



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANALOGÍA ECONOFÍSICA ENTRE LAS
DISTRIBUCIONES DE RIQUEZA GLOBAL PER
CAPITA EN CIUDADES Y ENERGÍA DE BOLTZMANN
EN GASES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A :

DIEGO BALAM SÁNCHEZ ESPINOSA

TUTOR

MARCELO DEL CASTILLO MUSSOT
CIUDAD UNIVERSITARIA, CD.MX., 2022





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradezco a mi familia, ya que siempre me han apoyado en todas las decisiones que he tomado. Gracias por todo.

Contents

1	Introducción	1
1.1	Motivación	1
1.2	Objetivo	2
1.3	Antecedentes	3
2	Marco Teórico	5
2.1	Econofísica	5
2.2	Distribución de Boltzmann	6
2.3	Mecánica Estadística de la Distribución del Dinero	10
2.4	Leyes de Potencias	11
2.5	Distribución Exponencial Estirada (SED)	13
3	Modelo	14
4	Base de Datos	16
4.1	Indicadores Económicos	16
4.2	OCDE	17
5	Resultados	20
5.1	Ciudades y Regiones pequeñas	20
5.2	Regiones	25
6	Conclusiones	32
	Referencias	34

<i>CONTENTS</i>	iii
Appendix A Tablas	37
A.1 Ciudades y Regiones más ricas	38
A.2 Ciudades y Regiones pequeñas	39
A.2.1 PCPIB-PPA Regiones pequeñas	39
A.2.2 PCPIB-PPA Ciudades	42
A.3 Regiones Grandes	45
A.3.1 PCPIB-PPA precios corrientes	45
A.3.2 PCPIB-PPA precios constantes	48
Appendix B Código Implementado	51

1 Introducción

1.1 Motivación

La distribución de la riqueza y la desigualdad son dos de los grandes problemas que se enfrentan en la economía mundial. Nuestro planeta tiene una cantidad limitada de recursos por lo que cómo se distribuyen estos, tiene un impacto en prácticamente todos los sectores, determinando las condiciones de vida de las personas. Actualmente, vivimos en un mundo muy desigual, la organización caritativa británica Oxfam asegura que 82% del dinero que se generó en el mundo en 2017 fue al 1% más rico de la población global. Este fenómeno está lejos de desaparecer, de hecho, la brecha entre ricos y pobres cada vez se hace más grande.

La desigualdad de riqueza también se observa entre países, los 20 miembros fundadores de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) representaban en 2010 el 40% del producto interno bruto (PIB) mundial y actualmente los países miembros y socios representan el 80% del PIB mundial. La desigual distribución de la riqueza en el planeta, tienen un impacto significativo en la calidad de vida de las personas; es por ello que es importante comprender la distribución de la riqueza.

La motivación de esta tesis surge del trabajo publicado en la revista *Physics A*, titulado *World per capita gross domestic product measured nominally and across countries with purchasing power parity: Stretched exponential or Boltzmann–Gibbs distribution* [1], que analiza la distribución de riqueza usando el PIB per cápita de los países reportados en la base de datos del Banco Mundial, en el periodo de 1990 a

2017.

EL propósito de esta tesis busca estudiar más a fondo la distribución de la riqueza. Es por ello que el análisis se realizara a nivel regional y de ciudades, no solo a nivel nacional. Lo que nos permite tener una muestra más grande y precisa; para ello se emplearan los datos de la OCDE.

1.2 Objetivo

Se estudiarán las distribuciones estadísticas de la riqueza per cápita en ciudades y regiones alrededor del mundo usando un enfoque de econofísica, la cual emplea herramientas de la física estadística. En un modelo muy exitoso de intercambio económico [2], se propone un sistema económico cerrado, en el cual el dinero al igual que la energía, se debe conservar. Por lo que se espera que la distribución de la riqueza siga una distribución de Boltzmann-Gibbs en gases de partículas, la cual denotaremos como DBG. El supuesto de conservación del dinero en períodos de tiempo no muy largos se basa en el hecho de que solo los bancos centrales pueden imprimir papel moneda, por lo que los agentes solo pueden intercambiar dinero entre ellos, más no aumentarlo.

En muchos países se ha demostrado que el ingreso de sus habitantes se ajusta bien a una distribución DBG, en particular para los individuos en la parte media de la distribución nacional de ingresos. Yakovenko y Silva encontraron que al analizar el ingreso en el caso de EEUU, había dos clases sociales: una abundante clase baja caracterizada por una distribución DBG y una clase alta que seguía una distribución de Pareto [3].

