

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA**



**SIMULACIÓN BASADA EN AGENTES PARA LA  
DISTRIBUCIÓN DEL BIENESTAR EN MÉXICO**

**POR**

**RUTH BERENICE CISNEROS ALDAMA**

**COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE  
MAESTRÍA EN CIENCIAS EN INGENIERÍA DE SISTEMAS**

**MARZO, 2017**

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA  
SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO**



**SIMULACIÓN BASADA EN AGENTES PARA LA  
DISTRIBUCIÓN DEL BIENESTAR EN MÉXICO**

**POR**

**RUTH BERENICE CISNEROS ALDAMA**

**COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE  
MAESTRÍA EN CIENCIAS EN INGENIERÍA DE SISTEMAS**

**MARZO, 2017**

**Universidad Autónoma de Nuevo León**  
**Facultad de Ingeniería Mecánica y Eléctrica**  
**Subdirección de Estudios de Posgrado**

Los miembros del Comité de Tesis recomendamos que la Tesis «Simulación basada en agentes para la distribución del bienestar en México.», realizada por el alumno Ruth Berenice Cisneros Aldama, con número de matrícula 1204077, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Ciencias en Ingeniería de Sistemas.

El Comité de Tesis

---

Dr. J. Arturo Berrones

Asesor

---

Dr. Vincent Boyer

Revisor

---

Dr. Javier Almaguer

Revisor

Vo. Bo.

---

Dr. Simón Martínez Martínez

Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León,

# ÍNDICE GENERAL

---

<b>1. Agradecimientos</b>	<b>13</b>
<b>2. Resumen</b>	<b>14</b>
<b>3. Introducción</b>	<b>19</b>
<b>4. Antecedentes.</b>	<b>24</b>
4.1. Modelos económicos . . . . .	26
4.2. Modelos estocásticos. . . . .	28
4.2.1. Modelos de simulación basados en agentes computacionales. . . . .	33
4.3. Modelos basados en el ajuste. . . . .	34
4.3.1. Características deseables de un modelo de distribución o proceso generado. . . . .	38
<b>5. Marco Teórico</b>	<b>41</b>
5.1. Función de densidad, acumulada y complemento. . . . .	42
5.1.1. Momentos y Coeficiente de Gini. . . . .	44
5.2. El modelo Beta Generalizado: un marco de análisis en común . . . . .	45

<i>ÍNDICE GENERAL</i>	5
5.3. MSBA: la teoría de excedente de producción de la estratificación social.	47
<b>6. Métodos de estimación de la distribución estadística.</b>	<b>52</b>
6.1. Estimación no paramétrica de la densidad. . . . .	53
6.1.1. Función de densidad acumulada empírica: estimación directa. .	53
6.1.2. Función de densidad empírica: estimación tipo núcleo o kernel.	55
6.1.3. Estadísticos descriptivos y coeficiente de Gini . . . . .	57
6.2. Estimación paramétrica de la densidad: estimación por máxima verosimilitud . . . . .	58
6.2.1. Condiciones de regularidad. . . . .	60
6.2.2. La matriz de información de Fisher. . . . .	61
6.2.3. Algoritmos tipo Newton . . . . .	62
6.2.4. Propiedades de los estimadores de $MV$ . . . . .	64
6.3. Pruebas de bondad de ajuste. . . . .	65
6.3.1. Criterios de error para estimaciones de densidad. . . . .	67
6.3.2. Prueba de plausibilidad: estimación por bootstrap. . . . .	69
6.3.3. Prueba de Vuong para modelos no anidados. . . . .	71
6.3.4. Estimación de la cola derecha de la distribución. . . . .	73
<b>7. Implementación y datos</b>	<b>75</b>
7.1. Algoritmos para la estimación y pruebas de ajuste. . . . .	75
7.2. Algoritmo para el MSBA. . . . .	80

<i>ÍNDICE GENERAL</i>	6
7.3. Datos empíricos: descripción de los datos . . . . .	81
<b>8. Resultados</b>	<b>84</b>
8.1. Análisis preliminar de las muestras de datos empíricos . . . . .	84
8.2. Ajuste de las distribuciones empíricas a un modelo de probabilidad . . . . .	87
8.2.1. Análisis del total de la distribución. . . . .	88
8.2.2. Análisis por región: ingresos altos . . . . .	94
8.2.3. Análisis por región: ingresos bajos y medios . . . . .	99
8.3. Resultados de la simulación del MSBA . . . . .	102
<b>9. Conclusiones.</b>	<b>112</b>
9.1. Contribuciones . . . . .	115
9.2. Trabajo futuro. . . . .	116
<b>A. Revisión bibliográfica</b>	<b>118</b>
<b>B. Información estadística para México</b>	<b>123</b>
B.1. Encuesta Ingreso Gasto de los Hogares en México . . . . .	123
<b>C. Análisis por región</b>	<b>133</b>
C.1. Resultados para la región total de la distribución del ingreso . . . . .	133
C.2. Resultados para la región de ingresos altos . . . . .	138
C.3. Resultados para la región de ingresos medios-bajos . . . . .	143

<i>ÍNDICE GENERAL</i>	7
<b>D. Análisis de la simulación</b>	<b>151</b>
D.1. Estabilidad del proceso de simulación del MSBA . . . . .	151
D.2. Estimaciones de los modelos sobre los datos simulados . . . . .	153
<b>E. Términos y conceptos útiles.</b>	<b>157</b>

# ÍNDICE DE FIGURAS

---

5.1. Familia Beta Generalizada . . . . .	48
8.1. Diagrama de caja para la distribución del ingreso en México: 1984-2014	86
8.2. Complemento de la función de densidad acumulada 2014 . . . . .	90
8.3. Prueba de plausibilidad para el total de la distribución de ingreso . .	91
8.4. Coeficientes de Gini para la totalidad de la distribución del ingreso .	92
8.5. Medidas de divergencia y cota inferior de ingreso para el modelo de Pareto para los datos de 2014 . . . . .	94
8.6. Complemento de la función de densidad acumulada para la región de ingresos altos 2014 . . . . .	95
8.7. Prueba de plausibilidad para la región de ingreso alto de la distribución de ingreso . . . . .	96
8.8. Coeficientes de Pareto e índice de Gibrat para la región de ingresos altos. . . . .	97
8.9. Complemento de la función de densidad acumulada para la región de ingresos bajos-medios de la distribución de ingresos de 2014 . . . . .	100

8.10. Prueba de plausibilidad para la región de ingresos bajos-medios de la distribución de ingresos . . . . .	101
8.11. Coeficientes de Gini para la región de ingresos bajos-medios de la distribución de ingresos . . . . .	101
8.12. Complemento de la función de distribución acumulada de los ingresos simulados y su ajuste con respecto al modelo Dagum y Pareto. . . . .	110
8.13. Prueba de plausibilidad para los datos generados por el modelo de simulación de agentes . . . . .	111
D.1. Promedio de la autocorrelación serial del ingreso por agente. . . . .	151

# ÍNDICE DE TABLAS

---

5.1. Distribuciones de probabilidad. . . . .	49
8.1. Estadísticos descriptivos de la variable ingreso personal . . . . .	85
8.2. Estimaciones de los modelos de probabilidad para la totalidad de la distribución de ingresos año 2014 . . . . .	89
8.3. Prueba razón de verosimilitud para la distribución del ingreso . . . .	93
8.4. Prueba razón de verosimilitud para la cola derecha de la distribución	98
8.5. Prueba razón de verosimilitud para la región de ingresos bajos-medios de la distribución de ingresos . . . . .	102
8.6. Estadísticos descriptivos de los datos simulados . . . . .	104
8.7. Intervalos para las estimaciones según datos empíricos y simulados. .	107
A.1. Resumen de referencias bibliográficas para el totalidad de la distribu- ción del ingreso . . . . .	120
A.2. Resumen de referencias bibliográficas para la región alta de la distri- bución del ingreso . . . . .	121
A.3. Resumen de referencias bibliográficas para la región baja de la distri- bución del ingreso . . . . .	122

B.1. Claves de la muestras provenientes de la ENIGH. . . . .	132
C.1. Estimaciones de los parámetros para los modelos de probabilidad: totalidad de la distribución de ingresos . . . . .	135
C.2. Medidas de divergencia y pruebas de plausibilidad: total de la distri- bución de ingresos . . . . .	138
C.3. Resultados de la selección de la cota inferior para el modelo de Pareto	139
C.4. Resultados de la selección de la cota inferior para el modelo Exponencial	140
C.5. Resultados de la selección de la cota inferior para el modelo Gamma .	142
C.6. Resultados de la selección de la cota inferior para el modelo Log normal	143
C.7. Estimaciones para la región de ingresos bajos-medios: modelo Pareto	144
C.8. Estimaciones para la región de ingresos bajos-medios: modelo Expo- nencial . . . . .	145
C.9. Estimaciones para la región de ingresos bajos-medios: modelo Gamma	146
C.10. Estimaciones para la región de ingresos bajos-medios: modelo Log normal . . . . .	147
C.11. Estimaciones para la región de ingresos bajos-medios: modelo Dagum	148
C.12. Estimaciones para la región de ingresos bajos-medios: modelo Singh- Maddala . . . . .	149
C.13. Criterios de convergencia para modelos Singh-Maddala . . . . .	150
D.1. Coeficientes de variación para los estadísticos descriptivos de la simu- lación del MSBA. . . . .	152
D.2. Estimaciones para la región total de los datos simulados . . . . .	154

D.3. Estimaciones para la región de ingresos altos de los datos simulados . 155

D.4. Estimaciones para la región de ingresos bajos y medios de los datos  
simulados . . . . . 156

## CAPÍTULO 1

# AGRADECIMIENTOS

---

Agradezco a mi familia y amigos por el apoyo y la comprensión que me han brindado.

Al Dr. Berrones le agradezco por compartir conmigo sus ideas, experiencia y conocimiento, además agradezco al comité de tesis por las observaciones y apoyo a mi trabajo. Por último, deseo agradecer a mis compañeros del posgrado PISIS, y en especial, a mis compañeros de equipo de investigación que siempre me ayudaron a resolver todas mis dudas. Por último a CONACyT por el apoyo económico recibido para la realización de este trabajo.

## CAPÍTULO 2

# RESUMEN

---

El bienestar social de una región es definido como el grado de acceso que sus habitantes tienen a servicios de educación, salud y disfrutar de una vida digna. A su vez, el bienestar social puede ser definido a través de la forma de la distribución del ingreso de la sociedad debido al impacto que tiene esta última sobre diferentes factores económicos como lo son el crecimiento económico, la eficiencia en el desarrollo de la política pública, conflictos e inseguridad social y limitar el acceso en el mercado de capitales. Por lo tanto, el análisis de cómo se origina, cómo se desarrolla a través del tiempo y qué variables son afectadas por la forma de la distribución del ingreso ha sido un tema de interés en varias ramas de la ciencia. Asimismo, este tema es de interés regional, porque México es un país que presenta altos grados de desigualdad no sólo en ingreso, sino también en aspectos educativos, de oportunidades y de consumo. Por otro lado, dentro de la teoría económica tradicional, herramienta fundamental para evaluar la política pública, el estudio de la distribución del ingreso se ha abordado a través de modelos de equilibrio general que suponen que los agentes económicos son homogéneos y racionales, y que existe información perfecta en el mercado. Empero, la sociedad como sistema presenta interacciones no lineales que son difíciles de modelar a través de este tipo de modelos, por lo que en este trabajo se presenta un modelo de simulación basado en agentes para estudiar la forma de la distribución en México. Además, para llevar a cabo este propósito se presenta una

metodología que identifica cuál modelo de probabilidad se adecua de mejor manera al caso mexicano, y por consecuencia, se determina si existe algún mecanismo que está detrás de la generación del ingreso en México.

Ruth Berenice Cisneros Aldama.

Candidato para obtener el grado de Maestría en Ciencias en Ingeniería de Sistemas.

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: SIMULACIÓN BASADA EN AGENTES PARA LA DISTRIBUCIÓN DEL BIENESTAR EN MÉXICO..

Número de páginas: 169.

MOTIVACIÓN Y JUSTIFICACIÓN:

OBJETIVOS Y MÉTODO DE ESTUDIO: Dos son los objetivos en los que se enfoca este trabajo, el primero es el análisis de las muestras de la distribución del ingreso de los agentes económicos en México en el periodo que comprende de 1982 a 2014. El análisis comprende tanto a la totalidad del rango de la distribución del ingreso como las regiones de ingresos medios-bajos y de ingresos altos. La metodología implementada consta de (i) estimaciones paramétricas y no paramétricas de la densidad de la distribución de la variable ingreso de los agentes económicos en México; (ii) una prueba que informa si un modelo de probabilidad es una hipótesis plausible para generar un conjunto de datos, la cual es desarrollada a través del método *bootstrap*; (iii) una búsqueda exhaustiva para obtener el valor del límite inferior óptimo de la cola derecha suponiendo que ésta se distribuya a un modelo de probabilidad; (iv) además de una prueba de bondad de ajuste que selecciona entre dos modelos cuál es el más cercano a la distribución verdadera de los datos.

El segundo objetivo, es poner a prueba un modelo de simulación basado en

agentes (*MSBA*) base, inspirado en la teoría económica y física, para generar una distribución del bienestar, medida a través del ingreso, que sea consistente con el análisis de los datos empíricos. El obtener un modelo con estas características permite estudiar/evaluar cómo la forma de la distribución del bienestar puede cambiar a través de ciertos mecanismos clave, de interés en políticas públicas, e inferir de que manera pueden afectar a otras variables macro económicas que dependan de ella.

CONTRIBUCIONES Y CONCLUSIONES: Gracias a la implementación del *MSBA* se aprendió la necesidad de evaluar si el modelo que se propone puede ser utilizado de manera eficaz para la de predicción del sistema en estudio, y con mayor cuidado si el modelo servirá para evaluar políticas públicas que tendrán un incidencia en gran parte de la población. Entonces, si bien el modelo puede llegar a una conclusión general aceptable, como lo es que el modelo tenga el potencial de caracterizar los hechos observados a través de datos empíricos, se debe dar un paso atrás para reflexionar si no existe algún otro modelo que sea mejor en representar la realidad y que aporte un mayor entendimiento de la forma en que se genera la distribución del ingreso.

Con base a esta conclusión, se vio la necesidad de antes de pasar a otro nivel en el desarrollo de un *MSBA*, que pueda ser útil en el análisis de la distribución de el ingreso de la región, de estudiar más a fondo la dinámica de los datos empíricos del ingreso monetario de las personas en México. Se concluye que existe evidencia para decir que el proceso generativo que gobierna alguna región de la distribución del ingreso no es el mismo. Por lo que las reglas que se implementen en un *MSBA* debe variar con respecto a estas conclusiones.

Asimismo, se detecto la necesidad de promover una sinergia entre varias disciplinas de las ciencias con el fin de generar avances interesantes en el la investigación de la forma de la distribución del ingreso, los factores que la afectan, y a su vez, los factores que dependen de ella.

A continuación se enumeran las principales aportaciones que tiene este estudio.

- Implementación computacional de una metodología que ayuda a identificar el modelo de probabilidad del que provienen una muestra agrupada a por un ponderador.
- Se propone que los modelos Dagum, Log normal y Gamma son plausibles hipótesis para describir el proceso que genera al total de región de la distribución de ingreso. De manera similar, para la región de ingresos medios y bajos los modelos que resultan plausibles son el Log normal y Dagum. Por su parte, el modelo de Pareto y Log normal, en menor medida, son plausibles para caracterizar la región de ingresos altos de la distribución de ingresos en México.
- Además, de entre los modelos plausibles, se identifica los modelos que tienen mejor ajuste con respecto a los datos empíricos. Los mejores modelos, por número de parámetros y región de la distribución de los ingresos en México para el periodo 1992-2014, son los que siguen: los modelos Dagum y Log normal, para el total de la distribución y región de ingresos medios y bajos, y el modelo de Pareto para la cola derecha de la distribución.
- Se implementa un modelo de simulación de agentes, con base en la teoría cinética de gases y que mantiene una interpretación económica.
- Se muestra que el *MSBA* tiende a generar distribuciones más inequitativas para las combinaciones en las que existen mayor probabilidad de perder un porcentaje de ingreso alto –sociedad con bajo desarrollo social- y donde la probabilidad de que el agente que gane el encuentro sea el de mayor ingreso –sociedad dominada por las clases ricas-.
- Se comprueba que, para algunas de las distribuciones que genera el *MSBA*, los modelos Dagum, Gamma, Log normal y Pareto son hipótesis plausibles para generar dichos datos
- Se concluye que, el *MSBA* tiene potencial para replicar la forma de la distribución total del ingreso en México, en especial, si se supone que el modelo

Gamma es el proceso que genera los datos en México, como se observo para las muestras del 2012 y 2014.

Firma del asesor: \_\_\_\_\_  
Dr. J. Arturo Berrones

## CAPÍTULO 3

# INTRODUCCIÓN

---

América latina es una región famosa por sus riquezas naturales pero también por mantener altos grados de desigualdad no sólo en ingreso, sino también en aspectos educativos, de oportunidades y de consumo, repercutiendo de manera negativa en gran parte de la población. Y México no es la excepción, el informe sobre Desarrollo Humano 2014 de las Naciones Unidas [91] muestra que en comparación con Nigeria, Argentina o Perú, México tiene un mayor ingreso per cápita, de 13.971 dólares al día, pero también una peor distribución del mismo. Por otro lado, durante el transcurso de las dos últimas décadas México no ha logrado avances significativos en la disminución de las disparidades con que cuenta el país, por ejemplo Székely (2005) [88] menciona que, no ha variado mucho la distribución del ingreso en el país y que los niveles de desigualdad acontecidos en 1984 son casi los mismos que en el 2002.

De acuerdo al punto de vista normativo se puede considerar que cierta parte de la sociedad tiene preferencia por la justicia social y cree que al menos algunos tipos de desigualdad son injustos, a menos que sea el resultado entre personas sujetas a las mismas restricciones, o por influencia de esfuerzos propios para mejorar resultados, o por la responsabilidad de las acciones de cada persona. Sin embargo, cuando estas cuestiones no son parte de la desigualdad a la que se enfrenta una sociedad puede tener costos en la economía [91], como lo son:

- Reducción de los posibles beneficios que el crecimiento económico puede llegar a dar a los pobres.
- Impedir el acceso a los mercados de capital físico y humano de algunos grupos de la población.
- Reduce los alcances de la política pública.
- Aumenta la inestabilidad social, crea tensiones y conflictos sociales.
- Reduce el crecimiento económico agregado de un país.

Dados las anteriores aseveraciones, el estudio de la desigualdad ha despertado el interés en varias ramas de la ciencia por más de un siglo. Los investigadores se han ocupado en describir cómo es la forma funcional de la desigualdad, cuáles mecanismos pueden alterarla y cuáles son las variables que dependen de ella. Sin embargo, sigue siendo un tema abierto porque no se ha llegado a un consenso para responder tales incógnitas.

La ausencia de datos sobre la distribución del bienestar ha forzado a los investigadores a usar variables similares en los estudios empíricos tales como lo son la riqueza, el ingreso, la educación y la tierra, medidos por los activos, el dinero, años de escolaridad, y hectáreas que poseen los individuos [2]. Por ejemplo, Székely y Hilgerth (2000) [89], dan algunas razones de el porqué es factible el estudio de la variable ingreso para representar al bienestar: (i) la distribución del ingreso es informativo en lo referente al acceso a oportunidades económicas, además permite saber qué tantos sectores de la población son beneficiados por el desarrollo, (ii) el ingreso puede ser visto como un índice que refleja la capacidad para consumir bienes y servicios de la sociedad.

En el campo de la economía, tradicionalmente el estudio de la distribución del bienestar se realiza a través de modelos de equilibrio general, que asumen supuestos como la hipótesis de agentes económicos homogéneos y racionales, y la información perfecta en el mercado, con el fin de encontrar manejable el modelo y así realizar

inferencia sobre el tema [5]. Empero, la sociedad como sistema presenta interacciones no lineales, las cuales pueden ser difíciles de modelar a través de modelos de equilibrio general. Además, se ha argumentado que para que un modelo sea robusto en la predicción se tienen que considerar los mecanismos micro económicos que no cambian en el tiempo, como lo son, las preferencias heterogéneas entre los individuos [57]. En consecuencia, recientemente se ha impulsado el desarrollo de modelos de simulación basados en agentes; éstos tienen respaldo en teoría física, por lo que se les ha denominado modelos econofísicos, y su uso es sido impulsado gracias a los recientes avances computacionales.

Entonces, los modelos de simulación basados en agentes (*MSBA*), que a diferencia de los modelos de equilibrio general, se diseñan considerando los siguientes supuestos: (*i*) estrategia de evaluación desde «abajo hacia arriba» con base en reglas de decisión sencillas pero relevantes; (*ii*) el comportamiento de agentes es no homogéneo, no está basado en completa racionalidad e información completa; (*iii*) permite modelar diferentes reglas de decisión: maximización de utilidad, toma de decisiones según comportamiento, entre otras; (*iv*) los agentes puedan estar fuera de equilibrio y establecer interacciones complejas entre ellos. Uno de los trabajos pioneros en este campo es el de Angle (1986) [3] quien realiza una simulación de agentes que, con base en la teoría del excedente de producción de la estratificación social llega a generar distribuciones de riqueza de tipo Gamma similares a las que se han estimado para distintas sociedades; a su vez este modelo, puede verse como un caso especial de los modelos generados por la econofísica que toman inspiración en la teoría cinética de gases.

Por su parte, uno de los pasos clave para la creación de un *MSBA*, que tenga como objeto ser herramienta de política pública, es el análisis preliminar de los datos empíricos del sistema a modelar. El análisis preliminar permite (*i*) identificar el proceso del que proviene el sistema y de esta manera ayuda a idear las reglas «sencillas pero básicas» que pueden entrar en el modelo, y (*ii*) una vez hecha la modelación, conocer el proceso detrás del sistema ayuda a calibrar el modelo para

replicar los hechos que se observaron en el sistema.

La identificación de un modelo de probabilidad que genere a los datos de ingreso o riqueza es un tema abierto en la literatura que analiza estos datos empíricos. Discernir si los datos de ingreso provienen de algún modelo de probabilidad no es una tarea fácil; casi cualquier modelo puede ser ajustado a un conjunto de datos y en consecuencia, se puede discutir la calidad del ajuste del modelo con respecto a los datos empíricos, así como evaluar esta calidad contra otros modelos teorizados, a través de un análisis visual o comparando medidas error en el ajuste. Sin embargo, este ejercicio no indica si es plausible que el modelo en realidad genere los datos observados, por lo que recientemente la literatura se enfoca a desarrollar metodologías que sirvan para identificar el proceso que posiblemente genere a los datos, a su vez que, sea el que mejor los caracterice.

### **Objetivo.**

Dos son los objetivos en los que se enfoca este trabajo. El primero es el análisis de las muestras, a través del tiempo, de la distribución del bienestar en México, medido por el ingreso monetario de los agentes económicos, en el periodo que comprende de 1982 a 2014. El estudio abarca a la totalidad del rango de la distribución del ingreso, la región de ingresos medios y bajos y la región de ingresos altos. Estas regiones serán llamadas, respectivamente, región total, región baja y región alta a lo largo del documento. Y en específico, se estudian los modelos Pareto, Exponencial, Gamma, Log normal, Singh-Maddala y Dagum, modelos comúnmente utilizados en la literatura que estudia la distribución del ingreso, para determinar a cuál de ellos es más probable que pertenece la distribución del ingreso en México, y además, que sea el modelo que mejor la caracterice.

El segundo objetivo, es poner a prueba un *MSBA* base para generar una distribución del bienestar, que sea consistente con el análisis de los datos empíricos. El obtener un modelo con estas características permite estudiar y evaluar cómo la forma de la distribución del bienestar puede cambiar a través de ciertos mecanismos

clave de interés en políticas públicas e inferir de qué manera afectan a otras variables macro económicas que dependan de ella.

### **Estructura de la tesis.**

Este trabajo se organiza como sigue: en el capítulo 4 se da un repaso de los principales trabajos que han aportado conceptos y modelos relevantes en el estudio de la distribución del ingreso, y se da especial énfasis, en presentar los alcances y limitaciones en las líneas de investigación de los modelos estocásticos y los modelos de probabilidad que han presentado un buena ajuste para caracterizar a la distribución del ingreso.

A continuación en el capítulo 5 se repasan conceptos para caracterizar a la distribución de una variable; se describe el modelo Beta Generalizado que es un marco en común para analizar a los modelos de probabilidad que se estudiarán en este documento para caracterizar al ingreso; y por último, se repasa la teoría del excedente de producción de la estratificación social que es base del *MSBA* estudiado en este trabajo. Luego, en los capítulos 6 y 3, se detallarán las metodologías utilizadas para analizar los datos empíricos en México y la descripción de la implementación de éstas y del *MSBA*.

Por último, en los capítulos 8 y 9 se detallan los resultados obtenidos en esta investigación y se procede a resumir las conclusiones generadas.

## CAPÍTULO 4

# ANTECEDENTES.

---

Los datos empíricos sobre el ingreso o riqueza muestran, a través de diferentes regiones geográficas y en el tiempo, la característica de que la gran mayoría de la población concentra un poco nivel de ingreso, mientras que, unos pocos llegan a obtener grandes sumas de ingreso; este fenómeno se traduce en que en la representación gráfica de la frecuencia de personas que reciben ingreso, por nivel, se observe una gran concentración de datos del lado izquierdo del gráfico, y a su vez se presenta, una cola derecha ancha y larga. Uno de los trabajos seminales que presenta estas similitudes fue el de Vilfredo Pareto en 1987 [66], en su investigación estudia, para diferentes momentos en el tiempo, la forma de la distribución del ingreso en distintas sociedades, como Alemania, Inglaterra, Irlanda, Perú y ciudades de Italia; concluye que las distribuciones podían ser descritas, con pocas diferencias, por una distribución tipo ley de potencia que actualmente se le conoce como ley de Pareto.

En específico, Vilfredo Pareto en 1897, observa que el ingreso puede ser descrito a través de una relación lineal decreciente entre el logaritmo del ingreso,  $x$ , y el logaritmo del número de personas que recibían un ingreso mayor que  $x$ ,  $N_x$ , especificándose de la siguiente forma:

$$\log N_x = A - \alpha \log x \quad (4.1)$$

en donde  $A$ ,  $\alpha > 0$ . De ahí que Pareto propone [66] que la distribución del ingreso

seguía la siguiente ley natural

$$P(X > x) = \begin{cases} \left(\frac{x}{x_0}\right)^{-\alpha} & \text{para } x \geq x_0 \\ 1 & \text{para } x < x_0 \end{cases} \quad (4.2)$$

Con un valor estimado de  $\alpha \approx 1.5$ . A esta expresión se le conoce como ley fuerte Pareto [59], y a su vez, se le llama ley de Pareto débil o estrecha cuando la cola derecha de la distribución se comporta como una ley de Pareto [59].

Este tipo de función de distribución tiene como característica que la varianza, e incluso la media, pueden no estar definidas matemáticamente, por lo que sería incorrecto hacer inferencia de los datos con base en estadística tradicional, que está basada en la varianza y desviación estándar. Entre los sistemas en los que se han documentado leyes de potencias se encuentran: terremotos, avalanchas, ventas comerciales, tamaños de guerras, caídas en los mercados de capitales, etcétera.

Por consiguiente, la aportación de Pareto crea gran expectativa por las siguientes razones: (i) impulsa un concepto de universalidad y de escala no variante en la distribución del ingreso; donde la uniformidad de la forma de las colas derechas de las distribuciones del ingreso observadas podrían implicar que un mismo proceso de distribución opera en todas las sociedades; (ii) además, respalda la idea de que la probabilidad de observar un evento de gran escala no es desdeñable y se contrapone a la inferencia con base en la teoría Gaussiana [84]; por último, (iii) da pie al desarrollo de uno de los temas más discutidos en economía que es la relación entre el crecimiento económico y la distribución del ingreso.

Asimismo, es importante reconocer si un fenómeno pertenece a una ley de potencia porque la distribución podría restringir el mecanismo detrás del sistema, así como crear grandes desviaciones que casi siempre controlan las dinámicas de largo plazo y organizan un sistema complejo [84]. En el campo de la economía la importancia de reconocer la forma de la distribución del ingreso o riqueza radica en su relación con variables que influyen en los mecanismos de la política económica, monetaria y fiscal, tales como el consumo, los ciclos económicos y la incidencia de la

carga impositiva [5]. Por lo tanto, si la cola de la distribución se comporta conforme a una ley tipo Pareto, esta relación se vuelve de gran interés, porque variaciones pequeñas en esta sección pueden ocasionar impactos de considerable tamaño en las variables macro económicas con las que se relacione.

Para describir las dinámicas del ingreso o riqueza de los individuos se pueden identificar tres líneas de investigación de acuerdo a la función de distribución de probabilidad que proponen. La primera línea desarrolla modelos basados en teoría económica que representan algún factor que determina a la forma de la distribución del ingreso; la segunda línea, propone modelos que explican la generación de la forma de la distribución del ingreso como resultado de un proceso estocástico; y por último, la tercera línea plantea modelos de probabilidad que obtienen un buen ajuste en datos empíricos, que sean fáciles de estimar y que sus parámetros tengan alguna explicación económica.

A continuación se dará una explicación de cada línea de investigación sobre la forma de la distribución del ingreso y se pondrá especial interés en los modelos de tipo estocástico y de ajuste que son los que estudian este trabajo.

## 4.1 MODELOS ECONÓMICOS

La economía es un sistema complejo, con inestabilidades endógenas que causan eventos críticos que pueden ser muy azarosos. En el análisis económico tradicional se suelen hacer ciertos supuestos para controlar las inestabilidades del sistema y de esta forma desarrollar modelos que ayuden a analizar ciertos fenómenos económicos. Los principales supuestos son los que siguen: *(i)* las acciones de los agentes generan precios que llevan al mercado al equilibrio donde la oferta es igual a la demanda, *(ii)* en los modelos se supone un agente representativo que engloba las preferencias colectivas y que toma decisiones racionales para optimizar el uso de recursos limitados, y *(iii)* mercados con información completa.

El concepto de equilibrio en la economía, desarrollado con base en el concepto del macro equilibrio de un gas, plantea que es el resultado de una multitud incesante de micro colisiones de las partículas de un gas [8]; de igual forma el equilibrio económico es resultado de la interacción de todos los agentes económicos que se encuentran relacionados entre sí [54]. Sin embargo, actualmente en el campo de la economía neoclásica al concepto de equilibrio se le da un rol normativo; forzando a que el concepto no describa el sistema económico como es, sino como debería de ser [84]. Dos modelos que, con base en los supuestos mencionados, desarrollan teorías de cómo se forma la distribución del ingreso o la riqueza son los que siguen:

1. Modelos de equilibrio general (*MEG*). Estos tipos de modelos se basan en la teoría neoclásica en la que los precios son el mecanismo que mantiene el equilibrio entre la oferta y la demanda del mercado. La forma en la que se genera la distribución del ingreso es explicada desde varios puntos de vista, como lo son la determinación de salarios en mercado laboral, el pago a los factores de producción y la acumulación de la riqueza [5]. Para una revisión más profunda de estos modelos puede consultar a Atkinson y Bourguignon (2014) [5].

2. Modelos de generaciones traslapadas (*MGT*): son modelos que estudian la distribución de recursos entre generaciones. El modelo supone a un agente representativo, que vive por dos o más periodos de tiempo, con características iniciales heterogéneas, como lo es la dotación de riqueza inicial, y que con base en expectativas racionales sobre su consumo en el futuro optimiza decisiones tales como invertir su tiempo o riqueza en estudiar o trabajar, ahorrar o gastar, entre otras. A diferencia de *MEG*, el estado estacionario no tiene que ser eficiente, se adapta para manejar varias teorías sobre la creación de la desigualdad en el ingreso como es la ineficiencia en el mercado de capitales, la apertura comercial y el crecimiento económico. Además justifican la introducción expectativas del futuro dinero como medio de intercambio [37, 67].

Sin embargo, la principal crítica de este tipo de modelos es sobre el supuesto

del agente representativo, el cual puede no ser adecuado para obtener predicciones de cómo se comporta un sistema cuando hay un cambio abrupto en las condiciones económicas o políticas del gobierno, ya que en la realidad, los agentes no siempre van a decidir con racionalidad ante un cambio. De igual forma, la crítica de Lucas [57], expone que los modelos macroeconómicos tienen la debilidad de no ser robustos porque no incorporan características micro económicas en sus modelos.

En respuesta a estas críticas se desarrollan los modelos de equilibrio general dinámico estocástico (*MEGD*) que tratan integrar bases de tipo microeconómicas, como lo es heterogeneidad en las preferencias de los agentes económicos, así como un dinamismo en el sistema. Sin embargo, no incluyen de manera completa la heterogeneidad en agentes porque enfrentan la disyuntiva entre un sistema manejable y la incorporación de la realidad para llegar a un análisis satisfactorio [7].

Un ejemplo de estos avances es el modelo evolutivo o de tipo Shumpeter. Éste es un modelo dinámico, que a diferencia del *MEGD* tradicional, analiza a la economía fuera del equilibrio. Entonces, este modelo utiliza ecuaciones diferenciales para integrar agentes heterogéneos y describir la dinámica de la economía. Sin embargo, usualmente solo se supone una característica diferente entre los agentes y la toma de decisión sigue procedimientos explícitos de optimización que se basan en las condiciones de mercado y las características homogéneas y heterogéneas de cada agente [1, 51]. Por lo que, se puede resumir que los modelos evolutivos terminan imponiendo la dinámica de los agentes a través de modelación de ecuaciones diferenciales [56].

## 4.2 MODELOS ESTOCÁSTICOS.

En la segunda línea de investigación se utilizan modelos estocásticos para describir las dinámicas del ingreso o riqueza de los individuos y así derivar su función de densidad de probabilidad. Tres trabajos precursores de esta línea de investigación,

además del de Pareto, son: el de Gibrat (1931) [39] concluye que la ley de Pareto no aplica a todo el rango de los datos de ingreso y que sólo se observa para la región de ingresos altos. Por lo cual, propone una ley de efecto proporcional en donde un pequeño cambio en una cantidad es independiente de la cantidad en sí misma, y por consecuencia, postula al modelo Log normal para describir a la gran mayoría de ingresos, bajos y medios. Además, Gibrat define un índice para medir la desigualdad en la distribución a partir del coeficiente de forma de la distribución. Asimismo, Champernowne en 1953, demuestra que bajo ciertos supuestos, un proceso de Markov podría aproximar a la distribución del ingreso de acuerdo a la ley de Pareto, sin importar la distribución inicial [5]. Por último, Mandelbrot en 1960 [59] presenta un modelo, con base en los avances de Champernowne, que llega a una distribución de Pareto para los ingresos altos de la distribución [59].

A partir de la década de los noventas se intensifica el interés por explicar la forma de la distribución del ingreso por métodos estocásticos. Este interés es promovido por la amplia disponibilidad de datos sobre ingreso de las sociedades y los avances tecnológicos en computación, que permiten estudiar los modelos estocásticos a través de los modelos de simulación basados en agentes computacionales [17]. El principal reto es construir una ecuación microscópica que gobierne la dinámica de la distribución del ingreso y que sea capaz de reproducir la forma de la distribución del ingreso empírica [13].

Una de estas líneas de investigación toma como inspiración la teoría cinética de colisión de gases de Boltzmann e introduce los modelos cinéticos de transferencia de dinero (*MCTD*). Esta metodología describe la manera en que los agentes transfieren dinero de un agente a otro, a través de transacciones en parejas, para generar distribuciones de equilibrio tipo Exponencial, Gamma, Log normal y Pareto. En específico, el modelo *MCTD* supone una economía cerrada, que más que describir la dinámica del mercado en términos de agentes inteligentes, tiene como objetivo predecir la evolución de la distribución de una variable, como el ingreso, al estudiar el proceso de flujos entre individuos. Entonces, la idea principal es llegar

a describir, de una manera simple, el comportamiento promedio de los agentes, por más complicadas que sean las reglas del intercambio del ingreso observada.

La justificación del uso de la analogía entre dinero y la teoría de gases para describir la generación de la distribución del ingreso o riqueza, es que a través del dinero se puede medir la riqueza y el ingreso de los agentes en un periodo dado de tiempo [13]. En la actualidad, el dinero es un bien que no tiene ningún valor intrínseco más el que la sociedad le otorga para representar el valor de los bienes y servicios que existen en la economía. Además el dinero tiene varias representaciones como lo son las monedas y billetes, cuentas corrientes o dinero bancario. De esta forma, el dinero es utilizado para los siguientes fines: (i) facilitar operaciones económicas como es el intercambio de bienes y servicios en la economía; (ii) medida del valor que hace posible la determinación de precios, costos, ganancias y pérdidas económicas; (iii) el medio que permite realizar pagos de deudas y la fijación del valor de futuras transacciones; y (iv) como medio para acumular riqueza a través del tiempo [75]. Asimismo, el dinero tiene la característica de que la cantidad circulando en la economía es regulada por un agente económico como lo es el gobierno y/o el banco central, éstos tienen como objetivo mantener la estabilidad económica a través del control de las reservas bancarias y la fijación de la tasa de interés [64]. Con base a esta descripción, en los modelos *MCTD* se utilizan la variable dinero para representar al ingreso o a la riqueza, de manera indistinta [13].

Entonces, en estos modelos el estado del agente está definido en términos del ingreso  $x_n$ ,  $n = 1, 2, \dots, N$ . La evolución del sistema, y su subsecuente distribución final de equilibrio, dependerá de la regla de transferencia designada a los agentes. El algoritmo se basa en una simple actualización de los resultados al aplicar la regla de transferencia en el tiempo  $t$ , cuando dos agentes  $i$  y  $j$  se enfrentan y es intercambiada una cantidad  $\Delta x$  del ingreso, tal que:

$$\begin{aligned} x_{i,t} &= x_{i,t-1} - \Delta x \\ x_{j,t} &= x_{i,t-1} + \Delta x, \end{aligned} \tag{4.3}$$

donde  $x_{i,t}$  y  $x_{j,t}$  es el ingreso de los agentes  $i$  y  $j$  después del encuentro, en tanto que,

$x_{j,t-1}$  y  $x_{i,t-1}$  es su similar antes del encuentro. Nótese que la cantidad  $x$  se conserva durante cada encuentro:  $x_{j,t} + x_{i,t} = x_{j,t-1} + x_{i,t-1}$ .

Cabe mencionar, que aunque este modelo se deriva del principio de máxima entropía, proveniente de la estadística mecánica, se demuestra que, el mismo modelo puede ser desarrollado a través del principio de la maximización de la utilidad, comúnmente utilizado en teoría económica, si se incorpora al modelo de intercambio una función de utilidad tipo Cobb-Douglas y precios relativos para el intercambio [80]. Es decir, el modelo puede obtener justificación tanto por teoría de la estadística mecánica como por la de tipo económica.

Uno de los trabajos pioneros en tomar esta idea fue de Angle (1986) [3]; después, los trabajos de Bouchaud y Mézard (2000) [10], Chakraborti y Chakrabarti (2000) [12] y Dragulescu y Yakovenko (2000) [28] dieron pie a una extensa línea de investigación basada en dicha teoría. Muchas reglas de transferencia para el modelo 4.3 son estudiadas y en consecuencia se desprenden diferentes conclusiones sobre la forma de la distribución de equilibrio generada.

Por ejemplo, Dragulescu y Yakovenko (2000) [28] realizan un modelo donde proponen que la cantidad  $\delta x$  es aleatoria y llegan a una distribución tipo exponencial Boltzmann; por su parte, Chakraborti y Chakrabarti (2000) [12] proponen un modelo en donde existe una propensión marginal constante para ahorrar entre los agentes y de acuerdo a este modelo se llega a una distribución estacionaria Gamma. Por su parte, Chatterjee et al. (2004) [16] llegan a una distribución de ingresos con una cola derecha conforme a la distribución de Pareto al presentar un modelo con propensiones ahorrar diferentes para cada agente,  $\lambda_i$ , donde la propensión se otorga de manera aleatoria entre los agentes, y además, agregan la regla de que se promedia la distribución estacionaria resultante de varias configuraciones de  $\lambda_i$ . Asimismo, Chatterjee y Chakrabarti (2005) [15] formulan una ecuación que caracteriza a la propensión ahorrar de acuerdo a diferentes modelos de probabilidad; analíticamente llegan a una distribución tipo Pareto en la cola, las limitaciones que encuentran son:

(i) no toma en cuenta con un incentivo para que los agentes entren en el intercambio, y (ii) su análisis se refiere al papel moneda. De manera similar, Gusman et al. (2005) [41] presentan un modelo en donde los agentes que pertenecen a la misma clase social pueden comerciar entre ellos; el modelo genera una distribución con marcada polarización y una disminución de la clase media, y comentan que la mejor manera de obtener la ley de Pareto es aumentando el promedio de conectividad. Por su parte Yarlagadda y Das (2005) [96] formulan una función de probabilidad para agentes con dos tipos de ingreso y asumen que aquellos agentes con un ingreso por debajo de un umbral de pobreza sólo interactúan entre ellos, en tanto que los agentes ricos sólo comercian con una entidad mayor; la función crea una distribución de tipo exponencial Boltzmann para la parte pobre y tipo Pareto para la parte de ingresos altos, además, analíticamente demuestran cómo obtener diferentes colas tipo Pareto y crean un enlace con teoría macroeconómica mediante una maximización de consumo y riqueza. Por último, se puede mencionar el trabajo de Lux (2005) [58] que crea un modelo en el que los agentes tratan de maximizar su función de utilidad al intercambiar bienes, asimismo, al vaciar el mercado mediante el precio de equilibrio, suponiendo los tipos de mercado monopolio y competencia perfecta. Lux (2005) [58] llega a una distribución de ingreso tipo Gamma sin crear una cola tipo Pareto, entonces concluyen que el intercambio de bienes hace la distribución del 90 % de población, empero discute que hace falta otro mecanismo para generar la cola tipo Pareto.

