrados



Fuente: Elaboración propia

En la primera gráfica de la figura 3.8, se muestra la tasa de clasificación incorrecta de la muestra MC_D y para todas las observaciones MC_CN de los 202 modelos, las líneas discriminan los peores modelos, dejando fuera los modelos construidos con más de 600 observaciones. Los mejores 63 modelos en el cuadrante inferior izquierdo también se visualizan en el segundo gráfico de la misma figura, destacando la precisión (TP) y tamaño del árbol (NF). En ambos gráficos destaca el modelo más eficiente considerando las diferentes medidas de ajuste (señalado con la flecha).

En la figura 3.7, se aprecia la relación entre tres las medidas anteriores para cada uno de los modelos que sobreviven al filtrado por frecuencia de variables. De tal manera, que si nos interesa el árbol más simple (intuitivo), seleccionaremos aquel con menor C_E , si la precisión es lo importante sería el que tenga menor PI, y si es la exactitud buscaríamos el correspondiente a la menor MC. En este trabajo se toma el modelo con el menor promedio de las tres medidas como el mejor árbol, es decir, la medida de eficiencia EF_{CART}

CAPÍTULO IV. ANÁLISIS DE RESULTADOS

En este apartado, se describen y analizan los resultados del mejor modelo tanto el perceptrón multicapa como el árbol de decisión clasificador, sobrevivientes del proceso de modelado descrito.

A partir de los cuales, se identifica el incremento en la tasa de fertilidad en adolescentes como el factor con mayor efecto sobre los aumentos de desigualdad, seguido de la inversión en investigación y desarrollo, el gasto militar y la esperanza de vida. Siendo los factores macroeconómicos internos mayormente importantes en la determinación de la desigualdad. Sin dejar de lado la importancia de los términos de intercambio, exportaciones e importaciones que se mantienen al filtrar las ocasiones que presentaron aumentos de globalización. El modelo CART encontrado identifica 14 reglas que muestran en que situaciones la desigualdad aumenta o no. En donde el primer criterio a considerar es la estructura del empleo de los países, así como factores comerciales como los términos de intercambio, el capital de gobierno, el crecimiento de la población y las exportaciones entre otros.

4.1. Modelo de perceptrón multicapa de RNA

Este modelo perceptrón multicapa considera $K = 42$ variables y $P = 376$ observaciones como datos de entrada denotadas por la matriz X y el $GINI$ como el valor de la salida. Una observación es una realización, es decir, un conjunto de 42 datos (uno por cada variable) para un país en un año determinado.

$$X_{PK} = \{GDP^{13}, POB_{URB}, \dots, US_{INT}\} = \begin{pmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{P1} & \cdots & x_{PK} \end{pmatrix}$$

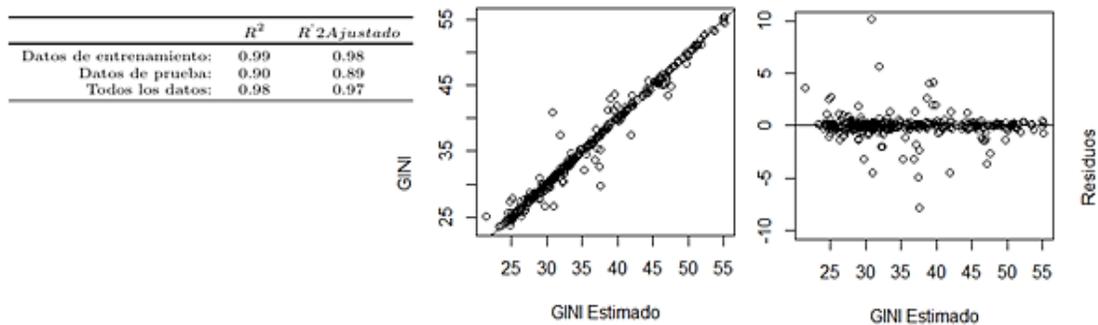
Para analizar la desigualdad, la mejor arquitectura estimada tiene dos capas, la capa oculta tiene $H = 6$ neuronas y utiliza una función de activación f_h tipo logística para cada una, la capa de salida solo una neurona $M = 1$ y una función identidad $f_m(x) = x$, en total la arquitectura del modelo tiene 289 parámetros incluyendo pesos w y sesgos b .

¹³ GDP es la variable $k=1$ y representa un vector con los n datos disponibles.

$$G = f_m \left(b_m + w_h * f_h \left(b_h + \sum_{i=1}^H w_{ik} x_{ki} \right) \right)$$

Como las redes neuronales artificiales no están basadas en supuestos estadísticos, su evaluación consta de pruebas de ajuste; se utiliza el método de Cross-Validation para evitar el principal problema de este tipo de modelos “el sobreajuste”, es decir, que se tenga muy buenos resultados con la observaciones con las cuales fueron estimados los pesos, pero que sean muy deficientes cuando se agregan datos desconocidos o fuera de la muestra.

Figura 4.1: Ajuste de la predicción y dispersión de errores.



Fuente: Elaboración propia con datos del SWIID y los estimados con RNA

En la tabla de la figura 4.1 se tienen los coeficientes de determinación $R^2 = 1 - \frac{E}{\sum_p \|y - \bar{y}\|^2}$ y ajustados –penalizado por el número de variables (n) $R_A^2 = 1 - (1 - R^2) \left[\frac{n-1}{n-K-1} \right]$ - para los datos de entrenamiento, para el 25% de los datos de prueba que no se incluyeron en el aprendizaje y para los 376 datos totales. Donde se aprecia que el coeficiente de determinación más bajo es 89 que corresponde al calculado con datos fuera de la muestra de entrenamiento.

4.2. Análisis de sensibilidad de la RNA

Los modelos RNA son altamente eficientes al estimar valores, además de que el mismo algoritmo de entrenamiento identifica la forma funcional de la relación. Dicha característica, permite estimar valores de desigualdad dados ciertos valores de las variables de entrada con alto grado de eficiencia, sin embargo, la relación funcional carece de interpretación práctica o directa, considerando que en el análisis de fenómenos sociales dicha interpretación resulta

clave para describir y proponer políticas públicas, se aplica un análisis de sensibilidad que permite identificar los factores con mayor impacto en la dinámica de la desigualdad.

Para conocer el impacto y dirección del efecto que tiene cada factor sobre la desigualdad a partir de un modelo con relaciones no lineales y que parecen no tener sentido interpretable, se aplican el método de perturbación y el del análisis de importancia relativa.

4.2.1. Método de perturbación

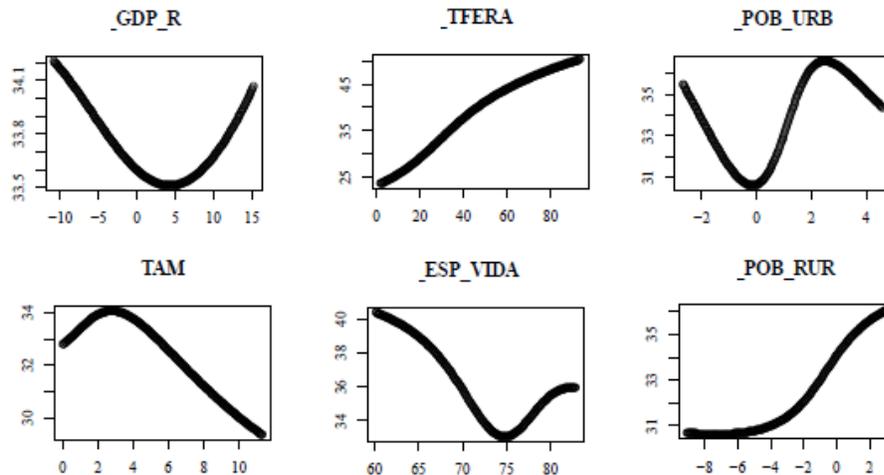
Para estimar la contribución de cada variable se utiliza el método de perturbación, que consiste en estimar el valor de salida, en este caso del Gini, a través del modelo de redes neuronales MLP con los mismos datos de entrada excepto el valor de la variable en cuestión; este método analiza el efecto en el coeficiente GINI ante cambios de una variable manteniendo las demás fija.

Se estima la desigualdad para los 376 datos manteniendo fijos los valores de todas las observaciones menos los de la variable que se está analizando $X'(x_i, \dots, x_{-p}, \dots, x_k)$. Los valores de x_j se estiman con base en el rango de los posibles valores que puede tomar el factor en cuestión, tomando como límites sus valores mínimo y máximo $\widehat{G}_k = f_M(\bar{x}_i, \dots, x_{-p}, \dots, \bar{x}_k)$. Para cada variable se estima y grafica su efecto con respecto al Gini, las gráficas se desarrollaron bajo el método de perturbación agregando ruido en cada variable y graficando las estimaciones bajo diferentes condiciones.

4.2.1.1. Factores macroeconómicos

Como se muestra en la primera gráfica de la figura 4.2, el crecimiento económico (GDP) tiene efectos mixtos; la desigualdad aumenta cuando la tasa es mayor al 5% y disminuye cuando son menores, sin embargo, la magnitud del cambio es muy pequeña (del 33.5 al 34.1) por lo que el efecto es casi nulo. El efecto es igualmente despreciable cuando los países abarcan menos del 2% del territorio mundial es importante considerar que pocos países son tan grandes, por lo que este no es un factor determinante.