Estudiaremos la distribución del ingreso en ciudades y regiones utilizando datos de la Organización para la Cooperación y Desarrollo Económico (OCDE), la cual reporta datos de 700 ciudades y regiones alrededor del mundo entre los años del 2001 hasta el 2017. Se usará el Producto Interno Bruto (PIB) per cápita para representar una medida promedio de la riqueza individual por ciudad y región para analizar anualmente la distribución correspondiente. Además, se ajustarán estos promedios

con un factor que representa muy bien el costo de vida en cada ciudad denominado Paridad de Poder de Adquisitivo (PPA) el cual convierte diferentes monedas a una moneda común proporcionando una descripción más realista entre ingreso económico y consumo. Por ejemplo, en Noruega el ingreso personal en general es muy alto, pero también es muy alto el costo de vida ahí.

Utilizando una metodología similar a un estudio econofísico muy completo del PIB per cápita ajustado con el PPA para todos los países del mundo desde 1990 hasta 2017 [1], compararemos las distribuciones anuales del PIB per cápita de las ciudades y regiones de la OCDE, empleando el PIB per cápita ajustado con el PPA, explorando ajustes a distribuciones exponenciales tipo DBG, así como con leyes de potencias y con la distribución de Pareto. De esta manera se demostrará bajo que condiciones se cumple la validez del modelo econofísico en el ingreso promedio en ciudades usando el modelo econofísico planteado.

1.3 Antecedentes

La distribución de riqueza es un problema central de la economía moderna. Existen diversos modelos propuestos para ajustar la distribución del PIB per capita. Di Guilmi et al. [4] Propone que el PIB per capita de los países que se encuentran entre los percentiles 30 y 85 percentiles de la distribución siguen una ley de potencias en el periodo de 1960 a 1997. Bustos-Guajardo y Moukazel estudian el periodo de 1960 a 2013, en ese periodo la distribución mundial del PIB per capita sigue un modelo de inversión comercial que encaja muy con los datos; sin embargo, encuentra que existe una sesgo estadístico en favor de los países desarrollados en las condiciones del comercio internacional y esto genera divergencia en los valores del PIB per capita de 1960 a 2008 [5]. Otros estudios muestran la distribución de los porcentajes de crecimiento del PIB. A. Williams et al. Encuentran que para 167 países el PIB en un periodo de 1950-2011; el porcentaje de crecimiento del PIB se pueden aproximar con una distribución de Laplace en la parte central y una distribución de ley de potencias en las colas [6].

Un enfoque distinto es proporcionado por la econofísica que busca aplicar teorías y métodos originalmente desarrollados para la física en el estudio de los fenómenos económicos. Dentro de este enfoque, podemos analizar los mercados y fenómenos económicos considerando los como un sistema termodinámico. Un primer modelo del intercambio económico, propone que el dinero, al igual que la energía se debe conservar, por lo que la distribución de la riqueza debe seguir una distribución Boltzmann-Gibbs (DBG) para la energía [2]. Silva y Yakovenko [3] encuentran que en USA tiene dos estructuras bien definidas. Por un lado, la población de bajos recursos sigue una distribución DBG. Por otro lado, la población más rica tiene una distribución de ley de potencias de Pareto. Además, la distribución DBG es muy estable a lo largo del tiempo, mientras que la distribución de la clase rica es más volátil siguiendo el stockmarket. Resultados similares se encuentran para distintos países como en Canadá [7], México [8], Italia [9], el Reino Unido [10] y la Unión Europea [11]. En el estudio de R. Kutner et al [12]. Se divide a la población en tres grupos de ingreso, bajo, medio y alto. En el caso de la población de bajo ingreso se sigue una distribución de Boltzmann-Gibbs, para la clase media una distribución de Pareto y para la clase alta una ley de Zipf. Sitabhra Sinha [13] encuentra para la India resultados similares. La distribución de rango resultante parece implicar una cola de ley de potencias para la distribución de riqueza, con un exponente de Pareto entre 0,81 y 0,92.

Marcelo et al. [1] encuentran que a pesar de que los datos de PIB per capita y PIB per capita en poder de paridad de compra son diferentes, se ajustan a dos distribuciones exponenciales estiradas (SED) una en la parte de los países más pobres y otra en el resto de la mayoría de los países.

2 Marco Teórico

2.1 Econofísica

La Econofísica busca aplicar métodos e ideas usualmente aplicados al estudio de la física en los fenómenos económicos. Si bien el concepto de econofísica surge en las últimas décadas del siglo XX, existieron acercamientos de la física con la economía mucho antes. Con el avance de la ciencia, se comenzó a intentar explicar los distintos fenómenos del universo por medio del raciocinio y la experimentación. La economía no fue ajena a este proceso, en el siglo XVIII Adam Smith y Walras buscaban encontrar leyes económicas que explicaran el comportamiento de los mercados, inspirados en las leyes de Newton.