Por otro lado, una de las principales críticas que reciben estos modelos es su falta de contenido sobre teoría económica, en específico, se tiene la percepción de que no aportan una teoría que explique la formación de la distribución del ingreso y que en general, dejan de lado la optimización de los individuos para la toma de decisión [63]. Sin embargo, Atkinson et al. (2014) [5] argumentan que esta crítica puede ser exagerada, porque no se encuentran diferencias fundamentales entre los modelos estocásticos y los modelos económicos; estos últimos, recurren a la aplicación de supuestos para llegar a modelos analíticos, y por ejemplo, asumen como un

error estocástico a todas aquellas variables exógenas que en su modelo no se pueden explicar; y por el contrario, mencionan que es una oportunidad para observar las fuerzas y limitaciones de las metodologías desarrolladas a través del tiempo y conseguir sinergias que ayuden a desarrollar la investigación sobre la distribución de ingreso [4].

En consecuencia, se ha iniciado esfuerzos por juntar de manera inclusiva las teorías de economía y la física; un primer intento fue una reunión en el instituto de Santa Fe en 1988 con resultados no muy fructíferos. Empero, poco a poco se encuentran varios acercamientos tanto de colaboraciones en investigación como publicaciones en revistas especializadas de ambos campos [94].

#### 4.2.1 MODELOS DE SIMULACIÓN BASADOS EN AGENTES COMPUTACIONALES.

Este tipo de modelos están diseñados para capturar la heterogeneidad en los agentes, que permiten modelar, diferentes preferencias y expectativas, que influyen en el proceso de decisión, así como diferentes tipos de agentes, como lo son consumidores, entidades del Estado, grandes firmas entre otros. Asimismo, la metodología permite llegar a desarrollar complejas interacciones de agentes, a través de reglas a nivel micro que sean sencillas pero relevantes para el sistema. Los modelos de simulación basados en agentes (*MSBA*) no están sujetos a condiciones de equilibrio; situaciones fuera de equilibrio pueden surgir como consecuencia de la interacción entre agentes o cambios en las condiciones del ambiente. Por lo tanto, esta metodología permite la interacción entre diversas disciplinas con el fin de definir las reglas de la simulación que se mantengan sencillas pero relevantes.

Asimismo, el modelo de simulación basado en agentes puede ser visto como una metodología que complementa a los modelos analíticos tradicionales. Por ejemplo, el *MSBA* puede aportar el equilibrio del sistema, si el modelo analítico es incapaz de

llegar a él; además, ayuda a explicar cómo surgen patrones de alto orden como lo son las leyes de potencia en la distribución del ingreso o del riesgo en el mercado de capitales, la segregación social persistente aun cuando existe una sociedad tolerante y estructuras de redes del internet; los *MSBA* también pueden ser útiles para identificar algún cambio interno o externos en el sistema que cause consecuencias extremas, y por lo tanto, permiten analizar cómo es la dependencia del sistema con el cambio, y a su vez, identificar la trayectoria en que el sistema regresa a su funcionalidad.

Las limitantes que se les encuentra son las siguientes: (i) es difícil de comprender de manera analítica cómo las reglas y elementos micro se transforman en comportamientos macro cuando se agregan los resultados de la interacción entre de varios agentes; (ii) los modelos pueden no ser comparables y robustos, esto es, el diseño de *MSBA* tiende a estar sesgado por las preferencias del modelador, además que, el desempeño del sistema depende de lo relevante de las reglas usadas y no es fácil generalizar las reglas aplicadas para utilizarlas en la predicción; y (iii) uno de los problemas más fuerte es la correcta calibración del modelo, es decir, se debe elegir aquellos parámetros que caractericen de mejor manera a los datos empíricos.

El conocimiento profundo de los mecanismos detrás del sistema, a través del análisis de datos empíricos, ayuda a elegir el tipo de *MSBA* que mejor se adecue a la realidad. Por lo tanto, es deseable tomar en cuenta los problemas metodológicos para estimar los parámetros que entraran al modelo; dichos problemas son la calidad de los datos, llegar a identificar el verdadero proceso generativo de datos, e identificar las condiciones del sistema en las que los parámetros son válidos.

### 4.3 MODELOS BASADOS EN EL AJUSTE.

Sin embargo, si se desea probar la veracidad de los modelos de estas tres líneas de investigación se debe tener una observación certera del verdadero proceso empírico de generación del ingreso [36, 83]. Actualmente, algunos de los estudios que conclu-

yen que un modelo de probabilidad es bueno para describir muestras empíricas de ingreso se basan en una relación espuria o sólo son válidos en un rango limitado [84]; las pruebas para discernir si datos empíricos provienen de algún modelo de probabilidad se basan en un examen visual del complemento de la función de densidad, el cual puede llegar a ser subjetivo, y no toman en consideración que casi cualquier distribución puede ser ajustada a datos empíricos, y por lo tanto, se debe distinguir cual es el valor del criterio de error que corresponde sólo al proceso de muestreo de aquel que indique que los datos provienen de otro proceso [19].

Dentro de esta línea de investigación, los modelos Pareto, Exponencial, Gamma, Log normal, Dagum y Singh-Maddala destacan por su buen ajuste empírico en el estudio de la distribución del ingreso. El estudio de las semejanzas entre estos modelos y datos empíricos es llevado para diferentes rangos en la distribución del ingreso y riqueza como lo son el total de la distribución, la cola derecha de la distribución y la región de ingresos pequeños y medios; diferentes definiciones de ingreso, regiones geográfica, y tiempo. Además, los métodos de evaluación varían entre el análisis visual, la comparación de medidas de error -como lo son la suma de cuadrados del error, el logaritmo de verosimilitud, el criterio de Akaike- pruebas de bondad de ajuste tal como lo son prueba de razón de verosimilitud y pruebas no paramétricas que utilizan la metodología *bootstrap*. Por otro lado, el método de estimación también cambia de investigación a investigación; se pueden encontrar estimaciones por mínimos cuadrados ordinarios, estimación con base en los momentos de la distribución, estimación por máxima verosimilitud y estimación bayesiana. Debido a esta variedad en los factores de estimación no se presentan un consenso en el valor coeficientes estimados de los modelos de probabilidad, así como en la preferencia entre el ajuste de un modelo a otro [6, 95]

En el apéndice A se encuentra una tabla con un resumen sobre la bibliografía en la que se basa este trabajo y se identifican las variantes que se acaban de mencionar. Además, a continuación se resume la información relevante sobre el ajuste sobre datos empíricos de los modelos que se analizan en este trabajo, y posteriormente,

en el capítulo 5 se revisan las expresiones que describen sus formas funcionales y estadísticos descriptivos relevantes para esta investigación.

La distribución Exponencial se utiliza en el ajuste de la totalidad del rango de la distribución y en la sección de ingresos bajos y medios. El interés en esta distribución radica en que fundamenta el desarrollo de modelos estocásticos, basados en teoría física, para analizar la dinámica del ingreso. En específico, se encuentra que el parámetro de la distribución exponencial, que denota al promedio de la distribución, es análogo al parámetro de temperatura en la distribución Boltzmann-Gibbs, utilizada en la teoría de gases [94]. Sin embargo, se observa que existen desviaciones en el ajuste de los extremos de la distribución que se atribuyen a omisiones en el reporte de ingresos muy bajos o sub declaración en ingresos muy altos [29, 30, 95]. Por otro lado, el modelo Log normal se usa para caracterizar a la mayor parte del rango de la distribución del ingreso, sin embargo, se muestra que la distribución log-normal no es estacionaria porque aumenta su ancho con el tiempo [74]. Además, McDonald (1984) y Bandourian et al. (2003) mencionan que para el rango total de la distribución, el modelo Log normal sobre estima el ingreso en la cola derecha y la media estimada es superior a la poblacional, es decir, que resulta en un modelo con mayor dispersión que los datos observados. Por su parte, el modelo Gamma es un modelo de dos parámetros que es utilizado para caracterizar datos empíricos de ingreso, presentado por Aman en 1925, y de manera más reciente propuesto por Salem y Mount en 1974 [74]. Los dos parámetros de la distribución Gamma pueden ser relacionados con las medidas de escala y asimetría estadística, por lo que también son asociados con la ley de proporcionalidad de Gibrat y a la desigualdad en la distribución. Con respecto al ajuste sobre datos empíricos de ingreso, el modelo Gamma reporta que la distribución tiende a exagerar la asimetría estadística [74].

Por su parte, los modelo Dagum y Singh Maddala son modelos de tres parámetros, propuestos en el estudio de la distribución del ingreso por Dagum (2008) [24] y Singh y Maddala (1976) [82]. Estas dos distribuciones están conectadas porque se pueden llevar resultados de una distribución a otra [53]. Entonces las distribuciones

Dagum y Singh-Maddala son introducidas al análisis de la distribución del ingreso al buscar un modelo que incorpore las colas anchas de la región alta del ingreso, como la distribución de Pareto, pero también que tengan una moda interior como la log normal [24]; por consiguiente, estos dos modelos tienen la característica de sus casos especiales son los modelos Pareto y Fisk.

Los resultados del análisis sobre qué modelo es mejor para ajustar los datos empíricos del ingreso no son conclusivos en ninguna de las regiones de la distribución. Por ejemplo, para la parte baja de la distribución, ingresos bajos y medios, se reporta como bueno el ajuste que la distribución exponencial sobre el ingreso personal y familiar para países como Estados Unidos y Reino Unido. [29, 30, 95]; de igual forma, estudios sobre la distribución Log normal presentan buen ajuste en la región de ingresos bajos para países como Estados Unidos, Reino Unido, Japón, Italia, Australia e India [20, 21, 26, 83, 85, 87].

En cuanto a la región de la cola derecha, ingresos altos, los estudios empíricos se enfocan, en su mayoría, al estudio de la distribución de Pareto y llegan a que ésta mantiene un buen ajuste visual sobre los datos para diversos países, diferentes definiciones de ingreso y lapsos de tiempo [20, 21, 26, 30, 35, 83]. Mientras que, unos pocos hacen pruebas más estrictas para saber si es plausible que un conjunto de datos empíricos provengan del modelo de probabilidad Pareto, sin embargo, los resultados que presentan son diversos. Por ejemplo, Clauzet et al. (2009) [19] reportan que no hay evidencia suficiente para aseverar que la riqueza neta individual de las personas más ricas del mundo, reportada por Forbes en 2003, provenga de un modelo tipo leyes de potencia. Por otro lado, Toda (2012) [90], a través de una prueba Kolmogorov-Smirnov y Anderson-Darling, no rechaza la hipótesis nula, en más del 80 % de sus muestras, de que el error de una ecuación de Mincer<sup>1</sup> sobre la cola derecha se distribuya de acuerdo a un modelo de Pareto. Por último, Brzezinski (2014) [11] afirma que, de acuerdo a la prueba de plausibilidad desarrollada por [19],

---

<sup>1</sup>El modelo de Mincer, desarrollado por Jacob Mincer, es utilizado para modelar el ingreso del salario de un individuo en función de escolaridad y experiencia

para el 28% de su muestra, compuesta por el ingreso personal de personas de Estados Unidos, China y Rusia, existe evidencia de que el modelo de Pareto sea una hipótesis plausible para sus datos.

Para finalizar, el análisis del rango total de la distribución incluye a todas las distribuciones mencionadas, anteriormente, y de igual forma los resultados son diversos. Por ejemplo, la distribución Dagum presenta menores criterios de error que los modelo Singh-Maddala, Log normal, Gamma, para muestras de diferentes países [6, 24, 53, 73]; sin embargo, también se reporta que modelo Singh-Maddala tienen mejores resultados para caracterizar al ingreso que el modelo Dagum [18] y los modelos Log normal y Gamma [60, 82]. Asimismo, hay evidencia de que el modelo Log normal es mejor que el modelo Gamma para ajustar los datos empíricos del ingreso [68], y por el contrario, existe preferencia por el modelo Gamma sobre el modelo Log normal [74]. Cabe mencionar, que también el modelo Exponencial es analizado, de manera aislada, en el estudio de la distribución del ingreso [29].

Para el caso mexicano, se concluye que para el rango total de la distribución, el modelo Dagum es el que obtiene mejor ajuste que los modelos Pareto y Log normal, sin embargo, dentro modelos de dos parámetros el modelo Log normal es el que mejor caracteriza a los datos [6, 73]. Cabe señalar, que hasta el momento de escribir este trabajo no se encuentran trabajos que estudien, mediante un modelo de probabilidad, a los datos empíricos de México por región de la distribución de ingresos.

#### 4.3.1 CARACTERÍSTICAS DESEABLES DE UN MODELO DE DISTRIBUCIÓN O PROCESO GENERADO.

Por último, la literatura enumera propiedades económicas, econométricas, estocásticas o matemáticas [3, 24] para identificar si un modelo es representativo de la distribución del ingreso. Algunas de las propiedades discutidas son las que siguen:

- El modelo puede derivarse con base en fundamentos lógicos y empíricos que sea acorde a la metodología desarrollada en ciencias y filosofía. Como lo es la hipótesis de Kusntez<sup>2</sup>
- Cumplir con la forma débil de la ley de Pareto [59], esto es, el modelo de probabilidad presenta como caso especial, para el grupo de ingreso alto, al modelo de Pareto.
- El modelo debe implicar un proceso que genera una distribución estable, es decir, que converge rápido a una distribución del ingreso personal empírica.
- Tener soluciones matemáticas explícitas y tratables que permitan estimar herramientas útiles para medir la desigualdad en la distribución como lo son el índice de Gini y curva de Lorenz.
- Buen desempeño del modelo para ajustar datos empíricos, en todos los niveles de ingreso, con el fin de ser base en el desarrollo de política pública.
- Preferencia por modelos y teorías con pocos supuestos y numero de parámetros; con gran poder explicativo, buen ajuste a datos empíricos y que se puedan estimar con métodos sencillos y eficientes.
- Comparación de las debilidades y fuerzas de distribuciones analizadas con anterioridad en la literatura.
- Flexibilidad en el modelo para manejar los siguientes puntos: (i) capaz de generar cambios en la forma de la distribución por medio de cambios en el valor de los parámetros; (ii) describir distribuciones que sea estrictamente decrecientes y uni modales; (iii) considerar, a través de sus parámetros, diferentes definiciones de ingreso como es la riqueza, el ingreso y el dinero, los cuales mantienen diferentes rangos.

---

<sup>2</sup>La hipótesis de Kuznets (1955) [55] establece una relación de refuerzo entre la desigualdad de la distribución del ingreso y el crecimiento económico, tal que, a menores niveles de ingreso la desigualdad en la distribución de la misma irá en aumento, pero al llegar a un punto de crecimiento del ingreso la desigualdad en la misma tenderá a caer.

De acuerdo a los criterios de selección de un modelo mencionados y con el objetivo de caracterizar la distribución del ingreso de los agentes económicos en México se analizan los modelos de probabilidad de Pareto (P), Exponencial (Exp), Gamma (G), Log normal (LN), Dagum (D) y Singh-Maddala (SM). Dichos modelos, como se acaba de exponer, se han destacado por su uso para caracterizar la distribución ingreso y son relevantes en el desarrollo de modelos estocásticos.

## CAPÍTULO 5

# MARCO TEÓRICO

---

En esta sección se introducen conceptos básicos sobre la distribución de una variable como son las definiciones de la función de densidad de probabilidad, función de densidad acumulada y su complemento. Asimismo, se describen expresiones cerradas para calcular estadísticos descriptivos de la distribución.

A continuación, para facilitar el análisis de los modelos de probabilidad propuestos, para caracterizar a los datos de ingreso en México, se toma como marco base el modelo Beta Generalizado (BG). McDonald (1984) [60] y McDonald y Xu [61] introducen el modelo BG con avances de Esteban [32,33]. El modelo BG es un modelo de 5 parámetros muy flexible porque más de 30 modelos de probabilidad pueden ser representados por él, como casos límite o especiales, y tal es el caso de los modelos Pareto, Exponencial, Gamma, Log normal, Singh-Maddala y Dagum. En especial, el modelo BG facilita la estimación de los parámetros y la comparación del ajuste de diferentes modelos a través de las formulas cerradas de los momentos de la distribución y medidas de desigualdad; además ayuda en el estudio de casos empíricos, como lo son la distribución del ingreso y el retorno de inversiones, que se caracterizan por presentar distribuciones de probabilidad sesgadas y con curtosis leptocúrtica<sup>1</sup> [61]. De ahí que, con base en el modelo BG en esta sección se establecerán, para las distri-

---

<sup>1</sup>En el anexo E puede encontrar definiciones estadísticas y probabilísticas que pueden ser de utilidad para la lectura de este documento.

buciones propuestas, sus formas funcionales, momentos de la distribución y medidas de desigualdad.

Por último, se presenta la teoría del excedente de producción de la estratificación social que es base del modelo *MSBA* que se presenta en este trabajo.

## 5.1 FUNCIÓN DE DENSIDAD, ACUMULADA Y COMPLEMENTO.

La función de distribución acumulada o función de densidad (FDA),  $F$ , relacionada con la variable  $X$ , asigna a el rango<sup>2</sup> de  $X$ ,  $\mathcal{R}_x$ , al dominio de probabilidad  $\mathcal{R}_X^\alpha$  o  $[0, 1]$ , y es tal que,

$$F(x) = Pr[X \leq x] = \alpha \quad x \in \mathcal{R}_X, \alpha \in \mathcal{R}_X^\alpha \quad (5.1)$$

Donde  $F(x)$  es la probabilidad que la variable  $X$  tome un valor menor o igual a  $x$ . La función  $F(x)$  es no decreciente y continua, que alcanza el valor de la unidad en el valor máximo de  $X$ .

Por su parte la función de densidad (FD) asociada a la variable  $X$ , o también conocida como función de densidad de probabilidad,  $f(x)$ , es la primera derivada de la FDA,  $F(x)$ , con respecto a  $x$ , si la derivada existe, se expresa como la siguiente expresión:

$$f(x) = \frac{d(F(x))}{dx} \quad (5.2)$$

Para una variable continua  $X$  el área bajo la curva de la función de densidad y dos puntos,  $x_L$  y  $x_U$ , en el rango de  $X$ , es igual a la probabilidad de que la variable  $X$ , tome un valor que esté entre  $x_L$  y  $x_U$ .

En algunos casos, como es el análisis de la distribución del ingreso, es útil trabajar con el complemento de la función de densidad acumulada (CFDA) [19,

---

<sup>2</sup>El rango de  $X$  se define como sigue: permita ser ser a  $\mathcal{R}_x$  ser el conjunto de todos los valores que la variable  $X$  puede tomar.

20], o también conocida como función de supervivencia. La CFDA, definida por la expresión (5.3), proporciona la probabilidad de que la variable  $X$  sea mayor a  $x$

$$S(x) = Pr[X > x] = 1 - F(x) \quad (5.3)$$

Por su parte, la caracterización de una variable es asociada a una FDA, en tanto que, algunos grupos de variables tienen funciones de distribución acumuladas que sólo se diferencian por el valor que toman sus parámetros. Si en una FDA generalizada los parámetros aparecen como símbolos, ésta corresponde a una familia de variables. Ejemplos de estas familias son la de las distribuciones Normal, Log normal, Gamma, Beta y Exponencial. La elección de cómo expresar los parámetros de una función de densidad puede ser arbitraria, sin embargo, éstos pueden ser examinados, por su sentido geométrico, de acuerdo a tres tipos básicos [34], que se describen como sigue:

*Parámetro de localización.* Es la abscisa del punto de localización del rango de la variable. Un efecto que tiene el parámetro de localización sobre la función de densidad es, por ejemplo, mover la función de densidad de lugar pero manteniendo la forma original de ésta. Algunos ejemplos del parámetro de localización son la media, la mediana y la moda.

*Parámetro de escala.* Determina la escala de medición del cuantil<sup>3</sup>  $x$ . En otras palabras, determina la escala o la dispersión estadística de una función de densidad de probabilidad; si el parámetro de escala es grande, entonces se espera que la distribución sea más extensa, mientras que, si el parámetro es pequeño la distribución estará más concentrada. Su recíproco es el parámetro de velocidad.

*Parámetro de forma.* Afecta a la forma de la función de distribución más que moverla, como el parámetro de localización, o alargarla/encogerla, como el parámetros de escala.

---

<sup>3</sup>Cuantil se define como puntos tomados, en intervalos regulares, de la función de densidad de una variable aleatoria. Es decir, para la variable aleatoria  $X$  permita al número real  $x$  denotar un elemento general del rango  $\mathcal{R}_x$ .

### 5.1.1 MOMENTOS Y COEFICIENTE DE GINI.

Además de las funciones recién descritas, existen otras funciones y cantidades relacionadas con una variable aleatoria, como lo son la función generadora de momentos (FGM) y el índice de Gini (G). Una de las ventajas de contar con expresiones cerradas del índice de Gini y de los primeros momentos de la distribución, como la media, es que permiten realizar pruebas de ajuste al comparar las estimaciones de éstos con sus contrapartes muestrales [24, 61].

Entonces, la FGM es una función de una variable auxiliar  $t$ , que su término general es de la forma  $\mu_r' t^r / r!$ , donde  $\mu_r'$  es el  $r$ -ésimo momento central<sup>4</sup>, que permite generar los momentos de la distribución de probabilidad si éstos existen en un entorno de  $t = 0$ .

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= 1 + \mu_1' t + \mu_2' t^2 / 2! + \dots + \mu_r' t^r / r! \end{aligned} \quad (5.4)$$

En tanto que, el índice de Gini, ideado por el estadístico italiano Corrado Gini, es una de las medidas más populares y útiles para medir cambios en la forma de la desigualdad de una distribución [24]. El estadístico toma un valor entre 0 y 1, en donde 0 corresponde a la perfecta igualdad y 1 corresponde a la perfecta desigualdad en la función de distribución de probabilidad.

El índice de Gini es definido a través de la curva de Lorenz. En específico, la curva de Lorenz fue introducida en 1905 por Max O. Lorenz [23], para ilustrar la desigualdad en el ingreso, y es una función de densidad acumulada que indica la fracción acumulada de  $X$  que es mantenida por el  $p\%$  de la población con un nivel más pequeño que el atributo. Por ejemplo, si toda la población tiene el mismo valor de la variable  $X$  entonces la curva de Lorenz es una línea de  $45^\circ$ . Para más información sobre la curva de Lorenz y el índice de Gini se puede consultar a Gastwirth(1972),

---

<sup>4</sup>La formula para obtener el  $r$ -ésimo momento centrado es  $\int_{-\infty}^{\infty} x^r f(x) dx$ . Por ejemplo la media se define como el primer momento centrado de la función de densidad.

Cowell (2008) y Handcock y Morris (1999) [23, 38, 43].

Por consiguiente, si la curva de Lorenz de la variable  $X$  se desvía de la curva de total igualdad significaría que existe desigualdad en la distribución de la variable  $X$ . Por lo tanto,  $G$  se define como la razón del área entre la línea de  $45^\circ$  y la curva de Lorenz, y el área debajo de la línea de  $45^\circ$ . Al área entre la línea de  $45^\circ$  y la curva de Lorenz se le llama área de concentración. Una forma de representar el coeficiente de Gini es a través de la diferencia media,  $\Delta$ , de la función de distribución acumulada de la variable  $X$ ,  $F(x)$  [38]; suponiendo que  $F(x)$  se incrementa sobre su soporte y que la media,  $\mu$ , de  $F(x)$  existe, entonces el índice de Gini de la curva de Lorenz  $L(p)$  generada por  $F(x)$ , es  $\Delta/(2\mu)$ . Matemáticamente se expresa como:

$$\begin{aligned} G &= \frac{1}{2\mu} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| dF(x)dF(y) \\ &= \frac{1}{\mu} \int_{-\infty}^{\infty} F(x)|1 - F(x)| d(x) \end{aligned} \quad (5.5)$$

De ahí que, el coeficiente de Gini expresa la desigualdad relativa de la variable  $X$  al ser la razón de una medida de dispersión, la diferencia media, y el valor medio.

## 5.2 EL MODELO BETA GENERALIZADO: UN MARCO DE ANÁLISIS EN COMÚN

La función de densidad de probabilidad del modelo  $BG$  puede representarse por la siguiente expresión:

$$BG(x; a, b, c, p, q) = \frac{|a|x^{ap-1}(1 - (1 - c)(x/b)^a)^{q-1}}{b^{ap}B(p, q)(1 + c(x/b)^a)^{p+q}} \quad \text{para } 0 < x^a < b^a/(1 - c), \quad (5.6)$$

y cero de otra forma, con  $0 \leq c \leq 1$ , siendo los parámetros  $b, p$  y  $q$  positivos. El parámetro  $b$  representa la escala, mientras que  $a, p$  y  $q$  son parámetros de forma, en particular,  $a$  define el pico de la distribución y  $q$  indica la asimetría estadística. Con la expresión (5.6) McDonald et al. (1995) [61] mostraron la relación de los modelos a estudiar con al marco de la familia  $BG$ . En concreto, el modelo  $BG$ , anida a

los modelos Beta Generalizado 1,  $BG1$ , y Beta Generalizado 2,  $BG2$ , los cuales son casos especiales de  $BG$  cuando el parámetro  $c$  toma los valores de  $c = 0$  y  $c = 1$ , respectivamente. A su vez,  $BG1$  y  $BG2$  tienen como caso límite el modelo Gamma Generalizado ( $GG$ ) que es definido por la expresión (5.7):

$$\begin{aligned} GG(x, a, \beta, p) &= \lim_{q \rightarrow \infty} BG(x; a, \beta q^{1/a}, c, p, q) \\ &= \frac{|a|x^{ap-1}e^{(x/b)^a}}{\beta^{ap}\Gamma(p)} \end{aligned} \quad (5.7)$$

para  $c = 0$  y  $c = 1$ .

De igual forma, los modelos Exponencial ( $E$ ), Gamma ( $G$ ) son casos especiales de  $GG$ , como se definen a continuación:

$$E(x, \beta) = GG(x; 1, \beta, 1) \quad (5.8)$$

$$G(x, \beta, p) = GG(x; 1, \beta, p) \quad (5.9)$$

en tanto que, el modelo Log normal ( $LN$ ) es un caso límite de  $GG$ , esto es,

$$LN(x, \mu, \sigma^2) = \lim_{a \rightarrow 0} GG(x; a, \beta^a = \sigma^2 a^2, p = (a\mu + 1)/\beta^a) \quad (5.10)$$

Por su parte, el modelo de Pareto ( $P$ ) es un caso especial del modelo  $BG1$ , mientras que los modelos Singh-Maddala (SM), o también conocida como Burr XII, y Dagum (D), o Burr III, son casos especiales del modelo  $BG2$  [61]. Estos modelos pueden ser expresados de la siguiente forma:

$$D(x, a, b, p) = BG(x, a, b, 1, p, 1) \quad (5.11)$$

$$SM(x, a, b, q) = BG(x, a, b, 1, 1, q) \quad (5.12)$$

$$P(x, b, p) = BG(x, -1, b, 0, p, 1) \quad (5.13)$$

De acuerdo a McDonald (1984) [60], las expresiones para derivar los momentos, funciones de densidad acumulada y algunas medidas de desigualdad son las que siguen:

$$I(y, h) = \int_0^y y^h f(y) dy \quad (5.14)$$

$$I^*(i, j) = \int_0^\infty y^i f(y) \int_0^y x^j f(x) dx dy \quad (5.15)$$

Con base a estas expresiones, los momentos, condicionados a su existencia, la función de distribución acumulada y el índice de Gini son definidas por las ecuaciones (5.16), (5.17) y (5.18), respectivamente.

$$E(x^h) = \lim_{y \rightarrow \infty} I(y, h) \quad (5.16)$$

$$F(x) = I(x, 0) \quad (5.17)$$

$$G = E(|x - y|)/2\mu = (1/\mu)(I^*(1, 0) - I^*(0, 1)) \quad (5.18)$$

en donde  $\mu = E(x)$ . En cada caso, las expresiones recién descritas son funciones de los parámetros  $(a, b, \beta, p, q, c)$  que definen a cada función de distribución que sea considerada. Para mayores detalles sobre las expresiones  $I(y, x)$  y  $I^*(i, j)$  refiérase a McDonald (1984) [60] y McDonald et al. (1995) [61].

Con base a estos resultados en la tabla 5.1 se muestran la función de densidad, función de densidad acumulada, la función generadora de momentos y el índice de Gini para los modelos Pareto, Exponencial, Gamma, Log normal, Dagum y Singh Maddala. Además, la figura 5.1, obtenida de McDonald et al. (1995) [61], presenta la flexibilidad de la familia  $BG$ .

### 5.3 MSBA: LA TEORÍA DE EXCEDENTE DE PRODUCCIÓN DE LA ESTRATIFICACIÓN SOCIAL.

El MSBA que en este trabajo estudia es el publicado por Angle en 1986 [3]. El modelo diseña una simulación de agentes, basada en la teoría de excedente de producción de la estratificación social, que explica por qué se crean distribuciones de riqueza de tipo estratificada. El modelo se basa en las siguientes premisas:

1. Existencia de un excedente de producción que no tiene que ser utilizado en

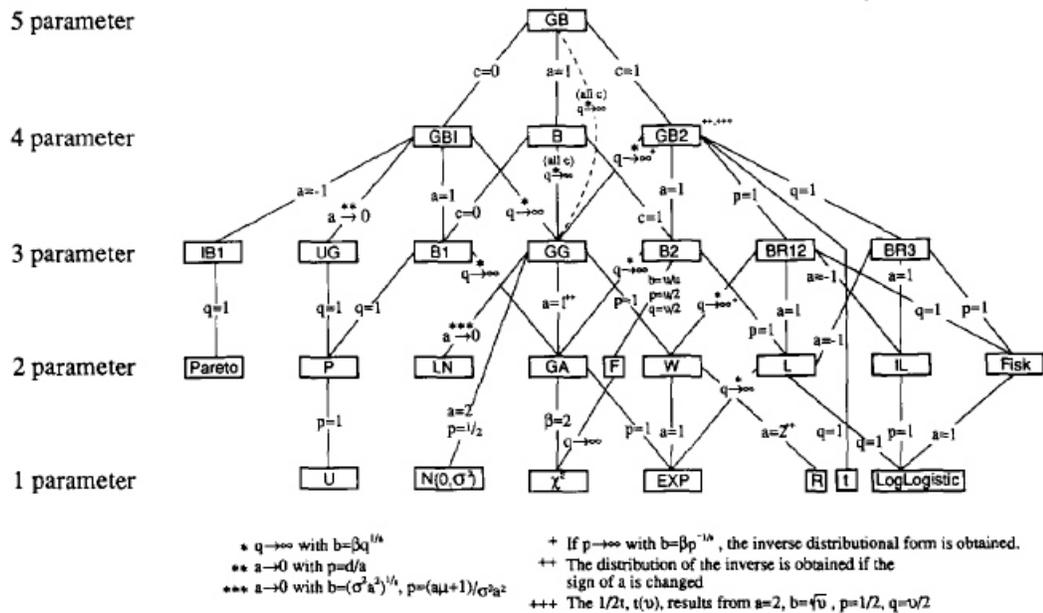


Figura 5.1: Familia Beta Generalizada. Casos especiales y límites de la Familia Beta Generalizada. Las distribuciones Exponencial, Pareto, Log normal, Gamma, Dagum y Singh-Maddala, son expresadas en la figura como EXP, Pareto, LN, GA, BR3 y BR12. Fuente: diagrama presentado en McDonald et al. (1995) [61].

Modelo	FD	FDA	Momentos	Gini	Rango
P	$\frac{pb^p}{x^{p+1}}$	$1 - (b/x)^p$	$pb^r / (p - r)$ $c > r$	$1 / (2p - 1)$	$b \leq x < \infty$
E	$\frac{1}{\beta} e^{-x/\beta}$	$1 - e^{-x/\beta}$	$\beta^r \Gamma(1 + r)$	$1/2$	$0 \leq x < \infty$
G	$\frac{(x/\beta)^{p-1} e^{-x/\beta}}{\beta \Gamma(p)}$	$\frac{(x/\beta)^p e^{-x/\beta}}{\beta \Gamma(p+1)} {}_1F_1 \left[ \begin{matrix} 1; x/\beta \\ p+1 \end{matrix} \right]$	$\frac{p^r \Gamma(p+r)}{\Gamma(p)}$	$\frac{\Gamma(p+1/2)}{\Gamma(p+1)\sqrt{\pi}}$	$0 \leq x < \infty$
LN	$\frac{1}{x\sigma\sqrt{2\pi}} e^{\left(\frac{-(\log x - \mu)^2}{2\sigma^2}\right)}$	$\frac{1}{2} + \frac{1}{2} \operatorname{erf} c^{\left(\frac{-(\log x - \mu)}{\sigma\sqrt{2}}\right)}$	$e^{r\mu + r^2\sigma^2/2}$	$2N[\sigma/\sqrt{2}; 0, 1] - 1$	$0 \leq x < \infty$
D	$\frac{ap^p y^{p-1}}{b^p (1+(x/b)^a)^{p+1}}$	$(1 + (x/b)^{-a})^{-p}$	$\frac{b^r \Gamma(p+r/a) \Gamma(1-r/a)}{\Gamma(p)}$	$\frac{\Gamma(p) \Gamma(2p+1/a)}{\Gamma(2p) \Gamma(p+1/a)} - 1$	$0 \leq x < \infty$
SM	$\frac{aqy^{q-1}}{b^q (1+(x/b)^a)^{1+q}}$	$1 - \frac{1}{(1+(x/b)^a)^q}$	$\frac{b^r B(1+r/a, q-h/a)}{B(1, q)}$	$1 - \frac{\Gamma(q) \Gamma(2q-1/a)}{\Gamma(q-1/a) \Gamma(2q)}$	$0 \leq x < \infty$

Distribuciones de probabilidad. Se presentan las expresiones para las formas funcionales de los modelos Pareto (P), Exponencial (E), Gamma (G), Log normal (LN), Dagum (D) y Singh-Maddala (SM). En las columnas se muestran la función de densidad (FD), función de densidad acumulada (FDA), función generadora de momentos (Momentos), el índice de Gini (Gini) y el rango de la distribución (Rango).  $B$ ,  $\Gamma$ ,  $\operatorname{erf} c$  y  ${}_pF_q$  denotan a la función Beta, Gamma, la función de error y series hipergeométricas generalizadas.

reversión. Además este excedente puede ser intercambiado entre los agentes por medio de actividades económicas.

2. El intercambio de riqueza no genera o destruye ingreso sino que solo se encarga de la distribución del mismo
3. Tendencia de que el flujo de ingreso fluya hacia aquellos que concentran ingreso o poder.
4. Mayor habilidad de trabajadores en sociedades desarrolladas, social y económicamente, para retener un mayor porcentaje de su ingreso que aquellos trabajadores de sociedades con tecnología más primitiva.

Entonces, el ingreso de los agentes  $i$  y  $j$  al inicio del encuentro, es  $x_{i,t-1}$  y  $x_{j,t-1}$ , asimismo,  $\delta$  es el parámetro que asigna la probabilidad para definir al ganador del intercambio de ingreso, en tanto que, el parámetro  $L$  define cuánto del ingreso inicial pierde (gana) el agente perdedor (ganador). Por lo tanto, el intercambio de los agentes en la iteración  $t$  se define como sigue:

$$x_{i,t} = x_{i,t-1} + dZx_{j,t-1} - (1-d)Zx_{i,t-1} \quad (5.19)$$

$$x_{j,t} = x_{j,t-1} + dZx_{i,t-1} - (1-d)Zx_{j,t-1} \quad (5.20)$$

En donde  $d$  denota quien gana el intercambio, 1 si lo gana el agente  $i$  y 0 si lo gana el agente  $j$ . Este resultado está en función del parámetro de concentración del poder,  $\delta$ , y puede ser expresado como:

$$d = \begin{cases} 1 & \text{con probabilidad } \delta \text{ si } x_{i,t-1} > x_{j,t-1}, \text{ de lo contrario la probabilidad será } (1 - \delta) \\ 0 & \text{de otra forma.} \end{cases} \quad (5.21)$$

Esto es, se toma una muestra, con reemplazo, de un elemento del conjunto  $s = 0, 1$  con probabilidad  $\delta$  o  $1 - \delta$  según los valores del ingreso de los agentes al inicio del intercambio.

Por su parte,  $Z$  es una función que calcula la proporción del ingreso del agente perdedor que será intercambiada en el encuentro y depende del parámetro sobre el desarrollo de la economía. Dicho argumento se expresa como:

$$Z = \frac{\sum_{i=1}^L U_i}{L} \quad (5.22)$$

donde  $U$  es un vector de tamaño  $L$  de  $L$  números aleatorios provenientes de una distribución uniforme con valores entre 0 y 1. Con la expresión (5.22), tipo series de potencia, se busca llegar a que la proporción de ingreso que es intercambiada por los agentes cumpla con el supuesto 3. Es decir, si el escenario en el que se hace el intercambio tiene un mayor desarrollo económico y social, mayor será el valor que obtenga  $U$ , por lo tanto, el porcentaje  $Z$  que se intercambia entre los agentes es menor que un escenario donde hay poco desarrollo en la economía.

## CAPÍTULO 6

# MÉTODOS DE ESTIMACIÓN DE LA DISTRIBUCIÓN ESTADÍSTICA.

---

Conocer la función de densidad de una variable aleatoria implica tener una descripción completa de la misma [62], por lo que la estimación de los parámetros que integran la función de densidad, a partir de los datos, se ha convertido en un problema fundamental de la estadística. La estimación de una distribución puede hacerse de forma paramétrica o no paramétrica. En la primera se supone que los datos siguen algún modelo de probabilidad y la estimación de los parámetros de los modelos descritos se pueden llevar a cabo por técnicas de estimación como lo son métodos de maximización de verosimilitud (MV), método de momentos o estimación bayesiana [60]. De manera contraria, se puede no suponer nada a *priori*, es decir que los datos no siguen ningún modelo de probabilidad en particular y éstos pueden adoptar cualquier forma, esto es, se permite que los datos determinen la forma funcional que los va a representar.

En esta investigación, se optó como método paramétrico de estimación de las funciones de densidad en estudio, en tanto que, para la estimación no paramétrica se utilizó una suavización de tipo núcleo y la estimación directa empírica para la función de densidad y de densidad acumulada, respectivamente. Por un parte, la selección del método de MV se eligió debido a que es uno de los procedimiento de estimación

con mayor uso en la estadística [97] debido a sus deseables propiedades asintóticas, como lo son, la consistencia, eficiencia, normalidad asintótica, entre otras, que serán explicadas en este capítulo; otra razón, fue que el método de MV ha sido utilizado en varios estudios dedicados a la forma de la distribución del ingreso, como se puede consultar en el ANEXO A, por lo que ayudará, al dejar de lado el factor de tipo de estimación en la variación en el error de la estimación, a comparar los resultados de este trabajo con aquellos publicados en la literatura. Por su parte, la estimación no paramétrica de tipo núcleo tiene entre sus bondades que, a comparación del método histograma, resulta en una mejor estimación de la función de densidad empírica [43], y por su parte, la estimación de la función de densidad empírica se prefirió debido a que es una herramienta estadística más fáciles de estimar.

Por otro lado, en esta sección también se presentan los métodos que se utilizaron para probar si un modelo de probabilidad puede generar cierto conjunto de datos. Los métodos que integran la prueba son los que siguen: (i) una prueba que informa si un modelo de probabilidad es una hipótesis plausible para generar un conjunto de datos, la cual es desarrollada a través del método *bootstrap*; (ii) una búsqueda exhaustiva para obtener el valor del límite inferior óptimo de la cola derecha suponiendo que ésta se distribuya a un modelo de probabilidad; y (iii), una prueba de bondad de ajuste que selecciona entre dos modelos cuál es el más cercano a la distribución verdadera de los datos.

## 6.1 ESTIMACIÓN NO PARAMÉTRICA DE LA DENSIDAD.

### 6.1.1 FUNCIÓN DE DENSIDAD ACUMULADA EMPÍRICA: ESTIMACIÓN DIRECTA.

La función de densidad acumulada empírica (*FDAE*) es una de las herramientas estadísticas más útiles y fáciles de estimar para analizar una muestra univaria-

da [42]. La  $FDAE$  es estimada de manera directa de los datos muestrales sin asumir cierta forma algebraica del modelo de distribución de probabilidad [34]. Permita a la  $FDAE$ , ser denotada por  $F_n$ , y definida como la proporción relativa de datos que no exceden al valor  $t \in [0, 1]$ , expresada por,

$$\begin{aligned}
 F_n(x) &= \frac{\text{número de elementos en la muestra } \leq x}{n} \\
 &= \frac{1}{n} \sum_{i=1}^n \mathcal{I}_{(-\infty, x]}(X_i) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{k}{n} & \text{si } X_{(k)} \leq x < X_{(k+1)} \\ 1 & \text{si } x \geq X_{(n)} \end{cases} \quad (6.1)
 \end{aligned}$$

y  $X_1, \dots, X_n$  son variables i.i.d. que corresponden a una muestra aleatoria de la variable  $X \sim F$ , además que  $X_{(1)}, \dots, X_{(n)}$  es la muestra ordenada. Por su parte,  $\mathcal{I}$  es la función indicadora.