Figura 4.2: Efecto de variables macroeconómicas sobre el coeficiente de Gini

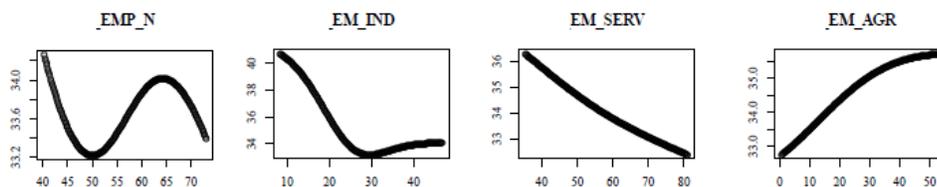


Fuente: Elaboración propia con el método de perturbación

La relación que muestra la red del Gini con respecto al crecimiento de la población urbana es un tanto más compleja, si las tasas son negativas o muy altas (mayor al 2%) podría reducir la desigualdad. El número de embarazos de adolescentes por cada 1000 mujeres en el país (TFERA) es según la red el factor determinante más importante, asimismo la esperanza de vida que según la gráfica capta el problema de las poblaciones viejas pues indica un aumento desigualdad a partir de 75 años.

El análisis del empleo (figura 4.3) indica una relación no lineal EMP_N aunque la magnitud del efecto se contrarresta debido a la cantidad de empleados en los diferentes sectores. El aumento de la desigualdad se debe a una mayor proporción de empleo en el sector agrícola y la disminución corresponde a la del sector industrial y de servicios.

Figura 4.3: Efecto del empleo sobre la desigualdad

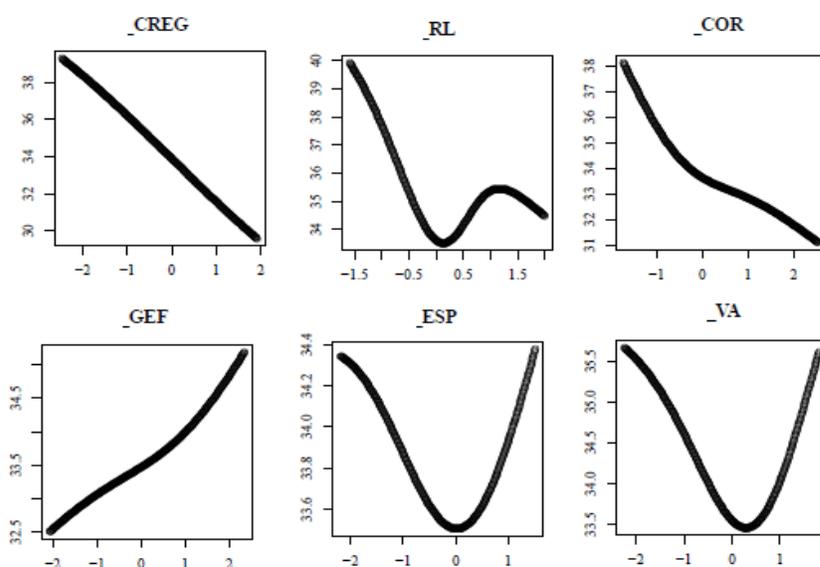


Fuente: Elaboración propia con el método de perturbación

Lo anterior es consistente con el argumento de Kuznets, referente al fenómeno de urbanización, así como el efecto “Skill Premium” si consideramos que el sector servicio demanda empleo altamente calificado.

Los índices de gobernanza corresponden a valores escalados a una desviación estándar están aproximadamente entre -2.5 (débil) y 2.5 (fuerte), donde los valores más altos corresponden a mejores niveles de gobernanza (Kaufmann et al., 2010). Los tres primeros índices en la figura 4.4 específicamente: la calidad de sus regulaciones, el estado de derecho y el control de la corrupción, indican que los países con mejores prácticas de gobernanza tienen menor desigualdad; aunque la dirección del efecto es adversa o combinada para los tres índices restantes, su magnitud es muy pequeña.

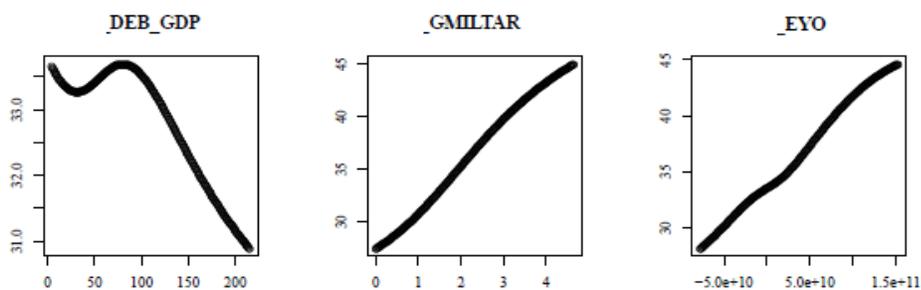
Figura 4.4: Efecto de los índices de gobernabilidad sobre el coeficiente de Gini



Fuente: Elaboración propia con el método de perturbación

El papel del gobierno también se valora a través de sus decisiones políticas y financieras; entre más dinero se gasta en el sector militar mayor es la desigualdad (figura 4.5). El efecto de la deuda externa es mínimo, pero muestra que niveles menores al 50 % del PIB ayudan a la desigualdad aunque aumenta con niveles mayores, y disminuye cuando la deuda sobrepasa el PIB situación poco probable.

Figura 4.5: Efecto de otras variables macroeconómicas sobre el coeficiente de Gini



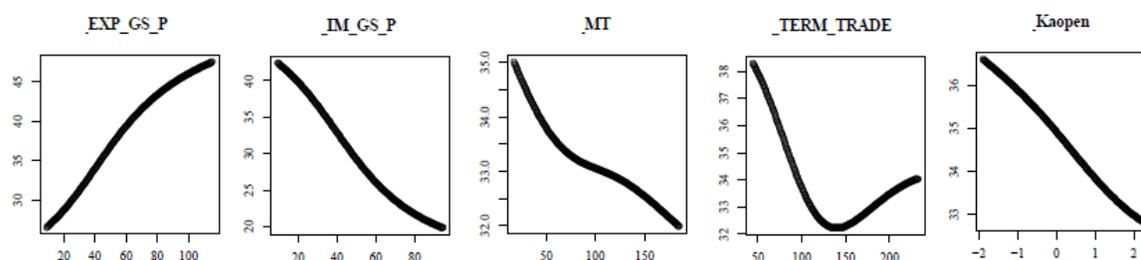
Fuente: Elaboración propia con el método de perturbación

En esta parte se incluye la cuenta “Errores y Omisiones” de la balanza de pagos, debido a que podría considerarse como evasión fiscal o falta de eficiencia del gobierno cuando las cantidades son considerables; el modelo capta una relación positiva.

4.2.1.2. Factores Comerciales

Un mayor porcentaje del PIB en exportaciones aumentan la desigualdad, las importaciones por el contrario la disminuyen, sin embargo, el efecto positivo gana por poco, como se comprueba al ver el efecto del intercambio de mercancías (MT); el Kaopen también disminuye la desigualdad.

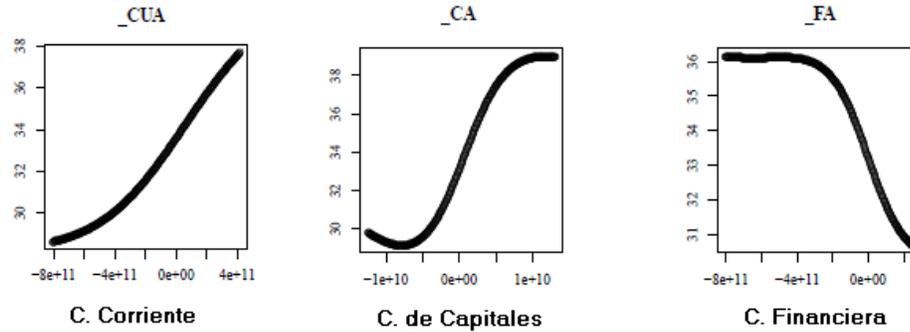
Figura 4.6: Efecto de los factores comerciales sobre la desigualdad



Fuente: Elaboración propia con el método de perturbación

Los términos de intercambio mide el poder adquisitivo de los habitantes de un país en comparación con los ciudadanos de otros países. Conforme las importaciones se vuelven más baratas que las exportaciones la desigualdad disminuye, aunque existe un punto en el cual el efecto cambia de dirección -cuando el índice es mayor a 150 como se muestra en la cuarta grafica de la figura 4.6.

Figura 4.7: Efecto de las cuentas generales de la balanza de pagos



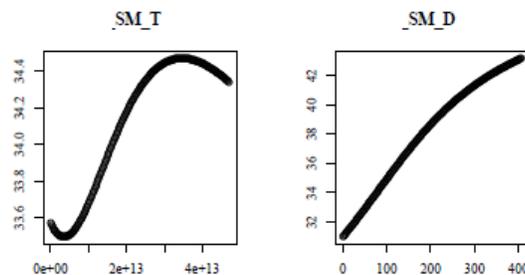
Fuente: Elaboración propia con el método de perturbación

En cuanto a la balanza de pagos una cuenta corriente positiva aumenta la desigualdad, considerando el análisis de las exportaciones e importaciones, este efecto puede deberse a los servicios, rentas o transferencias (ver figura 4.7). La cuenta de capitales positiva por el contrario aumenta la desigualdad, lo cual podría ser por la concentración de capital y deficiente distribución del ingreso. Mantenido un superávit en la cuenta financiera la desigualdad disminuye. En general se puede concluir que la globalización comercial reduce la desigualdad, aunque su efecto no es muy grade.