Propiedades de la  $FDAE$  son las que siguen: (i) tiene una forma escalonada; (ii) conforme a la ley de los grandes números existe una convergencia escalonada de la función de distribución empírica en la verdadera función de distribución de la variable aleatoria, en tanto que, un resultados más fuerte, conocido como el teorema Gilvenko-Cantelli, muestra una convergencia uniforme, que puede expresare como:

$$\lim_{n \rightarrow \infty} P\left\{ \sup_{-\infty < x < \infty} |F_n(x) - F(x)| > \epsilon \right\} = 0 \quad (6.2)$$

donde la norma del supremo es también conocida como la distancia Kolmogorov-Smirnov; (iii), la  $FDAE$  es un estimador insesgado para cada  $x$ , esto es, para una  $x$  fija la función indicadora  $\mathcal{I}$  es una variable aleatoria Bernoulli con parámetro  $p = F(x)$ , y por lo tanto,  $nF_n(x)$  es una variable aleatoria binomial con media  $nF_n(x)$  y varianza  $nF_n(x)(1 - F_n(x))$ .

### 6.1.2 FUNCIÓN DE DENSIDAD EMPÍRICA: ESTIMACIÓN TIPO NÚCLEO O KERNEL.

Con el objetivo de obtener una estimación no paramétrica de la función de densidad empírica se utilizó el estimador de tipo núcleo o kernel. El estimador de tipo núcleo se define con la siguiente expresión:

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (6.3)$$

donde  $K$  es la función núcleo, función kernel o función peso, que es una función densidad de probabilidad no negativa y en su mayoría es simétrica. La función  $K$  satisface las siguientes condiciones de regularidad.

$$\int_{-\infty}^{\infty} |K(x)| dx < \infty \quad (6.4)$$

$$\lim_{x \rightarrow \infty} |xK(x)| = 0 \quad (6.5)$$

$$\int_{-\infty}^{\infty} K(x) dx = 1 \quad (6.6)$$

Algunas de las funciones núcleo comúnmente usadas son la función Epanechnikov, Gauss, Triangular, Rectangular y Arco Coseno. Por su parte,  $h_n$  es el ancho de ventana o parámetro de suavización, y tiene como objetivo determinar la cantidad de suavización en la estimación.

Esta técnica se puede interpretar como una generalización del estimador histograma pero en la cual para cada punto,  $x$ , se estima una función kernel tomando como el centro al mismo valor. Luego, para crear la función de densidad de probabilidad se suman las funciones kernel, y posteriormente, se divide por el número de observaciones.

Una de las ventajas de la estimación tipo núcleo es que, a diferencia del histograma la cual es una función escalonada, al usar una buena función kernel sus propiedades sobre suavización y continuidad acercan de mejor manera a la estimación de la densidad empírica a la verdadera densidad [43]. Sin embargo, tiene como

deficiencia que tiene la tendencia de crear distorsiones en las colas de la distribución debido a que el parámetro de suavización es fijo sobre toda la muestra [81].

Las propiedades estadísticas del estimador tipo núcleo son las que siguen: (i) si se satisfacen que las constantes  $\lim_{n \rightarrow \infty} h_n = 0$ , y además, las propiedades (6.4), (6.5), (6.6), entonces la estimación es asintóticamente insesgada en todos los puntos de  $x$ , en los cuales la función de densidad es continua; y (ii), el estimador es consistente, es decir que el error de la estimación tiende a cero, cuando el número de observaciones tiende a infinito,  $n \rightarrow \infty$ , y si el límite del producto de las observaciones y el parámetro de suavización tiende a infinito,  $\lim_{n \rightarrow \infty} nh_n = \infty$ <sup>1</sup>. Para una demostración de estos resultados y sus consecuencias sobre la elección del ancho de ventana puede referirse a Miñarro (1998) [62].

Por otra parte, la estimación del ancho de ventana apropiado es el paso más difícil en la estimación de tipo núcleo; un ancho de ventana muy grande produce una desviación estándar grande en la ecuación (6.3), y en consecuencia, la probabilidad que se asigna a  $x_i$  será ampliamente distribuida en la vecindad de  $x_i$ . Por el contrario, valores muy pequeños de  $h_n$  presentan una desviación estándar pequeña y la probabilidad asignada a  $x_i$  será distribuida cerca de la misma. Comúnmente la determinación del ancho de ventana se realiza a través de la minimización de alguna medida de error, como la aproximación del error cuadrático medio integrado ( $MISE$ )<sup>2</sup>. Sin embargo, esta minimización depende de la densidad desconocida que se desea estimar por lo que se imposibilita la estimación directamente.

---

<sup>1</sup>El error cuadrático medio de la estimación es igual a la varianza de la estimación más el sesgo al cuadrado de la misma. Entonces el  $\lim_{n \rightarrow \infty} Var(F_n(x)) \rightarrow 0$  si  $\lim_{n \rightarrow \infty} nh_n = \infty$ .

<sup>2</sup>El error cuadrático medio integrado (MISE) es un criterio de error de la estimación que depende de la verdadera (desconocida) densidad, el estimador utilizado y el tamaño muestral; MISE es definido como el promedio de una función que es una realización particular de  $n$  puntos, tal que,  $MISE = \int E[F_n(x) - F(x)]^2$

## 6.1.3 ESTADÍSTICOS DESCRIPTIVOS Y COEFICIENTE DE GINI

Asimismo, expresiones de las cantidades muestrales para la media, desviación estándar, coeficiente de asimetría, curtosis, ajustadas por el ponderador muestral, se definen a continuación. Suponga que cada  $x_i$  tiene un ponderador entero,  $w_i$ , entonces el tamaño de la población será  $N = \sum_{i=1}^n w_i$ .

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n x_i w_i \quad (6.7)$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^n w_i (x_i - \bar{x})^2 \quad (6.8)$$

$$\eta_3 = \frac{1}{\sigma^{3/2} N} \sum_{i=1}^n w_i (x_i - \bar{x})^3 \quad (6.9)$$

$$\eta_4 = \frac{1}{\sigma^2 N} \sum_{i=1}^n w_i (x_i - \bar{x})^4 \quad (6.10)$$

Mientras que la moda se define como el cuantil en el que la *FDAE* tiene un máximo local y la mediana como el cuantil que es excedido con probabilidad 1/2.

Por su parte, un enfoque para la estimación no paramétrica del coeficiente de Gini es la utilización de la regla del trapecoide. La regla del trapecoide se utiliza para calcular el área de concentración, y en seguida dividir ésta por el área de debajo de la curva de completa igualdad, esto es:

$$\begin{aligned} Gini_n(x) &= 1 - 2 \sum_{i=1}^n (x_i - x_{i-1}) \left[ \frac{L(x_i) + L(x_{i-1})}{2} \right] \\ &= \sum_{i=1}^n [L(x_i)x_{i-1} - L(x_{i-1})x_i] \end{aligned} \quad (6.11)$$

en donde  $L(x)$  es la curva de Lorenz empírica y se estima de la siguiente forma: dada la variable aleatoria  $X$  ordenada de la forma  $x_1 \leq x_2 \leq \dots \leq x_n$ , su curva de Lorenz empírica es definida en los puntos  $W_i/N$ , donde  $W_i = w_1 + \dots + w_i$ , en tanto que  $w_i$  son los pesos de cada observación de  $X$  y  $N = \sum_{i=1}^n w_i$ . De ahí que  $L(0) = 0$  y  $L(w_i/N) = s_i/s_n$ , donde  $s_i = w_1 * x_1 + \dots + w_i * x_i$ .

Empero, la estimación del área de concentración por el método de trapecios genera un sesgo positivo en cada trapecio estimado, y por ende, el estimador del coeficiente de Gini siempre tendrá un sesgo negativo. Con el objetivo de obtener estimaciones más certeras del coeficiente de Gini se han desarrollado intervalos de confianza para la estimador del coeficiente de Gini [38, 40] y aproximaciones para datos agrupados en quintiles [40].

## 6.2 ESTIMACIÓN PARAMÉTRICA DE LA DENSIDAD: ESTIMACIÓN POR MÁXIMA VEROSIMILITUD

La idea central de la estimación por máxima verosimilitud (*MV*) es la función de verosimilitud, la cual se define en las siguientes líneas.

Sea  $X$  una variable aleatoria y  $x = (x_1, x_2, \dots, x_n)$  una muestra de  $n$  observaciones *i.i.d* que proviene de la función de densidad de probabilidad desconocida,  $f(\cdot)$ . Además, sea  $\theta$  el vector de parámetros desconocidos, tal que,  $\theta = (\theta_1, \dots, \theta_p)^T$ . Cabe mencionar que,  $\theta$  surge de un espacio de parámetros,  $\Omega$ , el cual contiene el conjunto de valores que a  $\theta$  le es posible tomar.

Entonces, sea  $f(x|\theta)$  la función densidad de probabilidad de la muestra  $x$  dado el vector de parámetros desconocidos. Por su parte, cada observación de la variable,  $x_i$ , puede ser tomada como una realización de alguna distribución de probabilidad con una función de densidad conjunta  $f(x_i|\theta)$ . Si se tiene una muestra con  $n$  observaciones, la función de densidad conjunta se puede expresar como  $f(x|\theta)$ , el producto de las funciones de densidad individuales. Asimismo, se supone que la función de densidad de la muestra pertenece a una familia de distribuciones, o modelo paramétrico, unido por  $\theta \in \Omega$ , esto es,  $\mathcal{M}_\theta = \{f(\cdot; \theta) : \theta \in \Omega\}$

De ahí que la estimación a través de *MV* se centra en la función de verosimilitud

(*FV*) de  $\theta$ , basada en la observación de  $X = x$ , que se expresa de la siguiente manera:

$$L(\theta|x) = f(x|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n L_i, \quad \theta \in \Omega \quad (6.12)$$

donde los datos son  $x = (x_1, \dots, x_n)^T$ .

Note que la expresión (6.12) proporciona el contraste entre dos cantidades: la primera, es la función de verosimilitud la cual es una función del parámetro  $\theta$ , dada la muestra  $x$ , mientras que la segunda cantidad indica la densidad que es una función de la muestra  $x$ , dado el parámetro  $\theta$ . En consecuencia, la *FV* indica, dado un modelo de probabilidad, la probabilidad de observar el conjunto de datos como función de los parámetros  $\theta$ . En resumen, la estimación a través de MV toma los datos como fijos para determinar el parámetro  $\theta$  que maximiza la probabilidad de obtener los datos observados.

Es común, continuar el análisis en una escala logarítmica porque al suponer que los datos son independientes, se puede sumar las contribuciones del logaritmo de las verosimilitudes. También, la suma de las verosimilitudes permite enunciar propiedades de muestras grandes para los estimadores de MV con base en la aplicación de la ley de los grandes números. Entonces, tomando el logaritmo de *FV* se obtiene la función logarítmica de verosimilitud (*FLV*),

$$\ell(\theta|x) = \sum_{i=1}^n \log(f(x_i|\theta)) = \sum_{i=1}^n \ell_i \quad (6.13)$$

Por consiguiente, la estimación por MV implica la maximización de *FLV*, o su equivalente,  $\ell$ , y se puede presentar como

$$\hat{\theta} = \arg \max_{\theta \in \Omega} \ell(\theta|x) \quad (6.14)$$

Sin embargo, la solución puede no ser única o no existir. Si  $\hat{\theta}$  se encuentra en la frontera, se obtendría  $\hat{\theta}$  resolviendo la primera derivada parcial de la *FV* para el conjunto  $\theta$  e igualando a cero, esto es,  $\partial \ell(\theta|x) / \partial \theta = 0$ . Por otra parte, cuando no se puede obtener una expresión cerrada del estimador de MV,  $\hat{\theta}$ , se recurre a métodos

iterativos de optimización numérica como lo son, por citar algunos, el método de optimización de Newton y el de método Nelder-Mead.

Asimismo, la notación para expresar el vector gradiente, la matriz hessiana y la matriz de información observada (MIO) se presentan en las ecuaciones (6.15), (6.16) y (6.17).

$$\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta}|x)}{\partial \boldsymbol{\theta}} \quad (6.15)$$

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \quad (6.16)$$

$$\mathbf{I}_O(\boldsymbol{\theta}) = -\mathbf{H}(\boldsymbol{\theta}) = -\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \quad (6.17)$$

### 6.2.1 CONDICIONES DE REGULARIDAD.

Para formalizar el método de  $MV$ , algunas propiedades sobre  $\mathcal{M}_\theta$  se deben de establecer. A estas condiciones se les llama condiciones regulares.

**Condicion I.** La dimensión de  $\boldsymbol{\theta}$  debe ser fija. El estimador de  $MV$  puede ser inconsistente si el número de parámetros crece junto con el tamaño de la muestra.

**Condicion II.** El parámetro  $\boldsymbol{\theta}$  es identificable.

**Condicion III.** Las distribuciones de  $\mathcal{M}_\theta$  tienen un soporte en común, esto es, que son independientes de  $\boldsymbol{\theta}$ .

**Condicion IV.**  $\Omega$  es un conjunto abierto de  $\mathbb{R}^p$ .

**Condicion V.** El valor verdadero,  $\boldsymbol{\theta}_*$ , se encuentra en  $\Omega$ .

**Condicion VI.** Las tres primeras derivadas de  $\ell$  existen en un conjunto abierto, llamado  $\mathcal{A}$ , que contiene a  $\boldsymbol{\theta}_*$ , y  $\partial^3 \log f(x, \boldsymbol{\theta}) / (\partial \boldsymbol{\theta}_s \partial \boldsymbol{\theta}_t \partial \boldsymbol{\theta}_u) \leq M(y)$  para  $\boldsymbol{\theta} \in \mathcal{A}$ , donde  $0 < E(M(x)) < \infty$ .

**Condicion VII.** Se puede intercambiar el orden de la doble diferenciación con respecto a  $\boldsymbol{\theta}$  y a la integración sobre  $S$ . Entonces para todo  $x \in X$  y  $\boldsymbol{\theta} \in \Omega$ ,  $\ell$  es dos veces diferenciable con

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_X f(x|\boldsymbol{\theta}) dx = \int_X \frac{\partial}{\partial \boldsymbol{\theta}} f(x|\boldsymbol{\theta}) dx$$

y

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \int_X f(x|\boldsymbol{\theta}) dx = \int_X \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f(x|\boldsymbol{\theta}) dx$$

### 6.2.2 LA MATRIZ DE INFORMACIÓN DE FISHER.

La información de Fisher es un concepto importante en la teoría de  $MV$  y es expresado por la matriz de información de Fisher o la matriz de información esperada ( $MIE$ ). Ésta mide el promedio de la información del parámetro  $\boldsymbol{\theta}$  sobre todas las posibles observaciones, no solo las observadas. Por ejemplo, si los datos proveen mucha información sobre  $\boldsymbol{\theta}$ , entonces el pico del estimador de  $MV$ ,  $\hat{\boldsymbol{\theta}}$ , será agudo porque el parámetro tiene una gran influencia en la función de verosimilitud. Por el contrario, si el pico es plano entonces el parámetro estimado muestra gran incertidumbre. La  $MIE$  es definida por:

$$\mathcal{I}_E(\boldsymbol{\theta}) = Var \left( \frac{\partial \ell}{\partial \boldsymbol{\theta}} \right) \quad (6.18)$$

En tanto que, las propiedades básicas de la matriz de información de Fisher son las que siguen:

1. Para observaciones independientes la  $MIE$  es aditiva, y para variables aleatorias i.i.d. se puede escribir como  $\mathcal{I}_E(\boldsymbol{\theta}) = n\mathcal{I}_{E1}(\boldsymbol{\theta})$ , donde  $\mathcal{I}_{E1}(\boldsymbol{\theta})$  es la  $MIE$  para la primera observación. Entonces, conforme el número de observaciones,  $n$ , aumenta también lo hace el valor de la información que hay sobre  $\boldsymbol{\theta}$ , y a su vez, implicaría que la incertidumbre en  $\hat{\boldsymbol{\theta}}$  debe decrecer, es decir, la  $Var(\hat{\boldsymbol{\theta}})$  decrece en un sentido matricial.

2. Es positiva semi-definida. En la practica  $MIE$  es positiva y definida sobre una gran parte de  $\Omega$ , aunque casos extremos de singularidad pueden ocurrir en la teoría de verosimilitud.
3. Puede cambiar a través de transformaciones. Bajo transformaciones monótonas la  $MIE$  se expresa de la siguiente manera: permita a  $g_j(\theta)$  ser un conjunto de  $p$  funciones invertibles que son diferenciables, entonces,

$$\mathcal{I}_E(\mathbf{g}) = \frac{\partial \theta^T}{\partial \mathbf{g}} \mathcal{I}_E(\theta) \frac{\partial \theta}{\partial \mathbf{g}^T} \quad (6.19)$$

donde  $\mathbf{g} = (g_1\theta, \dots, g_p\theta)^T$ .

4. Para algunos modelos con  $p > 1$ , es posible que el elemento  $(j, k)$  de la  $MIE$  sea igual a 0 ( $j \neq k$ ), y por lo tanto, se dice que  $\theta_j$  y  $\theta_k$  son ortogonales; es decir, implicaría independencia asintótica entre ellos. Una ventaja de esta característica es que el estimador de  $MV$  de un parámetro varía de manera lenta en relación con otro parámetro, y gracias a esto, es común que se prefieran expresiones con parámetros ortogonales que ayudan a establecer estabilidad, y computacionalmente, ayuda al algoritmo a converger más rápido con menos necesidad de memoria.

### 6.2.3 ALGORITMOS TIPO NEWTON

Para ciertas distribuciones, como es el caso de la distribución Dagum y Singh-Maddala, las ecuaciones de verosimilitud deben ser resueltas de manera numérica.

Uno de los algoritmos de optimización que se usa para resolver el problema de  $MV$  es el método Newton. Entonces, se procede a expandir a  $\ell$  en una serie de Taylor de primer orden sobre la actual estimación en la iteración  $a - 1$ :

$$\ell(\theta^{(a)}) \approx \ell(\theta^{(a-1)}) + (\theta^{(a)} - \theta^{(a-1)})^T \frac{\partial \ell(\theta^{(a-1)})}{\partial \theta}$$

Enseguida se toma la primera derivada,  $\partial \ell / \partial \theta$ , evaluada en  $\theta^{(a)}$

$$\frac{\partial \ell(\theta^{(a)})}{\partial \theta} = \frac{\partial \ell(\theta^{(a-1)})}{\partial \theta} + \frac{\partial^2 \ell(\theta^{(a-1)})}{\partial \theta \partial \theta^T} (\theta^{(a)} - \theta^{(a-1)})$$

$$= \mathbf{U}(\boldsymbol{\theta}^{(a-1)}) + \mathcal{H}(\boldsymbol{\theta}^{(a-1)})(\boldsymbol{\theta}^{(a)} - \boldsymbol{\theta}^{(a-1)}) \quad (6.20)$$

De manera ideal la iteración,  $a$ , podría ser óptima, y de ser así  $\boldsymbol{\theta}^{(a)}$  será el valor que tome el estimador de MV,  $\hat{\boldsymbol{\theta}}$ , y su vector gradiente,  $\mathbf{U}$ , será igual a 0. A continuación, el lado izquierdo de la ecuación se iguala a 0, y una vez reorganizando términos, se llega al paso Newton-Raphson:

$$\boldsymbol{\theta}^{(a)} = \boldsymbol{\theta}^{(a-1)} - \mathcal{H}(\boldsymbol{\theta}^{(a-1)})^{-1} \mathbf{U}(\boldsymbol{\theta}^{(a-1)}) \quad (6.21)$$

La iteración de (6.21) continua hasta que la diferencia,  $\|\boldsymbol{\theta}^{(a)} - \boldsymbol{\theta}^{(a-1)}\|$ , o la diferencia relativa,  $\|\boldsymbol{\theta}^{(a)} - \boldsymbol{\theta}^{(a-1)}\|/\|\boldsymbol{\theta}^{(a-1)}\|$ , es pequeña. Si en el vecindario donde se encontró el máximo, la función  $\ell$  es bien comportada, y los valores iniciales del algoritmo son cercanos a la solución, entonces, se esperaría que el algoritmo convergiera a una tasa cuadrática, esto es:

$$\lim_{a \rightarrow \infty} \frac{\|\boldsymbol{\theta}^{(a)} - \hat{\boldsymbol{\theta}}\|}{\|\boldsymbol{\theta}^{(a-1)} - \hat{\boldsymbol{\theta}}\|^2} = c \quad (6.22)$$

para una constante positiva  $c$ . En la práctica, una vez cerca de la solución, el número de dígitos correctos se dobla en cada iteración.

Sin embargo, el algoritmo Newton-Raphson adolece de los siguientes problemas: el primero es que requiere una inversión de memoria muy grande, porque guarda en memoria una matriz de orden  $p$ , que es  $O(p^3)$ , así como la programación de  $p(p+1)$  elementos únicos de  $\mathcal{H}$ ; y el segundo es que el algoritmo no garantiza una mejora en la verosimilitud en cada iteración. Una modificación que hace frente a los problemas mencionados es la propuesta por Fisher, conocida como el “tanteo de Fisher” («*Fisher scoring*»), que consiste en reemplazar la *MIO* por la *MIE*, de tal forma, que se obtendría el paso definido en (6.23).

$$\boldsymbol{\theta}^{(a)} = \boldsymbol{\theta}^{(a-1)} - \mathcal{I}_E^{-1}(\boldsymbol{\theta}^{(a-1)}) \mathbf{U}(\boldsymbol{\theta}^{(a-1)}) \quad (6.23)$$

y una tasa de convergencia lineal

$$\lim_{a \rightarrow \infty} \frac{\|\boldsymbol{\theta}^{(a)} - \hat{\boldsymbol{\theta}}\|}{\|\boldsymbol{\theta}^{(a-1)} - \hat{\boldsymbol{\theta}}\|} = c \quad (6.24)$$

para algún  $c$  tal que  $0 < c < 1$ .

Entonces, el uso de la MIE, en comparación MIO, beneficia al algoritmo debido a que para la mayoría de los modelos la MIE es definida y positiva sobre una larga porción del espacio de parámetros  $\Omega$ ; es decir, las  $n$  MIE son usualmente definidas y positivas, y en consecuencia, se puede asegurar una mejora en cada iteración si se usa un paso medio. Y por último, las MIE son más simples de estimar que las MIO [97].

#### 6.2.4 PROPIEDADES DE LOS ESTIMADORES DE $MV$

Bajo condiciones regulares, los estimadores de  $MV$  tienen buenas propiedades asintóticas. Con el fin de describirlas se expresará a los estimadores de  $MV$  como  $\theta_n$ , para denotar que están basados en una muestra de tamaño  $n$ , y al valor verdadero de  $\theta$  como  $\theta_*$ . Enseguida se enlistan las propiedades del estimador de  $MV$ .

1. *Consistencia.* para todo  $\epsilon > 0$  y  $\theta_* \in \Omega$ ,

$$\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n - \theta_*\|_{\infty} > \epsilon) = 0 \quad (6.25)$$

En otras palabras, la distribución de  $\hat{\theta}_n$  colapsa alrededor de  $\theta_*$ . En la ecuación (6.25) se muestra, con la máxima norma, que cada elemento de la secuencia  $\hat{\theta}_n$  converge en probabilidad al valor verdadero  $\theta_*$ . A la expresión (6.25) también se le conoce como consistencia débil.

2. *Normalidad.* Conforme el tamaño de la muestra crece, la distribución del estimador de  $MV$  tiene a la distribución normal con media  $\theta$  y matriz de covarianza igual a la inversa de la matriz de información de Fisher.
3. *Insesgado.* La esperanza del estimador de  $MV$  tiende al valor real cuando  $n$  crece.
4. *Eficiencia.* El estimador es eficiente si alcanza la cota de Crámer-Roa, y por lo tanto, es insesgado de mínima varianza. La cota Cramér-Rao es una cota

inferior para la varianza del estimador insesgado y se usa como referencia para comparar el desempeño de un estimador. Para el caso de un modelo de un solo parámetro, bajo condiciones de regularidad y observaciones *i.i.d.*, la cota Cramer-Rao se puede escribir como:

$$\frac{1}{n\mathcal{I}_{E1}(\boldsymbol{\theta})} = \mathcal{I}_E^{-1} \leq \text{Var}(\hat{\boldsymbol{\theta}}_n) \quad (6.26)$$

5. *Invariante.* Si se realiza una parametrización diferente  $g(\boldsymbol{\theta})$ , donde  $g$  es una función monótona de  $\boldsymbol{\theta}$ , el estimador de MV de  $g(\boldsymbol{\theta})$  es  $g(\hat{\boldsymbol{\theta}})$ . Es decir, la estimación por MV no varía ante alguna transformación de las observaciones.

Para una lectura más amplia de las demostraciones de las propiedades del estimador puede consultar a Boos y Stefanski (2013) [9].

### 6.3 PRUEBAS DE BONDAD DE AJUSTE.

Por otra parte, es deseable tener una prueba de bondad de ajuste que ayude a saber si es plausible que los datos empíricos puedan ser representados por el modelo de probabilidad seleccionado. Es común utilizar una prueba visual para identificar si un modelo se ajusta a una serie de datos; esta prueba consiste en verificar si se percibe un ajuste *bueno* entre los datos empíricos y los datos teóricos del modelo supuesto. Asimismo, es habitual considerar la prueba bondad de ajuste Kolmogorov-Smirnov, empero los resultados no serían válidos si los parámetros del modelo teorizado son dependientes de los datos de la muestra, porque, la distribución asintótica del estadístico depende del modelo subyacente de los datos, por lo que ésta no se puede aproximar si no se conoce la verdadera distribución de la población. De ahí que, los resultados de estos dos métodos pueden ser criticados, el primero porque depende del juicio del investigador, y el segundo, porque se puede cometer el error de utilizar los parámetros estimados con base en la muestra para realizar la prueba Kolmogorov-Smirnov.

Por otro lado, una de las problemáticas para probar si la cola derecha de la distribución se rige por un proceso tipo Pareto u alguna otra distribución es la estimación del valor mínimo,  $x_{min}$ , donde empieza la variable a distribuirse de esta forma. Si se desea que los estimadores de los parámetros de la distribución teorizada sean lo más preciso posible, también se necesitará, que los datos sólo tengan observaciones para los cuales el modelo teórico es válido. Por ejemplo, dos formas comúnmente utilizadas para estimar el límite inferior de la cola de una distribución, al suponer que la cola se distribuye conforme a una ley de potencia, son: (i) a partir de un examen visual del complemento de la función de densidad acumulada en escala logarítmica se busca identificar el valor de la variable donde la distribución se vuelve una línea recta; y (ii) construir un gráfico que muestre el parámetro de forma en función del límite inferior,  $x_{min}$ , y seleccionar aquel valor de la variable que presente un parámetro de forma estable. La crítica a estos métodos visuales, para estimar el límite inferior, se basa en que son subjetivos y sensibles a ruido o fluctuaciones en los datos.

Con el objetivo de obtener estimaciones del límite inferior de la cola derecha de la distribución,  $x_{min}$ , y tener una prueba de bondad de ajuste entre los modelos teóricos y los datos de ingreso se utilizó la metodología desarrollada en Clauset et al. (2009) [19].

Para la estimación del límite inferior de algún modelo que caracterice la cola derecha de la distribución se implementa una búsqueda exhaustiva que encuentra el valor de la variable que minimiza la distancia entre los datos y modelo de probabilidad propuesto.

Por su parte, se utiliza un método *bootstrap* para hacer una prueba de bondad de ajuste que calcula un *valor p* que cuantifica la plausibilidad de la hipótesis de que los datos provienen de algún modelo de probabilidad. La idea general del método *bootstrap* es realizar varias muestras con datos sintéticos, a partir de un modelo de densidad de probabilidad, y medir qué tanto las muestras se alejan de la misma

distribución, y por último, comparar estos resultados con la divergencia presentada en los datos empíricos. Entonces, si la medida de dispersión de los datos empíricos es mayor que aquella generada por los datos sintéticos se puede concluir que no es plausible que los datos empíricos provengan de la distribución de probabilidad que se está suponiendo. Esta prueba puede implementarse con respecto a la totalidad de la distribución, como también enfocarse a una región de interés en la distribución.

Ahora, la prueba de plausibilidad sólo concluye si un modelo es plausible o no, entonces, puede ser que otro modelo también sea plausible y que sea mejor para caracterizar los datos, entonces, para deducir qué modelo es mejor se utilizó la prueba de razón de verosimilitud de Vuong (1989) [92]. La idea principal de esta prueba es comparar el logaritmo de verosimilitud de cada modelo supuesto. Cabe señalar que esta prueba no asegura que alguna distribución es la correcta sino sólo identifica si una distribución es preferida a otras distribuciones.

### 6.3.1 CRITERIOS DE ERROR PARA ESTIMACIONES DE DENSIDAD.

Con el objetivo de evaluar el desempeño de los modelos de probabilidad para caracterizar a la distribución del ingreso se considerarán la divergencia Kullback-Leibler ( $KL$ ), la suma del cuadrado del error ( $SCE$ ) y la distancia Kolmogorov-Smirnov ( $KS$ ). Dichos criterios de error son comúnmente utilizados en la literatura sobre estimación paramétrica y no paramétrica de la función de densidad [22, 69].

La Divergencia Kullback-Leibler se puede interpretar como una medida de la entropía, o ineficiencia, que se produce al usar una aproximación  $F_n$  de la distribución verdadera  $F$  y no ésta misma. En específico, la distancia KL es siempre positiva y es igual a cero, si y sólo si, la distribución aproximada  $F_n$  es igual a  $F$  y se define como:

$$KL(F||F_n) = \int_{-\infty}^{\infty} f(x) \log(f(x)/f_n(x)) dx \quad (6.27)$$

Asimismo, dos desventajas de utilizar la divergencia  $KL$  para comparar distribuciones es que no es simétrica y no satisface la inequidad del triángulo por lo que no debería interpretarse como una distancia entre distribuciones. Además, para ciertos casos la divergencia  $KL$  es infinita, aun cuando haya absoluta continuidad en la densidad, es decir, dada una pobre suavización puede existir una nula densidad en la misma provocando que  $KL$  sea infinita, y en consecuencia, no se sabría si se mide la calidad del modelo, la suavización o ambos [65].

Por su parte, la  $SCE$  y la distancia  $KS$  miden la distancia entre la  $FDA$  de los datos y el modelo estimado. La  $SCE$  es ampliamente utilizada como medida de bondad de ajuste, en especial, en la optimización para la estimación de parámetros y en la selección de modelos [97]; la medida  $SCE$  da mayor peso a las discrepancias de mayor magnitud y es definida por la siguiente expresión:

$$SCE = (F_n(x) - F(x))^2 \quad (6.28)$$

Por otro lado, la distancia  $KS$  es considerada como la medida más común para evaluar la distancia entre dos distribuciones no normales [69]. Un punto cuestionado de la distancia  $KS$  es que puede ser insensible a diferencias en los extremos del rango entre las dos distribuciones, porque en estos puntos la  $FDA$  tiende a ser cero o uno; sin embargo, Clauset et al. (2009) [19] obtienen resultados similares en la comparación entre distribuciones al hacer uso de la distancia  $KS$  y otras métricas como la distancia  $KS$  ajustada o el estadístico Kuiper, los cuales mejoran la sensibilidad de la métrica en los límites de la distribución. Entonces, la descripción de la distancia  $KS$  es la que sigue: la distancia  $KS$  es definida como la máxima distancia entre la  $FDA$  empírica y la que resulta del modelo estimado, esto es:

$$D = \sup_x | F_n(x) - F(x) | \quad (6.29)$$

### 6.3.2 PRUEBA DE PLAUSIBILIDAD: ESTIMACIÓN POR BOOTSTRAP.

El *bootstrap* es una técnica que busca hacer inferencia sobre una población desconocida con base en el remuestreo de la muestra de la población. En específico, si la población es desconocida se asume que si la muestra es consistente de la población, y de manera similar, lo es el remuestreo de la muestra, entonces, se puede aproximar el error en la inferencia de la población por medio del error de la inferencia de la muestra al ser aproximada por la submuestra.

La técnica del *bootstrap* se recomienda para realizar inferencia, con base en cierto estadístico o alguna de sus características, cuando (i) el tamaño de la muestra es pequeño, (ii) si los datos se comportan de manera diferente a la distribución normal, (iii) o si la distribución teorizada es desconocida o complicada. Si los datos no se comportan de manera normal conlleva a que no se puede hacer suposiciones sobre el proceso detrás de los datos para realizar inferencia del tipo frecuentista. Asimismo, la distribución del estadístico en interés no podría ser aproximado por la distribución asintótica porque ésta depende de la distribución poblacional.

Entonces, la técnica del *bootstrap* permite estimar distribuciones de estadísticos sin la necesidad de suponer el modelo detrás de la distribución o tener que derivar la distribución muestral y que puede llegar, bajo ciertas condiciones, a obtener aproximaciones de la distribución al menos tan certeras como las que se obtienen con la teoría de distribución asintótica [48]. Por ejemplo, las condiciones básicas no se mantendrán en muestras donde la varianza de distribución de la población no es finita, como en las leyes de potencia, y en consecuencia los resultados pueden ser engañosos. Por último, el método *bootstrap* tiene como desventaja que llega a ser muy costoso computacionalmente.

Siguiendo a Horowitz (2001) [48], el método se puede ejemplificar de la siguiente manera: dada una muestra aleatoria,  $x_i$ , donde  $i = 1, \dots, n$ , de tamaño  $n$ , que proviene de una población desconocida con una función de densidad acumulada,

$F_0$ . Entonces, permita a  $F_0$  pertenecer a una familia de funciones de distribución  $\mathcal{F}$  finita, organizada por  $\theta$  y con parámetro poblacional  $\theta_0$ , entonces para denotar  $P(X \leq x)$  se puede escribir  $F_0(x, \theta_0)$ , o para un miembro general de la familia paramétrica como  $F_0(x, \theta)$ . Además, permita que un estadístico, o función de los datos, ser definido como  $T_n = T_n(x_1, \dots, x_n)$  y a  $G_n(\tau, F_0) \equiv P(T_n \leq \tau)$  ser la FDA finita y exacta de  $T_n$ . Asimismo, cuando los datos son una muestra que proviene de la distribución con FDA,  $F$ , deje ser a  $G_n(\cdot, F)$  la FDA exacta de  $T_n$ . En general, para diferentes distribuciones de  $F$ , habrán diferentes funciones  $G_n(\tau, F)$  de  $\tau$ , con excepción de si  $T_n$  es «pivotal», es decir, que no depende de  $F$ <sup>3</sup>. Si  $G_n(\cdot, F)$  depende de  $F$ , entonces  $G_n(\cdot, F_0)$ , o alguna de sus características, no puede ser calculada si  $F_0$  es desconocida.

Dos opciones para estimar  $G_n$  pueden ser utilizadas: la técnica bootstrap o a través de la teoría de distribución asintótica. Esta última es recurrida cuando se trabaja con estadísticos «pivotal» asintóticos<sup>4</sup> porque la distribución asintótica del estadístico,  $G_\infty(\cdot, F)$  no depende de  $F$  y por lo tanto, si el tamaño de muestra  $n$  es suficientemente largo, entonces  $G_n(\cdot, F_0)$  puede ser estimado por  $G_\infty(\cdot)$ , sin necesidad de conocer  $F_0$ . Sin embargo, si el tamaño de muestra es pequeño  $G_\infty$  será una pobre aproximación de  $G_n(\cdot, F_0)$ .

De ahí que, el método *bootstrap* ofrece una alternativa para la aproximación de la distribución del estadístico  $T_n$  al reemplazar la distribución desconocida  $F_0$  con un estimador  $F_n$ . Dos opciones para estimar  $F_n$  son las que siguen: (i) la función de densidad acumulada empírica, y (ii) un estimador paramétrico de  $F_0$ . De la segunda opción, si se supone que  $F_0(\cdot) = F(\cdot, \theta_0)$ , donde  $\theta_0$  es estimado consistentemente por  $\theta_n$ ; además si  $F(\cdot, \theta)$  es una función continua de  $\theta$ , en un vecindario de  $\theta_0$ , y si

---

<sup>3</sup>La distribución de la prueba Kolmogorov-Smirnov viene dada por una expresión que no depende de  $F$ , si la distribución es fija; de manera contraria, los valores críticos de la prueba son inválidos si los parámetros de  $F$  son determinados por los datos  $x$ .

<sup>4</sup>Un estadístico «pivotal» asintótico es aquel que, sin importar la distribución de los datos de la muestra, una vez que se ha centrado y normalizado, la distribución asintótica del estadístico es la distribución normal estándar o chi-cuadrada.

$n \rightarrow \infty$  en cada  $x$  entonces  $F(x, \theta_n) \rightarrow F(x, \theta_0)$ . Por lo tanto, la convergencia en probabilidad será casi segura de acuerdo a si  $\theta_n \rightarrow \theta_0$  en probabilidad. Y de ahí que, el estimador *bootstrap* de  $G_n(\cdot, F_0)$  es  $G_n(\cdot, F_n)$ , sin importar el método para estimar  $F_n$ .

Generalmente  $G_n(\cdot, F_n)$  no puede ser evaluada de forma analítica y se recurre a la simulación Montecarlo, en donde muestras aleatorias de  $F_n$  son estimadas. Los pasos del procedimiento general para estimar  $G_n(\tau, F_0)$  son los que siguen.

1. Generar una muestra *bootstrap*,  $x_i^*$  de tamaño  $n$  donde  $i = 1, \dots, n$ , de acuerdo al muestreo aleatorio de la distribución  $F_n$ . Si la  $F_n$  es la *FDAE*, entonces la muestra *bootstrap* puede ser producida al muestrear los datos de manera aleatoria con remplazo.
2. Calcular  $T_n^* \equiv T_n(x_1^*, \dots, x_n^*)$ .
3. Usar los resultados de muchas repeticiones de los pasos 1 y 2 para calcular la probabilidad empírica del evento  $T_n^* \leq \tau$ .

Dado que  $F_n$  y  $F_0$  son diferentes funciones, también lo serán sus distribuciones, y por lo tanto, su estimador es sólo una aproximación de la *FDA* de  $T_n$ . La calidad de la aproximación dependerá de (i) el error por utilizar la muestra  $x$  para representar la población y (ii) el error por no analizar exhaustivamente todas las muestras *bootstrap*. Asimismo, existen condiciones que aseguran la consistencia de  $G_n(\cdot, F_n)$ , para una descripción detallada de estos tópicos puede referirse a Heckman et al. (2007) y Davison et al. (1997) [25, 47].

### 6.3.3 PRUEBA DE VUONG PARA MODELOS NO ANIDADOS.

Vuong (1989) [92] propone un estadístico de razón de verosimilitud, basado en el criterio Kullback-Leibler, para probar la hipótesis nula de que dos modelos, que

compiten entre sí, están igual de cerca al verdadero proceso que genera los datos, contra la hipótesis alternativa que dice que un modelo está más cerca que el otro de la verdadera distribución. La prueba es adaptada para modelos anidados, traslapados y no anidados<sup>5</sup>. Sin embargo, la prueba no informa que tan bueno es el ajuste de modelo a los datos empíricos.

La idea básica de la prueba es que se puede definir el mejor modelo, entre un conjunto de modelos en competencia, como aquel que es más cercano a la distribución verdadera. Entonces, si  $F(x)$  es un modelo de probabilidad, su distancia a la densidad verdadera  $F_0$ , medida por el mínimo del criterio KL, es  $E[\log F_0(x)] - E[\log F(x)]$ , donde  $E[\cdot]$  es la esperanza con respecto a la verdadera distribución de  $x$ .

En este trabajo sólo se considera la versión de la prueba para modelos no anidados. Formalmente, dados dos modelo de densidad de probabilidad  $F(x)$  y  $G(x)$ , no anidados, la hipótesis nula propone que los dos modelos son equivalentes si  $E[\log F(x)] = E[\log G(x)]$ , en tanto que, la hipótesis alternativa 1 dice que  $F$  es mejor que  $G$  si  $E[\log F(x)] > E[\log G(x)]$  o de lo contrario, la hipótesis alternativa 2, mantiene que  $G$  es mejor que  $F$  si  $E[\log F(x)] < E[\log G(x)]$ .

Como la cantidad  $E[\log F(x)]$  es desconocida se puede aproximar por  $(1/n)$  veces el logaritmo de verosimilitud evaluado por el estimador de máxima verosimilitud. De manera similar, un estimador consistente para  $E[\log F(X)] - E[\log G(x)]$  es  $(1/n)$  veces logaritmo de la razón verosimilitud ( $LR$ ) de los datos con respecto a cada modelo. Por su parte, el logaritmo de la función de verosimilitud, dado un estimador de  $MV$  para el modelo  $F$ , viene dado por:

$$L_F = \sum_{i=1}^n \log f(x_i) \quad (6.30)$$

---

<sup>5</sup>Dos modelos estadísticos están anidados si el primer modelo puede ser transformado en el segundo modelo al imponer alguna restricción en el parámetro del primer modelo [92]. Por ejemplo, como se presentó en la sección 5.1 el modelo Beta Generalizado anida al modelo Gamma Generalizado, que a su vez, anida a los modelos Exponencial y Gamma; empero, los modelos Exponencial y Gamma no están anidados.

donde  $f(x_i)$  es la verosimilitud del punto  $x_i$  dentro de la distribución  $F$ . De ahí que el logaritmo de la razón de verosimilitud entre el modelo  $F$  y  $G$  estará dada por:

$$LR = L_F - L_G = \sum_{i=1}^n \log \frac{f(x_i)}{g(x_i)} \quad (6.31)$$

Así mismo, si  $n$  es el tamaño de la muestra y la variable  $x_i$  es *i.i.d.*, entonces, de acuerdo al teorema de limite central,  $1/n^{1/2}LR$  es centrado y normalizado y tiene como límite a la distribución normal estándar bajo la hipótesis nula. Entonces el estadístico Vuong para la prueba de modelos no anidados es el que sigue:

$$n^{1/2}LR/w \rightarrow N(0, 1) \quad (6.32)$$

En donde  $w$  es la varianza de LR. Las demostraciones de estos resultados pueden ser consultadas en Vuong (1989) [92].

Por consiguiente, se puede rechazar la hipótesis nula y concluir que el mejor modelo es  $F$  ( $G$ ), con un nivel de significancia estadística de  $\alpha$ , si el estadístico de Vuong es mayor que el valor positivo (es menor que el valor negativo) de  $c$ , el cuantil  $(1 - \alpha)$  de la distribución normal estándar. Por el contrario, si el valor absoluto del estadístico es menor o igual a  $c$  no se rechaza la hipótesis nula y no se puede discriminar entre los modelos que se están poniendo a prueba.

#### 6.3.4 ESTIMACIÓN DE LA COLA DERECHA DE LA DISTRIBUCIÓN.

La búsqueda exhaustiva es un método de optimización que consiste en enumerar todas las soluciones factibles de un problema determinado, para luego evaluar el desempeño de cada una de las soluciones, y por último, elegir la solución que haya dado lugar a la mejor solución. En general, la búsqueda exhaustiva es fácil de implementar y siempre lleva la solución óptima, sin embargo, no es eficiente porque dependiendo del tipo de problema el costo computacional podría crecer exponencialmente con el número de ítems que integra el problema [27]. Ejemplos de problemas difíciles de tratar son aquellos de tipo combinatorio, que pertenecen a la clase NP.

En específico, para estimar el límite inferior de la cola derecha de la distribución,  $x_{min}$ , se utiliza el método propuesto en Clauset et al. (2009) [19], que utiliza una búsqueda exhaustiva para encontrar el valor  $\hat{x}_{min}$  que haga lo más parecidas a la distribución de probabilidad de los datos y el mejor modelo paramétrico ajustado con base en los datos.

Para cuantificar las distancias entre los datos y los modelos paramétricos se usan las tres medidas de error descritas en la subsección 6.3.1. Además en Clauset et al. (2009) [19] muestran que el uso del estadístico KS es mejor que del criterio de información bayesiano (BIC) para la exploración exhaustiva de datos que mantienen un sesgo hacia la derecha.

## CAPÍTULO 7

# IMPLEMENTACIÓN Y DATOS

---

### 7.1 ALGORITMOS PARA LA ESTIMACIÓN Y PRUEBAS DE AJUSTE.

De acuerdo a los métodos descritos en el capítulo 6 en esta sección se presentan 4 rutinas que fueron implementadas, en el lenguaje de programación R [70], con el propósito de evaluar cuáles de los modelos de probabilidad propuestos se adecua mejor para caracterizar a la distribución del ingreso de los agentes económicos en México. La evaluación se lleva a cabo tanto en la totalidad de la distribución como en las regiones izquierda, de ingresos bajos-medios, y derecha, de ingresos altos, de la distribución. Además, las rutinas ponen especial énfasis en poner a prueba las siguientes hipótesis: *(i)* la cola derecha de la distribución sigue un proceso de tipo Pareto, *(ii)* la cola izquierda sigue un proceso de tipo Exponencial o Log normal, y por último, *(iii)* la totalidad de la distribución se distribuye de acuerdo a un modelo de probabilidad tipo Gamma.

El primer algoritmo tiene como objetivo, con base en los datos empíricos, estimar los coeficientes de los modelos estudiados y herramientas que serán de utilidad para poner a prueba el ajuste de los modelos a los datos empíricos. Entonces, dado un conjunto de datos muestrales, con sus respectivos factores de expansión, el

algoritmo 1 ajusta, a través de máxima verosimilitud, los modelos de probabilidad propuestos, calcula los coeficientes asociados a cada modelo, y por último, estima las medidas de divergencia entre los modelos y los datos empíricos.

**Algoritmo 1: Ajuste.**

1. Ordenar los datos según ingreso
2. Estimar, con base en la muestra, los estadísticos descriptivos y coeficiente de Gini.
3. Estimar de manera no paramétrica la función de distribución empírica.
  - Función acumulada y su complemento
  - Función de densidad
4. Para cada modelo de distribución de probabilidad que se está poniendo a prueba, estimar:
  - Los parámetros del modelo por MV.
  - Estadísticos teóricos: media y coeficiente de Gini
  - Funciones teóricas: función de densidad, de densidad acumulada y complemento
5. Con la información de 3 y 4, estimar las medidas de divergencia:
  - Kullback-Leibler.
  - Kolmogorov-Smirnov.
  - Suma de Cuadrados del error.
6. Guardar tabla de resultados.

Por su parte, los algoritmos 2 y 3 siguen la metodología propuesta por Clauset et al. (2009) [19] y presentada en la sección 6.3 . En específico, el algoritmo 2 implementa

una prueba que identifica si un modelo de probabilidad es una hipótesis plausible para generar los datos empíricos. En tanto que, la tercera rutina busca identificar la cota inferior,  $x_{min}$ , de la cola derecha de la distribución, que sea óptima para ajustar los datos empíricos a algún modelo de probabilidad en estudio. Cabe señalar que, ambos algoritmos son ajustados a los fines de este estudio, en específico, las dos rutinas consideraron las tres medidas de error y los seis modelos de probabilidad en estudio; además la rutina 3 se adecua para trabajar con datos muestrales, los cuales requieren que se tome en cuenta un factor de expansión de la muestra.

De ahí que, el algoritmo 2 supone que una muestra empírica se distribuye de acuerdo a un modelo de probabilidad  $F$  con parámetros  $\theta_n$ , y en consecuencia calcula la bondad de ajuste usando la técnica *bootstrap*. Si el *valor p* obtenido es mayor que 0.1, entonces el modelo de probabilidad propuesto es una posible hipótesis para los datos empíricos.

### Algoritmo 2: Prueba de plausibilidad.

1. Definir el modelo  $F$ , con parámetros  $\theta_n$ , y una medida de divergencia  $D$ . Los parámetros y la medida de divergencia pueden ser obtenidos a través del algoritmo 1.
2. Desde  $r = 1$  hasta  $R = 3,000^1$ , hacer los siguientes pasos:

a) Crear una sub muestra *bootstrap*  $x^*$  de tamaño  $n = 3,000^2$ , datos sintéti-

---

<sup>1</sup>Para obtener un *valor p* dentro de  $\epsilon$  del valor real, Clauset et al. (2009) [19] sugieren generar  $n = \frac{1}{4}\epsilon^{-2}$  conjuntos de datos sintéticos. En consecuencia, para obtener un *valor p* que sea exacto de dos decimales ( $\epsilon = 0.01$ ), se crearon 3,000 conjuntos de datos sintético).

<sup>2</sup>El número de datos sintéticos no se igualó al de tamaño de la muestra porque implicaría obtener submuestras de tamaño de más de 20 millones de observaciones, o en su defecto obtener un peso específico para cada observación. Entonces se decidió reducir el tamaño de la submuestra a 3,000 observaciones porque, como apuntan Clauset et al. (2009) [19], con un tamaño de muestra mayor a 100 es posible distinguir si un modelo aproxima pobremente a los datos empíricos; por el contrario con muestras menores o iguales a 100 la prueba falla en eliminar un modelo del que no provienen los datos. Cabe señalar que para la prueba de la región de ingresos altos de la distribución, no se

cos que siguen un modelo de probabilidad con  $FDA F(\cdot, \theta_n)$ .

- b) Para el conjunto de datos sintéticos estimar, a través de máxima verosimilitud, los coeficientes  $\theta_n^*$  del modelo de probabilidad  $F$ .
  - c) Con base en la sub muestra  $x^*$  y el modelo  $F(\cdot, \theta_n^*)$ , estimar y guardar los estadísticos de divergencia,  $D_n^*$ , entre los datos sintéticos y los teóricos.
3. Calcular *valor p*. El *valor p* es la proporción de las veces que la medida de divergencia,  $D_n^*$ , creada por el paso 2 fue mayor a la medida de divergencia,  $D$ , que entró en el inciso 1. De tal forma que *valor p* = (*contar los elementos que cumplan* ( $D_n^* > D$ ))/ $R$ .

Por su parte el algoritmo 3, al suponer que la región de ingresos altos de la distribución de los datos empíricos muestrales, se distribuyen de acuerdo a cierto modelo de probabilidad, identifica  $x_{min}$ , a través de una búsqueda exhaustiva. En la rutina ingresan los datos muestrales y sus factores de expansión, el modelo de probabilidad a ajustar y los percentiles mínimo y máximo que definirán el inicio y fin del vecindario de exploración. La búsqueda crea sub muestras de un subconjunto de datos y estima, a través de  $MV$ , los parámetros de la distribución y las medidas de divergencia de la sub muestra con respecto al modelo estimado.

**Algoritmo 3: Encuentra ingreso mínimo óptimo  $x_{min}$ .**

1. Definir vecindario de exploración,  $S$ , que es una sub muestra en la región derecha de la distribución total con inicio, en el percentil mínimo, y fin en el percentil máximo.
2. Explorar exhaustivamente el vecindario  $S$ : para cada  $x_i$ , en el vecindario  $S$ 
  - a) Definir a  $x_{min} = x_i$ .
  - b) Crear  $H$ , sub muestra de datos  $S$ , con inicio en  $x_{min}$ .

---

consideró el efecto de que la distribución se comporta de manera diferente antes y después del valor  $x_{min}$ .

- c) Estimar, a través de máxima verosimilitud, los parámetros que ajustan el modelo de probabilidad a la muestra  $\hat{\theta}_i$ .
  - d) Estimar PDF y CDF empírico de  $H$ .
  - e) Crear datos teóricos que siguen una distribución analizada de acuerdo a  $x_{min}$  y los coeficientes estimados  $\hat{\theta}_i$ .
  - f) Estimar estadísticos de divergencia (KS, KL y SCE).
  - g) Guardar en una matriz la siguiente información: el valor de  $x_{min}$ , los parámetros estimados, y los criterios de error KS, KL, y SCE, .
3. Para cada medida de divergencia KS, KL y SCE realizar lo que sigue:
- a) Encontrar el valor de  $x_{min}$ : definir el renglón de la matriz creada en el inciso anterior que mantiene la más pequeña medida de divergencia al utilizar la función «min», del paquete estadístico R [70], sobre la columna que específica a la medida de divergencia en turno.
  - b) Guardar en memoria el renglón identificado en el inciso anterior, esto es los valores correspondientes a  $x_{min}$ , los coeficientes estimados  $\hat{\theta}_i$  y la medida de divergencia.

Por último, si por medio del algoritmo 2 se obtiene que dos modelos de probabilidad son posibles hipótesis para los datos empíricos, entonces, se comparan los dos modelos conforme a la prueba de razón de verosimilitud propuesta por Vuong (1989) [92], tal como se describe en la sección ?? y que se implementa en el algoritmo 4. En la rutina ingresan los datos de la muestra de la variable, los factores de expansión y los modelos a comparar. Así, se genera la razón de verosimilitud normalizada ( $LR$ ), o estadístico de Vuong, y se calcula el *valor p* de la prueba.

**Algoritmo 4: Prueba de Vuong para modelos no anidados y datos muestrales.**

1. Para cada modelo elegido y para cada punto de la muestra, obtener la verosimilitud de ambos modelos, el factor muestral y el número de coeficientes.

2. Si se analiza la región de ingresos altos, y uno de los modelos a comparar es el de Pareto, entonces omitir los puntos que sean menores al ingreso mínimo del modelo de Pareto.
3. Obtener la razón de verosimilitud, ponderada por el factor de la muestra, y ajustada por el número de parámetros en los modelos.
4. Estimar el estadístico de Vuong al normalizar la razón de verosimilitud.
5. Con base en el la distribución normal estándar estimar los *valores p* unilaterales y bilaterales.

Cabe mencionar, que las librerías que se usaron para estimar diferentes secciones de las rutinas fueron las que siguen: para calcular las funciones empíricas  $FD$  y  $FDA$  se recurrió a las funciones «Ecdf», de la librería Hmisc [52], y «density», de la librería stats [70], respectivamente. En tanto que, de la librería VGAM [97] fueron utilizadas las funciones «vglm» para realizar las estimaciones, por medio de máxima verosimilitud, de los parámetros de los modelos de probabilidades, y las funciones de tipo «dpqr» para los cálculos las funciones teóricas  $FD$ ,  $FDA$  y  $CFDA$ . Además, para estimar el coeficiente de Gini con pesos muestrales se utilizó la librería reldist ([42]). Por último, la medida de divergencia Kullback-Leibler se calcula por medio de la función «KL.plugin» de la librería entropy [46].

## 7.2 ALGORITMO PARA EL MSBA.

El algoritmo para implementar el MSBA de la teoría del excedente de la estratificación social es como sigue: en la rutina entran como parámetros iniciales el número de agentes a crear,  $N$ , los parámetros estructurales de nivel de desarrollo social,  $L$ , y la concentración del riqueza,  $\delta$ , y el nivel de ingreso promedio al inicio de la simulación.

**Algoritmo 5: MSBA.**

1. Repetir cada iteración 12 veces
  - a) De forma secuencial, empezando con el agente uno, cada agente se enfrenta con cada uno de los demás agentes.
  - b) Los agentes intercambian riqueza de acuerdo a las expresiones (5.19) y (5.20). El agente que gana el intercambio y el monto del mismo está en función del ingreso de los agentes al inicio del enfrentamiento y de los parámetros  $L$  y  $\delta$ .
  - c) A partir de la iteración 3, guardar los ingresos resultantes del intercambio.
2. Crear la distribución final al juntar los ingresos de los agentes que fueron guardadas en el paso 1.

### 7.3 DATOS EMPÍRICOS: DESCRIPCIÓN DE LOS DATOS

Los datos empíricos, sobre el ingreso monetario que obtienen los agentes económicos en México, son obtenidos de las series de la Encuesta Ingreso Gasto de los Hogares (ENIGH) publicada por el Instituto Nacional de Estadística y Geografía (INEGI). La ENIGH tiene como objetivo ser fuente de información, con alto nivel de desagregación, para el estudio del monto, la estructura y la distribución del ingreso y gasto de los hogares en México, y en consecuencia, es una de las bases de datos más importantes para el estudio de la dinámica del ingreso y gasto de la población mexicana [77].

El diseño de la muestra de la ENIGH permite que los datos publicados se puedan generalizar a nivel nacional en donde la unidad de análisis es el hogar, la vivienda y los integrantes del hogar. Asimismo, la encuesta ha publicado datos para los años 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014, siendo el año 2014 la última publicación disponible al momento de la realización de este trabajo.

Con respecto a los ingresos, la encuesta recolecta información, por perceptor de ingresos en cada hogar, sobre el ingreso corriente monetario y no monetario obtenido en un trimestre del año en que se levanta la encuesta. Por su parte, los ingresos monetarios reportados en la ENIGH se definen en pesos mexicanos del año en que se levantó la encuesta y son ingresos netos después de impuestos, es decir, son los ingresos que obtienen los agentes económicos una vez que pagan contribuciones a seguridad social e impuestos.

Entonces, con el objetivo de comprobar la hipótesis sobre cuál modelo de probabilidad puede ajustar con mayor fidelidad a la distribución del ingreso para el modelo de agentes económicos en México, ya sea en todos los niveles de la distribución o en la región de ingresos altos, así como, comprobar si el modelo de probabilidad se mantiene a través del tiempo, se construye la variable *ingreso*.

En específico, la variable *ingreso* tiene el propósito de captar el ingreso monetario proveniente de la participación de los agentes económicos en alguna actividad productiva en un determinado periodo, pero dejando de lado las transferencias monetarias que reciben por parte de instituciones públicas, privadas o de otros hogares, o montos que modifican el patrimonio neto de los activos y pasivos del individuo. Por consiguiente, para construir una variable que sea afín a esta definición se seleccionaron, para los integrantes de los hogares, de las claves referentes al ingreso corriente monetario trimestral, los rubros que corresponden al ingreso provenientes del trabajo asalariado, del trabajo independiente y de los originados por la posesión de activos físicos y no físicos. Mayor especificación sobre la ENIGH y las claves seleccionadas en cada muestra pueden consultarse en el ANEXO B en la SECCIÓN B.1.

Por último, cabe señalar que la variable ingreso se especifica en pesos reales, con año base de 2010<sup>3</sup>, por perceptor de ingreso en el hogar y después de impuestos, que en el transcurso de este trabajo sólo será referida como *ingreso*. Es importante aclarar la definición del ingreso porque los resultados sobre el ajuste de los modelos

---

<sup>3</sup>Las muestras de la variable ingreso se llevaron a pesos del año 2010 con base en el índice nacional de precios al consumidor publicado por INEGI.

de probabilidad puede diferir de las conclusiones reportadas en otras publicaciones que utilizan otra definición de ingreso [6,95], es decir, trabajan con ingresos brutos , ingresos por hogar, ingresos en términos monetarios e incluso trabajan con la definición de riqueza. Para una muestra de las diferentes definiciones de ingreso utilizadas en el tema puede referirse a la revisión bibliográfica en el ANEXO A.

## CAPÍTULO 8

# RESULTADOS

---

### 8.1 ANÁLISIS PRELIMINAR DE LAS MUESTRAS DE DATOS EMPÍRICOS

Conforme al procedimiento de homologación de muestras, descrito en la SECCIÓN 7.3, resultaron 15 muestras anuales para la variable *ingreso* en México para el periodo 1984-2014. Las muestras indican que en el periodo estudiado, en promedio, se captaron 38 millones de perceptores de ingreso, incluidos perceptores menores de 12 años, que al trimestre obtuvieron 16,090 pesos en promedio con una desviación estándar de 31,575 pesos. Así mismo, los montos mínimo y máximo de ingreso reportados, en promedio, fueron 3 y 2,312,508 pesos respectivamente. Los estadísticos descriptivos de cada muestra se encuentran en la tabla 8.1.

En cuanto a la distribución de los datos, a través de la figura 8.1, que presenta un diagrama de caja para las muestras de ingreso, es posible observar dos características interesantes. La primera es que, para todos los años, la mayor parte de los ingresos, del percentil 25 al 75, se concentra en la parte de baja de la distribución, esto es, entre los niveles de ingreso de entre los 3,362 hasta casi los 20,000 pesos. En segunda, se observó que los datos empíricos pueden extenderse hasta valores de ingreso que rebasan la cifra de millones de pesos, aunque este conjunto representa

Año	n	Media	DesvStd	Mínimo	Máximo	Mediana	Moda	Asimetría	Curtois
1984	21,826,201	16,017	19,811	3	557,328	11,882	16,974	6	88
1989	24,650,911	17,125	51,083	10	5,766,076	10,829	12,930	82	8,996
1992	28,511,012	18,254	33,635	0	1,189,274	10,808	10,808	11	231
1994	30,648,217	18,832	36,832	3	2,254,069	11,048	9,207	21	987
1996	32,643,110	13,254	32,827	3	2,443,400	7,824	6,767	44	3,048
1998	35,029,794	14,347	29,586	3	1,925,473	8,246	7,258	26	1,412
2000	37,185,752	16,386	27,309	4	911,095	9,714	9,477	9	161
2002	39,350,223	15,799	21,861	4	523,099	10,179	8,483	5	42
2004	41,174,755	16,133	29,304	4	2,673,769	10,541	7,750	27	1,662
2005	41,979,955	16,826	40,444	4	4,966,716	10,708	7,514	31	1,565
2006	45,620,851	16,741	26,695	4	1,550,512	10,877	7,251	12	336
2008	45,023,276	17,202	41,327	3	4,756,823	10,773	13,269	62	6,618
2010	45,325,046	15,285	23,088	1	943,675	10,252	12,098	11	259
2012	50,212,864	14,547	24,634	1	726,933	9,228	10,904	10	192
2014	50,208,049	14,602	35,195	1	3,499,373	9,307	10,098	53	4,855
<b>Promedio</b>	37,959,334	16,090	31,575	3	2,312,508	10,148	10,053	27	2,030

Estadísticos descriptivos de la variable ingreso personal. Montos en pesos de 2010. Fuente: elaboración propia con información de INEGI.

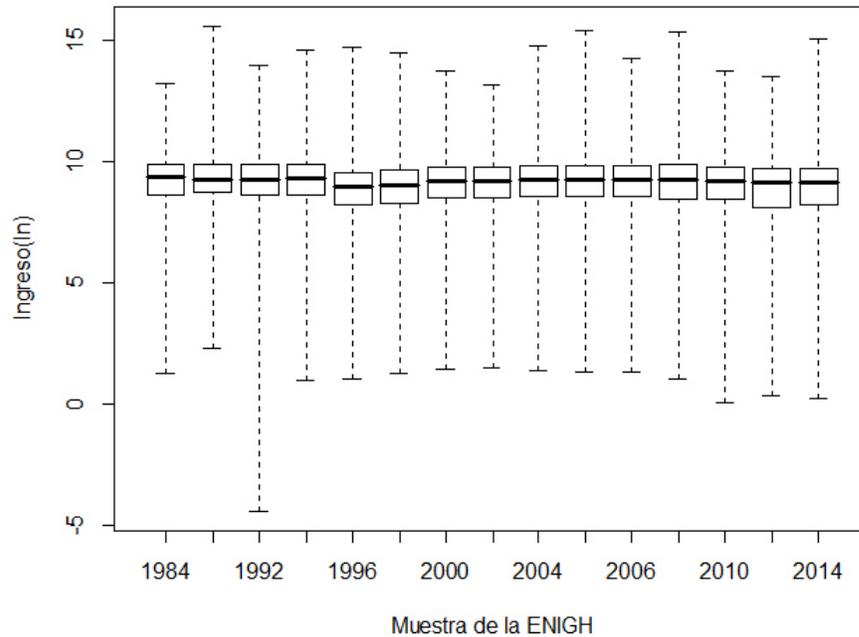


Figura 8.1: Diagrama de caja para la distribución del ingreso en México: 1984-2014. Fuente: elaboración propia con datos de la ENIGH, los ingresos están mostrados en logaritmos.

una pequeña proporción del total de las observaciones. Por su parte, las métricas de asimetría y curtosis, que ayudan a describir la forma de la distribución y que se presentan en la tabla 8.1, presentan signo positivo; por lo tanto, se sugiere que las distribuciones empíricas estudiadas presentan un sesgo positivo y una cola derecha de la distribución de tipo «ancha», es decir, se observa que la mayoría de las observaciones se concentran en los niveles bajo y medio del ingreso, y a su vez, la cola derecha de la distribución se extiende hasta valores muy altos del ingreso.

Además, la figura 8.1 sugiere que la mediana, el tamaño de la caja y la disposición de las mismas si bien no son iguales, éstas no variaron en gran manera a través de los años. Por último, conforme a los cálculos de la moda empírica, presentados en 8.1, se apunta a que los datos empíricos parecen seguir una distribución unimodal, en otras palabras, tienen una única moda.

Resumiendo, las características en común que se observaron, de manera preli-

minar, en las muestras del *ingreso* son que las distribuciones son unimodales, sesgadas a la derecha y mantienen una cola ancha en el lado derecho de la distribución. Estas propiedades son importantes porque también las presentan los modelos de probabilidad propuestos en este trabajo, y por la tanto, justifica su uso para caracterizar el caso mexicano.

## 8.2 AJUSTE DE LAS DISTRIBUCIONES EMPÍRICAS A UN MODELO DE PROBABILIDAD

Ya que se describió la variable a estudiar el contenido de esta sección es como sigue: en un principio se presentan, para la totalidad de la distribución del ingreso y los seis modelos de probabilidad propuestos, los resultados de las estimaciones de los parámetros de las distribuciones propuestas, las medidas de ajuste y las pruebas de ajuste que tienen como objetivo definir el modelo que mejor ajusta a la distribución del ingreso del caso mexicano.

Enseguida, se muestran resultados de la metodología descrita en la sección ?? para dividir en dos regiones a la distribución del ingreso: una que consta de los ingresos altos y la otra que engloba ingresos bajos y medios. Primero, se procede a estudiar la región alta de la distribución obteniéndose así, para cada modelo de distribución en estudio, las cotas mínimas de ingresos óptimos, y por último, se concluye cuál modelo se ajusta mejor a la cola derecha de la distribución en México. Posteriormente, suponiendo que la cota mínima del ingreso alto es la cota máxima de la región de ingresos bajos y medios, se procede a analizar esta última región y se define el mejor modelo para caracterizarla.

La división de la distribución permite constatar tres hipótesis que han cobrado interés en el ajuste empíricos de la distribución de ingresos y en el campo de la econofísica, donde son la base del desarrollo de modelos de simulación basados en agentes. La primera hipótesis afirma que la distribución del ingreso en todo su rango

puede ser descrita por un modelo Gamma [3, 12, 58]; la segunda hipótesis se refiere a que la región de ingresos altos sigue una distribución de tipo Pareto [15–17]; en tanto que, la última hipótesis pone énfasis en que en la región de ingresos bajos-medios se presenta una distribución tipo Exponencial [15, 28, 96], o Log normal [15, 21, 39, 85].

De acuerdo a estas hipótesis, para las regiones total y de ingresos medio-bajo se estiman los modelos Exponencial, Pareto, Gamma, Log Normal, Dagum y Singh-Maddala, mientras que, para la región de la cola derecha o ingresos altos se trabaja con las distribuciones Exponencial, Pareto, Log Normal y Gamma, las cuales son distribuciones útiles para caracterizar las cola derecha de la distribución ([19]).

### 8.2.1 ANÁLISIS DEL TOTAL DE LA DISTRIBUCIÓN.

Los modelos Pareto, Exponencial, Gamma, Log Normal, Dagum y Singh-Maddala se ajustan a las 15 muestras anuales del ingreso personal a través de máxima verosimilitud. Para cada modelo se estiman sus respectivos coeficientes, la media y el Gini teóricos y los criterios de divergencia entre la distribución empírica y los modelos. En la sección C.1 del apéndice C se presentan tablas de resultados para las 15 muestras y, como ejemplo, en la tabla 8.2 se reportan los resultados para el total de la distribución para la muestra del año 2014.

Entonces, en un primer análisis de la proximidad de la muestra de ingresos 2014 con los modelos estudiados se puede sugerir que, de entre todos los modelos, la distribución Dagum presenta mínimos criterios KS y SCE, que se basan en la función acumulada de densidad, en tanto que el criterio KL, basado en la información que logra captarse a través de la FD, favorece el desempeño en el ajuste del modelo Exponencial. Por su parte, las distribuciones que más se acercan a los datos empíricos, a través de la media y el coeficiente de Gini, son la distribución Exponencial y Singh-Maddala, respectivamente. Por último, conforme a la valoración visual de la FDA de la figura 8.2, la distribución Dagum, línea sólida más oscura, se aprecia

como la más parecida a la FDA de los datos empíricos.

<b>Datos Empíricos</b>									
<i>Media</i>	14,602								
<i>Gini</i>	0.56								
<b>Modelo</b>	$\alpha$	$\alpha 2$	$\beta$	$\mu$	<b>Gini</b>	<b>Media</b>	<b>KL</b>	<b>SCE</b>	<b>KS</b>
<i>Pareto</i>	0.12				-1.30	-	2.63	820.79	0.44
<i>Exponencial</i>			14,602		0.50	14,602	0.06	20.84	0.08
<i>Gamma</i>	0.78		18,701		0.55	14,602	0.17	16.03	0.07
<i>Log normal</i>			1.46	8.83	0.70	6,812	0.24	45.76	0.12
<i>Dagum</i>	2.40	0.33	19,652		0.55	13,736	0.11	2.72	0.04
<i>Singh-Maddala</i>	1.01	4.29	45,985		0.56	14,098	0.11	8.74	0.06

Tabla 8.2: Estimaciones de los modelos de probabilidad para la totalidad de la distribución de ingresos 2014. Las columnas presentan la siguiente información: ( $\alpha$ ) se refieren al parámetro de forma, en específico, para la distribución Dagum y Singh-Maddala es el parámetro de forma  $a$ ;  $\alpha 2$  es el parámetro de forma  $p$  para el modelo Dagum, en tanto que,  $q$  para el modelo Singh-Maddala;  $\beta$  es el parámetro de escala; por último,  $\mu$  se refiere al parámetro de localización. Las columnas Gini y Media presentan el índice de Gini y la Media teóricos que se desprenden de la estimación de los coeficientes estimados. Por último, se presentan los criterios de divergencia Kullback-Leibler, suma del cuadrado del error y la distancia Kolmogorov-Smirnov identificadas por las columnas KL, SCE y KS, respectivamente.

Sin embargo, los anteriores resultados sólo informan cuál de los modelos caracteriza mejor a los datos empíricos con base en un criterio, empero cualquier distribución puede utilizarse para ajustar los datos empíricos sin importar cuál sea la distribución de la cual provienen los mismos. Por consiguiente, al seguir la metodología descrita en la subsección 6.3.2, para probar cuantitativamente la hipótesis de que es plausible que los datos empíricos sigan alguna de los modelos de probabilidad propuestos, se obtienen los siguientes resultados.

Los resultados, presentados en la figura 8.3 muestran los *valores p* generados al implementar la prueba de plausibilidad conforme a las tres métricas de distancia estudiadas. La figura sugiere que sólo a través del criterio KL se puede encontrar

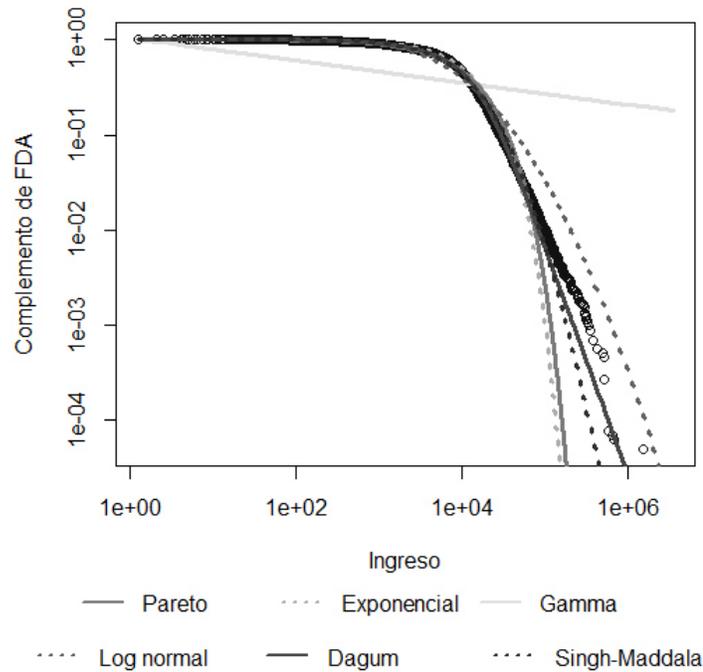


Figura 8.2: Complemento de la función de densidad acumulada 2014. Se muestra el complemento de la función de densidad acumulada de la muestra 2014, en puntos negros, y el ajuste de los modelos Pareto, Exponencial, Gamma, Log normal, Dagum y Singh-Maddala, en líneas sólidas y punteadas.

evidencia que respalda la hipótesis de que las distribuciones Dagum, Gamma, Log Normal y Singh Maddala son plausibles para describir los datos empíricos en ciertos años. En especial, la distribución Dagum resulta con *valores p* suficientemente grandes en el 80 % de las muestras, en tanto que, los porcentajes para las distribuciones Log Normal, Gamma y Singh-Maddala fueron de 60 %, 13 % y 7 % respectivamente.

Por su parte, los coeficientes estimados, de los modelos plausibles, fueron los que siguen: para el modelo Dagum los parámetros de forma  $a$  y  $p$  se ubican en los rangos de  $[2.02, 2.73]$  y  $[0.32, 0.53]$ . Para el modelo Log normal se presentan coeficientes de Gibrat<sup>1</sup> de entre 0.47 y 0.59, el modelo Gamma resulta con parámetros de forma

<sup>1</sup>El índice de Gibrat sirve para medir la desigualdad en la distribución de una variable y es definido como  $\beta \equiv 1/\sqrt{2\sigma^2}$ , donde  $\sigma$  es el parámetro de forma o la desviación estándar del logaritmo de la variable aleatoria [85].

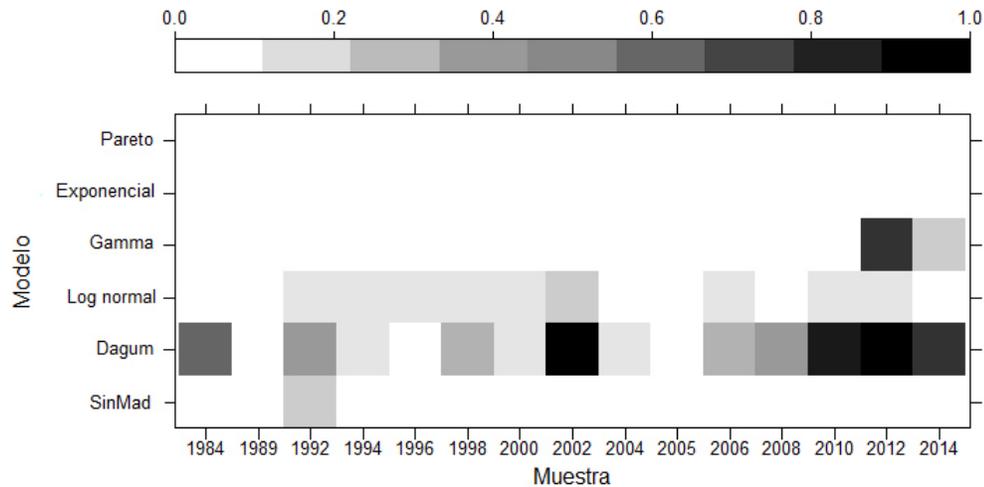


Figura 8.3: Prueba de plausibilidad para el total de la distribución de ingreso. *Valores p* para la prueba de plausibilidad por medida de error, modelo de probabilidad y año de la muestra. El panel muestra las estimaciones a través del criterio Kullback-Leibler con el que se obtuvieron *valor p* diferentes a cero.

en un rango de  $[0.76, 0.78]$ , por último para el modelo Singh-Maddala se estiman coeficientes de forma  $a$  y  $q$  de 1.36 y 1.83, respectivamente. Cabe señalar que, los coeficientes estimados se encuentran dentro de los rangos reportados en la literatura para diferentes países, y en especial para México, una muestra de la bibliografía consultada en este trabajo se presenta en la tabla A.1 del apéndice A. Sin embargo, los coeficientes estimados en este trabajo y los presentados en la revisión bibliográfica no son los mismos; las diferencias que existen entre los coeficientes estimados y los reportados en la literatura puede deberse a la discrepancia en factores como son la definición de ingreso, año y región estudiados y el método de estimación de los coeficientes [6, 95]. Para realizar un análisis de la desigualdad en la distribución de ingreso en el periodo estudiado, se procede a estimar coeficientes de Gini de manera paramétrica, con las ecuaciones presentadas en la sección 5.2, para los años y modelos que resultaron plausibles, además de estimaciones para los datos empíricos de la muestra de la ENIGH. Así, las estimaciones pueden consultarse en la figura 8.4. Los resultados apuntan a que en todo el periodo se ha incrementado la concentración del

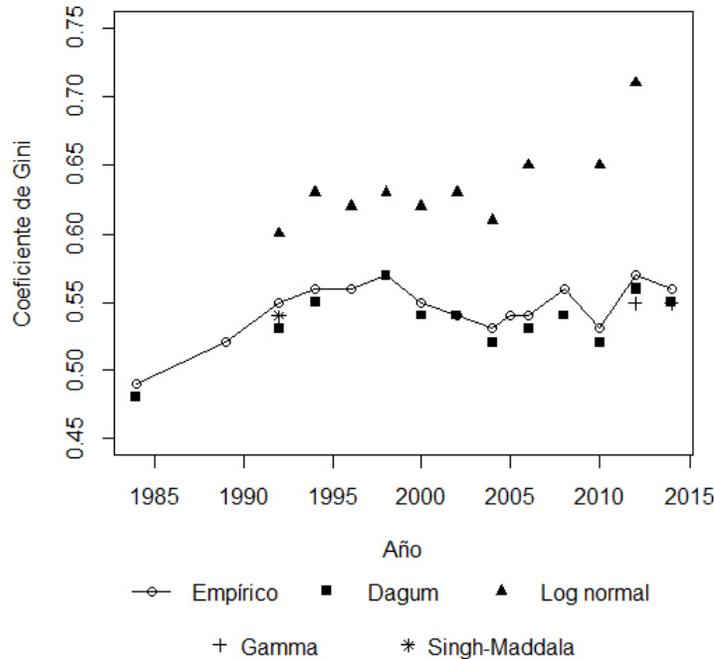


Figura 8.4: Coeficientes de Gini para la totalidad de la distribución del ingreso. Se presentan estimaciones para los datos de la muestra empírica, y estimaciones paramétricas con base en la asignación de los modelos de probabilidad Dagum, Log normal, Gamma y Singh-Maddala a los datos de ingreso.

ingreso en menos personas; en el periodo 1992 a 2014 el incremento fue de 16 %, de acuerdo a los datos de la muestra, y en un 14 %, con base en el modelo Dagum, en tanto que, si se reduce el periodo de 1992 a 2014 el incremento fue de 2.8 % y 3 % con respecto a la muestra y al modelo Dagum, respectivamente<sup>2</sup>.

Tomando como base los modelos que resultaron ser plausibles para originar a los datos de ingreso en un mismo año, se realiza una prueba de razón de verosimilitud para definir cuál de los modelos presenta mejor ajuste a los datos empíricos. En la tabla 8.3 se muestran los resultados de la prueba de razón de verosimilitud que compara al modelo Dagum con los modelos Gamma, Log normal y Singh-Maddala. De acuerdo a la tabla, en todas las pruebas se rechaza la hipótesis nula de que los dos

<sup>2</sup>En la literatura que estudia el ingreso en México se reporta que el índice de Gini, estimado por medio del modelo Dagum, se incrementó en 7.87 puntos porcentuales en el periodo 1998 a 2008 [73], en tanto que, del periodo 1984 a 1996 el incremento fue de 15 % [6].

(a) Dagum						
Año	Gamma		Log normal		Singh-Maddala	
	LR	valor <i>p</i>	LR	valor <i>p</i>	LR	valor <i>p</i>
1992	-	-	-447	<b>0</b>	-551	0
1994	-	-	-950	<b>0</b>	-	-
1998	-	-	-867	<b>0</b>	-	-
2000	-	-	-1028	<b>0</b>	-	-
2002	-	-	-1068	<b>0</b>	-	-
2004	-	-	-1136	<b>0</b>	-	-
2006	-	-	-1372	<b>0</b>	-	-
2010	-	-	-1377	<b>0</b>	-	-
2012	-568	<b>0</b>	-1558	<b>0</b>	-	-
2014	-326	<b>0</b>	-	-	-	-

(b) Log normal		
Año	Gamma	
	LR	<i>p value</i>
2012	-635	<b>0</b>

Tabla 8.3: Prueba razón de verosimilitud para la distribución del ingreso. En el panel (a) se compararán los modelos Gamma, Log normal y Singh-Maddala con respecto al modelo Dagum, en tanto que, en el panel (b) se comparará el modelo Gamma con respecto al modelo Log normal. LR es la razón de verosimilitud normalizada de acuerdo a  $\sqrt{n}R/\sigma$ , y los *valores p* que estadísticamente son significativos se muestran en negritas.

modelos comparados presentan el mismo ajuste para los datos empíricos, y por lo tanto, se sugiere que en general, el modelo Dagum es mejor que los modelos Log normal, Gamma y Singh-Maddala para caracterizar al ingreso de los agentes económicos en los años 1992, 1994, 1998, 2000, 2002, 2004, 2006, 2010, 2012 y 2014. Por su parte, los resultados de acuerdo al número de parámetros en el modelo favorecen al modelo Gamma sobre el Log normal, en la muestra de 2012, mientras que el modelo Dagum se muestra mejor que el Singh-Maddala en la muestra de 1992.

### 8.2.2 ANÁLISIS POR REGIÓN: INGRESOS ALTOS

Con base en la metodología descrita en la sección 6.3.4, se obtienen los niveles de ingreso mínimo óptimo para caracterizar la cola derecha de la distribución a través de los modelos de Pareto, Exponencial, Log Normal y Gamma. Las cotas mínimas de ingreso son obtenidas por medio de las medidas de error divergencia KL, distancia KS y la SCE. Los resultados son ejemplificados en la figura 8.5; la figura muestra, para el año 2014 y el modelo de Pareto, la variación de las medidas de divergencia KL, SCE y KS al variar la cota mínima de ingreso utilizada para ajustar a los datos empíricos el modelo de Pareto. Entonces, el gráfico sugiere que existe un diferente nivel de ingreso mínimo óptimo para cada medida de error y modelo de probabilidad en estudio, y en consecuencia, la correcta selección del ingreso mínimo dependerá de la calidad de la medida de divergencia [19].

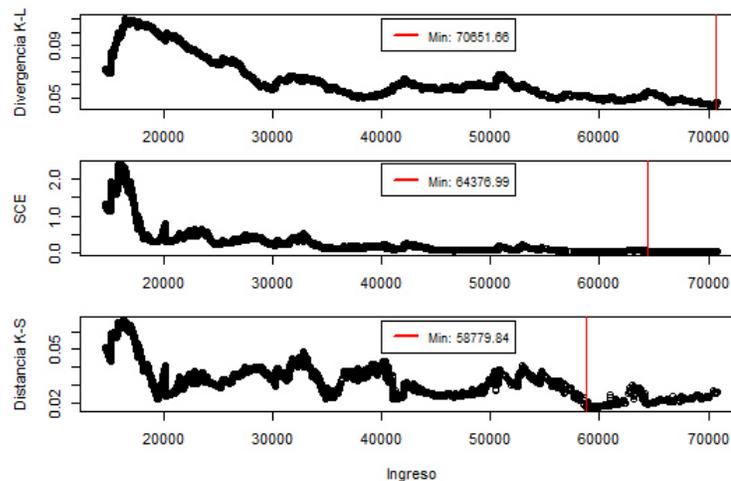


Figura 8.5: Medidas de divergencia y cota inferior de ingreso para el modelo de Pareto para los datos de 2014. Las medidas de divergencia KL, SCE y KS al variar la cota inferior del ingreso para el modelo de Pareto. Año de la muestra 2014

Por consiguiente, para el análisis de la caracterización de la región de ingresos altos, se trabaja con 12 sub muestras para cada una de las 15 muestras en estudio; estas submuestras provienen de la estimación de los niveles óptimos de ingreso mínimo. En el apéndice C en la sección C.2 se pueden consultar las tablas con los

niveles óptimos de ingreso mínimo estimados por año, modelo de probabilidad y medida de error; además se presentan el valor de la medida de divergencia asociada, el porcentaje de la población que obtiene ingresos mayores que el ingreso mínimo óptimo, y los coeficientes estimados. Por su parte, en la figura 8.6 se presentan, a manera de ejemplo visual, y conforme a las cotas mínimas obtenidas por medio de la divergencia KL, los resultados del ajuste entre los modelos de probabilidad -Gamma, Exponencial, Log normal y Pareto- y los datos empíricos del año 2014. La figura 8.6 sugiere que el modelo de Pareto, representado por la línea más oscura, es el que mejor se acerca al complemento de función de densidad acumulada empírica, representada por los puntos negros.