4.2.1.3. Factores financieros

Aunque el tamaño del mercado accionario no parece tener un efecto importante sobre el Gini como se muestra en la primera gráfica de la figura 4.8, el volumen de acciones negociadas nacionales con respecto al valor del total de acciones SM_D si tiene una importante participación, de tal forma que entre más bursátiles son las acciones domésticas mayor es el coeficiente de desigualdad.

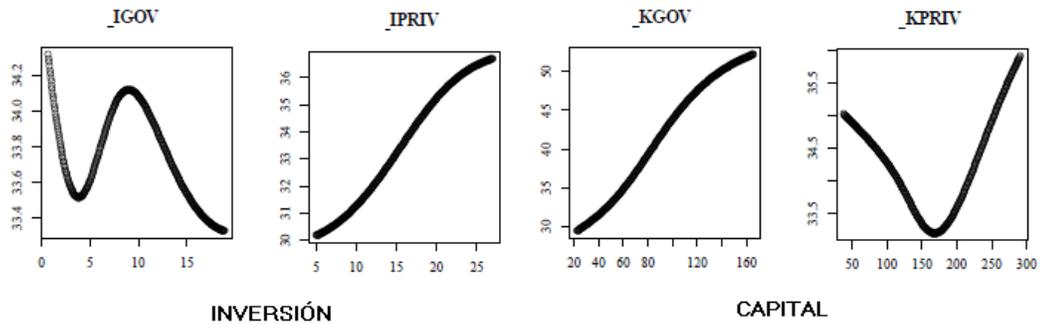
Figura 4.8: Efecto del mercado accionario



Fuente: Elaboración propia

En cuanto a la inversión, países con altos porcentajes de capital en el sector público o inversión en el sector privado tienen mayor nivel de desigualdad; la inversión en el sector público y el capital del sector privado tienen una relación compleja pero su efecto es pequeño.

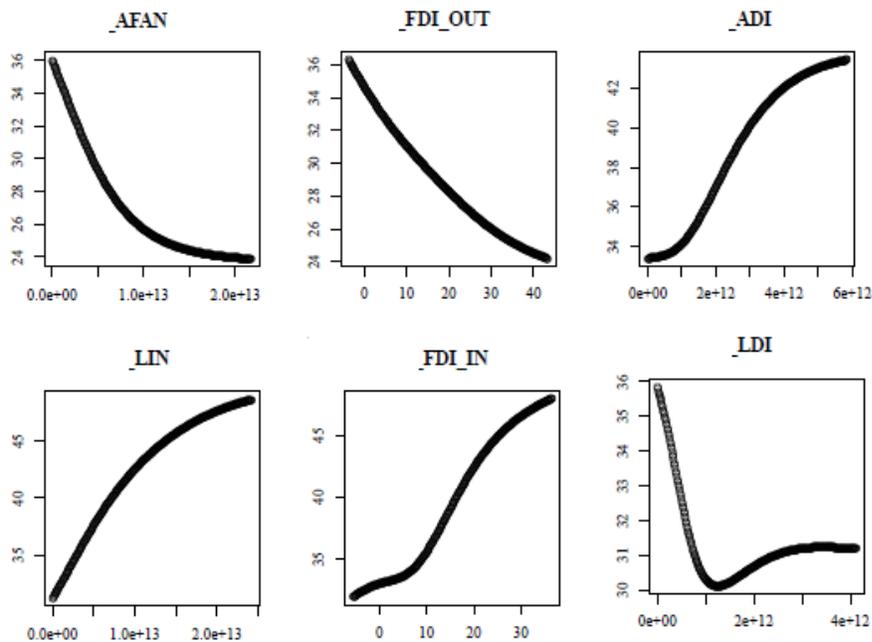
Figura 4.9: Efecto de la distribución del capital e inversión sobre la desigualdad



Fuente: Elaboración propia con el método de perturbación

La red neuronal también capta efectos de factores financieros (figura 4.10), la desigualdad aumenta en países con flujos de inversión entrantes, pero disminuye si los flujos salen del país (FDI_IN/OUT).

Figura 4.10: Efecto de las cuentas generales de la IIP sobre la desigualdad



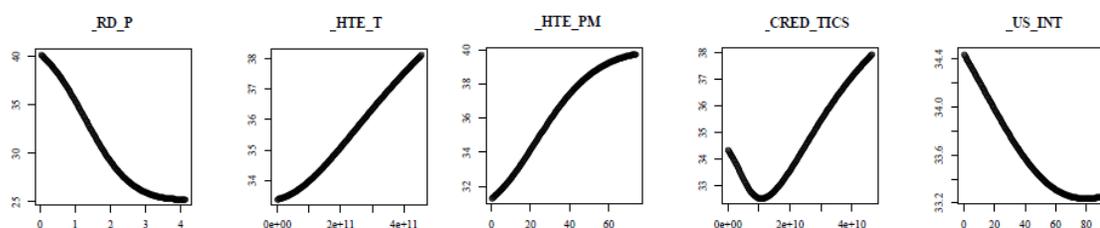
Fuente: Elaboración propia con el método de perturbación

Estos resultados son consistentes, puesto que sucede lo mismo al analizar la posición deudora o acreedora del país a través de las cuentas IIP medidas por los activos y pasivos netos (AFAN, LIN). Sin embargo, el efecto de los activos y pasivos de inversión directa es completamente opuesto, por un lado estas variables no incluyen activos especulativos a diferencia de las cuentas netas; por otro lado, la inversión extranjera directa FDI no contempla el capital y préstamos que las filiales hacen a sus matrices, cantidades que si contemplan las posiciones ADI y LDI (OECD, 2014).

4.2.1.4. Factores tecnológicos

Exportar alta tecnología es una señal de desarrollo, sin embargo, las dos variables utilizadas para medir este aspecto (HTE_T, HTE_PM) muestran un efecto negativo sobre la desigualdad, así mismo, grandes cantidades en créditos para tecnologías de información aumentan la desigualdad –consistente con el argumento de Jaumotte et al. Aunque el acceso a la información medida con el número de usuarios de internet efectivamente disminuye la desigualdad, su efecto es pequeño. Una de las variables que mayor efecto tienen es el porcentaje con respecto al PIB del gasto en inversión y desarrollo (RD_P) la cual reduce la desigualdad.

Figura 4.11: Efecto del mercado accionario sobre la desigualdad



Fuente: Elaboración propia con el método de perturbación

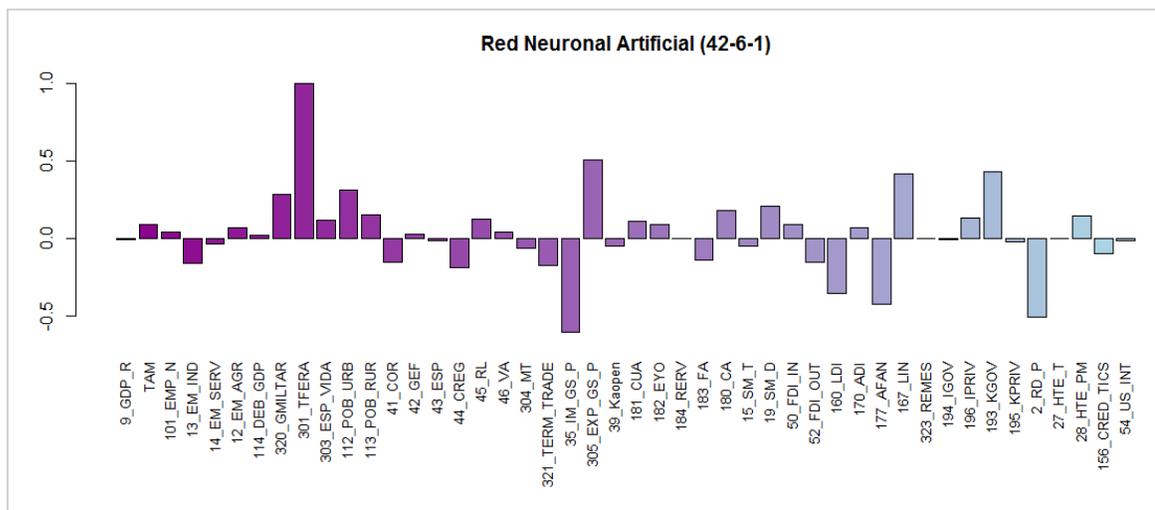
Lo anterior sugiere que el proceso de generar tecnología es mejor para la desigualdad que el proceso de producción y exportación, debido a que aumenta el salario y oferta de empleo calificado a su vez, empeora el salario y oferta del empleo no calificado o es sustituido en los procesos manufactureros.

4.1.2.2. Importancia de los factores que afectan la desigualdad en el modelo.

La metodología de importancia relativa de las variables se basa en la interacción entre los pesos sinápticos. La importancia relativa de la k-esima variables Q_k , se determina con base en la interacción entre los parámetros del modelo W y el porcentaje de influencia de una variable k sobre la de salida, como se muestra a continuación $Q_k = \frac{\sum_H R_{nk}}{\sum_N \sum_H R_{nk}}$, $R_{nk} = \frac{w_{nj}}{\sum_N w_{kj}} v_{nk}$ donde R_{ik} es el valor de importancia relativa de la variable k con respecto al valor total de entrada para la neurona n (Garson, 1991).

Para verificar la relación de los diferentes factores es necesario analizar los efectos en su conjunto, para lo cual se utiliza el método de importancia relativa que a su vez nos permitirá identificar las variables menos significativas según el MLP.

Figura 4.12: Análisis de sensibilidad



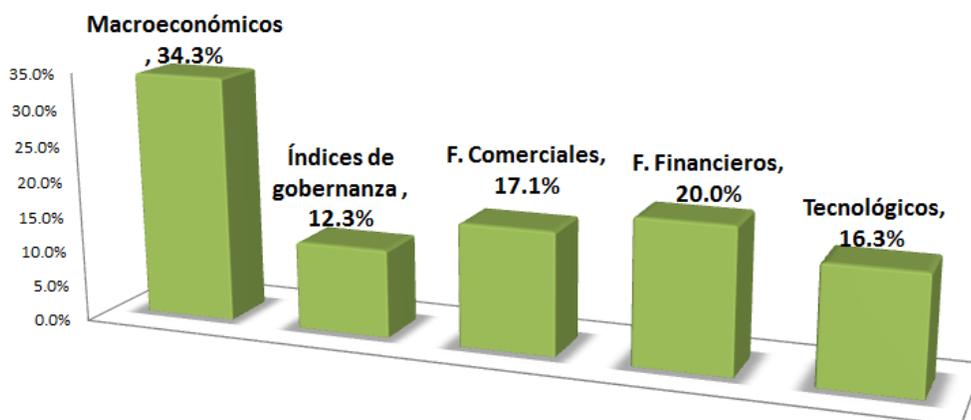
Fuente: Elaboración propia en R con base en Garson (1991)

En la figura 4.12 se aprecia el nivel de importancia de cada variable en el simplex y si su contribución es positiva o negativa con base en el MLP estimado.

El modelo sobrestima la importancia de la inversion extranjera directa según el promedio de los modelos de RNA estimados (figura 3.5). Sin embargo, la posición deudora o acreedora de los países es un factor determinante para la desigualdad, y efecto que se contrarresta con una balanza de pagos equilibrada.

En cuanto a los factores tecnológicos, se identifica claramente el gasto en inversión en investigación y desarrollo I&D como uno de los factores más importantes para reducir la desigualdad, así como, el crédito a las TICs aunque no en la misma magnitud. En cambio la exportación de altas tecnologías parece ser un fenómeno que aumenta la desigualdad.

Figura 4.13: Importancia de las variables para las dimensiones de la globalización



Fuente: Elaboración propia con base en los resultados del análisis de sensibilidad.

Es importante considerar que la relación de la globalización con la desigualdad depende de diversos factores macroeconómicos, es decir, de las características individuales de los países, en un 34%, y de los índices de gobernanza en 12%, así mismo la desigualdad debido a la globalización depende en un 20% la globalización financiera, en 17% de la comercial y en 16% de la tecnológica.

El incremento en la tasa de fertilidad en adolescentes es el factor con mayor efecto sobre los aumentos de desigualdad, seguido de la inversión en investigación y desarrollo, el gasto militar y la esperanza de vida, sin embargo, estos factores son características macroeconómicas internas de los países (o relacionada al progreso tecnológico como la inversión en RD) pero no son precisamente parte de los procesos de globalización. Otro conjunto de factores macroeconómicos importantes son los poblacionales, aunque la movilidad del empleo entre sectores y la población a zonas urbanas es incentivada por los cambios estructurales derivados de la globalización. Por otro lado, se identifica la importancia de factores meramente comerciales como los términos de intercambio, el intercambio de mercancías y las exportaciones e importaciones. De los últimos dos grupos

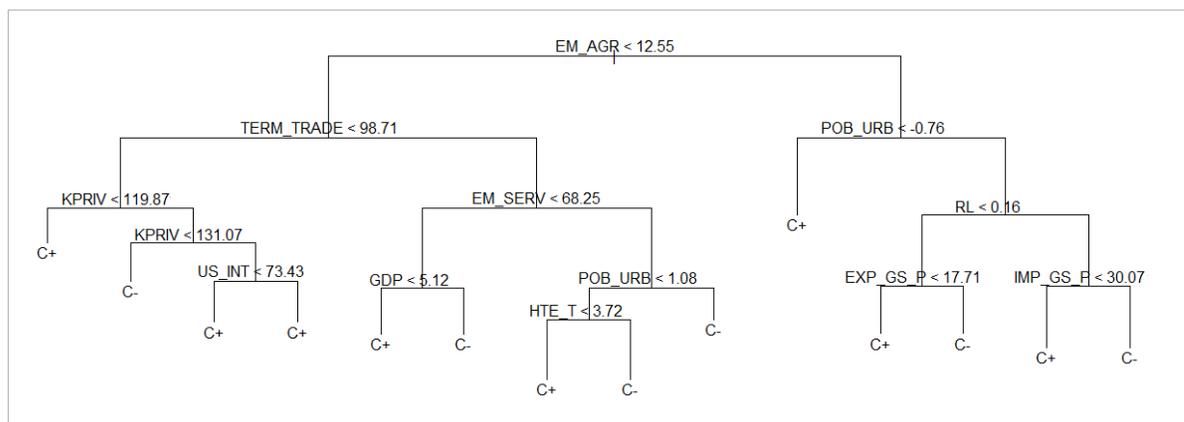
de factores mencionados algunos de ellos tiene efectos positivos y otros negativos sobre los cambios en la distribución del ingreso, para conocer la dinámica que genera estos efectos es necesario un análisis que involucre las relaciones entre los factores, como el que se muestra en el siguiente apartado.

4.3. Modelo árbol de decisión clasificador

Se utiliza esta herramienta, con el objetivo de analizar las interrelaciones entre los factores. A diferencia de la red neuronal descrita que permite un análisis de efectos individuales, los arboles describen situaciones en las que aumentos de globalización aumentan la desigualdad en países con características específicas y cuando el efecto por el contrario permite decrementos de desigualdad, además de indicar los rangos o parámetros que definen el punto de cambio.

De los modelos encontrados en el proceso anterior, el mejor modelo se aprecia en la figura 4.14, este contiene mayor número de variables comerciales. En cuanto a la estabilidad, el empleo en el sector agrícola es el criterio más importante en 77 de los 118 modelos construidos; su interacción con los términos de intercambio se presenta en 35% de los casos; los primeros nodos de decisión del árbol aparecen en más del 10% de los modelos; además de que los factores incluidos en el modelo también son considerados en los demás, es decir, cada variable aparece entre un 30 y 70 por ciento de los árboles.

Figura 4.14: Representación gráfica del árbol clasificador



Fuente: Elaboración propia en R

El árbol tiene 14 reglas o nodos finales, contruidos a partir de 27 nodos o criterios y 11 factores, de los cuales 4 son comerciales, 3 de población y empleo, 1 de inclusión y tecnología, 1 de crecimiento económico, 1 de inversión y 1 de gobernanza.

El modelo clasifica el 89% de los datos en la muestra de entrenamiento (semilla 3751) correctamente y el 84% en la muestra completa con una precisión del 79.61%, 70% fuera de la muestra y una MC considerando los 1241 datos de 36.74.

4.4. Extracción y análisis de las reglas de conocimiento del CART

A continuación se describen e interpretan los resultados del árbol construido en la sección anterior, de tal forma que, analizamos el conocimiento extraído del modelo. De manera general, en la cuadro 4.1 se presentan las condiciones bajo las cuales aumentos de apertura comercial, medidos con el índice estimado anteriormente (C_{IGC}), están relacionados a cambios en la desigualdad del ingreso dados por cambios en el coeficiente de Gini (C_{GINI}).

El factor discriminante más importante es la proporción de trabajadores en el sector agrícola con respecto al empleo total (EM_{AGR}). Este criterio separa los países con un valor mayor al 12.55 del lado derecho del árbol, entre los que se encuentran la mayoría de los países en el Sur de Asia y más de la mitad de las economías africanas y latinoamericanas (incluyendo a México), en total de 397 observaciones de 74 países. Del otro lado del árbol se encuentran las economías con poco empleo en el sector agrícola que conforman un grupo de 63 países y 377 datos e incluyen a más del 65% de los países europeos¹⁴. A pesar de que los países parecen dividirse según la región a la pertenecen, la variable *idRegión* no resulta un factor determinante, porque en ambos grupos se encuentran países de todas las regiones, además de que la estructura del empleo cambia en el tiempo. Conviene mencionar que las reglas no clasifican economías en cierto grupo de países con ciertas características porque suponemos que dichas características son dinámicas, de tal forma que en un país podría asociarse a una regla en cierto año y a asociarse a otra regla en otro año.

¹⁴ El análisis de los resultados en cada criterio se realiza filtrando el panel con datos nulos usando SQL Server, que si bien son 1241 observaciones en realidad depende que los registros donde las variables involucrados no tienen valores nulos. Por ejemplo la variable EM_{AGR} tiene 467 observaciones con algún valor nulo, entonces, el primer criterio se analiza con base en los 774 registros restantes.