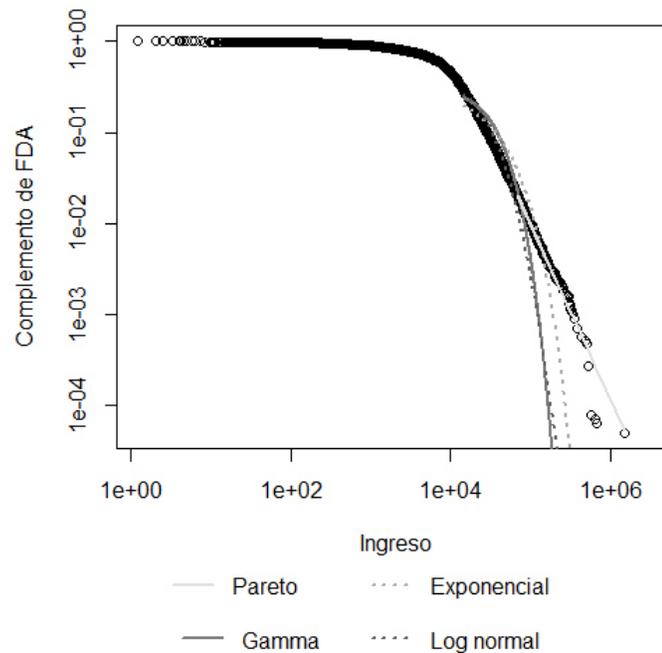


Figura 8.6: Complemento de la función de densidad acumulada para la región de ingresos altos 2014. Se muestra el complemento de la función de densidad acumulada de la muestra 2014 en puntos negros y el ajuste de los modelos Pareto, Exponencial, Gamma y Log normal en líneas sólidas. La divergencia KL fue la medida de error usada para obtener las cotas mínimas de ingreso para cada modelo.

Una vez que se obtienen las sub muestras asociadas a los ingresos óptimos mínimos, se proceden a realizar las pruebas de plausibilidad que tienen como hipótesis

nula que, dado un modelo de probabilidad, los datos empíricos provienen del modelo de probabilidad propuesto. En la figura 8.7, se presentan los *valores p* por modelo, año de la muestra y medida de divergencia para cada sub muestra; éstos apuntan a que, para algunos años, el modelo de Pareto y Log Normal son los únicos que presentaron una probabilidad suficientemente grande para no rechazar la hipótesis nula. En específico, el modelo de Pareto, mantiene *valores p* suficientemente grandes en todas las muestras generadas por el criterio KL; en 13 de las 15 muestras creadas por la SCE, y sólo en la muestra del año 2014 obtenida por la distancia KS. Por su parte, para el modelo Log normal no se puede rechazar la hipótesis de plausibilidad en las muestras de 1989 y 2005, obtenidas por el criterio KL. Asimismo, los valores

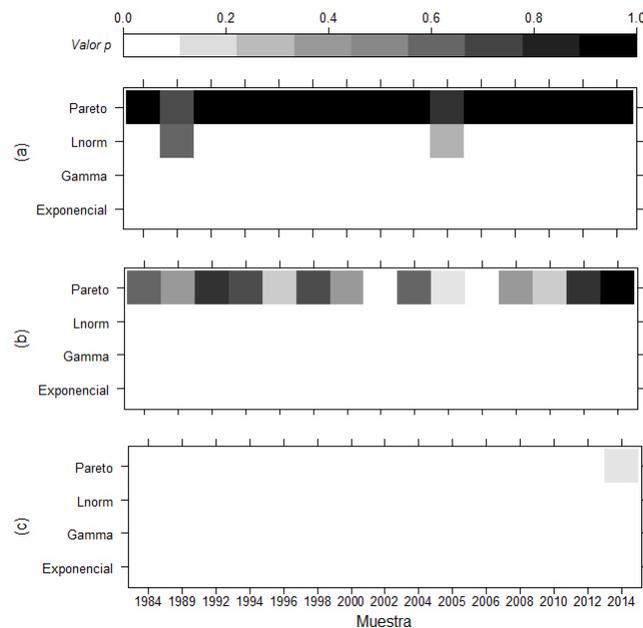


Figura 8.7: Prueba de plausibilidad para la región de ingreso alto de la distribución de ingreso. *Valores p* para la prueba de plausibilidad por modelo de probabilidad y año de la muestra. Cada panel corresponde a los resultados de las pruebas de plausibilidad según la medida de error que se utilizó para generar las submuestras y la misma prueba; el panel (a) divergencia Kullback-Leibler, panel (b) suma del cuadrado del error y en el panel (c), la distancia Kolmogórov-Smirnov.

estimados del coeficientes de Pareto y el índice de Gibrat que son significativos en la

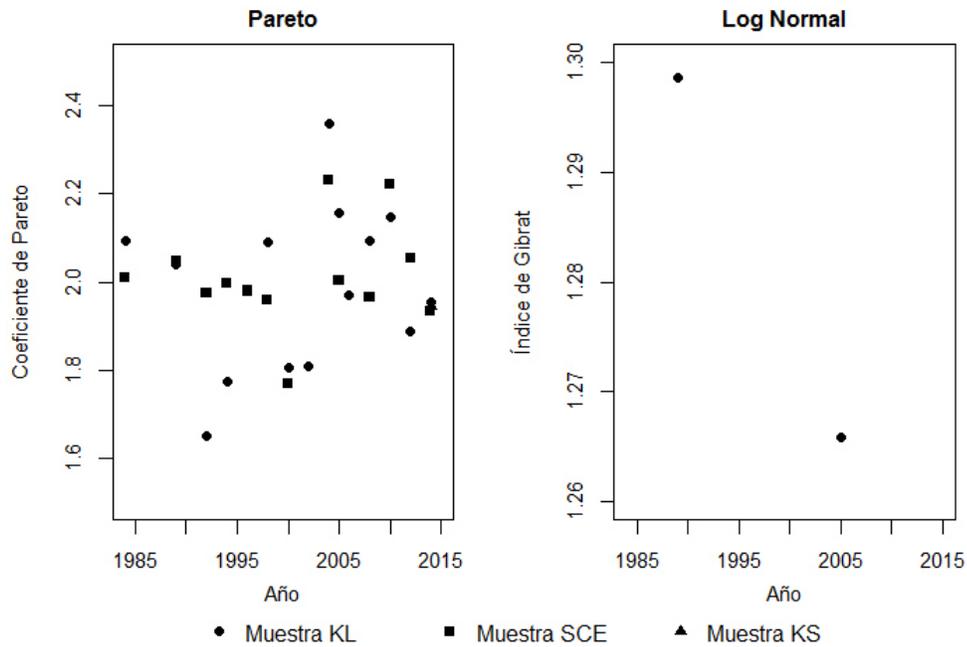


Figura 8.8: Coeficientes de Pareto e índice de Gibrat para la región de ingresos altos.

prueba de plausibilidad se muestran en la figura 8.8. En particular, los coeficientes que son reportados para el modelo de Pareto se mantienen entre los valores de 1.6 y 2.3, valores que entran en los rangos publicados en la literatura [71]<sup>3</sup>. En especial, el parámetro de forma de Pareto ayuda a describir el grado de desigualdad en una distribución porque asocia un mayor grado de desigualdad a un menor coeficiente de Pareto; en el caso mexicano, si bien no se observa una tendencia clara sobre la evolución del coeficiente de Pareto en el tiempo, sí se puede mencionar que los años con mayor desigualdad en la región de ingresos altos fue en 1992, de acuerdo a la divergencia KL, y 2000 de acuerdo a la SCE, mientras que, los años con menor desigualdad fue en los años 2004, para las muestras generadas por KL y SCE. Por último, cabe señalar que algunos valores estimados se encuentran por encima del rango,  $0 < \alpha < 2$ , por lo que se sugiere en que en esos años la distribución no es estable [34].

Por su parte, para los años 1989 y 2005 el coeficiente de Gibrat resulta ser

<sup>3</sup>En la tabla A.2 del apéndice A puede consultarse un resumen de la literatura empírica que se estudió en este trabajo.

de 1.29 y 1.26, respectivamente. De manera similar que el coeficiente de Pareto, el índice de Gibrat asocia un menor índice con una distribución más desigual [21, 85]; de ahí que, de acuerdo al modelo Log normal, se observa que la distribución del ingresos altos en México es más desigual en año 2005 que en el año 1989.

Año	LR	Valor $p$
<b>1989</b>	-152	<b>0</b>
<b>2005</b>	-326	<b>0</b>

Tabla 8.4: Prueba razón de verosimilitud para la cola derecha de la distribución de ingreso entre los modelos de Pareto y Log normal. LR es la razón de versomilitud normalizada de acuerdo a  $\sqrt{n}R/\sigma$ , y los *valores p* estadísticamente significativos en negritas.

Dado que la prueba de plausibilidad sugiere que los datos empíricos, para las sub muestras generadas por la divergencia KL en los años 1989 y 2005, pueden ser provenientes de los modelos de Pareto o Log normal, se procede a realizar pruebas de razón de verosimilitud con el objetivo de obtener cuál modelo presenta mejor ajuste a los datos empíricos. En la tabla 8.4, se reportan el estadístico de razón de verosimilitud y el *valor p* de la prueba; la tabla muestra que la prueba favorece el ajuste del modelo de Pareto para ambos años.

El análisis descrito en esta sección apunta a que al menos para el caso mexicano existe evidencia suficiente para concluir que la parte alta de la distribución en México sigue un modelo de Pareto. Este resultado respalda la hipótesis de que la parte de la distribución de ingresos altos en México pudiese ser generada por un modelo estocástico multiplicativo al que se añade ruido aditivo y/o cotas restrictivas con interacciones entre agentes [26]. Además, dada esta conclusión, en la siguiente subsección se supondrá que las regiones de ingresos bajos-medios de la distribución de ingresos en México terminan en la cotas mínimas del modelo de Pareto.

### 8.2.3 ANÁLISIS POR REGIÓN: INGRESOS BAJOS Y MEDIOS

De manera similar, que en las secciones del análisis de la distribución total y de la región alta de ingresos, las estimaciones del ajuste de los modelos de probabilidad con respecto a los datos empíricos de los ingresos bajo y medio, se pueden consultar en el apéndice C.3<sup>4</sup> y por medio de la figura 8.9 se presentan, a manera de ejemplo, los datos empíricos para la sub muestra 2014, generados por el criterio KL, y el ajuste que cada modelo presentó hacia éstos.

De acuerdo a las pruebas de plausibilidad para la región de ingresos bajos-medios, se muestra que sólo las sub muestras originadas por el criterio KL presentan suficiente evidencia estadística para sostener que es plausible que, para ciertos años, los modelos Log normal, Dagum y Singh-Maddala generan a los datos empíricos, y por lo contrario, se pueden descartar del análisis a los modelos Exponencial, Pareto y Gamma. En específico, como se presenta en la figura 8.10 el modelo Dagum presenta valores  $p$  suficientemente grandes en 12 de las 15 muestras, los años con *valores p* por debajo del criterio de aceptación son los años 1984, 1992 y 2000, y los parámetros estimados para las 12 muestras plausibles se posicionan en los rangos  $[2.5, 8.5]$  y  $[0.1, 0.4]$  para los coeficientes de forma  $a$  y  $p$ , respectivamente. Por su parte, para el modelo Log normal no se descartan en 5 de las 15 muestras, que representan a los años 1996, 1998, 2002, 2008 y 2014, que mantuvieron índices de Gibrat de entre 0.5

---

<sup>4</sup>Cabe señalar que la estimación del modelo Singh-Maddala a través del algoritmo de VGAM no convergen para algunas muestras. En específico, para obtener estimaciones se ponen a prueba dos métodos que se evalúan a través de una prueba de razón de verosimilitud. El primer método consiste en bajar el criterio de convergencia hasta obtener resultados, mientras que el segundo método, sugerido por Yee (2015) [97], resulta de usar un modelo más simple, el modelo Fisk, para proveer valores iniciales a un modelo más complejo, el modelo Singh-Maddala. Como sea, en el segundo método también necesita bajar el criterio de convergencia para poder obtener estimaciones por lo cual se sugiere tomar estos resultados con precaución ya que ante un fallo en obtener estimaciones es recomendable abandonar el modelo Yee (2015) ([97]). Los resultados de la prueba de verosimilitud, los criterios de convergencia y el método de estimación usados para el ajuste de las muestras que no convergieron, mediante los criterios base, se encuentran en la tabla C.13 del apéndice C.3.

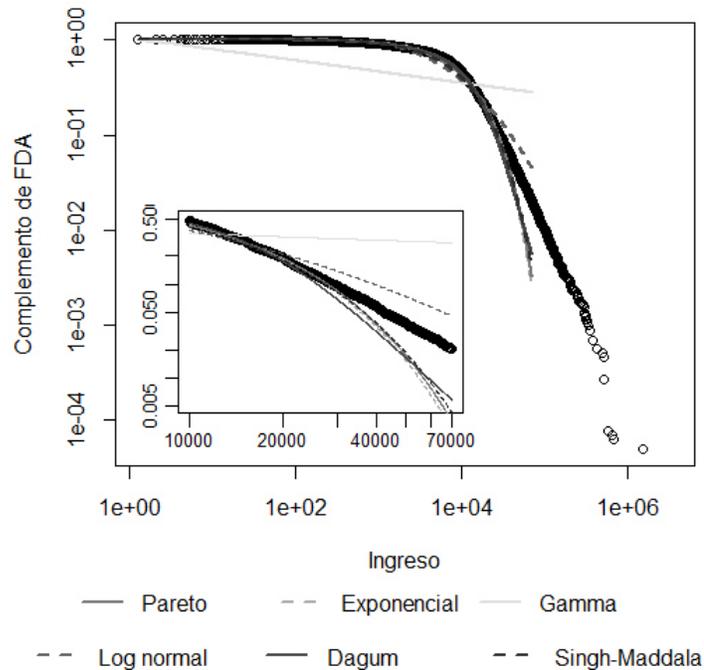


Figura 8.9: Complemento de la función de densidad acumulada para la región de ingresos bajos-medios 2014. Se muestra el complemento de la función de densidad acumulada de la muestra 2014 en puntos negros y el ajuste de los modelos Pareto, Exponencial, Gamma, Log normal, Dagum y Singh-Maddala. La divergencia KL fue la medida de error usada para obtener la división de la distribución del ingreso en dos regiones.

a 0.6. Por último, el modelo Singh-Maddala obtuvo suficiente evidencia estadística en la muestra del año 2000 con coeficientes de forma  $a$  y  $p$  de 1.15 y 13.7.

Asimismo, en la figura 8.11 se presentan los coeficientes de Gini para la región de ingresos bajos y medios que se estimaron de manera paramétrica, a través de los coeficientes estimados que resultaron plausibles, y de manera no paramétrica con base en los datos empíricos. La figura refiere que para algunos años, como lo son 1996 y 1998, para el modelo Log normal, y 1992 y 2002, para el modelo Dagum, los coeficientes de Gini muestrales y los estimados de manera paramétrica llegaron a ser casi idénticos. Además, la figura 8.11 sugiere que la desigualdad en la parte baja de la distribución se ha incrementado en el periodo analizado.

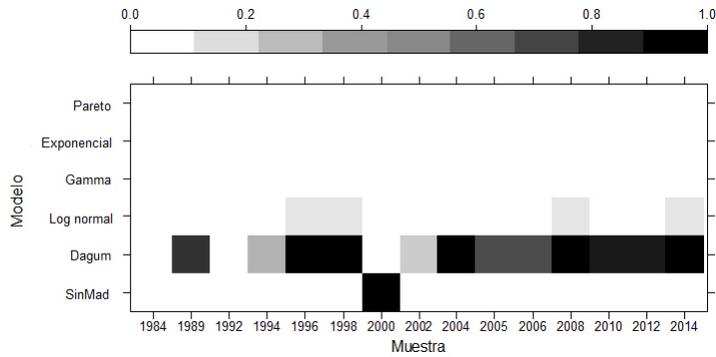


Figura 8.10: Prueba de plausibiliad para la región de ingresos bajos-medios de la distribución de ingresos *Valores p* para la prueba de plausibilidad por modelo de probabilidad y año de la muestra. El panel muestra las estimaciones a través del criterio Kullback-Leibler con el que se obtuvieron *valor p* diferentes a cero.

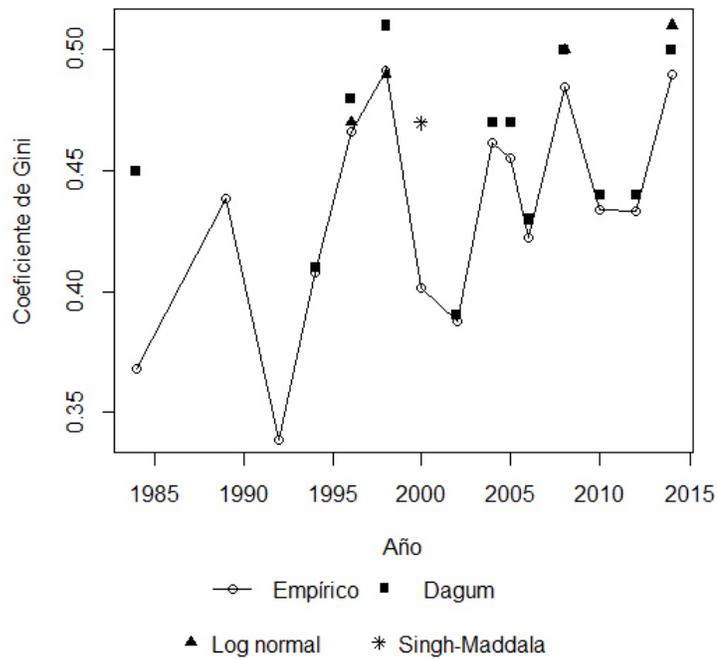


Figura 8.11: Coeficientes de Gini para la región de ingresos bajos-medios de la distribución de ingresos. Se presentan estimaciones para los datos de la muestra empírica, y estimaciones paramétricas con base en la asignación de los modelos de probabilidad Dagum, Log normal y Singh-Maddala a los datos de ingreso.

Para finalizar, en la tabla 8.5, se encuentran los resultados de las pruebas de razón de verosimilitud entre los modelos Dagum y Log normal que, para un mismo año, no fueron descartados en el anterior proceso. Es así que la tabla 8.5, sugiere que en los años 1996, 1998, 2008 y 2014, el modelo Dagum caracteriza de mejor manera a los datos empíricos. Es decir, existe suficiente evidencia para sugerir que los datos provienen de una distribución Dagum, y que además, el modelo Dagum mantuvo superioridad en el ajuste en comparación de los modelos que también pueden ser compatibles con los datos, tal es el caso de los modelos Singh-Maddala y Log normal. Por último cabe señalar, que el análisis rechaza la hipótesis de que los datos de ingresos medio y bajo en México resulten de la distribución Exponencial, empero, sí se presenta suficiente evidencia que apoya a la hipótesis de que la distribución Log normal puede generar los datos de la región analizada [20, 21, 26, 83, 87].

<b>Lnorm</b>		
<b>Año</b>	<i>LR</i>	<i>valor p</i>
<b>1996</b>	-1039	<b>0</b>
<b>1998</b>	-960	<b>0</b>
<b>2008</b>	-1664	<b>0</b>
<b>2014</b>	-1895	<b>0</b>

Tabla 8.5: Prueba razón de verosimilitud para la región de ingresos bajos-medios de la distribución de ingresos entre el modelo Log normal y el modelo Dagum. LR es la razón de verosimilitud normalizada de acuerdo a  $\sqrt{n}R/\sigma$ , y se presentan los *valores p* estadísticamente significativos en negritas.

### 8.3 RESULTADOS DE LA SIMULACIÓN DEL MSBA

De acuerdo al modelo presentado en la sección 5.3 y la implementación descrita en la sección 7.2 en esta parte del trabajo se presentan el análisis de las distribuciones generados por medio de la implementación computacional. En particular, la implementación tiene por objeto identificar si la distribución generada a través del MSBA replica a alguna de las regularidades empíricas identificadas en la sección 8.2

para caracterizar las regiones de la distribución ingreso de los agentes económicos en México.

La simulación considera 18 escenarios, generados por la combinación de los parámetros de desarrollo social ( $L = 1, 2, 3, 4, 5, 6$ ) y de grado de concentración de poder conforme al nivel de ingreso ( $\delta = 0.4, 0.5, 0.6$ ). Además, se trabaja con 200 agentes computacionales, cada uno con la misma cantidad de riqueza al inicio de la simulación, la cual fue la media del ingreso observado en el año 2012 en México. En cada iteración los agentes interactúan, en pares, con cada uno de los demás agentes para intercambiar riqueza en función de la riqueza de los dos agentes y el valor de los parámetros estructurales  $\delta$  y  $L$ .

En total, se llevan a cabo 12 iteraciones<sup>5</sup>; las primeras dos son descartadas para quitar «efectos de esquina», en tanto que las 10 restantes se unen para formar la distribución final, de ahí que se genera una distribución de 2000 observaciones.

---

<sup>5</sup>El uso de las 12 iteraciones se hizo siguiendo a Angle (1986) [3], quien señala que después de 10 iteraciones es posible comprobar que la distribución es estacionaria dado que las observaciones del mismo agente es independiente al ser la correlación serial es casi nula.

Tabla 8.6: Estadísticos descriptivos de los datos simulados

$\delta$	$L$	$n$	<i>Media</i>	<i>DesvStd</i>	<i>Moda</i>	<i>Mediana</i>	<i>Mínimo</i>	<i>Máximo</i>	<i>Asimetría</i>	<i>Curtosis</i>	<i>Gini</i>
0.4	1	2,000	14,547	18,043	2,129	8,532	0	137,087	2	8	0.58
	2	2,000	14,547	12,912	4,783	10,982	11	94,319	2	3	0.46
	3	2,000	14,547	10,899	6,119	12,244	53	94,472	1	4	0.40
	4	2,000	14,547	9,711	11,565	12,532	243	73,675	1	3	0.35
	5	2,000	14,547	8,701	10,814	12,669	612	66,916	1	2	0.32
	6	2,000	14,547	8,002	11,244	13,148	828	54,959	1	2	0.30
0.5	1	2,000	14,547	20,370	1,606	6,964	0	177,073	3	12	0.63
	2	2,000	14,547	14,890	3,513	9,724	0	101,054	2	5	0.51
	3	2,000	14,547	12,669	5,295	10,903	9	101,369	2	4	0.45
	4	2,000	14,547	10,890	6,760	11,869	74	80,644	1	3	0.40
	5	2,000	14,547	10,321	8,799	12,220	39	79,640	1	3	0.38
	6	2,000	14,547	9,841	8,376	12,318	328	61,488	1	2	0.36
0.6	1	2,000	14,547	22,130	1,582	5,052	0	165,592	3	10	0.67
	2	2,000	14,547	18,371	2,517	7,807	2	177,852	3	10	0.59
	3	2,000	14,547	14,962	3,597	9,827	6	120,670	2	5	0.51
	4	2,000	14,547	13,737	3,862	10,628	77	99,987	2	4	0.48
	5	2,000	14,547	12,839	4,737	10,976	93	99,338	2	4	0.45
	6	2,000	14,547	11,713	5,458	11,682	47	91,001	2	4	0.42

Tabla 8.6: Estadísticos descriptivos de los datos simulados. Montos en pesos de 2010.

Para saber si existe independencia en los datos se mide la correlación del ingreso del agente a través de las iteraciones y de ahí que se concluye que el ingreso del agente en la iteración  $t$  esta solo correlacionado con el ingreso en la iteración inmediata anterior. Asimismo, se presenta que son mínimos los coeficientes de variación de los estadísticos descriptivos de los datos generados en 100 simulaciones del proceso, por lo que se sugiere que existe estabilidad del MSBA. La información con los datos que respaldan la estabilidad del proceso y la independencia del ingreso entre agentes se encuentran en la tabla D.1 y en la figura D.1 del apéndice D.

Los estadísticos descriptivos de los datos para cada escenario, se muestran en la tabla 8.6. La asimetría estadística y el exceso de curtosis de todas las muestras presentan signos positivos, sugiriéndose así que las distribuciones tienen una cola derecha más larga y ancha que la cola izquierda, y a su vez que, el decaimiento de la cola es más disperso que la distribución normal. Además, se muestra que las combinaciones de los parámetros que resultan con mayor coeficiente de Gini, distribuciones más inequitativas, asimetría estadística y curtosis postivias son aquellas que se sesgan hacia el desarrollo social bajo y concentración del poder en los agentes con mayor riqueza.

De la misma manera que con los datos empíricos, cada distribución que se genera a través del MSBA se divide en dos regiones; la región de ingresos altos y de ingresos bajos y medios. Para la división se supone que la cola derecha de la distribución es generada por un proceso tipo Pareto, y se estiman los ingresos mínimos óptimos a través del criterio  $KL$ . Enseguida, para cada región de las distribución se calculan los coeficientes de los modelos de probabilidad, que en la sección 8.2 resultaron ser plausibles para caracterizar a los datos empíricos en México, y los estadísticos de divergencia entre los datos simulados y el modelo de probabilidad estimado. Por último, se realizan pruebas de plausibilidad sobre los modelos recién ajustados.

En la figura 8.12 se encuentra una representación gráfica del  $CFDA$  de los datos

simulados para los escenarios de nivel de desarrollo medio, ( $L = 3$ ), con variación en el parámetro de concentración de poder de acuerdo al grado de ingreso,  $\delta = 0.4, 0.5, 0.6$ . Además, el gráfico muestra el ajuste de la distribución con respecto al modelo Dagum, para las regiones de ingresos bajo y total, y del modelo de Pareto, para la región de ingresos altos. Una inspección visual de los tres paneles sugiere que el decaimiento de la cola derecha de los datos simulados no es tan disperso como el que ajustan los modelo Dagum y Pareto, y como el de los datos empíricos de la sección 8.2.

Por su parte, en las figura 8.13 se presentan los resultados para para la prueba de plausibilidad por región de la distribución del ingreso, medida de error, modelo de probabilidad supuesto y escenario del MSBA. En el total de la distribución los resultados indican que, mediante todos los criterios de error analizados, el modelo Gamma obtiene la suficiente evidencia estadística par afirmar que los datos del modelo MSBA pueden ser generados por éste modelo en 13 de los 18 escenarios. Mientras que, los modelos Dagum y Log normal, obtienen respaldo en la prueba de plausibilidad, implementada por el criterio  $KL$ , en el 61% y 22% de las muestras, respectivamente. Por su parte en la sección de ingresos medios y bajos las muestras que pasaron la prueba de plausibilidad, bajo el criterio  $KL$ , son 3 suponiendo el modelo Log normal y 7 en el modelo Dagum. Para la cola derecha de las distribuciones, el modelo de Pareto genera un *valor p* lo suficientemente grande para no descartarlo, mediante el criterio  $SCE$ , en 15 de las 18 distribuciones.

Para identificar si los datos generados por el MSBA replican a los modelos estimados de los datos empíricos, se analiza si los intervalos de los valores de los coeficientes de los modelos plausibles ajustados a los datos simulados contienen a los intervalos afines provenientes de los modelos plausibles de los datos empíricos. Esto se propone con el fin de saber si al calibrar los parámetros del *MSBA* se pueda llegar a una distribución que se acercarse a los modelos provenientes de los datos

empíricos<sup>6</sup>. La tabla 8.7 presenta los resultados del enfoque recién descrito <sup>7</sup>.

Tabla 8.7: Intervalos para las estimaciones según datos empíricos y simulados.

Región	Modelo	Datos	Criterio	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	Gini
Alto	Pareto	D			[1.7, 2.4]	[18,014, 95,902]*		[0.3, 0.4]
		S	SCE		[2.9, 6.8]	[29,033, 73,633]*		[0.1, 0.2]
Bajo	Log normal	D		[1.2, 1.4]			[8.7, 9.0]	[0.5, 0.5]
		S	KL	[1.8, 2.7]			[7.8, 8.5]	[0.8, 0.9]
	Dagum	D		[2.5, 8.2]*	[0.1, 0.4]*	[14,801, 23,488]		[0.4, 0.5]*
		S	KL	[3.3, 5.5]*	[0.1, 0.2]*	[23,765, 44,710]		[0.4, 0.6]*
Total	Gamma	D			[0.8, 0.8]**	[18,701, 19,138]**		[0.5, 0.6]**
		S	KS-SCE-KL		[0.4, 2.2]**	[6,620, 36,628]**		[0.4, 0.7]**
	Log normal	D		[1.2, 1.5]			[8.8, 9.2]	[0.6, 0.7]
		S	KL	[1.7, 2.7]			[7.9, 8.7]	[0.8, 0.9]
Dagum	D		[2.0, 2.7]*	[0.3, 0.5]*	[14,148, 21,921]*		[0.5, 0.6]**	
	S	KL	[2.2, 3.3]*	[0.1, 0.4]*	[20,360, 41,427]*		[0.4, 0.7]**	

Tabla 8.7: Intervalos para las estimaciones según datos empíricos y simulados conforme región de la distribución y modelo plausible. Se presenta la siguiente información:  $a/\sigma$  es el parámetro de forma  $a$  para las distribuciones Dagum y el parámetro  $\sigma$  para el modelo Log normal;  $p/q$  es el parámetro de forma  $p$  para el modelo Dagum y Gamma;  $\beta$  es el parámetro de escala;  $\mu$  es el parámetro de localización para el modelo Log normal; y por último, *Gini* representan el índice de Gini estimado de manera paramétrica. Los intervalos marcados con \*/\*\* indican si el intervalo generado por datos simulados incluyó manera parcial/total al intervalo estimado de los datos empíricos.

Conforme a los resultados de la tabla 8.7 se pueden extraer los siguientes puntos: para el total de la distribución se encuentra que el conjunto de coeficientes e índices de Gini estimados para el modelo Dagum, con base en los datos simuladas, no

<sup>6</sup>Se realizaron pruebas Kolmogorov-Smirnov, por región de la distribución, con la hipótesis nula de que el modelo de probabilidad que generó a los datos de la simulación es alguno de los modelos que resultaron como hipótesis plausible de los datos empíricos; para todas las distribuciones generados por el MSBA se rechaza la hipótesis nula y se confirma que ninguna de las distribuciones replican de manera exacta a los modelos basados en los datos empíricos.

<sup>7</sup>En la tabla 8.7 de la sección D.2 del apéndice D se pueden encontrar las estimaciones de los modelos de probabilidad para la simulación de los MSBA.

contiene en su totalidad al conjunto de parámetros que se desprendieron del análisis de los datos empíricos, es decir, las estimaciones con datos simulados arrojan coeficientes de forma  $a$  y  $b$  mayores que los mínimos observados en la estimación sobre datos empíricos, y coeficientes de forma  $p$ , menores que los estimados por los datos empíricos. Para el modelo Log normal, los intervalos de los coeficientes estimados por los datos empíricos y por los datos simulados no contienen ningún elemento en común. En contraste, los intervalos de los coeficientes estimados del ajuste de los datos simulados a la distribución Gamma incluyen en su totalidad a los intervalos provenientes de los datos empíricos.

Por su parte, los modelos Dagum y Log normal, ajustados en la región de ingresos bajos y medios, obtienen resultados similares al ajuste descrito para la región total de la distribución; la estimación del modelo Dagum con respecto a los datos simulados logra contener de manera parcial al conjunto de coeficientes estimados de los datos empíricos, por el contrario, ninguno de los modelos Log normal ajustados a los datos simulados logra acercarse a los modelos Log normal estimados por los datos empíricos.

Para finalizar, los intervalos de los coeficientes de los modelos de Pareto ajustados al MSBA no logran abarcar a los intervalos de los modelos de Pareto estimados para los datos empíricos, en particular, los coeficientes de Pareto estimados para los datos empíricos son más pequeños que los estimados para los datos simulados, que significaría que las colas derechas de las distribuciones generadas por la simulación son menos anchas que las presentadas en los datos empíricos lo cual confirma el análisis visual del CFDA.

En resumen, se pone en evidencia que el MSBA de la teoría del excedente de la estratificación social tiene potencial para usarse como base para modelar política económica con el fin de cambiar la distribución del ingreso de los agentes en México. En específico, se muestra que el modelo Gamma es útil para caracterizar a la totalidad de la distribución del ingreso de los agentes económicos en México para

años recientes, años 2012 y 2014. Sin embargo, si se deseará un modelo que incorpore información sobre las variables, que a través de los años pudieran cambiar a la distribución del ingreso de los agentes económicos de México, se recomienda suponer que los datos de la simulación, para las regiones total y de ingresos medios y bajos de la distribución, provienen del modelo Dagum, y en consecuencia, se debe calibrar los parámetros del MSBA para obtener un modelo base de agentes que replique, en promedio, a las condiciones observadas en México a partir de 1984. Por último, si lo que se desea es poner especial énfasis en el estudio de la cola derecha de la distribución se recomienda utilizar otro tipo de modelos de simulación de agentes, como lo son aquellos que se basan en la teoría cinética, que se enfocan en generar distribuciones con colas derechas de tipo leyes de potencia y para la cola izquierda modelos tipo log normal y exponencial.

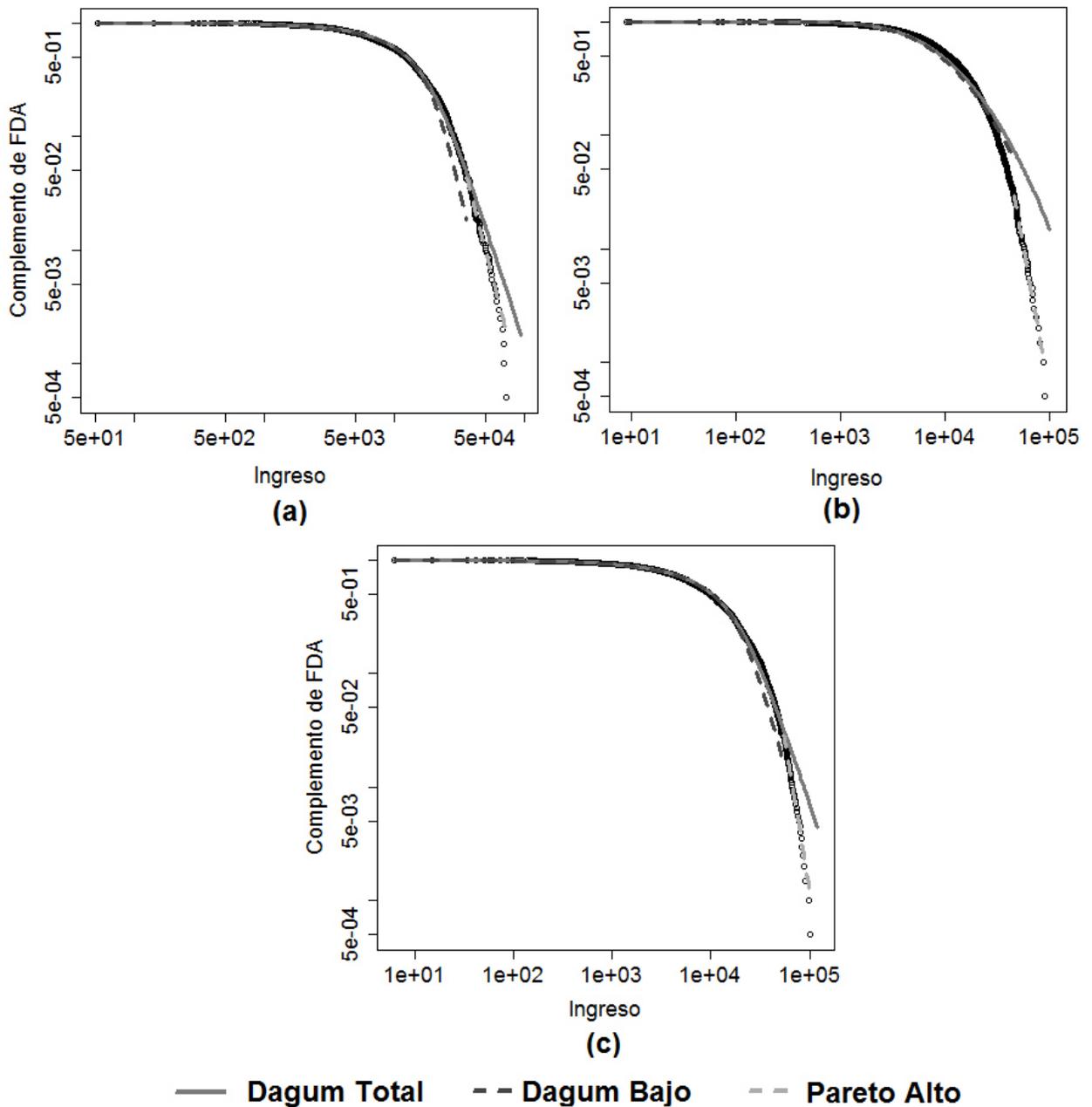


Figura 8.12: Complemento de la función de distribución acumulada de los ingresos simulados para los escenarios de nivel de desarrollo social medio,  $L = 3$ , con variación en el parámetro de concentración de poder de acuerdo al grado de ingreso  $\delta = 0.4, 0.5, 0.6$ , así como el ajuste que tuvieron sobre el modelo Dagum, para la región total y de ingresos bajo, y Pareto, para la región de ingresos altos. En el panel (a) se muestra el escenario ( $L = 3, \delta = 0.4$ ), en el panel (b) el escenario ( $L = 3, \delta = 0.5$ ), y en el panel (c) el escenario ( $L = 3, \delta = 0.6$ ).

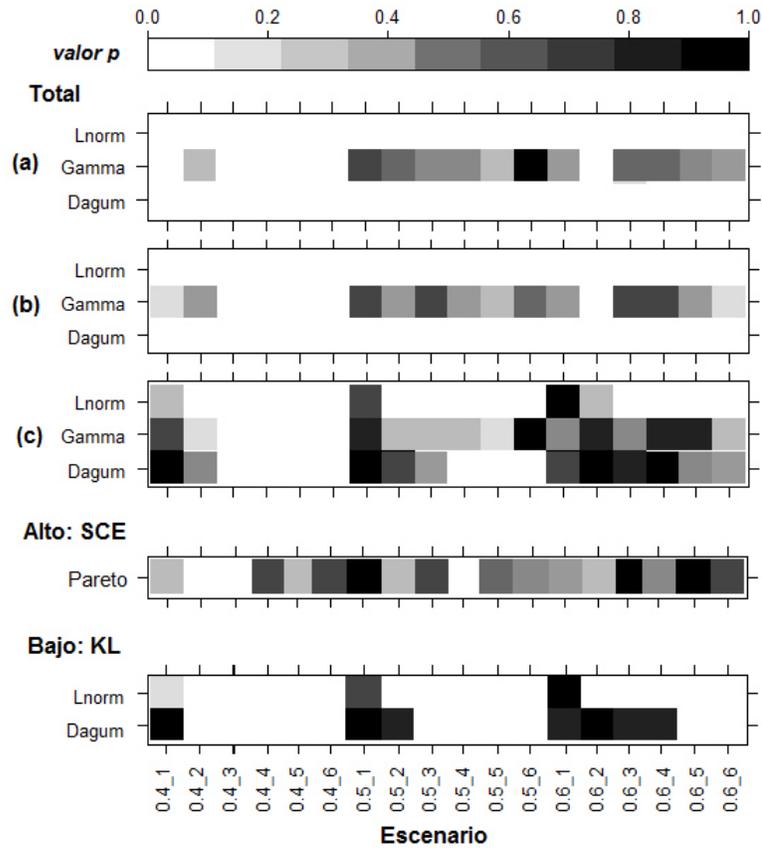


Figura 8.13: *Valores p* para la prueba de plausibilidad por región del ingreso, medida de error, modelo de probabilidad supuesto y escenario del MSA. En específico, para la región total de la distribución de ingresos en el panel (a) se muestran las estimaciones con base a la divergencia Kullback-Leibler, en el panel (b) de la suma del cuadrado del error y en el panel (c), la distancia Kolmogórov-Smirnov. En tanto que, para las regiones alta y baja se menciona el criterio con el que se obtuvo el *valor p*; no se presentan los criterios que generaron *valores p* iguales a cero.

## CAPÍTULO 9

# CONCLUSIONES.

---

Contar con un análisis confiable de la forma de la distribución del ingreso monetario en México, en un primer análisis, puede ayudar en las siguientes cuestiones. Por un lado se tiene una guía para identificar, de manera más eficiente, cuáles modelos estocásticos pueden ser los más adecuados para modelar el caso Mexicano; si es mas conveniente utilizar un modelo que considera un proceso multiplicativo o aditivo. Asimismo, una vez diseñado el modelo de simulación es de gran utilidad conocer los parámetros esperados de la distribución de equilibrio para calibrar el modelo, y así, tener más confiabilidad en la inferencia que se desprende de los resultados de la simulación. Y por otro lado, el conocer el tipo de proceso que genera una distribución puede dar una idea de las posibles consecuencias de una implementación en alguna política pública, en el cambio de la forma de la distribución así como en las variables macro económicas que pueden estar relacionadas con la distribución del ingreso.