Cuadro 4.1: Reglas de conocimiento que llevan a cambios de desigualdad cuando aumenta la globalización comercial

REGLA	CONDICIÓN	C_GINI	PROB.
R-1	EM_AGR<12.55 & TERM_TRADE<98.7197 & KPRIV<119.876	C+	1.00
R-2	EM_AGR <12.55 & TERM_TRADE<98.7197 & KPRIV>119.876 & KPRIV<131.077	C-	0.70
R-3	EM_AGR<12.55 & TERM_TRADE<98.7197 y KPRIV>=131.077 & US_INT<73.435	C+	0.90
R-4	EM_AGR<12.55 & TERM_TRADE<98.7197 & KPRIV>=131.077 & US_INT>=73.435	C-	0.60
R-5	EM_AGR<12.55 & TERM_TRADE>=98.7197 & EM_SERV < 68.25 & GDP<5.128	C+	0.90
R-6	EM_AGR<12.55 & TERM_TRADE>=98.7197 & EM_SERV < 68.25 & GDP>=5.128	C-	0.80
R-7	EM_AGR<12.55 & TERM_TRADE>=98.7197 & EM_SERV>=68.25 & POB_URB<1.08183 & HTE_T<3.72539	C+	1.00
R-8	EM_AGR<12.55 & TERM_TRADE>=98.7197 & EM_SERV>=68.25 & POB_URB<1.08183 & TE_T>=3.72539	C-	1.00
R-9	EM_AGR<12.55 & TERM_TRADE>=98.7197 v EM_SERV>=68.25 & POB_URB>=1.08183	C-	0.90
R-10	EM_AGR>=12.55 & POB_URB< -0.76	C+	1.00
R-11	EM_AGR>=12.55 & POB_URB> -0.76 & RL < 0.1631 & EXP_GS_P<17.7135	C+	0.60
R-12	EM_AGR>=12.55 & POB_URB> -0.76 & RL < 0.1631 & EXP_GS_P>=17.7135	C-	0.90
R-13	EM_AGR>=12.55 & POB_URB> -0.76 & RL>=0.1631 & IMP_GS_P<30.0741	C+	0.90
R-14	EM_AGR>=12.55 & POB_URB> -0.76 & RL>=0.1631 & IMP_GS_P>=30.0741	C-	0.80

Fuente: Elaboración propia con base en el modelo CART.

En el cuadro 4.1, se presentan las 14 reglas del modelo, una regla es una combinación de condiciones que subdividen las observaciones, cada regla corresponde a una ruta que

termina en un nodo final en el árbol de la figura 4.14 y cada una tiene una probabilidad asociada al factor resultado ya sea aumentos o reducción de desigualdad, es decir, $\{C+, C-\}$.

Partiendo de la regla 10, en los países con una proporción de empleo en sector agrícola mayor al 12% ($EM_AGR > 12.55\%$) y donde la población en zonas urbanas está decreciendo a una tasa menor que el 0.76%, la desigualdad aumenta con una alta probabilidad (**R-10**). Como sucedió en Georgia en el 2001 y 2007, donde -con más del 50% de su empleo en el sector agrícola- su desigualdad en el ingreso aumentó, al igual que en Lituania (2001-2006) que a pesar de que la mayor parte del empleo se concentra en el sector servicios tiene reducciones importantes en el crecimiento en la población urbana asociado a la desindustrialización.

Por el contrario, si la tasa de crecimiento de la población es positiva, el estado de derecho, medido con el índice de globalización RL (imperio de la ley) resulta ser un factor determinante. El índice se muestra en valores estandarizados entre -2.5 y 2.5, por lo que valores positivos revelan una buena calificación en cuanto a la concentración del poder jurídico. Entonces, si este indicador es prácticamente negativo ($RL < 0.16$) y el porcentaje de sus exportaciones con respecto al ingreso nacional es pequeño ($EXP < 17.7$), la desigualdad aumenta (**R-11**). Esta es la regla menos confiable del modelo, ya que indica aumentos de desigualdad con una probabilidad del 60%, clasificando a Colombia, Egipto e India correctamente algunos años e incorrectamente en otros, de tal manera que existe otro fenómeno no contemplado (o que desapareció en el proceso de podado del árbol) que evita que la desigualdad en estos países cambie cuando la globalización aumenta. Sin embargo, este criterio de clasificación es importante para la regla 12 que tiene una probabilidad de clasificar acertadamente del 0.9, además de que 156 observaciones correspondientes a 38 países tienen dichas características en algunos años. Entonces la desigualdad tiende a disminuir con la globalización comercial, en países que tienen una proporción de empleo en el sector agrícola mayor al 12.5, un crecimiento constante en zonas urbanas, y aunque un deficiente imperio de la ley, exportan más del 17.7% de su producto interno bruto (**R-12**). No es que el deterioro del estado de derecho propicie el bienestar económico, por el contrario, esta condición podría indicar, que es la globalización y no el gobierno lo que ha permitido

dichas mejoras a través de los salarios en estas economías con potencial de crecimiento o en procesos de industrialización.

El 24% de los países se han beneficiado del comercio vía exportaciones, en algún momento, especialmente los latinoamericanos (casi el 40%), como fue el caso de Chile, Ecuador, El Salvador, Guatemala, Nicaragua, Panamá, Colombia -que con problemas logra exportar más del 17% para el 2008, y México a partir de 1998 (a excepción del periodo entre el 2004-2007) con una estructura promedio del empleo por sectores del 18% en el agrícola, 25% en el industrial y 57% en sector servicios. Entre los países asiáticos con estas características y consistentes con la regla 12 destacan Armenia, Tailandia y Camboya (exporta más del 65 de su producto), además de Moldavia, Turquía, y Egipto, con más del 50% de su producción en exportaciones a partir del 2006.

Al igual que Vietnam (2001-2006), China es una excepción a la regla, este país ha tenido incrementos de desigualdad desde 1983 hasta el 2010 a pesar de que sus exportaciones superaron 17.7% del PIB en 1993; con más del 40% del empleo concentrado en el sector agrícola y un alto -aunque decreciente- crecimiento de la población en zonas urbanas, sus exportaciones no fueron suficientes para reducir la desigualdad.

Por otro lado, los países con buena calificación en el imperio de la ley suelen ser países con instituciones confiables, buenas políticas públicas que a su vez tienen altos niveles de educación y, por ende, primas salariales mayores, dando lugar a que se formen círculos virtuosos (Acemoglu & Robinson, 2008) que por definición tienden a reducir la desigualdad. En nuestro modelo los cambios en la desigualdad en países con un estado de derecho eficiente que tienen más del 12% del empleo en el sector agrícola, además de crecimiento urbano, también dependen de la globalización comercial, específicamente de la proporción de importaciones con respecto a los ingresos nacionales. El modelo indica que la desigualdad disminuye en países con un estado de derecho fuerte, aunado a aumentos de globalización por vía de las importaciones con niveles mayores al 30% (**R-14**), como es el caso de Bolivia antes del 2009, Costa Rica, Honduras, Filipinas, entre otros, la mayoría asiáticos; pero aumenta en países con poca iteración comercial $IMP < 30\%$ (**R-13**), como es el caso de Bangladesh 2000-2003, Pakistán (que también tienen bajos porcentajes de exportaciones) e Indonesia (que además ha venido reduciendo sus exportaciones de manera importante).

Como excepción a esta regla (R-13), Brasil ha reducido su desigualdad desde 1988 a pesar de que sus importaciones no alcanzan ni el 15 por ciento de su producto, lo cual podría estar más relacionado con la eficiencia del gobierno y políticas públicas enfocadas a la educación que incrementan la oferta de trabajo en el sector servicios.

Con respecto a las reglas en las que el empleo en el sector agrícola es menor al 12.55%, correspondiente al lado izquierdo del modelo CART (figura 3), el deterioro de los términos de intercambios es una característica determinante para el aumento de la desigualdad. Este criterio divide las observaciones, según si se han deteriorado ($TERM_TRADE < 98.8$) o no los términos de intercambio ($TERM_TRADE \geq 98.8$).

En la primera parte del árbol (a la derecha), encontramos que en países con poco empleo en el sector agrícola ($EMP_AGR < 12.5\%$) y en los cuales el capital privado es menor que el producto interno bruto ($KPRIV < 120\%$), el deterioro de los términos de intercambio viene acompañado de aumentos en la desigualdad en el ingreso (R-1), este fue el caso de Bolivia en 1992 y 2001, Lituania 2008-2013 y Suiza 2005-2013; pero si la inversión en el sector privado supera el PIB por más del 30% ($KPRIV \in (120, 130)$), el coeficiente de Gini podría disminuir (R-2). Entonces, una forma de evitar que aumente la desigualdad cuando los términos de intercambio se deterioran es aumentando el capital pero solo hasta cierto punto ($119.9 < KPRIV < 131.0$), esto para economías que han dejado atrás la producción agrícola, situación en Los Países Bajos 2007, Estonia 2005-2006, Grecia 2004, 2007, Irlanda, Republica de Mauricio África y Eslovaquia 2010, aunque no para todos los años.

Si el capital privado es lo suficientemente grande ($KPRIV > 131$), entonces el efecto de la globalización comercial sobre el coeficiente de Gini –aunado al deterioro de los términos de intercambio- depende de la inclusión tecnológica específicamente en términos de comunicación, medido en este caso con el número de usuarios de internet por cada cien. De tal manera que, si más del 27% de la población carece de acceso a internet, la desigualdad se incrementa (R-3) y por el contrario se reduce en países con más de 73% (R-4).

Austria es un excelente ejemplo del efecto de la inclusión digital sobre la desigualdad determinada en las reglas 3 y 4, con una tasa del 39 usuarios por cada 100 en el 2001 experimentó aumentos de desigualdad, entonces fue incrementando el acceso a internet hasta que en el 2007 paso el límite de la condición ($US_INT > 73$) y la medida de desigualdad comenzó

a decrecer. Otro aspecto interesante, es que, la mayoría de los años en los que se cumple la R-3 coinciden con periodos de crisis –a la crisis tecnológica (2002-2005) o a la crisis financiera (2007-2010)¹⁵; mientras que del otro lado del nodo, en la R-4 se muestra que Suiza, Austria, Finlandia y E.U., presentaron reducciones de desigualdad después de que aumentaran el acceso a internet, incluyendo el periodo entre el 2007-2010. Lo cual sugiere que la inclusión digital ayuda a evitar que la desigualdad aumente en tiempo de crisis. Además de que, podría existir una relación entre las crisis y la desigualdad, que estaría siendo afectada por la inclusión tecnológica, de ser así, estaríamos omitiendo un factor de crisis importante.

Regresando a la importancia de los términos de intercambio, en aquellas economías en las que el valor de sus exportaciones valen más que sus importaciones, el empleo en el sector agrícola es menor al 12%, el empleo en el sector servicios menor al 68% y, por lo tanto, el empleo en el sector industrial mayor al 20%, el efecto de una mayor integración económica sobre la desigualdad en el ingreso se determina por la tasa de crecimiento de su producción. Como se describe en las reglas 5 y 6, a menos que dicha tasa supere el cinco por ciento (**R-6**), la desigualdad aumenta con una alta probabilidad (**R-5**), razón por la cual países como Australia (2004-2006), Bulgaria (2010-2013), Croacia (2008), Alemania (2001-2004) Portugal y Rusia, entre otros, presentaron incrementos de desigualdad. Por consiguiente, en este modelo la tasa de crecimiento económico actúa como discriminante; el modelo indica que es posible reducir la desigualdad sin alcanzar el 68% del empleo en el sector servicios, si el sector industrial es lo suficientemente eficiente para generar un crecimiento mayor al 5%, con términos de intercambio aceptables. Lo que sugiere una relación entre crecimiento económico y desigualdad en el ingreso, cuando aumenta la globalización comercial aunada a las condiciones establecidas en la regla 6.

Para las economías que tienen buenos términos de intercambio ($TERM_TRADE > 98.7$) y una estructura del mercado laboral en la que los trabajadores en el sector agrícola constituyen menos del 12%, los del sector industrial menos del 32%, pero sobre todo en la

¹⁵ Alemania (2005-2006), Luxemburgo (2005), Portugal (2003-2005, 2012), Hong Kong (2005), Japón (2005, 2011-2013), Corea (2001-2004), Singapur (2002-2006), Suiza (2005), EU (2005), Francia (2008-2009), Italia (2008-2011), Irlanda (2008-2010), España (2008-2012), Israel (2001-2008)

que los trabajadores en el sector servicios representan más del 68% de la fuerza laboral, la desigualdad en el ingreso aumenta debido al desacelerado crecimiento de la población urbana ($POB_URB < 1.08\%$) aunado al desaprovechamiento de la educación de su fuerza de trabajo para exportar más del 3% en alta tecnología (R-7). Las economías que sufrieron aumentos de desigualdad, por este tipo de estancamiento fueron: Australia en el 2007, Dinamarca desde 2006, Hong Kon en el cambio de siglo, Nueva Zelanda (2011, 2013), Uruguay 1995-2002 e incluso Estados Unidos y el Reino Unido en el 2003.

Si la población en zonas urbanas sigue creciendo a una tasa de al menos el 1%, es común que la desigualdad decrezca (R-9), de lo contrario esta decrecerá solo si exporta alta tecnología en un porcentaje mayor al 3.72% del total de las exportaciones (R-8). Los países que se han beneficiado de la globalicen comercial, por su crecimiento urbano aunado a las características de su mercado laboral son: Argentina, Perú, Canadá, Los Países Bajos, Nueva Zelanda, Noruega, Sudáfrica, Suiza y El Reino Unido (R-9); y por exportar más del 3.7% de bienes y servicios considerados de alta tecnología son: Suecia en 2001-2002, UK en el 2002 y Bélgica en 2004-2007 (R-8). Estados Unidos es una excepción a la regla 9, ya que, a pesar del rápido crecimiento de su población urbana, su desigualdad dejó de aumentar hasta 1993 cuando sus exportaciones en alta tecnología superaron el 1.89%, igual que Argentina hasta el 2001 cuando empieza a exportar más 0.05%.

CONCLUSIONES

Los efectos de la globalización sobre la desigualdad dependen principalmente de las características macroeconómicas internas de cada país, aquellos con políticas de bienestar social que reducen el número de embarazos en adolescentes, fomentan la investigación e incrementan la esperanza de vida son menos susceptibles a aumentos de desigualdad. Sin embargo, cuando se perciben incrementos de globalización es importante describir las situaciones que propician cambios en la distribución del ingreso, considerando en primera instancia la estructura del mercado de trabajo, así como su desempeño tecnológico y comercial.

Como menciona Ravallion (2003) no es posible generalizar este fenómeno debido a la heterogeneidad de los países y a los cambios temporales, sin embargo, dada la información disponible, el proceso de modelado propuesto identifica los factores más importantes, así como las condiciones específicas en las que se generan cambios de desigualdad, ya sean positivos o negativos.

Por una parte, el modelo PML de redes neuronales artificiales tiene una alta eficiencia al tener un error cuadrático muy pequeño, mostrando que la dinámica de las relaciones es no lineal y, aunque su relación funcional (que involucra 289 pesos o parámetros) no tiene una interpretación práctica o directa, el análisis de sensibilidad aplicado aporta bastante información.

El modelo de redes neuronales más estable, según el índice de frecuencia, identifica como factor de mayor impacto sobre la desigualdad a la tasa de fertilidad en adolescentes. El análisis de variabilidad entre la importancia de los factores que se aprecia en la figura 3.6, nos muestra que la dinámica de la desigualdad depende en mayor medida de factores macroeconómicos demográficos y de la estructura de trabajo o de las políticas implementadas para aumentar la inversión en ciencia y tecnología, más que de factores de globalización comercial o financiera.

Según el modelo, la globalización comercial puede reducir la desigualdad, dependiendo del intercambio de mercancías y las importaciones y exportaciones. En cambio la globalización financiera puede tener distintos efectos, dependiendo de la posición

acreedora o deudora del país, así como del tamaño de su mercado accionario. Con respecto a la globalización tecnológica, no hay duda de que los créditos a las tecnologías de información y el gasto en investigación y desarrollo son los factores más importantes para reducir la desigualdad. Las variables macroeconómicas más importantes para reducirla son el empleo en el sector industrial y de servicios, el control de la corrupción y la calidad de las regulaciones. Por el contrario, el presupuesto destinado al gasto militar, el crecimiento de la población en zonas urbanas, la esperanza de vida y principalmente la tasa de fertilidad en adolescentes aumentan la desigualdad.

Con base en la magnitud y dirección de las variables más importantes en el MPL, podemos afirmar que los efectos de las dimensiones comerciales y financieros sobre la desigualdad se contrarrestan si se mantiene una estabilidad en la balanza de pagos, es decir, entre las exportaciones e importaciones, los flujos de inversión extranjera directa, el total de pasivos y activos financieros.

Por otra parte, un análisis de los factores interrelacionados y su efecto en conjunto y no individual se realiza a través del modelo CART. Gracias a la intuitiva interpretación del modelo CART presentado se identifican circunstancias y factores clave para analizar el efecto de la globalización comercial sobre la desigualdad en el ingreso de los hogares dentro de un país. La estructura del mercado de trabajo refleja los efectos de la industrialización y el progreso tecnológico en la distribución del ingreso de la población, consistente con Kuznets (1955) y Milanovic (2016), sin embargo, es necesario especificar bajo qué condiciones el resultado es positivo o no para la desigualdad.

En países no agrícolas, en los cuales el capital privado es menor a su nivel de producción, el deterioro de los términos de intercambio viene acompañado de aumentos de desigualdad, en este caso la inversión privada y la inclusión digital juegan un papel importante pues pueden reducir la desigualdad, si más de 27% de la población carece de acceso a internet el efecto sobre la desigualdad es positivo.

El árbol clasificador también indica que en los países con altos niveles de empleo en el sector servicios y sin deterioro en los términos de intercambio, en su mayoría economías desarrolladas, la desigualdad en ingreso puede aumentar en periodos de estancamiento. Por el contrario, este tipo de países reducen su desigualdad con tasas de crecimientos superiores

al 5%. Entonces, para que el bienestar económico de la población de estos países mejore, es importante que no se estanquen y aprovechen el potencial de su alta concentración de su mano de obra calificada exportando alta tecnología. Lo cual es consistente con un periodo de ajuste en la estructura del mercado de trabajo, en el que la sociedad paga un costo de aprendizaje, pero una vez fortalecida la mano de obra en el sector servicios se generan rendimientos suficientes que posteriormente reducirán su desigualdad.

El cambio de la estructura del mercado de trabajo en el que se permite una mayor concentración de trabajadores en el sector servicios, es determinante para la distribución del ingreso, debido a la prima salarial y las diferencias salariales entre sectores. Como se enfatiza en el análisis, las economías agrícolas (con más del 12% del empleo en este sector) pero en procesos de industrialización (reflejado en el crecimiento de la población urbana positivo características de países en desarrollo) presentan reducciones en la desigualdad ante aumentos de globalización comercial.