El análisis empírico por región de la distribución presentado en este trabajo se enfoca a implementar una metodología más formal, que aquellas que se basan en el análisis visual, para identificar si un modelo de probabilidad es una hipótesis plausible para generar un conjunto de datos. En especial, se lleva a cabo un análisis no sólo para el rango total de la distribución sino que se diferencia el estudio de la región de ingresos bajos-medios y la región de ingreso altos, esto con el propósito, de determinar si la cola derecha de la distribución de ingresos sigue un modelo tipo ley

de Pareto, y si, el resto de la distribución sigue un proceso exponencial o log normal.

Entonces, los resultados arrojan, por un lado, que se podía suponer a ciertos modelos como hipótesis plausibles de los datos, y a su vez, si dos o mas modelos resultaban plausibles para un mismo año, y por consiguiente, se determino cuál de ellos se acercaba más a la distribución empírica observada.

En específico, los resultados son los que siguen:

- Modelos plausibles. Para la totalidad del ingreso se distingue a los modelos Dagum, Singh-Maddala, Log normal y Gamma; para la región de ingresos bajos y medios es plausible suponer que los datos provienen de los modelos Dagum, Singh-Maddala o Log normal; en tanto que, para la cola derecha de la distribución el modelo de Pareto y Log normal son hipótesis plausibles de los datos observados.
- Mejor ajuste. En la región total los modelos que mejor caracterizan a los datos, considerando el número de parámetros, son los modelos Dagum y Log normal, mientras, que el modelo Dagum es el que presentó mejor ajuste entre todos. De igual manera, en al región de ingresos bajos y medios, el modelo Dagum resulta con mejor ajuste que el modelo Log normal. Por último, el modelo de Pareto obtiene mejor ajuste para la región de ingresos altos.

Por consiguiente, estos resultados ofrecen una guía para seleccionar un modelo estocastico que sea el que mejor caracterice al caso mexicano. Si por ejemplo, se decidiera que la manera más adecuada de representar al ingreso en México es a través de un modelo log normal en la parte baja de la distribución, y un modelo tipo Pareto en la parte alta, entonces, el plausible proponer un modelo multiplicativo, como el de la ley Gibrat, con un efecto limite, en donde se agrega la restricción de que el ingreso no puede ser menor a cierto umbral, con la intención de derivar una ley de potencia para la parte baja de la distribución.

Por otra parte, siguiendo el supuesto de que el ingreso de los agentes económicos

en México sigue un modelo Log normal y Pareto, en la región de ingresos bajos-medios y altos, respectivamente, podría inferirse de manera preliminar cómo sería la afectación de una política pública a la forma del ingreso de la población en ambas regiones. Por ejemplo, en la región baja podría producir un cambio proporcional, sin importar el nivel inicial del ingreso de los agentes y tampoco los cambios de otros factores que tengan alguna relación con el ingreso, y en consecuencia, la incidencia de la política pública no generará grandes cambios en la forma de la distribución ni en las variables macro económicas con las que esté relacionada. Por otro lado, un cambio, aunque sea mínimo, en la zona de ingresos altos, se traducirá en grandes cambios en el ingreso, así como, en los factores económicos que sean dependientes de la relación como lo son la efectividad de la política impositiva, el crecimiento y desarrollo económico, el alcance a los mercados de capital, la estabilidad social, entre otros. Sin embargo, cabe señalar que para saber exactamente cuáles son las consecuencias se debe estudiar la manera en que están relacionadas las variables macro económicas con la forma de la desigualdad.

Asimismo, este análisis se abre a la discusión de la práctica de desdeñar los datos extremos observados en la distribución del ingreso. Por un lado, cuando se realiza esa práctica es posible llegar a utilizar estadística tradicional y de obtener relaciones promedio acordes con teoría económica. Empero, por otro lado, dejar de lado estos datos significa pasar por alto una fuente de información, que aunque sea compleja y difícil de estudiar, puede ayudar a comprender más a fondo la importancia de un fenómeno como lo es la desigualdad en la distribución del ingreso.

Por último, gracias a la implementación del *MSBA* se pone énfasis en la necesidad de ser más rígido a la hora de evaluar, si el modelo que se propone, puede ser utilizado para idear políticas públicas destinadas a una región. Si bien el modelo puede ser una hipótesis plausible para generar los datos observados, se debe dar un paso atrás para reflexionar si el principio estocástico en el que se basa el modelo es el mejor para caracterizar a la realidad que se pretende estudiar, y de manera similar, tratar de resolver si en realidad el modelo es capaz de replicar una distribución tal

cual se observa en la realidad.

## 9.1 CONTRIBUCIONES

- Implementación computacional de una metodología que ayuda a identificar el modelo de probabilidad del que provienen datos empíricos que provienen de una muestra agrupada a través de un ponderador: (i) prueba de plausibilidad para comprobar si un modelo de probabilidad genera un conjunto de datos observados; (ii) prueba de bondad de ajuste Kolmogorov-Smirnov; (iii) prueba de bondad de ajuste de Vuong; (iv) estimación del valor de límite inferior óptimo que la variable en estudio si se supone que se distribuye conforme a cierto modelo de probabilidad.
- Se comprueba que los modelos Dagum, Log normal y Gamma son hipótesis plausibles para describir el proceso que genera al total de región de la distribución de ingreso, de manera similar, para la región de ingresos medios y bajos los modelos que resultaron plausibles fueron el Dagum y Log normal, por su parte, el modelo de Pareto y Log normal, en menor medida, resultaron plausible para caracterizar la región de ingresos altos de la distribución de ingresos.
- Además, de entre los modelos plausibles, se identifica cuales modelos obtuvieron mejor ajuste con respecto a los datos empíricos. Los mejores modelos por número de parámetros y región de la distribución de los ingresos en México para el periodo 1992-2014, son los que siguen: los modelos Dagum y Log normal, para el total de la distribución y región de ingresos medios y bajos, y el modelo de Pareto para la cola derecha de la distribución.
- Se presenta un modelo de simulación de agentes, con base en la teoría cinética de gases, que incluye dos parámetros, los cuales deciden el monto de intercambio entre los agentes y que tienen una interpretación económica.

- Se muestra que el *MSBA* tiende a generar distribuciones más inequitativas para las combinaciones en las que existe mayor probabilidad de perder un porcentaje de ingreso alto –sociedades con bajo desarrollo- y donde la probabilidad de que el agente que gane el encuentro sea el de mayor ingreso –sociedad dominada por las clases ricas-.
- Se comprueba que, para algunas de las distribuciones que genera el *MSBA*, los modelos plausibles para generar dichos datos son los que siguen: para el total de la distribución los modelos Dagum, Gamma y Log normal –en menor medida-; para la región de ingresos bajos los modelos Log normal y Dagum; mientras que, para la región alta de la distribución se puede emplear el modelo Pareto.
- Se concluye que, el *MSBA* tiene potencial para replicar, la forma de la distribución total del ingreso en México, en especial, si se supone que el modelo Gamma es el proceso que genera los datos en México, como se observa para la muestra del 2012 y 2014.

## 9.2 TRABAJO FUTURO.

De acuerdo a los resultados que se obtuvieron en esta investigación, el desarrollo del estudio de la formación de la distribución del ingreso en México que englobe el punto de vista tanto de la teoría económica como de física puede seguir las siguientes líneas:

- Afinar la prueba de plausibilidad para datos muestrales que considere la probabilidad de observar una muestra en la cual los datos que pertenecen a los dos regiones de la distribución se comportan de acuerdo a dos modelos de probabilidad diferentes.
- A partir de un método heurístico calibrar los parámetros del *MSBA* con el

objetivo de que el modelo sea capaz de replicar a los modelos de probabilidad que resultaron generar a los datos en México.

- Desarrollar un *MSBA* que sea más afín los resultados presentados en el análisis empírico de la distribución del ingreso. Por ejemplo, si se considera los modelos que resultaron plausibles y con mayor parsimonia, entonces, el *MSBA* que se desarrolle debe ser capaz de generar, para la región de ingresos bajos y medios, un proceso multiplicativo tipo Log normal y, para la región de ingresos altos, una distribución tipo Pareto.
- Caracterizar en el *MSBA* base, a través de reglas sencillas, algunas de las teorías económicas que explican la formación de la desigualdad en la distribución del ingreso. Estas teorías son: las imperfecciones en el mercado de capital, desigualdad en la distribución de la educación, apertura comercial, mercado laboral.
- Desarrollar un modelo en el que puedan hacer sinergia un modelo económico con el *MSBA*.
- Estudiar la relación de la forma de la distribución del ingreso empírico con eventos económicos que ocurrieron en el periodo de estudio, como lo fueron, las crisis económicas de 1994-95 y 2008, la apertura comercial a través del TLC, el cambio de partido político en la presidencia. Con base en estos resultados probar la capacidad de predicción del *MSBA* sobre cambios en la distribución del ingreso por producto de un cambio en la economía.
- Definir si existe una distribución estable en el periodo analizado

## APÉNDICE A

# REVISIÓN BIBLIOGRÁFICA

---

Para las regiones total, alta y baja de la distribución del ingreso, a continuación se presentan tres tablas con información bibliográfica acerca del ajuste de modelos de probabilidad en datos empíricos. Las columnas de las tablas muestran el autor del estudio «Autor», la región del mundo «Región» y el periodo de tiempo que se estudió «Per», así como la definición de la variable que se trabajó «Var», el método con que se estimó el modelo «Met», los modelos de probabilidad que se estudiaron «Mod», el criterio de selección del mejor modelo «Criterio», el mejor modelo que ajustó al estudio «Mej», y por último, los coeficientes estimados «Coeficientes».

Las claves para la columna «Var»: IF, ingreso del hogar; IF (B), ingreso del hogar bruto proveniente de cualquier trabajo; IF (L), ingreso del hogar libre de impuestos proveniente de cualquier trabajo; IF (2), ingreso del hogar (dos personas) y por persona; RP, riqueza personal; RP (L), riqueza personal libre de impuestos; IP, ingreso personal; IP (H), ingreso personal de sueldos y salarios de hombres, condicional a la escolaridad y la experiencia; IP (J), ingreso del jefe del hogar por sueldos y salarios condicional a la escolaridad y la experiencia; IP (F), ingreso personal del hogar. Claves para la columna «Met»: MCO, mínimos cuadrados ordinarios; MV, máxima verosimilitud; BAY, estimación Bayesiana.

Claves para la columna «Mod»: FBG, familia Beta Generalizada; E, Exponen-

cial; MG, mezcla de Gammas; LN, Log normal; P, Pareto; G, Gamma; W, Weibull; D, Dagum; SM, Singh-Maddala; PD, Pareto Doble; BG II, Beta Generalizada II; PDLN Pareto doble Log normal; B II, Beta II; EE, Estiramiento exponencial; PCE, Pareto con corte exponencial.

Claves para la columna «Crit»: LR, prueba de razón de verosimilitud; Log L, logaritmo de verosimilitud; M, comparación media muestral; SCE, suma del cuadrado del error; SEA, suma del error absoluto; RMCE, Raíz de la media del cuadrado del error; KS, prueba Kolmogorov Smirnov; AD, prueba Anderson-Darling; [19], metodología desarrollada por [19]; G-G, similitud entre índice de Gini y de Gilbrat; RG, razón de concentración de Gini; G, comparación índice de Gini muestral; AIC, criterio de información de Aikaike; CL, criterio de concentración de Lorenz. Los coeficientes  $\alpha$  y  $\beta$  se refieren a los parámetros de forma y al índice de Gibrat.

Autor	Región	Per	Var	Met	Mod	Crit	Mej	Coefficientes
[6]	23 países	datos quin- quenales a partir de 1970	IF (B)	MV	FBG	LR, Log L, M, SCE, SEA	(i) W en el 62 %, D 84 % y GB II en el 100 %	LN: $\beta \in [0.34, 1.49]$ ; G: $\alpha \in [0.7, 4.7]$ ; D: $\alpha_a \in [2, 7.8]$ , $\alpha_p \in [0.08, 0.93]$ , SM: $\alpha_a \in [0.9, 3.2]$ , $\alpha_q \in [1, 1E6]$
[6]	México	1984, 1989, 1992, 1994, 1996	IF (B)	MV	FBG	LR, Log L, M, SCE, SEA	(i) LN en 60 %, Dagum en 80 % y GB II en 100 %	LN: $\beta \in [0.65, 0.76]$ ; G: $\alpha \in [1.1, 1.4]$ ; D: $\alpha_a \in [2, 2.5]$ , $\alpha_p \in [0.52, 0.72]$ ; SM: $\alpha_a \in [1.3, 1.6]$ , $\alpha_q \in [1.4, 2.63]$
[18]	Canada	1979	IF (B)	BAY	MG, LN, B II, SM, D	RMCE, KS	MG y SM	
[24]	USA, Argentina, Sri-Lanka, Canada	1959- 1973	IF	MCO	LN, G, SM y D	SCE, RG	Para USA, años 1969 y 1970 D	D: $\alpha_a \in [2.1, 4.3]$ , $\alpha_p \in [0.34, 0.37]$
[29]	USA	1996 y 1993	IF (2) y IP	MCO	E	Visual	E	

[60]	USA	1970, 1975 y 1980	IF	MV	FBG	LR, Log L, M, SCE, SEA	SM, BG	LN: $\beta\epsilon[0.91, 1.02]$ ; G: $\alpha\epsilon[1.56, 1.6]$ ; SM: $\alpha_a\epsilon[1.69, 1.96]$ , $\alpha_q\epsilon[2.93, 8.36]$
[68]	189 países	1976- 2006	IF	MCO	LN, G, W	G	LN	
[73]	México	1998, 2002, 2008	NA	BAY	P, LN, D	Aik	D	D: $\alpha_a\epsilon[2]$ , $\alpha_p\epsilon[0.29, 0.35]$
[74]	USA	1960- 1969	IP (F)	MV	G, LN	CL	G	G: $\alpha\epsilon[1.9, 2.5]$
[82]	USA	1960- 1972	IF	MCO	LN, G, SM		SM	SM: $\alpha_a\epsilon[1.97, 2.1]$ , $\alpha_q\epsilon[2.5, 5.6]$
[90]	USA	2000- 2009	IP (J)	MV	E, G, LN, W, PD, BG II, PDLN	AIC	DPLN	P: $\alpha_{alto}\epsilon[2.5]$ , $\alpha_{bajo}\epsilon[1]$
[90]	USA	1968- 1993		MV	E, G, LN, W, PD, BG II, PDLN	AIC	DPLN	P: $\alpha_{alto}\epsilon[3, 4]$ , $\alpha_{bajo}\epsilon[1.5, 2.5]$

Tabla A.1: Resumen de referencias bibliográficas para el totalidad de la distribución del ingreso

Autor	Región	Per	Var	Met	Mod	Crit	Mej	Coefficientes
[11]	Mundo	1996- 2002	IP	MV	P, E, LN, EE, PCE	[19]	(i) P en 28%	P $\bar{\alpha}$ : 1.5
[11]	USA	1988- 2012	IP	MV	P, E, LN, EE, PCE	[19]	(i) P en 29%	P $\bar{\alpha}$ : 1.4
[11]	China	2006- 2012	IP	MV	P, E, LN, EE, PCE	[19]	(i) P en 83%	P $\bar{\alpha}$ : 1.7

[11]	Rusia	2004-2011	IP	MV	P, E, LN, EE, PCE	[19]	(i) P en 25 %	P $\bar{\alpha}$ : 0.9
[19]	USA	2003	RP (L)	MV	P, E, LN, EE, PCE	[19]	Se rechaza hipótesis	P $\alpha$ : 1.3
[21]	Italia	1977-2002	IP	MCO	P	Visual	P	P: $\alpha \in [1.09, 2.3]$
[20]	USA, UK y Alemania	1980-2002	IF (L)	MCO	P	P	P: $\alpha \in [0.01, 4.76]$	
[26]	Australia	1993-1997			P		P	P: $\alpha \in [1.2, 2.6]$
[30]	UK	1996			P	Visual	P	P: $\alpha \in [.9]$
[30]	UK	1994-1998	IP	MCO	P	Visual	P	P: $\alpha \in [1, 1.3]$
[30]	USA	1998	IP	MCO	P	Visual	P	P: $\alpha \in [.8, 1.2]$
[35]	Japón	1987-2000	IP	MCO	P	Visual	P	P: $\alpha \in [1.8, 2.1]$
[83]	India	2002-2004	RP (L)	MCO	P		P	P: $\alpha \in [0.8, 0.9]$
[83]	India	1997	IF	MCO	P		P	P: $\alpha \in [1.51]$
[87]	Japón	1900-2000	IF (L)	NA	P		Pareto	P: $\alpha \in [1.5, 2.5]$
[90]	USA	1968-1993	IP (H)	MV	P	pruebas KS y AD	(i) P en 90 %	P: $\alpha \in [3, 4]$
[90]	USA	2000-2009	IP (J)	MV	pruebas KS y AD P	[19]	(i) P en 100 %	P: $\alpha \in [1.5, 2.5]$
[95]	USA	1983-2001	IP	NA	P	Visual	P	P: $\alpha \in [1.4, 1.8]$

Tabla A.2: Resumen de referencias bibliográficas para la región alta de la distribución del ingreso

Autor	Región	Per	Var	Met	Mod	Crit	Mej	Coefficientes
[21]	Italia	1977-2002	IP	MV	LN	Visual	LN	LN: $\beta \in [2.08, 2.7]$
[20]	USA, UK y Alemania	1980-2002	IF (L)	MCO	LN		LN	LN: $\beta \in [1, 2.73]$
[26]	Australia	1993-1997	IF	ND	LN		LN	LN: $\beta \in [1.17, 1.41]$

[30]	UK	1996	RP	MCO	E	Visual	E	
[30]	UK	1994-1998	IP	MCO	E	Visual	E	
[30]	USA	1998	IP	MCO	E	Visual	E	
[83]	India	1997	IF	MCO	LN		LN	
[87]	Japón	1950-2000	IF (L)		LN		LN	LN: $\beta \in [2.2, 3]$
[95]	USA	1983-2001	IP		E	Visual	E	

Tabla A.3: Resumen de referencias bibliográficas para la región baja de la distribución del ingreso

## APÉNDICE B

# INFORMACIÓN ESTADÍSTICA PARA MÉXICO

---

## B.1 ENCUESTA INGRESO GASTO DE LOS HOGARES EN MÉXICO

La ENIGH es una fuente de información sobre los ingresos y egresos de los hogares mexicanos que puede encontrar como primer antecedente, en 1956, a la encuesta de Ingresos y Egresos de la Población en México por parte de la Dirección General de Estadística. Sin embargo, a partir de 1984, se desarrolló Encuesta Nacional de Ingreso-Gasto de los Hogares que buscó homogeneizar la metodología, tomar en cuenta recomendaciones internacionales y necesidades de diferentes usuarios.

Además de proporcionar datos de calidad y desagregación sobre los ingresos y egresos de la población mexicana, la ENIGH recaba información sobre «las características ocupacionales y sociodemográficas de los integrantes del hogar, así como de las características de la infraestructura de la vivienda y el equipamiento del hogar» [50]. Enseguida se desglosa la cobertura temática la ENIGH.

- Ingreso corriente total (monetario y no monetario) de los hogares.

- Gasto corriente total (monetario y no monetario) de los hogares.
- Percepciones financieras y de capital de los hogares y sus integrantes.
- Erogaciones financieras y de capital de los hogares.
- Características de la vivienda.
- Residentes e identificación de hogares en la vivienda.
- Características sociodemográficas de los residentes de la vivienda.
- Condición de actividad y características ocupacionales de los integrantes del hogar de 12 y más años.
- Equipamiento del hogar, servicios.

El marco muestral la encuesta utiliza la información demográfica y cartográfica proveniente del Censo de Población y Vivienda del 2000, y su unidad de muestreo y de observación es la vivienda y el hogar respectivamente. Asimismo, el esquema de muestreo es probabilístico y el diseño es estratificado, bietápico y por conglomerados. A partir de este diseño los resultados obtenidos de la ENIGH pueden generalizarse hacia el total de la población donde la unidad de análisis es el hogar, la vivienda y los integrantes del hogar.

La encuesta está diseñada para tener una cobertura geográfica a nivel nacional con un desglose geográfico para áreas rurales (localidades con menos de 2,500 habitantes) y áreas urbanas (localidades de 2,500 habitantes y más). Dado que la ENIGH es una encuesta muestral se cuenta con un ponderador factor de expansión que se interpreta como «la cantidad de viviendas (hogares o personas) en la población que representa una vivienda en la muestra» [50]. LA ENIGH cuenta con dos ponderadores, uno asociado a la vivienda y otro al hogar. Entonces, para obtener la información representativa nivel poblacional se procedió multiplicar la variable en estudio por el factor de expansión de hogares para cada registro, quedando así una variable representativa, a nivel nacional, por integrante del hogar en México.

Para la realización de este trabajo se tomaron en consideración claves referentes al ingreso corriente monetario. La ENIGH define al ingreso corriente como los recursos, ya sean monetarios o en especie, que el agente económico puede obtener de manera regular, que están disponibles para el consumo de bienes y servicios, y no incluye montos que modifican el patrimonio neto de los activos y pasivos del hogar. Por su parte, el ingreso corriente monetario está constituido por percepciones provenientes del pago al trabajo o por participación en alguna actividad productiva, indemnizaciones y transferencias.

De ahí que, las claves seleccionadas para la construcción de la variable ingreso son las que refieren a ingresos monetarios por trabajo subordinado, trabajo independiente y por posesión de activos físicos y no físicos. Cabe señalar que la codificación de las muestras ha cambiado en el tiempo por lo que las tablas que continuación se presentan agrupan a las muestras con igual codificación.

<i>Tipo</i>	<b>Claves de ENIGH que constituyen la variable ingreso</b>
<i>Años 1984 y 1989</i>	
	<b>Ingresos netos por remuneraciones al trabajo</b>
<b>Trabajo sub ordinado</b>	P001 Sueldos, salarios y horas extras
	P002 Comisiones y propinas
	P003 Aguinaldo, gratificaciones y premios
	P004 Primas vacacionales
	P005 Reparto de utilidades
	<b>Ingresos netos de negocios propios</b>
<b>Trabajo independiente</b>	P008 Negocios industriales
	P009 Negocios comerciales
	P010 Prestación de servicios
	P011 Procesamiento de materias primas
	P012 Producción agrícolas
	P013 Producción pecuarios
	P022 Ingresos provenientes de cooperativas de producción
	<b>Ingresos netos por renta de la propiedad</b>
<b>Posesión de activos físicos y no físicos</b>	P015 Alquiler de tierras y terrenos
	P016 Alquiler de casas, edificios, locales y otros
	P017 Intereses provenientes de inversiones a plazo fijo
	P018 Intereses provenientes de cuentas de ahorro
	P019 Intereses provenientes de préstamos a terceros
	P020 Intereses provenientes de acciones, bonos y cédulas
	P021 Alquiler de marcas, patentes y derechos de autor

<i>Año 1992</i>	
	<b>Ingresos netos por remuneraciones al trabajo</b>
<b>Trabajo sub or- dinado</b>	P001 Sueldos, salarios y horas extras
	P002 Comisiones y propinas
	P003 Aguinaldo, gratificaciones y premios
	P004 Primas vacacionales
	P005 Reparto de utilidades
	P006 Ingresos provenientes de cooperativas de producción
	<b>Ingresos netos de negocios propios</b>
<b>Trabajo inde- pen- diente</b>	P007 Negocios industriales
	P008 Negocios comerciales
	P009 Prestación de servicios
	P010 Procesamiento de materias primas
	P011 Producción agrícola
	P013 Producción pecuaria y forestal
	<b>Ingresos netos por renta de la propiedad</b>
<b>Posesión de ac- tivos físicos y no físicos</b>	P015 Alquiler de tierras y terrenos
	P016 Alquiler de casas, edificios, locales y otros
	P017 Intereses provenientes de inversiones a plazo fijo
	P018 Intereses provenientes de cuentas de ahorro
	P019 Intereses provenientes de préstamos a terceros
	P020 Intereses provenientes de acciones, bonos y cédulas
	P021 Alquiler de marcas, patentes y derechos de autor
<i>Años 1994 y 1996</i>	
	<b>Ingresos netos por remuneraciones al trabajo</b>
<b>Trabajo sub or- dinado</b>	P001 Sueldos, salarios, jornal y horas extras
	P002 Comisiones, propinas y destajo
	P003 Aguinaldo, gratificaciones, premios y recompensas adicionales
	P004 Primas vacacionales y otras prestaciones en efectivo
	P005 Reparto de utilidades
	<b>Ingresos netos de negocios propios</b>
<b>Trabajo inde- pen- diente</b>	P006 Negocios industriales
	P007 Negocios comerciales
	P008 Prestación de servicios
	P009 Producción agrícola
	P010 Producción pecuaria y derivados
	P011 Producción forestal
	P012 Recolección de flora y fauna
	P013 Acuicultura y pesca
	P014 Sueldos o salarios
	P015 Ganancias o utilidades
	<b>Ingresos netos por renta de la propiedad</b>
<b>Posesión de ac- tivos físicos y no físicos</b>	P016 Alquiler de tierras y terrenos
	P017 Alquiler de casa, edificios, locales y otros inmuebles

- P018 Intereses provenientes de inversiones a plazo fijo
- P019 Intereses provenientes de cuentas de ahorro
- P020 Intereses provenientes de préstamos a terceros
- P021 Intereses provenientes de acciones, bonos y cédulas
- P022 Alquiler de marcas, patentes y derechos de autor

*Años 1998, 2000 y 2002*

	<b>Ingresos netos por remuneraciones al trabajo</b>
	P001 Sueldos, salarios o jornal
	P002 Destajo
<b>Trabajo sub or- dinado</b>	P003 Comisiones y propinas
	P004 Horas extras
	P005 Crédito al salario
	P006 Incentivos, gratificaciones o premios
	P007 Aguinaldo, bono, percepción adicional o sobresueldo
	P008 Primas vacacionales y otras prestaciones en efectivo
	P009 Reparto de utilidades
	<b>Ingresos netos de negocios propios</b>
	P010 Negocios industriales
	P011 Negocios comerciales
<b>Trabajo inde- pen- diente</b>	P012 Prestación de servicios
	P013 Producción agrícola
	P014 Producción pecuaria y derivados
	P015 Producción forestal
	P016 Recolección de flora, productos forestales y caza
	P017 Acuicultura y pesca
	P018 Sueldos o salarios
	P019 Ganancias o utilidades
<b>Posesión de ac- tivos físicos y no físicos</b>	P020 Alquiler de tierras y terrenos
	P021 Alquiler de casas, edificios, locales y otros inmuebles
	P022 Intereses provenientes de inversiones a plazo fijo
	P023 Intereses provenientes de cuentas de ahorro
	P024 Intereses provenientes de préstamos a terceros
	P025 Intereses provenientes de acciones, bonos y cédulas
	P026 Alquiler de marcas, patentes y derechos de autor
	P027 Otros ingresos por renta de la propiedad no considerados en los anteriores

*Años 2004, 2005 y 2006*

	<b>Ingresos por remuneraciones al trabajo asalariado</b>
	P001 Sueldos, salarios o jornal
	P002 Destajo
<b>Trabajo sub or- dinado</b>	P003 Comisiones y propinas
	P004 Horas extras
	P005 Aguinaldo
	P006 Incentivos, gratificaciones o premios
	P007 Bono, percepción adicional o sobresueldo

	P008 Primas vacacionales y otras prestaciones en efectivo
	P009 Reparto de utilidades
<hr/>	
	<b>Ingresos netos de negocios propios</b>
	P010 Negocios industriales
	P011 Negocios comerciales
	P012 Negocios de prestación de servicios
	P013 Negocios agrícolas
	P014 Negocios de cría, explotación y productos derivados de animales
	P015 Negocios de reproducción, recolección de productos forestales y tala de árboles
	P016 Negocios de pesca, caza y captura de animales
	<b>Ingresos por cooperativas</b>
	P017 Sueldos o salarios
	P018 Ganancias o utilidades
	<b>Ingresos por sociedades</b>
	P019 Sueldos, salarios o jornal
	P020 Destajo
	P021 Comisiones y propinas
	P022 Horas extras
	P023 Aguinaldo
<b>Trabajo</b>	P024 Incentivos, gratificaciones o premios
<b>inde-</b>	P025 Bono, percepción adicional o sobresueldo
<b>pen-</b>	P026 Primas vacacionales y otras prestaciones en efectivo
<b>diente</b>	P027 Reparto de utilidades
	P028 Ganancias o utilidades
	<b>Ingresos de empresas que funcionan como sociedades</b>
	P029 Sueldos, salarios o jornal
	P030 Destajo
	P031 Comisiones y propinas
	P032 Horas extras
	P033 Aguinaldo
	P034 Incentivos, gratificaciones o premios
	P035 Bono, percepción adicional o sobresueldo
	P036 Primas vacacionales y otras prestaciones en efectivo
	P037 Reparto de utilidades
	P038 Ganancias o utilidades
<hr/>	
	<b>Ingresos netos por renta de la propiedad</b>
	P039 Alquiler de tierras y terrenos, dentro y fuera del país
<b>Posesión</b>	P040 Alquiler de casas, edificios, locales y otros inmuebles que están dentro del país
<b>de ac-</b>	P041 Alquiler de casas, edificios, locales y otros inmuebles que están fuera del país
<b>tivos</b>	P042 Intereses provenientes de inversiones a plazo fijo
<b>físicos</b>	P043 Intereses provenientes de cuentas de ahorro
<b>y no</b>	P044 Intereses provenientes de préstamos a terceros
<b>físicos</b>	P045 Rendimientos provenientes de acciones o dividendos, bonos y cédulas
	P046 Alquiler de marcas, patentes y derechos de autor
	P047 Otros ingresos por renta de la propiedad no considerados en los anteriores
<hr/>	

Año 2008

<b>Trabajo sub or- dinado</b>	<b>Ingresos del trabajo principal para subordinados</b>
	P001 Sueldos, salarios o jornal
	P002 Destajo
	P003 Comisiones y propinas
	P004 Horas extras
	P005 Incentivos, gratificaciones o premios
	P006 Bono, percepción adicional o sobresueldo
	P007 Primas vacacionales y otras prestaciones en efectivo
	P008 Reparto de utilidades del ejercicio 2007
	P009 Aguinaldo del ejercicio 2007
<b>Ingresos del trabajo secundario para subordinados</b>	
P018 Total de ingresos monetarios	
P019 Reparto de utilidades y aguinaldo	

<b>Trabajo inde- pen- diente</b>	<b>Ingresos del trabajo principal provenientes de las cooperativas, sociedades y empresas que funcionan como sociedades</b>
	P011 Sueldos o salarios
	P012 Ganancias/utilidades
	P013 Otros ingresos
	<b>Ingresos del trabajo secundario provenientes de las cooperativas, sociedades y empresas que funcionan como sociedades</b>
	P015 Sueldos o salarios
	P016 Ganancias/utilidades
	P017 Otros ingresos
	<b>Ingresos por trabajo de personas menores de 12 años</b>
	P063 Ingreso por trabajo de personas menores de 12 años
	<b>Ingresos por negocio propio</b>
	<i>Trabajo principal</i>
	P067 Por negocios con tipo de actividad industrial
	P068 Por negocios con tipo de actividad comercial
	P069 Por negocios prestadores de servicios
	P070 Por negocios con actividades agrícolas
	P071 Por negocios con actividades de cría y explotación de animales
	P072 Por negocios con actividades de recolección, reforestación y tala de árboles
	P073 Por negocios con actividades de pesca, caza y captura de animales
	<i>Trabajo secundario</i>
P074 Por negocios con tipo de actividad industrial	
P075 Por negocios con tipo de actividad comercial	
P076 Por negocios prestadores de servicios	
P077 Por negocios con actividades agrícolas	
P078 Por negocios con actividades de cría y explotación de animales	
P079 Por negocios con actividades de recolección, reforestación y tala de árboles	
P080 Por negocios con actividades de pesca, caza y captura de animales	

**Ingresos por renta de la propiedad**

**Posesión  
de ac-  
tivos  
físicos**

- P023 Alquiler de tierras y terrenos, dentro y fuera del país
- P024 Alquiler de casas, edificios, locales y otros inmuebles que están dentro del país
- P025 Alquiler de casas, edificios, locales y otros inmuebles que están fuera del país
- P026 Intereses provenientes de inversiones a plazo fijo
- P027 Intereses provenientes de cuentas de ahorro
- P028 Intereses provenientes de préstamos a terceros
- P029 Rendimientos provenientes de bonos o cédulas
- P030 Alquiler de marcas, patentes y derechos de autor
- P031 Otros ingresos por renta de la propiedad no considerados en los anteriores

---

*Año 2010*

---

**Ingresos del trabajo principal para subordinados**

**Trabajo  
sub or-  
dinado**

- P001 Sueldos, salarios o jornal
- P002 Destajo
- P003 Comisiones y propinas
- P004 Horas extras
- P005 Incentivos, gratificaciones o premios
- P006 Bono, percepción adicional o sobresueldo
- P007 Primas vacacionales y otras prestaciones en efectivo
- P008 Reparto de utilidades del ejercicio 2007
- P009 Aguinaldo del ejercicio 2009

**Ingresos del trabajo secundario para subordinados**

- P014 Monto recibido en el trabajo secundario
- P015 Reparto de utilidades del ejercicio 2009
- P016 Aguinaldo del ejercicio 2009

**Ingreso por trabajo de personas menores de 12 años**

- P067 Ingreso por trabajo de personas menores de 12 años
- 

**Ingresos del trabajo principal provenientes de las cooperativas, sociedades y empresas que funcionan como sociedades**

- P011 Sueldos o salarios
- P012 Ganancias/utilidades
- P013 Otros ingresos

**Ingresos del trabajo secundario provenientes de las cooperativas, sociedades y empresas que funcionan como sociedades**

- P018 Sueldos o salarios
- P019 Ganancias/utilidades
- P020 Otros

**Ingresos por negocio propio**

*Trabajo principal*

**Trabajo  
inde-  
pen-  
diente**

- P068 Por negocios con tipo de actividad industrial
- P069 Por negocios con tipo de actividad comercial
- P070 Por negocios prestadores de servicios
- P071 Por negocios con actividades agrícolas
- P072 Por negocios con actividades de cría y explotación de animales
- P073 Por negocios con actividades de recolección, reforestación y tala de arboles
- P074 Por negocios con actividades de pesca, caza y captura de animales

*Trabajo secundario*

- P075 Por negocios con tipo de actividad industrial
- P076 Por negocios con tipo de actividad comercial
- P077 Por negocios prestadores de servicios
- P078 Por negocios con actividades agrícolas
- P079 Por negocios con actividades de cría y explotación de animales
- P080 Por negocios con actividades de recolección, reforestación y tala de arboles
- P081 Por negocios con actividades de pesca, caza y captura de animales

---

**Ingresos por renta de la propiedad**

**Posesión  
de ac-  
tivos  
físicos  
y no  
físicos**

- P023 Alquiler de tierras y terrenos, dentro y fuera del país
- P024 Alquiler de casas, edificios, locales y otros inmuebles que están dentro del país
- P025 Alquiler de casas, edificios, locales y otros inmuebles que están fuera del país
- P026 Intereses provenientes de inversiones a plazo fijo
- P027 Intereses provenientes de cuentas de ahorro
- P028 Intereses provenientes de préstamos a terceros
- P029 Rendimientos provenientes de bonos o cédulas
- P030 Alquiler de marcas, patentes y derechos de autor
- P031 Otros ingresos por renta de la propiedad no considerados en los anteriores

---

*Años 2012 y 2014*

---

**Ingresos del trabajo principal para subordinados**

**Trabajo  
sub or-  
dinado**

- P001 Sueldos, salarios o jornal
- P002 Destajo
- P003 Comisiones y propinas
- P004 Horas extras
- P005 Incentivos, gratificaciones o premios
- P006 Bono, percepción adicional o sobresueldo
- P007 Primas vacacionales y otras prestaciones en dinero
- P008 Reparto de utilidades
- P009 Aguinaldo

**Ingresos del trabajo secundario para subordinados**

- P014 Monto recibido en el trabajo secundario
- P015 Reparto de utilidades
- P016 Aguinaldo

**Ingreso por trabajo de personas menores de 12 años**

- P067 Ingreso por trabajo de personas menores de 12 años

---

**Ingresos por negocios del hogar, trabajo principal**

- P011 Sueldos o salarios
- P012 Ganancias/utilidades
- P013 Otros ingresos

**Ingresos del trabajo secundario provenientes de las cooperativas, sociedades y empresas que funcionan como sociedades**

- P018 Sueldos o salarios
- P019 Ganancias/utilidades
- P020 Otros

**Ingresos por negocio propio**

	<i>Trabajo principal</i>
	P068 Por negocios con tipo de actividad industrial
	P069 Por negocios con tipo de actividad comercial
	P070 Por negocios prestadores de servicios
	P071 Por negocios con actividades agrícolas
	P072 Por negocios con actividades de cría y explotación de animales
	P073 Por negocios con actividades de recolección, reforestación y tala de árboles
	P074 Por negocios con actividades de pesca, caza y captura de animales
	<i>Trabajo secundario</i>
	P075 Por negocios con tipo de actividad industrial
	P076 Por negocios con tipo de actividad comercial
	P077 Por negocios prestadores de servicios
	P078 Por negocios con actividades agrícolas
	P079 Por negocios con actividades de cría y explotación de animales
	P080 Por negocios con actividades de recolección, reforestación y tala de árboles
	P081 Por negocios con actividades de pesca, caza y captura de animales
	<b>Ingresos por renta de la propiedad</b>
	P023 Alquiler de tierras y terrenos, dentro y fuera del país
<b>Posesión de ac- tivos físicos y no físicos</b>	P024 Alquiler de casas, edificios, locales y otros inmuebles que están dentro del país
	P025 Alquiler de casas, edificios, locales y otros inmuebles que están fuera del país
	P026 Intereses provenientes de inversiones a plazo fijo
	P027 Intereses provenientes de cuentas de ahorro
	P028 Intereses provenientes de préstamos a terceros
	P029 Rendimientos provenientes de bonos o cédulas
	P030 Alquiler de marcas, patentes y derechos de autor
	P031 Otros ingresos por renta de la propiedad no considerados en los anteriores ingresos

Tabla B.1: Claves de la muestras provenientes de la ENIGH.

## APÉNDICE C

# ANÁLISIS POR REGIÓN

### C.1 RESULTADOS PARA LA REGIÓN TOTAL DE LA DISTRIBUCIÓN DEL INGRESO

Año	Modelo	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	Gini	Media	KL	SCE	KS
1984	Pareto		0.13			-1.34	-	2.23	246.15	0.46
	Exponencial			16,017		0.5	16,017	0.05	3.97	0.06
	Gamma		1.09	14,676		0.48	16,017	0.03	2.69	0.06
	Log normal	1.17			9.16	0.59	18,735	0.11	9.87	0.11
	Dagum	2.73	0.37	21,158		0.48	15,564	0.01	0.51	0.04
	Singh-Maddala	1.27	3.35	34,621		0.49	15,808	0.02	1.79	0.06
1984	Pareto		0.14			-1.41	-	1.71	451.59	0.46
	Exponencial			17,125		0.5	17,125	0.07	13.64	0.09
	Gamma		1.04	16,505		0.49	17,125	0.06	13.12	0.09
	Log normal	1.11			9.16	0.57	18,148	0.15	10.88	0.10
	Dagum	2.28	0.53	16,745		0.5	16,229	0.04	1.01	0.04
	Singh-Maddala	1.45	1.87	18,619		0.5	16,393	0.05	1.98	0.05
1992	Pareto		0.07			-1.17	-	6.28	573.29	0.52
	Exponencial			18,254		0.5	18,254	0.05	12.41	0.10
	Gamma		0.94	19,488		0.51	18,254	0.07	12.53	0.10
	Log normal	1.19			9.16	0.6	19,960	0.11	11.32	0.10
	Dagum	2.14	0.53	17,357		0.53	17,362	0.01	1.11	0.04
	Singh-Maddala	1.36	1.83	18,988		0.54	17,622	0.02	2.22	0.05
1994	Pareto		0.12			-1.32	-	2.2	312.35	0.46
	Exponencial			18,832		0.5	18,832	0.04	11.72	0.11
	Gamma		0.88	21,471		0.53	18,832	0.08	10.34	0.10
	Log normal	1.26			9.16	0.63	21,328	0.1	8.5	0.09

	Dagum	2.07	0.51	18,401		0.55	18,220	0.02	0.71	0.03
	Singh-Maddala	1.26	2.07	22,432		0.55	18,285	0.03	1.51	0.04
1996	Pareto		0.13			-1.34	-	Inf	494.74	0.46
	Exponencial			13,254		0.5	13,254	0.04	13.45	0.10
	Gamma		0.89	14,967		0.52	13,254	0.08	12.04	0.09
	Log normal	1.24			9.16	0.62	14,749	0.12	10.85	0.09
	Dagum	2.09	0.51	13,077		0.55	12,821	0.03	0.62	0.03
	Singh-Maddala	1.26	2.12	16,262		0.55	12,787	0.03	1.6	0.04
1998	Pareto		0.13			-1.35	-	2.16	429.05	0.45
	Exponencial			14,347		0.5	14,347	0.04	13.99	0.10
	Gamma		0.85	16,862		0.53	14,347	0.09	10.79	0.08
	Log normal	1.28			9.16	0.63	16,309	0.12	9.28	0.09
	Dagum	2.02	0.5	14,148		0.57	14,103	0.03	0.57	0.03
	Singh-Maddala	1.21	2.2	18,368		0.56	13,956	0.03	1.38	0.04
2000	Pareto		0.13			-1.35	-	2.04	380.55	0.45
	Exponencial			16,386		0.5	16,386	0.05	9.35	0.09
	Gamma		0.91	18,098		0.52	16,386	0.08	8.88	0.09
	Log normal	1.25			9.16	0.62	18,734	0.1	9.64	0.10
	Dagum	2.13	0.49	16,689		0.54	15,918	0.02	0.73	0.04
	Singh-Maddala	1.26	2.16	20,850		0.54	16,000	0.02	1.65	0.05
2002	Pareto		0.13			-1.36	-	1.82	435.99	0.45
	Exponencial			15,799		0.5	15,799	0.03	7.57	0.08
	Gamma		0.92	17,132		0.52	15,799	0.05	7.5	0.08
	Log normal	1.27			9.16	0.63	18,825	0.09	13.07	0.10
	Dagum	2.19	0.45	17,694		0.54	15,695	0.02	0.92	0.04
	Singh-Maddala	1.19	2.66	26,587		0.54	15,664	0.02	2.29	0.05
2004	Pareto		0.13			-1.35	-	1.95	490.59	0.46
	Exponencial			16,133		0.5	16,133	0.05	9.4	0.08
	Gamma		0.95	16,960		0.51	16,133	0.07	9.7	0.08
	Log normal	1.23			9.16	0.61	18,576	0.12	14.57	0.10
	Dagum	2.29	0.45	18,021		0.52	15,573	0.02	1.07	0.04
	Singh-Maddala	1.25	2.52	24,889		0.53	15,710	0.03	2.7	0.05
2005	Pareto		0.13			-1.35	-	Inf	778.06	0.45
	Exponencial			16,826		0.5	16,826	0.07	15.4	0.09
	Gamma		0.92	18,361		0.52	16,826	0.11	16.34	0.09
	Log normal	1.25			9.16	0.62	19,349	0.18	26.4	0.12
	Dagum	2.36	0.43	19,086		0.52	15,767	0.05	2.34	0.05
	Singh-Maddala	1.23	2.65	27,072		0.53	16,034	0.06	5.38	0.06
2006	Pareto		0.13			-1.35	-	2.2	806.75	0.44
	Exponencial			16,741		0.5	16,741	0.05	13.22	0.09
	Gamma		0.89	18,730		0.52	16,741	0.09	14.07	0.08
	Log normal	1.31			9.16	0.65	20,622	0.15	32.59	0.12
	Dagum	2.34	0.4	20,246		0.53	16,094	0.04	3.06	0.05
	Singh-Maddala	1.15	3.1	33,683		0.54	16,376	0.05	7.01	0.06
	Pareto		0.12			-1.33	-	2.69	966.34	0.44

2008	Exponencial			17,202	0.5	17,202	0.06	23.11	0.08
	Gamma	0.82		21,050	0.54	17,202	0.15	20.28	0.07
	Log normal	1.41		9.16	0.68	22,676	0.22	51.38	0.12
	Dagum	2.35	0.36	21,921	0.54	16,293	0.09	3.15	0.04
	Singh-Maddala	1.06	3.72	44,136	0.55	16,621	0.09	9.19	0.06
2010	Pareto		0.12		-1.33	-	2.69	966.34	0.44
	Exponencial			15,285	0.5	15,285	0.04	11.5	0.07
	Gamma		0.9	16,893	0.52	15,285	0.06	12.82	0.06
	Log normal	1.34		9.16	0.65	19,541	0.13	40.42	0.11
	Dagum	2.45	0.37	19,942	0.52	14,824	0.03	2.35	0.03
2012	Singh-Maddala	1.12	3.82	40,227	0.53	15,016	0.03	7.01	0.05
	Pareto		0.12		-1.31	-	Inf	460.44	0.44
	Exponencial			14,547	0.5	14,547	0.02	11.33	0.07
	Gamma		0.76	19,138	0.55	14,547	0.11	6.75	0.06
	Log normal	1.51		9.16	0.71	20,774	0.16	26.04	0.13
2014	Dagum	2.38	0.32	20,558	0.56	13,995	0.07	1.67	0.04
	Singh-Maddala	0.96	5.2	62,354	0.57	14,276	0.06	5.1	0.07
	Pareto		0.12		-1.3	-	2.63	820.79	0.44
	Exponencial			14,602	0.5	14,602	0.06	20.84	0.08
	Gamma		0.78	18,701	0.55	14,602	0.17	16.03	0.07
1984	Log normal	1.46		9.16	0.7	19,853	0.24	45.76	0.12
	Dagum	2.4	0.33	19,652	0.55	13,736	0.11	2.72	0.04
	Singh-Maddala	1.01	4.29	45,985	0.56	14,098	0.11	8.74	0.06

Tabla C.1: Estimaciones de los parámetros para los modelos de probabilidad: totalidad de la distribución de ingresos. La columnas presentan la siguiente información:  $a/\sigma$  es el parámetro de forma  $a$  para las distribuciones Dagum y Singh-Maddala y el parametro  $\sigma$  para el modelo Log normal;  $p/q$  es el parámetro de forma  $p$  para el modelo Dagum, Gamma y Pareto, mientras que,  $q$  para el modelo Singh-Maddala;  $\beta$  es el parámetro de escala; y  $\mu$  el parametro de localización para el modelo Log normal. Las columnas Gini y Media presentan el índice de Gini y la Media que se desprenden de la estimación de los coeficientes estimados. Por último, se presentan los criterios de divergencia Kullback-Leibler, suma del cuadrado del error y la distancia Kolmogorov-Smirnov identificadas por las columnas KL, SCE y KS, respectivamente.

Año	Modelo	KL	Valor p	SCE	Valor p	KS	Valor p
1984	Pareto	2.23	0.00	246.15	0.00	0.46	0.00
	Exponencial	0.05	0.00	3.97	0.00	0.06	0.00
	Gamma	0.03	0.00	2.69	0.00	0.06	0.00

	Log normal	0.11	0.07	9.87	0.00	0.11	0.00
	Dagum	0.01	<b>0.51</b>	0.51	0.00	0.04	0.00
	Singh-Maddala	0.02	0.00	1.79	0.00	0.06	0.00
	Pareto	1.71	0.00	451.59	0.00	0.46	0.00
	Exponencial	0.07	0.00	13.64	0.00	0.09	0.00
1989	Gamma	0.06	0.00	13.12	0.00	0.09	0.00
	Log normal	0.15	0.01	10.88	0.00	0.10	0.00
	Dagum	0.04	0.00	1.01	0.00	0.04	0.00
	Singh-Maddala	0.05	0.01	1.98	0.00	0.05	0.00
	Pareto	6.28	0.00	573.29	0.00	0.52	0.00
	Exponencial	0.05	0.00	12.41	0.00	0.10	0.00
1992	Gamma	0.07	0.00	12.53	0.00	0.10	0.00
	Log normal	0.11	<b>0.10</b>	11.32	0.00	0.10	0.00
	Dagum	0.01	<b>0.46</b>	1.11	0.00	0.04	0.00
	Singh-Maddala	0.02	<b>0.29</b>	2.22	0.00	0.05	0.00
	Pareto	2.20	0.00	312.35	0.00	0.46	0.00
	Exponencial	0.04	0.00	11.72	0.00	0.11	0.00
1994	Gamma	0.08	0.02	10.34	0.00	0.10	0.00
	Log normal	0.10	<b>0.18</b>	8.50	0.00	0.09	0.00
	Dagum	0.02	<b>0.15</b>	0.71	0.00	0.03	0.00
	Singh-Maddala	0.03	0.03	1.51	0.00	0.04	0.00
	Pareto		0.00	494.74	0.00	0.46	0.00
	Exponencial	0.04	0.00	13.45	0.00	0.10	0.00
1996	Gamma	0.08	0.01	12.04	0.00	0.09	0.00
	Log normal	0.12	<b>0.11</b>	10.85	0.00	0.09	0.00
	Dagum	0.03	0.07	0.62	0.00	0.03	0.00
	Singh-Maddala	0.03	0.02	1.60	0.00	0.04	0.00
	Pareto	2.16	0.00	429.05	0.00	0.45	0.00
	Exponencial	0.04	0.00	13.99	0.00	0.10	0.00
1998	Gamma	0.09	0.06	10.79	0.00	0.08	0.00
	Log normal	0.12	<b>0.14</b>	9.28	0.00	0.09	0.00
	Dagum	0.03	<b>0.31</b>	0.57	0.00	0.03	0.00
	Singh-Maddala	0.03	0.01	1.38	0.00	0.04	0.00
	Pareto	2.04	0.00	380.55	0.00	0.45	0.00
	Exponencial	0.05	0.00	9.35	0.00	0.09	0.00
2000	Gamma	0.08	0.00	8.88	0.00	0.09	0.00
	Log normal	0.10	<b>0.18</b>	9.64	0.00	0.10	0.00
	Dagum	0.02	<b>0.17</b>	0.73	0.00	0.04	0.00
	Singh-Maddala	0.02	0.06	1.65	0.00	0.05	0.00
	Pareto	1.82	0.00	435.99	0.00	0.45	0.00
	Exponencial	0.03	0.00	7.57	0.00	0.08	0.00
2002	Gamma	0.05	0.01	7.50	0.00	0.08	0.00
	Log normal	0.09	<b>0.28</b>	13.07	0.00	0.10	0.00
	Dagum	0.02	<b>0.92</b>	0.92	0.00	0.04	0.00
	Singh-Maddala	0.02	0.02	2.29	0.00	0.05	0.00

2004	Pareto	1.95	0.00	490.59	0.00	0.46	0.00
	Exponencial	0.05	0.00	9.40	0.00	0.08	0.00
	Gamma	0.07	0.00	9.70	0.00	0.08	0.00
	Log normal	0.12	<b>0.10</b>	14.57	0.00	0.10	0.00
	Dagum	0.02	<b>0.11</b>	1.07	0.00	0.04	0.00
	Singh-Maddala	0.03	0.00	2.70	0.00	0.05	0.00
2005	Pareto	Inf	0.00	778.06	0.00	0.45	0.00
	Exponencial	0.07	0.00	15.40	0.00	0.09	0.00
	Gamma	0.11	0.00	16.34	0.00	0.09	0.00
	Log normal	0.18	0.03	26.40	0.00	0.12	0.00
	Dagum	0.05	0.01	2.34	0.00	0.05	0.00
	Singh-Maddala	0.06	0.00	5.38	0.00	0.06	0.00
2006	Pareto	2.20	0.00	806.75	0.00	0.44	0.00
	Exponencial	0.05	0.00	13.22	0.00	0.09	0.00
	Gamma	0.09	0.00	14.07	0.00	0.08	0.00
	Log normal	0.15	<b>0.11</b>	32.59	0.00	0.12	0.00
	Dagum	0.04	<b>0.34</b>	3.06	0.00	0.05	0.00
	Singh-Maddala	0.05	0.00	7.01	0.00	0.06	0.00
2008	Pareto	2.69	0.00	966.34	0.00	0.44	0.00
	Exponencial	0.06	0.00	23.11	0.00	0.08	0.00
	Gamma	0.15	0.06	20.28	0.00	0.07	0.00
	Log normal	0.22	0.05	51.38	0.00	0.12	0.00
	Dagum	0.09	<b>0.50</b>	3.15	0.00	0.04	0.00
	Singh-Maddala	0.09	0.00	9.19	0.00	0.06	0.00
2008	Pareto	2.69	0.00	966.34	0.00	0.44	0.00
	Exponencial	0.04	0.00	11.50	0.00	0.07	0.00
	Gamma	0.06	0.01	12.82	0.00	0.06	0.00
	Log normal	0.13	<b>0.14</b>	40.42	0.00	0.11	0.00
	Dagum	0.03	<b>0.85</b>	2.35	0.00	0.03	0.00
	Singh-Maddala	0.03	0.00	7.01	0.00	0.05	0.00
2012	Pareto	Inf	0.00	460.44	0.00	0.44	0.00
	Exponencial	0.02	0.10	11.33	0.00	0.07	0.00
	Gamma	0.11	<b>0.78</b>	6.75	0.00	0.06	0.00
	Log normal	0.16	<b>0.11</b>	26.04	0.00	0.13	0.00
	Dagum	0.07	<b>1.00</b>	1.67	0.00	0.04	0.00
	Singh-Maddala	0.06	0.04	5.10	0.00	0.07	0.00
2014	Pareto	2.63	0.00	820.79	0.00	0.44	0.00
	Exponencial	0.06	0.00	20.84	0.00	0.08	0.00
	Gamma	0.17	<b>0.21</b>	16.03	0.00	0.07	0.00
	Log normal	0.24	0.04	45.76	0.00	0.12	0.00
	Dagum	0.11	<b>0.80</b>	2.72	0.00	0.04	0.00
	Singh-Maddala	0.11	0.00	8.74	0.00	0.06	0.00

Tabla C.2: Medidas de divergencia y pruebas de plausibilidad para el total de la distribución de ingresos. Se muestran las medidas de divergencia KL, SCE y KS y el *valor p* asociado a sus respectivas pruebas de plausibilidad. En negritas se presentan los *valores p* que estadísticamente son significativos.

## C.2 RESULTADOS PARA LA REGIÓN DE INGRESOS

### ALTOS

Submuestra	Año	$x_{min}$	%	$p$	D	Valor p
KL	1984	25,462	15.2 %	2.09	0.027	<b>1.00</b>
KS	1984	19,238	25.9 %	2.02	0.028	0.00
SCE	1984	19,143	26.0 %	2.01	0.065	<b>0.58</b>
KL	1989	74,347	2.3 %	2.04	0.059	<b>0.67</b>
KS	1989	37,928	7.9 %	1.91	0.032	0.00
SCE	1989	73,269	2.3 %	2.05	0.080	<b>0.46</b>
KL	1992	18,014	29.3 %	1.65	0.035	<b>1.00</b>
KS	1992	25,069	17.3 %	1.65	0.031	0.00
SCE	1992	79,891	2.8 %	1.98	0.043	<b>0.79</b>
KL	1994	38,361	10.2 %	1.77	0.045	<b>1.00</b>
KS	1994	36,187	11.2 %	1.76	0.038	0.00
SCE	1994	95,902	2.1 %	2.00	0.057	<b>0.65</b>
KL	1996	53,569	3.1 %	1.98	0.041	<b>1.00</b>
KS	1996	32,142	7.5 %	1.85	0.036	0.00
SCE	1996	53,569	3.1 %	1.98	0.105	<b>0.30</b>
KL	1998	72,583	2.2 %	2.09	0.043	<b>0.96</b>
KS	1998	32,965	8.4 %	1.82	0.033	0.00
SCE	1998	57,210	3.3 %	1.96	0.052	<b>0.70</b>
KL	2000	32,221	10.8 %	1.80	0.038	<b>1.00</b>
KS	2000	30,642	11.5 %	1.77	0.024	0.01
SCE	2000	30,642	11.5 %	1.77	0.081	<b>0.44</b>
KL	2002	25,423	15.7 %	1.81	0.033	<b>1.00</b>
KS	2002	23,752	16.9 %	1.72	0.042	0.00
SCE	2002	72,046	2.8 %	2.55	0.286	0.02
KL	2004	75,243	2.3 %	2.36	0.037	<b>0.95</b>
KS	2004	77,504	2.2 %	2.24	0.038	0.00
SCE	2004	77,504	2.2 %	2.23	0.068	<b>0.56</b>
KL	2005	73,890	2.4 %	2.16	0.049	<b>0.79</b>
KS	2005	75,143	2.2 %	2.01	0.038	0.00

SCE	2005	75,143	2.2 %	2.00	0.131	<b>0.19</b>
KL	2006	42,781	7.0 %	1.97	0.042	<b>1.00</b>
KS	2006	40,969	7.4 %	1.91	0.039	0.00
SCE	2006	64,851	3.3 %	2.11	0.176	0.09
KL	2008	84,589	2.0 %	2.09	0.045	<b>0.92</b>
KS	2008	49,150	5.6 %	1.97	0.019	0.06
SCE	2008	48,608	5.7 %	1.97	0.085	<b>0.42</b>
KL	2010	43,560	5.6 %	2.15	0.038	<b>0.99</b>
KS	2010	57,467	3.2 %	2.21	0.026	0.00
SCE	2010	57,467	3.2 %	2.22	0.104	<b>0.29</b>
KL	2012	27,260	12.4 %	1.89	0.039	<b>1.00</b>
KS	2012	26,351	12.8 %	1.84	0.025	0.00
SCE	2012	54,818	3.5 %	2.05	0.043	<b>0.80</b>
KL	2014	70,652	2.0 %	1.95	0.044	<b>0.12</b>
KS	2014	58,780	2.9 %	1.94	0.018	<b>0.12</b>
SCE	2014	64,377	2.4 %	1.93	0.028	<b>0.95</b>

Tabla C.3: Resultados de la selección de la cota inferior para el modelo de Pareto. Conforme a cada submuestra generada por las medidas de error KL, SCE y KS y año, se muestra los niveles óptimos de ingreso mínimo ( $x_{min}$ ), el porcentaje de la población que obtiene ingresos mayores e iguales que el ingreso mínimo óptimo (%), la medida de divergencia asociada (D), el coeficiente de Pareto ( $p$ ) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Submuestra	Año	$x_{min}$	%	$\beta$	D	Valor p
KL	1984	96,189	1.0 %	142,166	0.026	0.00
SCE	1984	94,335	1.2 %	134,027	1.658	0.00
KS	1984	17,314	30.0 %	34,172	0.397	0.00
KL	1989	19,346	25.0 %	42,569	0.034	0.00
SCE	1989	111,922	1.0 %	231,972	3.278	0.00
KS	1989	16,491	29.9 %	38,530	0.348	0.00
KL	1992	135,788	1.0 %	252,125	0.031	0.00
SCE	1992	136,903	1.0 %	254,723	2.258	0.00
KS	1992	18,020	27.2 %	44,683	0.332	0.00
KL	1994	126,820	1.2 %	239,115	0.036	0.00
SCE	1994	107,410	1.8 %	199,487	3.050	0.00
KS	1994	18,426	27.0 %	47,197	0.323	0.00
KL	1996	61,340	2.3 %	123,534	0.044	0.00
SCE	1996	99,524	1.0 %	187,531	3.063	0.00
KS	1996	12,708	29.0 %	31,721	0.330	0.00

KL	1998	80,144	1.7 %	157,417	0.038	0.00
SCE	1998	107,867	1.0 %	200,209	2.887	0.00
KS	1998	13,523	30.0 %	34,161	0.327	0.00
KL	2000	57,216	3.9 %	114,098	0.037	0.00
SCE	2000	127,936	1.0 %	208,202	3.374	0.00
KS	2000	15,412	30.0 %	38,112	0.333	0.00
KL	2002	76,839	2.4 %	119,389	0.038	0.00
SCE	2002	106,034	1.2 %	148,239	4.273	0.00
KS	2002	17,001	26.0 %	39,309	0.351	0.00
KL	2004	25,376	14.2 %	54,917	0.044	0.00
SCE	2004	96,880	1.3 %	170,922	5.291	0.00
KS	2004	15,507	30.0 %	36,601	0.346	0.00
KL	2005	16,290	30.0 %	38,424	0.037	0.00
SCE	2005	112,101	1.0 %	236,672	4.778	0.00
KS	2005	16,290	30.0 %	38,424	0.346	0.00
KL	2006	77,021	2.3 %	138,784	0.045	0.00
SCE	2006	108,175	1.2 %	177,147	7.244	0.00
KS	2006	18,149	25.4 %	41,978	0.351	0.00
KL	2008	16,918	30.0 %	39,953	0.039	0.00
SCE	2008	115,774	1.0 %	231,946	12.860	0.00
KS	2008	16,918	30.0 %	39,997	0.345	0.00
KL	2010	85,747	1.4 %	147,362	0.048	0.00
SCE	2010	95,659	1.1 %	160,381	12.686	0.00
KS	2010	15,794	29.4 %	34,885	0.364	0.00
KL	2012	99,769	1.0 %	183,127	0.036	0.00
SCE	2012	99,953	1.0 %	183,335	3.519	0.00
KS	2012	14,630	30.0 %	34,165	0.348	0.00
KL	2014	14,685	30.0 %	34,164	0.044	0.00
SCE	2014	99,721	1.0 %	201,212	6.881	0.00
KS	2014	15,232	27.9 %	35,608	0.348	0.00

Tabla C.4: Resultados de la selección de la cota inferior para el modelo Exponencial. Conforme a cada submuestra generada por las medidas de error KL, SCE y KS y año, se muestra los niveles óptimos de ingreso mínimo ( $x_{min}$ ), el porcentaje de la población que obtiene ingresos mayores e iguales que el ingreso mínimo óptimo (%), la medida de divergencia asociada (D), el coeficiente de escala ( $\beta$ ), y por último, el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Submuestra	Año	$x_{min}$	%	<i>p</i>	$\beta$	D	Valor p
KL	1984	96,189	1.0 %	8.15	17,444	0.01	0.00

SCE	1984	96,189	1.0 %	8.15	17,444	0.40	0.00
KS	1984	49,981	3.9 %	5.59	15,787	0.14	0.00
KL	1989	16,581	29.8 %	2.25	17,185	0.03	0.00
SCE	1989	111,922	1.0 %	2.17	106,966	2.23	0.00
KS	1989	19,397	23.0 %	2.31	19,359	0.19	0.00
KL	1992	99,075	1.8 %	3.47	54,432	0.04	0.00
SCE	1992	136,903	1.0 %	3.87	65,827	0.98	0.00
KS	1992	36,087	9.0 %	2.54	33,242	0.17	0.00
KL	1994	135,520	1.0 %	3.89	65,843	0.04	0.00
SCE	1994	130,426	1.1 %	3.80	63,616	1.14	0.00
KS	1994	108,847	1.6 %	3.64	58,436	0.17	0.00
KL	1996	12,708	29.0 %	2.13	14,922	0.05	0.00
SCE	1996	99,524	1.0 %	2.69	69,781	1.93	0.00
KS	1996	13,537	26.4 %	2.17	15,487	0.18	0.00
KL	1998	89,595	1.3 %	3.51	50,371	0.04	0.00
SCE	1998	107,867	1.0 %	3.66	54,652	1.06	0.00
KS	1998	14,559	27.1 %	2.20	16,468	0.17	0.00
KL	2000	126,512	1.0 %	6.01	34,586	0.04	0.00
SCE	2000	127,936	1.0 %	6.02	34,603	1.05	0.00
KS	2000	66,397	3.0 %	4.19	30,974	0.14	0.00
KL	2002	85,039	1.8 %	10.18	13,087	0.02	0.00
SCE	2002	108,154	1.1 %	12.98	11,827	0.49	0.00
KS	2002	44,182	6.1 %	5.34	15,327	0.13	0.00
KL	2004	16,719	28.2 %	2.50	15,167	0.05	0.00
SCE	2004	106,568	1.0 %	3.49	54,972	3.11	0.00
KS	2004	31,066	10.2 %	3.09	21,224	0.17	0.00
KL	2005	16,290	30.0 %	2.18	17,615	0.03	0.00
SCE	2005	112,101	1.0 %	1.99	119,195	5.29	0.00
KS	2005	18,795	23.7 %	2.23	19,668	0.19	0.00
KL	2006	110,881	1.1 %	5.45	34,358	0.05	0.00
SCE	2006	108,766	1.1 %	5.35	33,697	1.86	0.00
KS	2006	80,530	2.1 %	4.75	30,563	0.16	0.00
KL	2008	16,918	30.0 %	2.27	17,599	0.04	0.00
SCE	2008	115,774	1.0 %	2.78	83,416	6.43	0.00
KS	2008	16,918	30.0 %	2.27	17,612	0.18	0.00
KL	2010	91,227	1.2 %	5.08	30,618	0.06	0.00
SCE	2010	95,659	1.1 %	5.14	31,215	3.54	0.00
KS	2010	72,742	1.9 %	4.69	27,413	0.17	0.00
KL	2012	99,769	1.0 %	4.39	41,711	0.05	0.00
SCE	2012	99,953	1.0 %	4.39	41,729	1.02	0.00
KS	2012	99,953	1.0 %	4.39	41,729	0.17	0.00
KL	2014	14,685	30.0 %	2.23	15,300	0.05	0.00
SCE	2014	100,984	1.0 %	2.71	75,589	3.33	0.00
KS	2014	15,865	26.0 %	2.28	16,273	0.19	0.00

Tabla C.5: Resultados de la selección de la cota inferior para el modelo Gamma. Conforme a cada submuestra generada por las medidas de error KL, SCE y KS y año, se muestra los niveles óptimos de ingreso mínimo ( $x_{min}$ ), el porcentaje de la población que obtiene ingresos mayores e iguales que el ingreso mínimo óptimo (%), la medida de divergencia asociada (D), los coeficientes de forma ( $p$ ) y de escala ( $\beta$ ), y por último, el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Submuestra	Año	$x_{min}$	%	$\mu$	$\sigma$	D	Valor p
KL	1984	63,371	2.5 %	11.48	0.362	0.01	0.00
SCE	1984	96,189	1.0 %	11.80	0.330	0.35	0.00
KS	1984	54,013	3.4 %	11.35	0.385	0.12	0.00
KL	1989	19,346	25.0 %	10.42	0.545	0.01	0.51
SCE	1989	111,922	1.0 %	12.11	0.512	1.06	0.00
KS	1989	16,459	30.0 %	10.32	0.552	0.14	0.00
KL	1992	103,218	1.7 %	12.04	0.487	0.02	0.00
SCE	1992	136,903	1.0 %	12.31	0.478	0.76	0.00
KS	1992	39,696	7.9 %	11.21	0.547	0.13	0.00
KL	1994	135,029	1.0 %	12.31	0.451	0.03	0.00
SCE	1994	130,426	1.1 %	12.26	0.455	0.68	0.00
KS	1994	19,457	25.8 %	10.54	0.590	0.13	0.00
KL	1996	12,687	30.4 %	10.08	0.586	0.02	0.00
SCE	1996	99,524	1.0 %	11.94	0.478	0.95	0.00
KS	1996	17,086	18.9 %	10.38	0.559	0.13	0.00
KL	1998	92,215	1.2 %	11.96	0.460	0.02	0.00
SCE	1998	107,867	1.0 %	12.06	0.452	0.61	0.00
KS	1998	14,559	27.1 %	10.25	0.584	0.13	0.00
KL	2000	87,265	2.0 %	11.86	0.421	0.02	0.00
SCE	2000	94,299	1.7 %	11.93	0.410	0.78	0.00
KS	2000	73,856	2.6 %	11.72	0.439	0.12	0.00
KL	2002	85,039	1.8 %	11.75	0.300	0.01	0.00
SCE	2002	108,154	1.1 %	11.90	0.264	0.35	0.00
KS	2002	48,220	5.4 %	11.28	0.397	0.11	0.00
KL	2004	16,719	28.2 %	10.33	0.552	0.02	0.00
SCE	2004	106,568	1.0 %	12.01	0.453	1.81	0.00
KS	2004	31,066	10.2 %	10.92	0.492	0.12	0.00
KL	2005	16,306	30.0 %	10.31	0.559	0.01	0.36
SCE	2005	112,101	1.0 %	12.10	0.578	2.86	0.00
KS	2005	18,967	23.5 %	10.45	0.549	0.14	0.00
KL	2006	110,881	1.1 %	12.05	0.384	0.03	0.00

SCE	2006	108,766	1.1 %	12.01	0.388	1.14	0.00
KS	2006	80,530	2.1 %	11.78	0.414	0.12	0.00
KL	2008	16,918	30.0 %	10.36	0.565	0.01	0.05
SCE	2008	115,774	1.0 %	12.16	0.493	3.46	0.00
KS	2008	16,918	30.0 %	10.36	0.565	0.13	0.00
KL	2010	91,227	1.2 %	11.85	0.395	0.04	0.00
SCE	2010	95,659	1.1 %	11.88	0.393	2.10	0.00
KS	2010	15,794	29.4 %	10.27	0.529	0.13	0.00
KL	2012	99,769	1.0 %	12.00	0.446	0.03	0.00
SCE	2012	99,953	1.0 %	12.00	0.445	0.75	0.00
KS	2012	17,268	23.4 %	10.38	0.544	0.13	0.00
KL	2014	14,685	30.0 %	10.20	0.563	0.02	0.00
SCE	2014	100,984	1.0 %	12.03	0.509	2.11	0.00
KS	2014	16,199	25.2 %	10.30	0.555	0.14	0.00

Tabla C.6: Resultados de la selección de la cota inferior para el modelo Log normal. Conforme a cada submuestra generada por las medidas de error KL, SCE y KS y año, se muestra los niveles óptimos de ingreso mínimo ( $x_{min}$ ), el porcentaje de la población que obtiene ingresos mayores e iguales que el ingreso mínimo óptimo (%), la medida de divergencia asociada (D), los coeficientes para la media ( $\mu$ ) y la desviación estándar ( $\sigma$ ) en logaritmos, y por último, el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

### C.3 RESULTADOS PARA LA REGIÓN DE INGRESOS

#### MEDIOS-BAJOS

Submuestra	Año	%	<i>p</i>	Gini	Media	D	Valor <i>p</i>
KL	1984	84.8 %	0.13	-1.35	-	2.22	0.00
SCE	1984	74.0 %	0.13	-1.36	-	181.25	0.00
KL	1989	97.8 %	0.15	-1.41	-	1.66	0.00
SCE	1989	97.7 %	0.15	-1.41	-	443.07	0.00
KL	1992	70.7 %	0.08	-1.17	-	6.28	0.00
SCE	1992	97.2 %	0.07	-1.17	-	561.77	0.00
KL	1994	89.8 %	0.12	-1.32	-	2.16	0.00
SCE	1994	97.9 %	0.12	-1.32	-	307.25	0.00
KL	1996	96.9 %	0.13	-1.35	-	1.90	0.00
SCE	1996	96.9 %	0.13	-1.35	-	482.63	0.00

KL	1998	97.8 %	0.13	-1.36	-	2.08	0.00
SCE	1998	96.7 %	0.13	-1.36	-	420.15	0.00
KL	2000	89.2 %	0.13	-1.37	-	1.99	0.00
SCE	2000	88.5 %	0.13	-1.37	-	343.18	0.00
KL	2002	84.3 %	0.14	-1.36	-	1.79	0.00
KL	2004	97.7 %	0.13	-1.35	-	1.86	0.00
SCE	2004	97.8 %	0.13	-1.35	-	478.33	0.00
KL	2005	97.6 %	0.13	-1.35	-	1.97	0.00
SCE	2005	97.8 %	0.13	-1.35	-	763.15	0.00
KL	2006	93.0 %	0.13	-1.35	-	2.05	0.00
KL	2008	98.0 %	0.13	-1.34	-	2.45	0.00
SCE	2008	94.3 %	0.13	-1.34	-	891.00	0.00
KL	2012	87.7 %	0.12	-1.31	-	2.51	0.00
SCE	2012	96.5 %	0.12	-1.31	-	449.40	0.00
KL	2014	98.0 %	0.12	-1.31	-	2.36	0.00
SCE	2014	97.6 %	0.12	-1.31	-	803.19	0.00
KS	2014	97.1 %	0.12	-1.31	-	0.44	0.00

Tabla C.7: Estimaciones para la región de ingresos bajos-medios: modelo Pareto. Las columnas muestran, el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%), el coeficiente de Pareto ( $p$ ), estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Submuestra	Año	%	$\beta$	Gini	Media	D	Valor p
KL	1984	84.8 %	1.04E+04	0.50	10,428	0.11	0.00
SCE	1984	74.0 %	8,780	0.50	8,780	18.52	0.00
KL	1989	97.8 %	1.40E+04	0.50	14,008	0.07	0.00
SCE	1989	97.7 %	1.40E+04	0.50	13,961	14.99	0.00
KL	1992	70.7 %	8,074	0.50	8,074	0.14	0.00
SCE	1992	97.2 %	1.43E+04	0.50	14,274	9.72	0.00
KL	1994	89.8 %	1.16E+04	0.50	11,647	0.06	0.00
SCE	1994	97.9 %	1.53E+04	0.50	15,296	3.24	0.00
KL	1996	96.9 %	1.03E+04	0.50	10,265	0.04	0.00
SCE	1996	96.9 %	1.03E+04	0.50	10,265	5.30	0.00
KL	1998	97.8 %	1.16E+04	0.50	11,567	0.04	0.00
SCE	1998	96.7 %	1.10E+04	0.50	10,979	2.87	0.00
KL	2000	89.2 %	1.02E+04	0.50	10,225	0.08	0.00
SCE	2000	88.5 %	1.00E+04	0.50	10,041	16.14	0.00
KL	2002	84.3 %	9,146	0.50	9,146	0.08	0.00

KL	2004	97.7%	1.34E+04	0.50	13,402	0.05	0.00
SCE	2004	97.8%	1.34E+04	0.50	13,442	6.99	0.00
KL	2005	97.6%	1.36E+04	0.50	13,572	0.05	0.00
SCE	2005	97.8%	1.37E+04	0.50	13,685	12.28	0.00
KL	2006	93.0%	1.18E+04	0.50	11,786	0.06	0.00
KL	2008	98.0%	1.42E+04	0.50	14,175	0.03	0.00
SCE	2008	94.3%	1.23E+04	0.50	12,324	18.83	0.00
KL	2010	94.4%	1.15E+04	0.50	11,472	0.05	0.00
SCE	2010	96.8%	1.24E+04	0.50	12,409	13.05	0.00
KL	2012	87.7%	8,805	0.50	8,805	0.05	0.00
SCE	2012	96.5%	1.14E+04	0.50	11,357	3.95	0.00
KL	2014	98.0%	1.19E+04	0.50	11,943	0.02	0.00
SCE	2014	97.6%	1.17E+04	0.50	11,731	5.53	0.00
KS	2014	97.1%	1.15E+04	0.50	11,485	0.06	0.00

Tabla C.8: Estimaciones para la región de ingresos bajos-medios: modelo Exponencial. Las columnas muestran el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%), el coeficiente de escala ( $\beta$ ), estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Tipo	Año	%	<i>p</i>	$\beta$	Gini	Media	D	Valor p
KL	1984	84.8%	1.55	6,740	0.42	8,780	7.56	0.00
SCE	1984	74.0%	1.60	5,476	0.41	13,961	2.20	0.00
KL	1989	97.8%	1.36	1.03E+04	0.44	14,274	1.94	0.00
SCE	1989	97.7%	1.36	1.02E+04	0.44	15,296	1.96	0.00
KL	1992	70.7%	1.77	4,558	0.40	10,265	1.23	0.00
SCE	1992	97.2%	1.25	1.14E+04	0.46	10,979	1.05	0.00
KL	1994	89.8%	1.37	8,517	0.44	10,041	3.83	0.00
SCE	1994	97.9%	1.10	1.39E+04	0.48	13,442	2.34	0.00
KL	1996	96.9%	1.16	8,812	0.47	13,685	4.71	0.00
SCE	1996	96.9%	1.16	8,812	0.47	12,324	12.85	0.00
KL	1998	97.8%	1.06	1.09E+04	0.49	12,409	7.11	0.00
SCE	1998	96.7%	1.11	9,874	0.48	11,357	5.26	0.00
KL	2000	89.2%	1.39	7,335	0.44	11,731	7.35	0.00
SCE	2000	88.5%	1.41	7,123	0.44	10,428	0.05	0.00
KL	2002	84.3%	1.41	6,490	0.44	14,008	0.02	0.00
KL	2004	97.7%	1.17	1.15E+04	0.47	8,074	0.07	0.00
SCE	2004	97.8%	1.17	1.15E+04	0.47	11,647	0.02	0.00
KL	2005	97.6%	1.17	1.16E+04	0.47	10,265	0.01	0.00

SCE	2005	97.8%	1.16	1.18E+04	0.47	11,567	0.02	0.00
KL	2006	93.0%	1.22	9,654	0.46	10,225	0.02	0.00
KL	2008	98.0%	0.98	1.44E+04	0.50	9,146	0.03	0.00
SCE	2008	94.3%	1.08	1.14E+04	0.48	13,402	0.02	0.00
KL	2010	94.4%	1.16	9,857	0.47	13,572	0.02	0.00
SCE	2010	96.8%	1.10	1.13E+04	0.48	11,786	0.03	0.00
KL	2012	87.7%	1.03	8,542	0.49	14,175	0.03	0.00
SCE	2012	96.5%	0.92	1.23E+04	0.52	11,472	0.02	0.00
KL	2014	98.0%	0.94	1.27E+04	0.51	8,805	0.04	0.00
SCE	2014	97.6%	0.95	1.23E+04	0.51	11,943	0.04	0.03
KS	2014	97.1%	0.96	1.19E+04	0.51	11,485	0.07	0.00

Tabla C.9: Estimaciones para la región de ingresos bajos-medios: modelo Gamma. Las columnas muestran el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%), los coeficientes de forma ( $p$ ) y de escala ( $\beta$ ), estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Tipo	Año	%	$\sigma$	$\mu$	Gini	Media	D	Valor p
KL	1984	84.8%	1.05	8.74	0.54	10,662	15.56	0.00
SCE	1984	74.0%	1.04	9.13	0.54	16,033	12.52	0.00
KL	1989	97.8%	1.05	9.12	0.57	17,017	13.52	0.00
SCE	1989	97.7%	1.05	9.11	0.60	18,712	9.63	0.00
KL	1992	70.7%	1.00	8.75	0.59	12,470	13.75	0.00
SCE	1992	97.2%	1.12	8.79	0.61	13,587	12.08	0.00
KL	1994	89.8%	1.12	8.82	0.57	12,497	17.91	0.00
SCE	1994	97.9%	1.20	9.02	0.59	16,452	16.92	0.00
KL	1996	96.9%	1.17	9.04	0.60	17,092	30.51	0.00
SCE	1996	96.9%	1.17	8.89	0.65	17,666	68.96	0.00
KL	1998	97.8%	1.22	8.91	0.63	16,741	49.25	0.00
SCE	1998	96.7%	1.20	8.71	0.70	17,310	31.80	0.00
KL	2000	89.2%	1.11	8.76	0.68	17,254	52.95	0.00
SCE	2000	88.5%	1.11	8.90	0.42	12,725	0.17	0.00
KL	2002	84.3%	1.12	9.14	0.44	16,080	0.09	0.02
KL	2004	97.7%	1.17	8.69	0.40	9,768	0.20	0.00
SCE	2004	97.8%	1.17	8.95	0.44	14,518	0.12	0.03
KL	2005	97.6%	1.19	8.75	0.47	12,470	0.09	0.13
SCE	2005	97.8%	1.19	8.82	0.49	14,223	0.10	0.16
KL	2006	93.0%	1.22	8.83	0.44	12,724	0.14	0.02

KL	2008	98.0%	1.37	8.73	0.44	11,560	0.15	0.01
SCE	2008	94.3%	1.33	9.02	0.47	16,410	0.11	0.09
KL	2010	94.4%	1.26	9.03	0.47	16,971	0.13	0.05
SCE	2010	96.8%	1.28	8.91	0.46	15,564	0.15	0.05
KL	2012	87.7%	1.40	8.97	0.50	20,090	0.13	0.17
SCE	2012	96.5%	1.45	8.86	0.47	15,541	0.15	0.06
KL	2014	98.0%	1.42	8.53	0.49	13,449	0.17	0.08
SCE	2014	97.6%	1.41	8.77	0.51	17,528	0.14	0.16
KS	2014	97.1%	1.41	8.75	0.68	16,929	0.13	0.00

Tabla C.10: Estimaciones para la región de ingresos bajos-medios: modelo Log normal. Las columnas muestran el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%), los coeficientes para la media ( $\mu$ ) y la desviación estándar ( $\sigma$ ) en logaritmos, estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Tipo	Año	%	$\alpha$	$p$	$\beta$	Gini	Media	D	Valor p
KL	1984	84.8%	250.62	0.00	1.90E+04	0.36	8,975	0.83	0.00
SCE	1984	74.0%	2.86	0.40	1.85E+04	0.45	14,189	0.64	0.00
KL	1989	97.8%	2.81	0.38	1.96E+04	0.47	14,499	0.54	0.00
SCE	1989	97.7%	2.52	0.40	2.02E+04	0.50	15,656	0.46	0.00
KL	1992	70.7%	2.77	0.35	1.48E+04	0.48	10,544	0.39	0.00
SCE	1992	97.2%	2.71	0.35	1.61E+04	0.49	11,350	0.50	0.00
KL	1994	89.8%	5.19	0.18	1.91E+04	0.40	10,193	0.58	0.00
SCE	1994	97.9%	2.87	0.34	1.96E+04	0.47	13,604	0.57	0.00
KL	1996	96.9%	2.99	0.32	2.07E+04	0.47	13,805	1.44	0.00
SCE	1996	96.9%	4.14	0.19	2.40E+04	0.45	12,418	1.23	0.00
KL	1998	97.8%	3.41	0.25	2.16E+04	0.47	12,527	1.14	0.00
SCE	1998	96.7%	3.49	0.21	2.21E+04	0.49	11,465	0.89	0.00
KL	2000	89.2%	3.22	0.24	2.12E+04	0.49	11,810	1.40	0.00
SCE	2000	88.5%	16.81	0.05	2.26E+04	0.37	10,692	0.02	0.01
KL	2002	84.3%	2.84	0.41	1.85E+04	0.45	14,235	0.01	0.70
KL	2004	97.7%	22.56	0.04	1.64E+04	0.34	8,171	0.01	0.00
SCE	2004	97.8%	4.50	0.21	2.14E+04	0.41	11,834	0.01	0.38
KL	2005	97.6%	2.77	0.35	1.48E+04	0.48	10,544	0.01	0.91
SCE	2005	97.8%	2.49	0.38	1.57E+04	0.51	11,976	0.02	0.92
KL	2006	93.0%	4.88	0.19	1.92E+04	0.41	10,379	0.02	0.06
KL	2008	98.0%	8.18	0.11	1.93E+04	0.39	9,314	0.02	0.27
SCE	2008	94.3%	2.89	0.34	1.96E+04	0.47	13,565	0.01	0.93

KL	2010	94.4 %	3.03	0.31	2.07E+04	0.47	13,697	0.02	0.65
SCE	2010	96.8 %	4.41	0.20	2.22E+04	0.43	11,859	0.02	0.65
KL	2012	87.7 %	2.97	0.28	2.35E+04	0.50	14,274	0.03	0.99
SCE	2012	96.5 %	4.18	0.20	2.17E+04	0.44	11,576	0.03	0.83
KL	2014	98.0 %	8.00	0.09	2.07E+04	0.44	8,883	0.06	0.88
SCE	2014	97.6 %	3.10	0.25	2.11E+04	0.50	12,015	0.04	1.00
KS	2014	97.1 %	3.38	0.23	2.13E+04	0.49	11,570	0.03	0.00

Tabla C.11: Estimaciones para la región de ingresos bajos-medios: modelo Dagum. Las columnas muestran el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%), los coeficientes de forma ( $a$ ) y ( $p$ ) y de escala ( $\beta$ ), estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Tipo	Año	%	M	Crit	$a$	$q$	$\beta$	Gini	Media	D	Valor p
KL	1984	84.8 %	M2	1.E-01	1.16	4.E+07	1.E+12	-	-	0.06	0.00
SCE	1984	74.0 %	M1	1.E-01	1.17	21.38	2.E+05	0.46	12,138	12.88	0.04
KL	1989	97.8 %			1.33	5.66	5.E+04	0.44	14,003	0.01	0.00
SCE	1989	97.7 %			1.33	5.94	5.E+04	0.44	13,955	1.51	0.00
KL	1992	70.7 %	M1	1.E-02	1.42	1.E+11	5.E+12	-	-	0.10	0.00
SCE	1992	97.2 %			1.24	7.14	7.E+04	0.46	14,267	1.45	0.00
KL	1994	89.8 %	M1	1.E-01	1.16	734.56	3.E+06	-	-	0.03	0.06
SCE	1994	97.9 %			1.16	6.15	7.E+04	0.49	15,297	1.18	0.00
KL	1996	96.9 %			1.13	21.24	2.E+05	0.47	10,263	0.02	0.00
SCE	1996	96.9 %			1.13	21.24	2.E+05	0.47	10,263	1.07	0.00
KL	1998	97.8 %			1.10	8.83	8.E+04	0.49	11,570	0.03	0.00
SCE	1998	96.7 %			1.07	162.31	1.E+06	-	10,979	1.05	0.00
KL	2000	89.2 %	M1	1.E-02	1.15	13.69	1.E+05	0.47	13,394	0.02	<b>0.94</b>
SCE	2000	88.5 %	M2	1.E-01	1.17	4.E+06	3.E+09	-	-	56.75	0.00
KL	2002	84.3 %	M2	1.E-01	1.07	2.E+14	3.E+18	-	-	0.07	0.00
KL	2004	97.7 %			1.15	13.69	1.E+05	0.47	13,394	0.02	0.00
SCE	2004	97.8 %			1.16	12.57	1.E+05	0.47	13,433	1.90	0.00
KL	2005	97.6 %			1.12	66.97	6.E+05	-	13,562	0.03	0.01
SCE	2005	97.8 %			1.13	29.51	3.E+05	0.47	13,674	3.92	0.00
KL	2006	93.0 %	M1	1.E-03	1.06	6.E+54	2.E+55	-	-	2.59	0.00
KL	2008	98.0 %	M2	1.E-03	0.97	5.E+18	2.E+23	-	-	0.06	0.02
SCE	2008	94.3 %	M1	1.E-03	0.96	9.E+79	2.E+83	-	-	4679.31	0.00
KL	2010	94.4 %	M2	1.E-02	1.03	3.E+55	2.E+56	-	-	24.76	0.00
SCE	2010	96.8 %	M1	1.E-02	1.17	21.38	2.E+05	0.46	12,138	46.57	0.00

KL	2012	87.7%	M1	1.E-01	0.95	4.E+08	7.E+12	-	-	0.18	0.00
SCE	2012	96.5%	M1	1.E-03	1.33	5.94	5.E+04	0.44	13,955	2235.74	0.00
KL	2014	98.0%	M2	1.E-02	0.92	2.E+07	1.E+12	-	-	0.08	0.00
SCE	2014	97.6%	M2	1.E-02	1.24	7.E+00	7.E+04	0.46	14,267	10.68	0.00
KS	2014	97.1%	M2	1.E-02	0.92	6.E+06	2.E+11	-	-	0.11	0.00

Tabla C.12: Estimaciones para la región de ingresos bajos-medios: modelo Singh-Maddala. Las columnas muestran el porcentaje de la población que obtiene ingresos menores que la cota inferior de la región alta de la distribución (%); el método de estimación (M) donde M1 consiste en bajar el criterio de convergencia del algoritmo VGAM, mientras que M2 usa un modelo más simple, el modelo Fisk, para proveer valores iniciales al modelo más complejo, Singh-Maddala; el criterio de convergencia (Crit). Si los espacios de las columnas M y Crit están en blanco es porque se utilizaron los criterios convencionales del algoritmo VGAM. Asimismo, se presentan los coeficientes de forma ( $a$ ) y ( $q$ ) y de escala ( $\beta$ ), estimaciones paramétricas del índice de Gini (Gini) y la media (Media), la medida de divergencia asociada (D) y el *valor p* de la prueba de plausibilidad (*Valor p*). Los *valores p* que estadísticamente son significativos se presentan en negritas.