Este argumento se cumple para al menos 38 países, entre los que se encuentra México, ya que tiene entre el 18 y 13 por ciento del empleo en el sector agrícola, una tasa de crecimiento de la población en zonas urbanas positiva y que a partir de 1995 exporta más del 20% de su producto, actualmente exporta más del 35%. Considerando su bajo desempeño en el imperio de la ley según los índices de gobernanza, es la globalización y no el gobierno lo que ha permitido reducciones de desigualdad a través de los salarios. Como en la mayoría de los países latinoamericanos con potencial de crecimiento o en procesos de industrialización.

Por lo anterior, se recomienda de forma general el fortalecimiento del sector industrial, los índices de gobernanza y el impulso del progreso tecnológico a través de créditos o apoyo a la investigación y desarrollo, sin dejar de lado las políticas de bienestar social y de control demográfico, considerando que la mayor parte de la desigualdad tiene que ver con las características macroeconómicas de un país, este tipo de factores determinan más del 45 por ciento del modelo (incluyendo los índices de gobernabilidad).

A través de este tipo de análisis es posible concentrarse en una economía para hacer algunas recomendaciones de políticas públicas específicas para cada país según sus características macroeconómicas.

REFERENCIAS

- Acemoglu, D., & Robinson, J. A. (2008), "Persistence of power, elites, and institutions", *American Economic Review*, 98(1), 267-93.
- Aghion, P., Caroli, E., & Garcia-Penalosa, C. (1999), "Inequality and economic growth: The perspective of the new growth theories", *Journal of Economic literature*, 37(4), 1615-1660.
- Aguilar Gutiérrez, G. (2000), *Desigualdad y pobreza en México, ¿son inevitables?*
- Akerman, A., Helpman, E., Itskhoki, O., Muendler, M. A., & Redding, S. (2013), "Sources of wage inequality", *American Economic Review*, 103(3), 214-19.
- Alderson, A. S., & Nielsen, F. (2002), "Globalization and the great U-turn: Income inequality trends in 16 OECD countries", *American Journal of Sociology*, 107(5), 1244-1299.
- Angeles-Castro, G. (2007), "Factors driving changes in income distribution in post-reform, Mexico", *Department of Economics Discussion Paper, University of Kent*.
- Asteriou, D., Dimelis, S., & Moudatsou, A. (2014). "Globalization and income inequality: A panel data econometric approach for the eu27 countries", *Economic Modelling*. 36(7):592-599.
- Atkinson, A. B. (1970), "On the Measurement of Inequality", *Journal of Economic Theory*.
- Atkinson, A. B. (2003), "Income inequality in OECD countries: Data and explanations", *CESifo Economic Studies*, 49(4), 479-513.
- Atkinson, A. B. (2015). *Inequality: what can be done?* Harvard University Press, ISBN 9780-674-28-7037, 398 p.
- Atkinson, A. B., Piketty, T., & Saez, E. (2011), "Top incomes in the long run of history", *Journal of economic literature*, 49(1), 3-71.
- Atkinson, A. B., Smeeding, T. M., & Brandolini, A. (2001). *Producing Time Series Data for Income Distribution: Sources, Methods, and Techniques*. I. Becker, N. Ott, & G. Rolf (Eds.). Maxwell School of Citizenship and Public Affairs, Syracuse University.

- Attanasio, O., Goldberg, P. K., & Pavcnik, N. (2004), "Trade reforms and wage inequality in Colombia", *Journal of development Economics*, 74(2), 331-366.
- Azzimonti, M., De Francisco, E., & Quadrini, V. (2014), "Financial globalization, inequality, and the rising public deb", *The American Economic Review*, 104(8), 2267-2302.
- Balan, F., Torun, M., & Kilic, C. (2015), "Globalization and Income Inequality in G7: A Bootstrap Panel Granger Causality Analysis", *International Journal of Economics and Finance*, 7(10), 192.
- Bhagwati, J. (1994), "Free trade: old and new challenges", *The Economic Journal*, 104(423), 231-246.
- Bluestone, B., & Harrison, B. (2001), *Growing prosperity: The battle for growth with equity in the twenty-first century*. Univ of California Press.
- Bourguignon, F. (2017), *The globalization of inequality*, Princeton University Pres.
- Chapra, S. C., & Canale, R. P. (2010), *Numerical methods for engineers*. Boston: McGraw-Hill Higher Education,
- Chen, J. H., & Tsai, W. C. (2012, "A comparison of international income inequality: an ordered probit model analysis", *Applied Economics*, 44(13), 1701-1716.
- Chiquiar, D. (2008), "Globalization, regional wage differentials and the Stolper–Samuelson Theorem: Evidence from Mexico", *Journal of International Economics*, 74(1), 70-93.
- Cornia, A. G. (2004). "Inequality, growth, and poverty in an era of liberalization and globalization", *Oxford University Press for UNU-WIDE*
- Cowell, F. (2011). *Measuring inequality*. Oxford University Press.
- Cynamon, B. Z. and Fazzari, S. M. (2015), "Inequality, the great recession and slow recovery", *Cambridge Journal of Economics*. 40(2), 373-399
- Deininger, K., & Squire, L. (1996), "A new data set measuring income inequality", *The World Bank Economic Review*, 10(3), 565-591.

- De Oña, J., & Garrido, C. (2014), “Extracting the contribution of independent variables in neural network models: a new approach to handle instability”, *Neural Computing and Applications*, 25(3-4), 859-869.
- Dollar, D., & Kraay, A. (2001), “Trade, growth, and poverty”, *World Bank*, Development Research Group, Macroeconomics and Growth.
- Dollar, D. (2004), “Globalization, poverty, and inequality since 1980”, The World Bank.
- Dreher, A. (2006), “Does Globalization Affect Growth? Empirical Evidence from a new Index”, *Applied Economics* 38, 10: 1091-1110.
- Dreher, A., Gaston, N., & Martens, P. (2008), “Measuring globalisation gauging its consequences”, *New York: Springer*.
- Easterly, W. R. (2004), “Globalization, inequality and development: the big picture”, *Institute for Monetary and Economic Studies, Bank of Japan*.
- Faraway, J. J. (2016), *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC.
- Fisher, J. D., Johnson, D. S., & Smeeding, T. M. (2013), “Measuring the trends in inequality of individuals and families: Income and consumption”, *American Economic Review*, 103(3), 184-88.
- FitzGerald E. V. K. (1996), *The New Trade Regime: Macroeconomic Behaviour and Income Distribution in Latin America*, in Victor Bulmer-Thomas, eds. London: Macmillan Press LTD
- Garson, G.D. (1991), “Interpreting neural network connection weights”, *Artificial Intelligence Expert*. 6(4):46–51.
- Gereffi, G. (2008), “Development models and industrial upgrading in China and Mexico”, *European Sociological Review*, 25(1), 37-51.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003), “Review and comparison of methods to study the contribution of variables in artificial neural network models”, *Ecological Modelling*, 160(3), 249-264.

- Goldberg, P. K., & Pavcnik, N. (2007), "Distributional effects of globalization in developing countries", *Journal of economic Literature*, 45(1), 39-82.
- Ha, E. (2012), "Globalization, government ideology, and income inequality in developing countries", *The journal of Politics*, 74(2), 541-557.
- Han, J., Liu, R., & Zhang, J. (2012), "Globalization and wage inequality: Evidence from urban China", *Journal of international Economics*, 87(2), 288-297.
- Harrison, A., & Hanson, G. (1999), "Who gains from trade reform? Some remaining puzzles", *Journal of development Economics*, 59(1), 125-154.
- Haykin, S. (2005) *Neural Networks, A comprehensive foundation*. Person Prentice Hall 2^oEd.
- Heathcote, J., Storesletten, K., & Violante, G. L. (2017), "Optimal tax progressivity: An analytical framework", *The Quarterly Journal of Economics*, 132(4), 1693-1754.
- Heckscher, E. F., & Ohlin, B. G. (1991). *Heckscher-Ohlin trade theory*. The MIT Press.
- Helpman, E., Itskhoki, O., & Redding, S. (2010), "Inequality and unemployment in a global economy", *Econometrica*, 78(4), 1239-1283.
- Ietto-Gillies, G. (1992), *International production: Trends, theories, effects*. Cambridge: Polity Press.
- International Monetary Fund [IMF] (2007) "Global financial stability report", *World Economic and Financial Surveys*, 74-76.
- Jaumotte, F., Lall, S., & Papageorgiou, C. (2013), "Rising income inequality: technology, or trade and financial globalization?" *IMF Economic Review*, 61(2), 271-309.
- Jeon, K., & Kabukcuoglu, Z. (2018), "Income inequality and sovereign default", *Journal of Economic Dynamics and Control*, 95, 211-232.
- Kaldor, N. (1957), "A model of economic growth", *The economic journal*, 67(268), 591-624.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). *Governance matters VIII: aggregate and individual governance indicators, 1996-2008*. The World Bank.