Año	KL			SCE			KS		
	Criterio	Método	LR	Criterio	Método	LR	Criterio	Método	LR
<b>1984</b>	1E-01	M2	<b>-4897</b>	1E-01	M1	<b>282357</b>	1E-01	M1	<b>281406</b>
<b>1989</b>							1E-03	M1	<b>1636</b>
<b>1992</b>	1E-02	M1	<b>9132</b>				1E-02	M1	<b>6051</b>
<b>1994</b>	1E-01	M1	<b>3522</b>				1E-02	M1	-3503
<b>1996</b>							1E-02	M1	<b>171606</b>
<b>1998</b>							1E-02	M1	<b>1901</b>
<b>2000</b>	1E-02	M1	0	1E-01	M2	<b>-4144</b>	1E-01	M2	<b>-4144</b>
<b>2002</b>	1E-01	M2	<b>-2697</b>				1E-01	M2	<b>-812</b>
<b>2004</b>									
<b>2005</b>									
<b>2006</b>	1E-03	M1	<b>8448</b>	1.E-02	M1	<b>4399</b>	1.E-02	M2	<b>-7374</b>
<b>2008</b>	1E-03	M2	<b>-762</b>	1.E-03	M1	<b>8499</b>	1.E-02	M2	<b>-8137</b>
<b>2010</b>	1E-02	M2	<b>-8113</b>	1.E-02	M1	<b>5508</b>	1.E-02	M1	<b>5508</b>
<b>2012</b>	1E-01	M1	<b>205531</b>	1.E-03	M1	<b>8279</b>	1.E-01	M1	<b>209310</b>
<b>2014</b>	1E-02	M2	<b>-175</b>	1.E-02	M2	<b>-324</b>	1.E-02	M2	<b>-5361</b>

---

Tabla C.13: Criterios de convergencia para modelos Singh-Maddala. Las columnas muestran el estadístico de razón e verosimilitud normalizada de acuerdo a  $\sqrt{n}R/\sigma$  ( $LR$ ) para la prueba entre el método 1 -bajar el criterio de convergencia del algoritmo VGAM- y el método 2 -usar un modelo más simple, el modelo Fisk, para proveer valores iniciales al modelo más complejo Singh-Maddala-; el método con el que se estimó el modelo Singh-Maddala de acuerdo a la prueba de verosimilitud; y por último, el criterio de convergencia para el algoritmo VGAM. Los  $LR$  que estadísticamente son significativos se presentan en negritas. Los renglones en blanco Para los modelos que no se obtuvieron.

## APÉNDICE D

# ANÁLISIS DE LA SIMULACIÓN

---

### D.1 ESTABILIDAD DEL PROCESO DE SIMULACIÓN DEL MSBA

$\delta$	L	Rezago					
		0	1	2	3	4	5
0.4	1	1	0.441	0.1538	0.0162	-0.0593	-0.084
	2	1	0.4933	0.199	0.024	-0.0496	-0.0873
	3	1	0.5031	0.2232	0.0479	-0.056	-0.0978
	4	1	0.5135	0.2159	0.0358	-0.0306	-0.0775
	5	1	0.5668	0.2721	0.0981	-0.0186	-0.0887
	6	1	0.5544	0.2863	0.1138	0.0106	-0.0567
0.5	1	1	0.4828	0.1989	0.0512	-0.0434	-0.0962
	2	1	0.5273	0.2555	0.0869	-0.015	-0.0693
	3	1	0.5672	0.3041	0.1376	0.0155	-0.0607
	4	1	0.5908	0.3227	0.1459	0.0231	-0.0633
	5	1	0.5937	0.3262	0.1434	0.0053	-0.068
	6	1	0.6074	0.3334	0.1509	0.0249	-0.0657
0.6	1	1	0.5223	0.2651	0.0893	-0.0044	-0.0661
	2	1	0.5738	0.2875	0.1186	0.0013	-0.0724
	3	1	0.5925	0.3055	0.1102	-0.0092	-0.0714
	4	1	0.6271	0.3519	0.1505	0.0193	-0.0701
	5	1	0.6203	0.3508	0.1707	0.0431	-0.0405
	6	1	0.5905	0.3372	0.1686	0.0403	-0.0595

Figura D.1: Promedio de la autocorrelación serial del ingreso por agente.

Tabla D.1: Coeficientes de variación para los estadísticos descriptivos de la simulación del MSBA.

<b>Estadístico</b>	0.4 <sub>2</sub>	0.4 <sub>3</sub>	0.4 <sub>4</sub>	0.4 <sub>5</sub>	0.5 <sub>2</sub>	0.5 <sub>3</sub>	0.5 <sub>4</sub>	0.5 <sub>5</sub>	0.6 <sub>2</sub>	0.6 <sub>3</sub>	0.6 <sub>4</sub>	0.6 <sub>5</sub>
Media	14,547	14,547	14,547	14,547	14,547	14,547	14,547	14,547	14,547	14,547	14,547	14,547
Var	12,259	11,353	9,863	8,879	15,153	12,787	10,887	10,654	17,712	15,549	14,297	12,723
(CV)	0.13	0.19	0.23	0.17	0.16	0.16	0.09	0.15	0.21	0.27	0.16	0.20
Asimetría	1.25	1.49	1.22	1.01	2.20	1.80	1.21	1.37	2.27	2.20	2.17	1.47
(CV)	0.29	0.34	0.51	0.30	0.26	0.23	0.24	0.17	0.35	0.32	0.44	0.41
Curtosis	4.61	5.97	5.28	4.18	9.87	7.88	4.57	5.25	10.02	9.80	10.57	5.80
(CV)	0.38	0.38	0.53	0.28	0.50	0.35	0.32	0.25	0.62	0.54	0.80	0.60
Mediana	11,431	12,013	12,710	12,654	10,177	10,909	12,073	11,607	8,364	9,680	10,040	11,129
(CV)	0.07	0.08	0.06	0.04	0.11	0.09	0.06	0.06	0.06	0.10	0.08	0.15
Gini	0.45	0.40	0.36	0.33	0.50	0.44	0.40	0.38	0.57	0.51	0.48	0.45
(CV)	0.06	0.07	0.09	0.07	0.05	0.07	0.03	0.07	0.04	0.08	0.05	0.07

Tabla D.1: Coeficientes de variación para los estadísticos descriptivos de la simulación del MSBA.

## D.2 ESTIMACIONES DE LOS MODELOS SOBRE LOS DATOS SIMULADOS

Tabla D.2: Estimaciones para la región total de los datos simulados

Modelo	$\delta$	L	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	Gini	Media	
	0.4	1		0.63	23,006		0.59	14,547	
		2		1.20	12,145		0.47	14,547	
		3		1.71	8,495		0.40	14,547	
		4		2.25	6,453		0.36	14,547	
		5		2.85	5,104		0.32	14,547	
		6		3.27	4,450		0.30	14,547	
	Gamma	0.5	1		0.51	28,491		0.63	14,547
			2		0.91	15,923		0.52	14,547
			3		1.28	11,383		0.45	14,547
			4		1.70	8,572		0.40	14,547
			5		2.00	7,285		0.38	14,547
			6		2.20	6,620		0.36	14,547
	0.6	1		0.40	36,628		0.68	14,547	
		2		0.67	21,868		0.58	14,547	
		3		0.94	15,427		0.51	14,547	
		4		1.13	12,902		0.48	14,547	
		5		1.31	11,124		0.45	14,547	
		6		1.49	9,793		0.43	14,547	
	0.4	1	1.83			8.62	0.81	29,674	
		2	1.16			9.11	0.59	17,725	
		3	0.90			9.27	0.47	15,827	
		4	0.75			9.35	0.40	15,211	
		5	0.65			9.40	0.35	14,896	
		6	0.60			9.42	0.33	14,843	
	Log normal	0.5	1	2.19			8.35	0.88	46,049
			2	1.39			8.95	0.67	20,142
			3	1.09			9.15	0.56	17,063
			4	0.90			9.26	0.48	15,851
			5	0.80			9.31	0.43	15,305
			6	0.75			9.34	0.41	15,129
	0.6	1	2.73			7.93	0.95	114,665	
		2	1.71			8.67	0.77	24,999	
		3	1.33			8.97	0.65	18,914	
		4	1.16			9.08	0.59	17,149	
		5	1.04			9.16	0.54	16,306	
		6	0.98			9.21	0.51	16,234	
	1	2.63	0.21	30,141		0.59	15,119		

Tabla D.2: (continued)

Modelo	$\delta$	L	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	Gini	Media
Dagum	0.4	2	3.04	0.31	22,953		0.47	15,017
		3	3.53	0.35	21,155		0.40	14,876
		4	3.54	0.45	18,243		0.36	14,779
		5	3.60	0.56	16,563		0.33	14,798
		6	4.16	0.49	17,498		0.30	14,720
		1	2.44	0.19	33,023		0.64	15,274
	2	2.83	0.27	25,451		0.52	15,109	
	3	3.07	0.32	22,319		0.46	15,015	
	4	3.45	0.35	21,018		0.40	14,927	
	5	3.24	0.46	18,242		0.39	14,935	
	6	3.30	0.49	17,748		0.37	14,950	
	1	2.50	0.14	41,427		0.68	15,488	
	2	2.19	0.28	24,902		0.61	15,803	
	3	2.57	0.31	23,071		0.53	15,315	
	4	2.76	0.34	21,957		0.49	15,176	
	5	2.79	0.38	20,360		0.47	15,189	
	6	3.34	0.33	21,933		0.43	14,938	

Tabla D.2: Estimaciones para la región total de los datos simulados. Se presenta la siguiente información:  $a/\sigma$  es el parámetro de forma  $a$  para las distribuciones Dagum y el parámetro  $\sigma$  para el modelo Log normal;  $p/q$  es el parámetro de forma  $p$  para el modelo Dagum y Gamma;  $\beta$  es el parámetro de escala; y  $\mu$  el parametro de localización para el modelo Log normal. Las columnas *Gini* y *Media* presentan el índice de Gini y la Media que se desprenden de los coeficientes estimados.

Tabla D.3: Estimaciones para la región de ingresos altos de los datos simulados

Modelo	$\delta$	L	\$ %	$p/q$	$b/\beta$	Gini	Media
Pareto	0.4	1	5 %	3.03	49,162	0.20	73,436
		2	12 %	3.17	29,457	0.19	43,054
		3	4 %	4.44	35,991	0.13	46,452
		4	2 %	6.82	41,349	0.08	48,457
		5	6 %	5.05	30,157	0.11	37,601
		6	6 %	5.63	29,033	0.10	35,299
	1	2 %	3.69	73,633	0.16	101,051	
	2	4 %	4.19	48,161	0.14	63,237	
	3	3 %	5.14	45,955	0.11	57,049	
	4	5 %	4.73	35,945	0.12	45,580	
	5	2 %	6.49	42,242	0.08	49,932	
	6	9 %	4.66	29,313	0.12	37,323	

Tabla D.3: (continued)

Modelo	$\delta$	L	%	$p/q$	$b/\beta$	Gini	Media
		1	4 %	3.28	65,413	0.18	94,059
		2	8 %	2.93	42,119	0.21	63,985
	0.6	3	3 %	5.21	54,638	0.11	67,631
		4	6 %	3.99	40,701	0.14	54,325
		5	4 %	4.86	44,569	0.11	56,106
		6	2 %	5.34	45,801	0.10	56,352

Tabla D.3: Estimaciones para la región de ingresos altos de los datos simulados. Se presenta la siguiente información: el porcentaje de la población que obtiene ingresos mayores e iguales que el ingreso mínimo óptimo (%), el coeficiente de Pareto ( $p/q$ ), los niveles óptimos de ingreso mínimo ( $b/\beta$ ), el coeficiente de Gini (*Gini*) y Media (*Media*) estimados de manera paramétrica .

Tabla D.4: Estimaciones para la región de ingresos bajos y medios de los datos simulados

Modelo	$\delta$	L	%	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	Gini	Media
		1	95 %	5.01	0.10	32,121		0.53	12,054
		2	88 %	8.28	0.10	23,825		0.39	11,366
	0.4	3	96 %	5.32	0.21	22,914		0.36	13,499
		4	98 %	4.24	0.36	19,176		0.34	14,021
		5	94 %	4.87	0.38	17,626		0.29	13,434
		6	94 %	6.20	0.30	18,833		0.27	13,521
		1	98 %	3.40	0.13	34,941		0.60	13,076
		2	96 %	4.34	0.16	27,322		0.47	13,000
Dagum	0.5	3	97 %	4.09	0.23	23,930		0.42	13,708
		4	95 %	5.26	0.21	22,823		0.36	13,437
		5	98 %	3.75	0.38	19,067		0.37	14,173
		6	91 %	5.52	0.26	19,308		0.32	12,716
		1	96 %	5.50	0.06	44,710		0.62	12,146
		2	92 %	3.94	0.15	25,725		0.53	10,963
	0.6	3	97 %	3.27	0.24	24,412		0.49	13,594
		4	94 %	4.36	0.20	23,765		0.43	12,687
		5	96 %	3.74	0.27	22,158		0.43	13,548
		6	98 %	4.16	0.25	23,252		0.40	14,069
		1	95 %	1.78			8.48	0.79	23,606
		2	88 %	1.08			8.91	0.56	13,313
	0.4	3	96 %	0.86			9.20	0.46	14,343
		4	98 %	0.73			9.31	0.39	14,459
		5	94 %	0.60			9.33	0.33	13,581

Tabla D.4: (continued)

Modelo	$\delta$	$L$	%	$a/\sigma$	$p/q$	$b/\beta$	$\mu$	<i>Gini</i>	<i>Media</i>
Log normal	0.5	6	94 %	0.56			9.36	0.31	13,631
		1	98 %	2.16			8.27	0.87	39,928
		2	96 %	1.35			8.86	0.66	17,438
		3	97 %	1.06			9.09	0.55	15,650
		4	95 %	0.87			9.19	0.46	14,270
		5	98 %	0.78			9.28	0.42	14,582
	6	91 %	0.69			9.23	0.38	12,953	
	0.6	1	96 %	2.69			7.77	0.94	87,964
		2	92 %	1.63			8.47	0.75	17,831
		3	97 %	1.29			8.91	0.64	17,014
		4	94 %	1.10			8.97	0.56	14,480
		5	96 %	1.00			9.09	0.52	14,697
6		98 %	0.96			9.18	0.50	15,311	

Tabla D.4: Estimaciones para la región de ingresos bajos y medios de los datos simulados. Se presenta la siguiente información:  $a/\sigma$  es el parámetro de forma  $a$  para las distribuciones Dagum y el parámetro  $\sigma$  para el modelo Log normal;  $p/q$  es el parámetro de forma  $p$  para el modelo Dagum y Gamma;  $\beta$  es el parámetro de escala; y  $\mu$  el parámetro de localización para el modelo Log normal. Las columnas *Gini* y *Media* presentan el índice de Gini y la Media que se desprenden de la estimación de los coeficientes estimados.

## APÉNDICE E

# TÉRMINOS Y CONCEPTOS ÚTILES.

---

- Experimento probabilístico. Es aquel que experimento con ocurrencias como lo son tirar una moneda, tirar un dado, o la observación de la caída de lluvia en un día particular donde el contexto complejo natural conlleva a una probabilidad de resultado.
- Espacio muestral. El conjunto de todos los posibles resultados de un experimento probabilístico. También se le es conocido como muestra o evento.
- Variable aleatoria. Es una función que define eventos definidos en un espacio muestral a un conjunto de valores. Diferentes variables aleatorias pueden ser definidas con base en un mismo experimento. Por ejemplo, en el caso de tirar dos monedas, el número de cabezas observadas es una variable aleatoria.
- Variante. En el estudio de las las funciones de distribuciones conveniente trabajar en terminso de variantes. Ésta es una generalización de la idea de variable aleatoria, y en un sentido mantiene similares propiedades probabilísticas, pero es definida sin alguna referencia particular al experimento probabilístico. Por ejemplo, el numero de cabezas o colas observadas en un juego independiente de tirar monedas son elementos de la misma variante, dado que son los mismos factores probabilísticos los que gobiernan al resultado .
- Rango. El rango de  $X$  se define como sigue: permita ser ser a  $\mathcal{R}_x$  ser el conjunto

de todos los valores que la variable  $X$  puede tomar.

- **Cuantil.** Se define como puntos tomados, en intervalos regulares, de la función de densidad de una variable aleatoria. Es decir, para la variable aleatoria  $X$  permita al número real  $x$  denotar un elemento general del rango  $\mathcal{R}_x$ .
- **Asimetría estadística.** Es una medida de asimetría de la distribución de probabilidad de una variable aleatoria con respecto a su media. La interpretación de cualitativa de la asimetría es complicada y no sugiere, de manera estricta, una relación entre media y mediana. Por ejemplo, si una distribución uni modal tiene sesgo negativo entonces la cola izquierda de la distribución es más larga o ancha que la del lado derecho pero no distingue la forma, y de manera contraria si la asimetría es positiva. Sin embargo, la asimetría a veces no sigue una regla simple, por ejemplo, en el caso en que una cola sea larga y la otra ancha, si el sesgo es cero se refiere a una distribución balanceada, como puede ser una distribución simétrica pero también aplica a una distribución asimétrica.
- **Curtosis.** Del griego *kyrtos* o *kurtos*, que significa curvo. Es una medida sobre la cola de la función de densidad de una variable aleatoria con el objeto de describir la forma de la función (igual que la asimetría) y esta basada en el cuarto momento de la distribución. Para estimarla, se puede hacer de acuerdo a distribuciones teóricas o por la muestra de una población (igual que la asimetría). La interpretación de la misma depende de la medida de curtosis que se use, pero en general, se trata de describir el peso de la cola, el ancho de la cola, y la existencia de valores atípicos en el datos. La medida estándar de la curtosis es el que desarrollo Karl Pearson, basado en el cuarto momento de la distribución, y se enfoca en medir las colas pesadas pero no los picos. Si se presenta un estadístico de Pearson grande significaría que gran parte de la varianza es resultado de raras desviaciones extremas en contraste con frecuentes desviaciones de modesto tamaño. La curtosis para la distribución normal es 3 por lo que es común comparar la curtosis observada con este valor. Una distribución con curtosis menor que 3 se dice que es platicúrtica, que puede

describirse con colas delgadas, ejemplo de una distribución de este tipo es la distribución uniforme. Por otro lado, distribuciones con una curtosis mayor a 3 se hacen llamar leptocúrtica, como por ejemplo la distribución de Laplace, este tipo de distribuciones tienden a presentar colas mas anchas que la normal. Es común usar una versión ajustada de la curtosis de Pearson llamada exceso de curtosis, que es la curtosis de Pearson menos 3.

# BIBLIOGRAFÍA

---

- [1] AGHION, P., «Schumpeterian growth theory and the dynamics of income inequality», *Econometrica*, **70**(3), págs. 855–882, 2002.
- [2] AGHION, P. y J. G. WILLIAMSON, *Growth, inequality, and globalization: theory, history, and policy*, Cambridge University Press, 1998.
- [3] ANGLE, J., «The surplus theory of social stratification and the size distribution of personal wealth», *Social Forces*, **65**(2), págs. 293–326, 1986.
- [4] ATKINSON, A. B. y F. BOURGUIGNON, «Introduction: Income distribution and economics», *Handbook of income distribution*, **1**, págs. 1–58, 2000.
- [5] ATKINSON, A. B. y F. BOURGUIGNON, *Handbook of Income Distribution SET vols. 2A-2B*, Elsevier, 2014.
- [6] BANDOURIAN, R., J. McDONALD y R. S. TURLEY, «A comparison of parametric models of income distribution across countries and over time», *Estadística*, **55**, págs. 135–152, 2003.
- [7] BERTOLA, G., R. FOELLMY y J. ZWEIMÜLLER, *Income distribution in macroeconomic models*, Princeton University Press, 2014.
- [8] BOLTZMANN, L., *Lectures on gas theory*, Courier Corporation, 2012.
- [9] BOOS, D. D. y L. A. STEFANSKI, *Essential Statistical Inference*, Springer-Verlag New York, New York, USA, 2013.

- [10] BOUCHAUD, J.-P. y M. MÉZARD, «Wealth condensation in a simple model of economy», *Physica A: Statistical Mechanics and its Applications*, **282**(3), págs. 536–545, 2000.
- [11] BRZEZINSKI, M., «Do wealth distributions follow power laws? Evidence from ‘rich lists’», *Physica A: Statistical Mechanics and its Applications*, **406**, págs. 155–162, 2014.
- [12] CHAKRABORTI, A. y B. K. CHAKRABARTI, «Statistical mechanics of money: how saving propensity affects its distribution», *The European Physical Journal B-Condensed Matter and Complex Systems*, **17**(1), págs. 167–170, 2000.
- [13] CHAKRABORTI, A., I. M. TOKE, M. PATRIARCA y F. ABERGEL, «Econophysics review: II. Agent-based models», *Quantitative Finance*, **11**(7), págs. 1013–1041, 2011.
- [14] CHAMPEMOWNE, D., «A Model of Income Distribution», *The Economic Journal*, **63**(250), págs. 318–351, 1953.
- [15] CHATTERJEE, A. y B. K. CHAKRABARTI, «Ideal-gas like markets: effect of savings», en *Econophysics of Wealth Distributions*, Springer, págs. 79–92, 2005.
- [16] CHATTERJEE, A., B. K. CHAKRABARTI y S. MANNA, «Pareto law in a kinetic model of market with random saving propensity», *Physica A: Statistical Mechanics and its Applications*, **335**(1), págs. 155–163, 2004.
- [17] CHEN, S., «Agent-based modeling for wealth and income distributions», , 2011.
- [18] CHOTIKAPANICH, D. y W. E. GRIFFITHS, «Estimating income distributions using a mixture of gamma densities», en *Modeling Income Distributions and Lorenz Curves, Economic Studies in Equality, Social Exclusion and Well-Being*, tomo 5, Springer, 2008.
- [19] CLAUSET, A., C. R. SHALIZI y M. E. NEWMAN, «Power-law distributions in empirical data», *SIAM review*, **51**(4), págs. 661–703, 2009.

- [20] CLEMENTI, F. y M. GALLEGATI, «Pareto's law of income distribution: evidence for Germany, the United Kingdom, and the United States», en A. Chatterjee, S. Yarlagadda y B. K. Chakrabarti (editores), *Econophysics of wealth distributions: Econophys-Kolkata I*, Springer Science & Business Media, 2005.
- [21] CLEMENTI, F. y M. GALLEGATI, «Power law tails in the Italian personal income distribution», *Physica A: Statistical Mechanics and its Applications*, **350**(2), págs. 427–438, 2005.
- [22] COVER, T. M. y J. A. THOMAS, «Entropy, relative entropy and mutual information», *Elements of Information Theory*, **2**, págs. 1–55, 1991.
- [23] COWELL, F., «Measuring inequality, LSE perspectives in economic analysis», , 2008.
- [24] DAGUM, C., «A new model of personal income distribution: specification and estimation», en C. Duangkamon (editor), *Modeling Income Distributions and Lorenz Curves, Economic Studies in Equality, Social Exclusion and Well-Being*, tomo 5, Springer, 2008.
- [25] DAVISON, A. C. y D. V. HINKLEY, *Bootstrap methods and their application*, tomo 1, Cambridge university press, 1997.
- [26] DI MATTEO, T., T. ASTE y S. HYDE, «Exchanges in complex networks: income and wealth distributions», *arXiv preprint cond-mat/0310544*, 2003.
- [27] DÍAZ, A., F. GLOVER, H. M. GHAZIRI, J. GONZÁLEZ, M. LAGUNA, P. MOSCATO y F. T. TSENG, «Optimización heurística y redes neuronales», , 2000.
- [28] DRAGULESCU, A. y V. M. YAKOVENKO, «Statistical mechanics of money», *The European Physical Journal B-Condensed Matter and Complex Systems*, **17**(4), págs. 723–729, 2000.
- [29] DRĂGULESCU, A. y V. M. YAKOVENKO, «Evidence for the exponential distribution of income in the USA», *The European Physical Journal B-Condensed Matter and Complex Systems*, **20**(4), págs. 585–589, 2001.

- [30] DRĂGULESCU, A. y V. M. YAKOVENKO, «Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States», *Physica A: Statistical Mechanics and its Applications*, **299**(1), págs. 213–221, 2001.
- [31] DRAGULESCU, A. A., «Applications of physics to economics and finance: money, income, wealth, and the stock market», *arXiv preprint cond-mat/0307341*, 2003.
- [32] ESTEBAN, J., «Income-share elasticity, density functions and the size distribution of income», Mimeographed manuscript, University of Barcelona, 1981.
- [33] ESTEBAN, J., «Income-share elasticity and the size distribution of income», *International Economic Review*, págs. 439–444, 1986.
- [34] FORBES, C., M. EVANS, N. HASTINGS y B. PEACOCK, *Statistical distributions*, John Wiley & Sons, 2011.
- [35] FUJIWARA, Y., «Pareto-Zipf, Gibrat's laws, detailed-balance and their breakdown», en A. Chatterjee, S. Yarlagadda y B. K. Chakrabarti (editores), *Econophysics of Wealth Distributions: Econophys-Kolkata I*, Springer, 2005.
- [36] FUJIWARA, Y., W. SOUMA, H. AOYAMA, T. KAIZOJI y M. AOKI, «Growth and fluctuations of personal income», *Physica A: Statistical Mechanics and its Applications*, **321**(3), págs. 598–604, 2003.
- [37] GALOR, O. y J. ZEIRA, «Income distribution and macroeconomics», *The review of economic studies*, **60**(1), págs. 35–52, 1993.
- [38] GASTWIRTH, J. L., «The estimation of the Lorenz curve and Gini index», *The Review of Economics and Statistics*, págs. 306–316, 1972.
- [39] GIBRAT, R., «Les Inégalités Economiques (Recueil Sirey, Paris)», en *On economic inequalities*, Macmillan New York, 1931.

- [40] GOLDEN, J., «A simple geometric approach to approximating the Gini coefficient», *The Journal of Economic Education*, **39**(1), págs. 68–77, 2008.
- [41] GUSMAN, S. R., M. LAGUNA y J. IGLESIAS, «Wealth distribution in a network with correlations between links and success», en *Econophysics of Wealth Distributions*, Springer, págs. 149–158, 2005.
- [42] HANDCOCK, M. S., *Relative Distribution Methods*, Los Angeles, CA, version 1.6-4. Project home page at [urlhttp://www.stat.ucla.edu/handcock/RelDist](http://www.stat.ucla.edu/handcock/RelDist), 2015, URL <http://CRAN.R-project.org/package=reldist>.
- [43] HANDCOCK, M. S. y M. MORRIS, *Relative Distribution Methods in the Social Sciences*, Springer, New York, iSBN 0-387-98778-9, 1999, URL <http://www.stat.ucla.edu/~handcock/RelDist>.
- [44] HANSEN, P., N. MLADENOVIC y J. A. M. PÉREZ, «Variable neighbourhood search», *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, **19**, págs. 77–92, 2003.
- [45] HARRELL, F. E., «Hmisc: A library of miscellaneous S functions», Available from [biostat.mc.vanderbilt.edu/s/Hmisc](http://biostat.mc.vanderbilt.edu/s/Hmisc), 2005.
- [46] HAUSSER, J. y K. STRIMMER, *entropy: Estimation of Entropy, Mutual Information and Related Quantities*, r package version 1.2.1, 2014, URL <http://CRAN.R-project.org/package=entropy>.
- [47] HECKMAN, J. J. y E. LEAMER, *Handbook of econometrics*, tomo 6, Elsevier, 2007.
- [48] HOROWITZ, J. L., «The bootstrap», *Handbook of econometrics*, **5**, págs. 3159–3228, 2001.
- [49] INEGI, «Encuesta Ingreso Gasto de los Hogares», <http://www.inegi.org.mx/est/contenidos/Proyectos/encuestas/hogares/regulares/enigh/>, microdatos de la encuesta para los años 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012, 2014.

- [50] INEGI, *Encuesta Nacional de Ingresos y Gastos de los Hogares: ENIGH 2014: descripción de la base de datos*, 2015.
- [51] JONES, C. I. y J. KIM, «A Schumpeterian model of top income inequality», *Informe técnico*, National Bureau of Economic Research, 2014.
- [52] JR, F. E. H., WITH CONTRIBUTIONS FROM CHARLES DUPONT y MANY OTHERS., *Hmisc: Harrell Miscellaneous*, r package version 3.15-0, 2015, URL <http://CRAN.R-project.org/package=Hmisc>.
- [53] KLEIBER, C., «A guide to the Dagum distributions», en C. Duangkamon (editor), *Modeling Income Distributions and Lorenz Curves, Economic Studies in Equality, Social Exclusion and Well-Being*, tomo 5, Springer, 2008.
- [54] KRUGMAN, P., *The self-organizing economy.*, 338.9 KRU 1996. CIMMYT., 1996.
- [55] KUZNETS, S., «Economic growth and income inequality», *The American economic review*, **45**(1), págs. 1–28, 1955.
- [56] KWASNICKI, W., «25 Schumpeterian modelling», *Elgar Companion to Neo-Schumpeterian Economics*, pág. 389, 2007.
- [57] LUCAS, R. E., «Econometric policy evaluation: A critique», en *Carnegie-Rochester conference series on public policy*, tomo 1, North-Holland, págs. 19–46, 1976.
- [58] LUX, T., «Emergent statistical wealth distributions in simple monetary exchange models: a critical review», en *Econophysics of wealth distributions*, Springer, págs. 51–60, 2005.
- [59] MANDELBROT, B., «The Pareto-Levy law and the distribution of income», *International Economic Review*, **1**(2), págs. 79–106, 1960.
- [60] McDONALD, J. B., «Some generalized functions for the size distribution of income», *Econometrica: journal of the Econometric Society*, págs. 647–663, 1984.

- [61] McDONALD, J. B. y Y. J. XU, «A generalization of the beta distribution with applications», *Journal of Econometrics*, **66**(1), págs. 133–152, 1995.
- [62] MIÑARRO, A., «Estimación no paramétrica de la función de densidad», *Documento de Trabajo. Universidad de Barcelona. Barcelona, España*, 1998.
- [63] MINCER, J., «Investment in human capital and personal income distribution», *The journal of political economy*, págs. 281–302, 1958.
- [64] MISHKIN, F. S., J. G. M. ARAIZA, A. S. VARGAS, I. KATZ y A. R. PORRAS, *Moneda, banca y mercados financieros*, Pearson Educación México D. F, 2008.
- [65] MURGUE, T. y C. DE LA HIGUERA, «Distances between distributions: Comparing language models», en *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, Springer, págs. 269–277, 2004.
- [66] PARETO, V., «Manual of political economy», , 1971.
- [67] PERSSON, T. y G. TABELLINI, «Is inequality harmful for growth?», *The American Economic Review*, págs. 600–621, 1994.
- [68] PINKOVSKIY, M. y X. SALA-I MARTIN, «Parametric estimations of the world distribution of income», *Informe técnico*, National Bureau of Economic Research, 2009.
- [69] PRESS, W. H., *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge university press, 2007.
- [70] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014, URL <http://www.R-project.org/>.
- [71] RICHMOND, P., S. HUTZLER, R. COELHO y P. REPETOWICZ, *A review of empirical studies and models of income distributions in society*, Wiley-VCH: Berlin, Germany, 2006.

- [72] RIMOLDINI, L., «Weighted skewness and kurtosis unbiased by sample size and Gaussian uncertainties», *Astronomy and Computing*, **5**, págs. 1–8, 2014.
- [73] RIVERA, F. Y. M., P. P. RODRÍGUEZ y S. P. ELIZALDE, «Ajuste del ingreso en México con un enfoque bayesiano», *Estudios Económicos*, págs. 273–293, 2012.
- [74] SALEM, A. y T. MOUNT, «A convenient descriptive model of income distribution: the gamma density», *Econometrica: journal of the Econometric Society*, págs. 1115–1127, 1974.
- [75] SAMUELSON, P. A., W. D. NORDHAUS y E. RABASCO, *Macroeconomía: con aplicaciones a México*, McGraw-Hill, 1998.
- [76] SCAFETTA, N., S. PICOZZI y B. J. WEST, «A trade-investment model for distribution of wealth», *Physica D: Nonlinear Phenomena*, **193**(1), págs. 338–352, 2004.
- [77] SHCP, «Distribución del pago de impuestos y recepción del gasto público por deciles de hogares y personas para el año 2012», [http://www.hacienda.gob.mx/INGRESOS/ingresos\\_distribucion\\_pago/doc\\_Ingreso-Gasto\\_2012\\_enero2015.pdf](http://www.hacienda.gob.mx/INGRESOS/ingresos_distribucion_pago/doc_Ingreso-Gasto_2012_enero2015.pdf), 2012.
- [78] SHEATHER, S. J. y M. C. JONES, «A reliable data-based bandwidth selection method for kernel density estimation», *Journal of the Royal Statistical Society. Series B (Methodological)*, págs. 683–690, 1991.
- [79] SHEATHER, S. J. *et al.*, «Density estimation», *Statistical Science*, **19**(4), págs. 588–597, 2004.
- [80] SILVER, J., E. SLUD y K. TAKAMOTO, «Statistical equilibrium wealth distributions in an exchange economy with stochastic preferences», *Journal of Economic Theory*, **106**(2), págs. 417–435, 2002.
- [81] SILVERMAN, B. W., *Density estimation for statistics and data analysis*, tomo 26, CRC press, 1986.

- [82] SINGH, S. K. y G. MADDALA, «A function for size distribution of incomes», en C. Duangkamon (editor), *Modeling Income Distributions and Lorenz Curves, Economic Studies in Equality, Social Exclusion and Well-Being*, tomo 5, Springer, 2008.
- [83] SINHA, S., «Evidence for power-law tail of the wealth distribution in India», *Physica A: Statistical Mechanics and its Applications*, **359**, págs. 555–562, 2006.
- [84] SORNETTE, D., «Physics and financial economics (1776–2014): puzzles, Ising and agent-based models», *Reports on Progress in Physics*, **77**(6), pág. 062 001, 2014.
- [85] SOUMA, W., «Physics of personal income», en *Empirical science of financial fluctuations*, Springer, págs. 343–352, 2002.
- [86] SOUMA, W., Y. FUJIWARA y H. AOYAMA, «Complex networks and economics», *Physica A: Statistical Mechanics and its Applications*, **324**(1), págs. 396–401, 2003.
- [87] SOUMA, W. y M. NIREI, «Empirical study and model of personal income», en A. Chatterjee, S. Yarlagadda y B. K. Chakrabarti (editores), *Econophysics of Wealth Distributions: Econophysics-Kolkata I*, Springer, 2005.
- [88] SZÉKELY, M., «Pobreza y desigualdad en México entre 1950 y 2004», *El Trimestre Económico*, págs. 913–931, 2005.
- [89] SZÉKELY, M. y M. HILGERT, «What drives differences in inequality across countries?», , 2000.
- [90] TODA, A. A., «The double power law in income distribution: Explanations and evidence», *Journal of Economic Behavior & Organization*, **84**(1), págs. 364–381, 2012.
- [91] UN DEVELOPMENT PROGRAMME (UNDP), «Human Development Report 2014», disponible en

- <http://www.undp.org/content/undp/es/home/librarypage/hdr/2014-human-development-report.html>, 2014.
- [92] VUONG, Q. H., «Likelihood ratio tests for model selection and non-nested hypotheses», *Econometrica: Journal of the Econometric Society*, págs. 307–333, 1989.
- [93] WILLIS, G. y J. MIMKES, «Evidence for the independence of waged and unwaged income, evidence for Boltzmann distributions in waged income, and the outlines of a coherent theory of income distribution», *arXiv preprint cond-mat/0406694*, 2004.
- [94] YAKOVENKO, V. M. y J. B. ROSSER JR, «Colloquium: Statistical mechanics of money, wealth, and income», *Reviews of Modern Physics*, **81**(4), pág. 1703, 2009.
- [95] YAKOVENKO, V. M. y A. C. SILVA, «Two-class structure of income distribution in the USA: Exponential bulk and power-law tail», en A. Chatterjee, S. Yarlagadda y B. K. Chakrabarti (editores), *Econophysics of wealth distributions: Econophys-Kolkata I*, Springer Science & Business Media, 2005.
- [96] YARLAGADDA, S. y A. DAS, «A Stochastic Trading Model of Wealth Distribution», en *Econophysics of Wealth Distributions*, Springer, págs. 137–148, 2005.
- [97] YEE, T. W., *Vector Generalized Linear and Additive Models: With an Implementation in R*, Springer, New York, USA, 2015.