- Kuznets, S. (1955), "Economic growth and income inequality", *The American economic review*, 1-28.
- Lawrence, R. Z., Slaughter, M. J., Hall, R. E., Davis, S. J., & Topel, R. H. (1993), "International trade and American wages in the 1980s: giant sucking sound or small hiccup?" *Brookings papers on economic activity*. Microeconomics, 1993(2), 161-226.
- Ledolter, J. (2013). *Data mining and business analytics with R*. Hoboken, NJ: Wiley.
- *McCulloch, W. S., & Pitts, W. (1943), "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics*, 5(4), 115-133.
- McLachlan, G. J. (1999), "Mahalanobis distance", *Resonance*, 4(6), 20-26.
- Menezes-Filho, N. A., Muendler, M. A., & Ramey, G. (2008), "The structure of worker compensation in Brazil, with a comparison to France and the United States", *The Review of Economics and Statistics*, 90(2), 324-346.
- Milanovic, B. (2002), *Can we discern the effect of globalization on income distribution? Evidence from household budget surveys*, The World Bank.
- Milanovic, B. (2007), "Globalization and inequality", *Global Inequality*, Cambridge, 26-49.
- Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press.
- Mills, M. (2008), "Globalization and inequality", *European sociological review*, 25(1), 1-8.
- Montaño, J. J., & Palmer, A. (2002), "Artificial Neural Networks, opening the black box", *Metodología de las Ciencias del Comportamiento*, 4(1), 77-93.
- Nissanke, M., & Thorbecke, E. (2006), "Channels and policy debate in the globalization–inequality–poverty nexus", *World development*, 34(8), 1338-1360.
- North, D. C. (1990) "A transaction cost theory of politics", *Journal of Theoretical Politics*, 2(4), 355-367.

- Obstfeld, M., Shambaugh, J. C., & Taylor, A. M. (2005), "The trilemma in history: tradeoffs among exchange rates, monetary policies, and capital mobility", *Review of Economics and Statistics*, 87(3), 423-438.
- Olson, D. L., & Delen, D. (2008), "Advanced data mining techniques", *Springer Science & Business Media*.
- Palma, J. G., & Stiglitz, J. E. (2016), "Do nations just get the inequality they deserve? The "Palma Ratio" re-examined", *In Inequality and Growth: Patterns and Policy* (pp. 35-97). Palgrave Macmillan, London.
- Piketty, T. (2014). *Capital in the twenty first century*. Harvard University Press.
- *Ravallion, M. (2003), "The debate on globalization, poverty and inequality: why measurement matters", *International Affairs*, 79(4), 739-753.
- *Rhee, D. E., & Kim, H. (2018), "Does income inequality lead to banking crises in developing countries? Empirical evidence from cross-country panel data", *Economic Systems*, 42(2), 206-218.
- Ricardo, D. (1817). *On the Principles of Political Economy and Taxation*, London.
- Rodrik, D. (1997), "Has globalization gone too far?" *California Management Review*, 39(3), 29-53.
- Rokach, L., & Oded Z. M. (2008), "Data mining with decision trees: theory and applications", *World scientific*. (Vol. 69).
- Rosenblatt, F. (1958), "The perceptron: a probabilistic model for information storage and organization in the brain", *Psychological Review*, 65(6), 386.
- Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995), "Backpropagation: The basic theory", *Theory, architectures and applications*, 1-34.
- Sala-i-Martin, X. (1992), "La desigualdad global desaparece a medida que crece la economía global", *Convergence*, 100(2), 223-251.

- Shahbaz, M. (2012), “Does trade openness affect long run growth? Cointegration, causality and forecast error variance decomposition tests for Pakistan”, *Economic Modelling*, 29(6), 2325-2339.
- Solt, F. (2016), “The standardized world income inequality database”, *Social science quarterly*, 97(5), 1267-1281.
- Solt, F.. 2019. “Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database.”
- Stiglitz, J. (2015). *The great divide*. Penguin UK
- Stockhammer, E. (2015), “Rising inequality as a cause of the present crisis”, *Cambridge Journal of Economics*, 39(3), 935-958.
- Stolper, W. F., & Samuelson, P. A. (1941), “Protection and real wages”, *The Review of Economic Studies*, 9(1), 58-73
- Thorbecke, E., & Charumilind, C. (2002),”Economic inequality and its socioeconomic impact”, *World Development*, 30(9), 1477-1495.
- Ultsch, A. (1993), “Self-organizing neural networks for visualisation and classification. In Information and classification” *Springer*, Berlin, Heidelberg. 307-313
- Williamson, J. G. (1997), “Globalization and inequality, past and present”, *The World Bank Research Observer*, 12(2), 117-135.
- Wood, A., & Ridao-Cano, C. (1999), “Skill, trade, and international inequality”, *Oxford economic papers*, 51(1), 89-119.

APÉNDICE A. Niveles y cambios de desigualdad

Para el análisis con árboles de decisión se requiere de una variable dependiente tipo factor o categórica, acorde con el objetivo de esta tesis esta variable debe indicar si la desigualdad aumentó o disminuyó en el último año. Para lo cual, se calcula el cambio porcentual por país del coeficiente suavizado de Gini neto (R_{GINI}). Aunque se tienen 5082 observaciones, el número de datos anuales por país varían desde 5 hasta 18 registros por lo que se calcula el cambio porcentual en el coeficiente de Gini de un año a otro, de tal forma que se pierde un año por país, dejando 4892 registros.

$$R_{GINI} = \text{Log} \left(\frac{GINI_{Año}}{GINI_{Año-1}} \right)$$

Si este cambio porcentual es positivos entonces se produjo un aumento de desigualdad del año anterior al actual, a partir de lo cual se genera la siguiente variable dependiente categórica

$$CAMBIO_{GINI} = f \begin{cases} C-, & R_{GINI} < 0 \\ SC, & R_{GINI} = 0 \\ C+, & R_{GINI} > 0 \end{cases}$$

Donde C+ representa un cambio positivo en el coeficiente de Gini o aumento de desigualdad, C- un cambio negativo y SC significa sin cambio. De los 4233 registros, el 43% (1820) corresponde a cambios positivos, el 41.4% (1751) a cambios decrecientes y solo el 15.6% (662) a cambios nulos.

Otra opción para categorizar el coeficiente de Gini, es etiquetar los registros (país-año) como una instancia con distintos niveles de desigualdad analizando la distribución de los datos. Para agrupar los más de cinco mil valores del coeficiente Gini en cuatro niveles A: Alto, MA: Medio Alto, MB: Medio Bajo y B: Bajo, se utiliza la mediana con base en la cual se definen los límites.

$$L2 = \text{mediana}(x) , \quad L1 = L2/2 , \quad L3 = L2 + (\max(x) - L2)/2,$$

$$N_{GINI} = \begin{cases} \mathbf{B: Baja} & , & 0 \leq Gini < L1 \\ \mathbf{MB: Media Baja}, & L1 \leq Gini < L2 \\ \mathbf{MA: Media Alta}, & L2 \leq Gini < L3 \\ \mathbf{A: Alta} & , & L3 \leq Gini \end{cases}$$

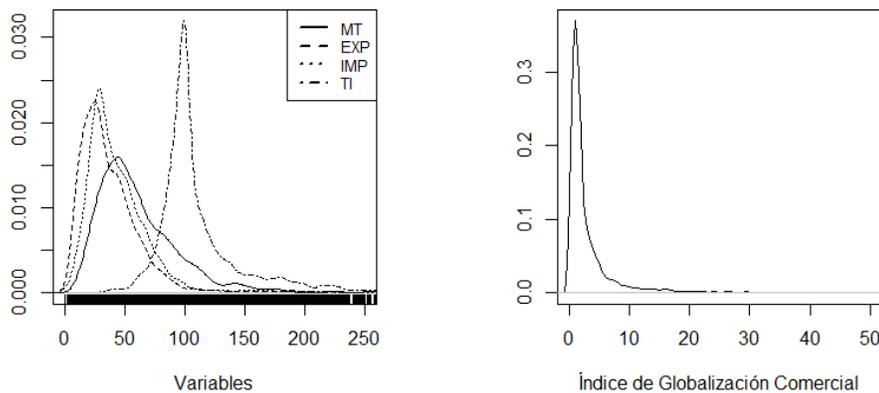
APÉNDICE B. Índice de globalización

Para medir que tan globalizado está un país se propone el cálculo de un índice utilizando la distancia de Mahalanobis entre las principales variables que definen la apertura de los países a mercados globales. Esta medida es ampliamente usada en diversos campos para clasificación y reconocimiento de patrones estadísticos en relaciones multivariante; considera las desviaciones y covariancia entre las variables como se muestra en la siguiente ecuación.

$$D_M(X) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$$

Donde $(x - \mu)^T$ es una matriz transpuesta que contiene las diferencias de los valores con respecto a su media μ , y Σ^{-1} es la matriz de varianzas-covarianzas inversa de los variables involucradas (McLachlan, 1999), y X es la matriz de atributos a partir de los cuales se estima el índice, el cual incluye el volumen de importaciones y exportaciones con respecto al PIB, los términos de intercambio y el intercambio de mercancías finales. El *Índice de Globalización Comercial* (I_{GC}) se estima con 4,234 observaciones $I_{GC} = D_M (EXP_{GSP}, IMP_{GSP}, MT, TI,)$.

Distribución del Índice de Globalización Comercial



Fuente: Elaboración propia en R

Como se puede apreciar en la figura 1, tanto el índice como las variables que lo conforman tienen valores extremos, correspondientes a países con alto volumen de comercialización como Aruba, Guinea y Luxemburgo. Los países con exportaciones e importaciones mayores al 150% de su producto interno bruto son Singapur, Hong Kong y Luxemburgo.