A principios del siglo XX Louis Bachelier (1870-1940) en su tesis doctoral encuentra que los activos financieros siguen una caminata aleatoria; esto lo realiza cinco años antes de trabajos de Albert Einstein relacionados con caminatas aleatorias. Los trabajos de Bachelier fundaron la teoría financiera moderna y que servirán para en futuro dar paso a uno de los modelos financieros más importantes Black-Scholes. Fischer Black y Myron Scholes resolvieron el problema de encontrar una fórmula para valorar opciones que tengan aplicación en el cambio de acciones. Una de sus ideas fue aplicar la ecuación de calor aplicada en la física [14].

El término econofísica emergió con Stanley et al. en una conferencia en Calcuta en 1995 haciendo referencia a una rama de la física en el área de sistemas complejos, que busca estudiar las propiedades estadísticas de los mercados financieros aprovechando la gran cantidad de datos existentes [15].

Desde la convención de 1995 en Calcuta, el término econofísica ha ido ganando popularidad y el número de trabajos referentes a la econofísica han ido aumentando. La econofísica se basa principalmente en la física estadística, donde se busca obtener propiedades promedio de sistemas macroscópicos a través del estudio de las dinámicas microscópicas del sistema. Por lo que se busca estudiar los fenómenos económicos o sociales, a través de la física estadística, como sistemas complejos formados por muchas partículas o agentes que interactúan entre sí [15].

La econofísica a pesar de ser una rama de la física enfrenta varios problemas, como el hecho de que no pueden llevarse a cabo experimentos para reproducir sus resultados, ya que el objeto de estudio nunca está en las mismas condiciones iniciales. Por esta razón de no reproducibilidad de resultados, se puede pensar en la econofísica como pseudo ciencia. Sin embargo, hay disciplinas como la astrofísica o meteorología que presentan el mismo problema de reproducibilidad de resultados. Para ello, se hace uso de simulaciones computacionales y tratamiento estadístico y probabilísticos de los datos, para con ello darle mayor certidumbre y confianza a las predicciones.

2.2 Distribución de Boltzmann

Para el desarrollo de la distribución de Boltzmann consideraremos un estado en equilibrio con N partículas cada una con masa puntual m . Si el sistema tiene paredes rígidas e impermeables entonces el sistema tiene un volumen V y una energía U fijas. Cada partícula está descrita por 6 coordenadas, tres de posición y 3 de momento.

La idea de Boltzmann es no estudiar la dinámica individual de cada partícula, sino, estudiar la dinámica del grupo de partículas. Para ello fijamos un punto en el tiempo y podemos subdividir nuestro sistema en pequeñas regiones, las cuales tendrán un nivel de energía ϵ_i . Tendremos como hipótesis que las partículas son distinguibles, por lo que podremos numerarlas del 1 al N .

Cada una de las celdas tendrá n_i partículas, las cuales por el supuesto podremos distinguir entre sí. A cada uno de estos arreglos lo llamaremos microestado. Sin embargo, dada un volumen V , una energía U y un número de partículas N , el

microestado que cumple con estas condiciones no es único. Si suponemos que la celda i con energía ϵ_i tiene n_i partículas. Entonces el número de microestados que corresponden a esa distribución que denotamos por W_{n_j} y que llamaremos un macroestado está dado por:

$$W_{n_j} = \frac{N!}{\prod n_j!} \quad (2.1)$$

Sin embargo, W_{n_j} es el número de microestados que cumplen con la distribución $(n_1, n_2, \dots, n_j, \dots)$. Pero hay otras distribuciones que cumplen con las condiciones iniciales. Denotamos el número total de microestados que cumplen con las condiciones iniciales como $\Omega(N, U, V)$, él cuál se obtiene sumándolos W_{n_j} sobre todas las distribuciones posibles.

Boltzmann reconoció que las únicas distribuciones relevantes son aquellas consistentes con la información disponible al observador, es decir las que obedecen la condición de que N y U están fijas, por lo que se cumple:

$$n_1 + n_2 + \dots + n_j + \dots = \sum n_i = N \quad (2.2)$$

$$\epsilon_1 n_1 + \epsilon_2 n_2 + \dots + \epsilon_j n_j + \dots = \sum \epsilon_i n_i = U \quad (2.3)$$

Por lo que $\Omega(N, U, V)$ representara el número de macroestados consistentes con las condiciones iniciales del sistema macroscópico, con lo que se obtiene la siguiente observación:

$$\Omega(N, U, V) = \sum_{n_j} W_{n_j} = \sum \frac{N!}{\prod n_j!} \quad (2.4)$$

Una hipótesis de las ecuaciones anteriores es que todos los microestados tienen la misma probabilidad. Posteriormente procedemos a encontrar los valores de n que determinen el macroestado máximo. Para ello maximizamos la siguiente expresión:

$$t(n) = \frac{N!}{\prod n_j!} \quad (2.5)$$

Para pequeñas variaciones de δn_i de los números n_i tendríamos:

$$\delta t(n) = \sum \left(\frac{\partial t}{\partial n_i} \right)_{i \neq j} \delta n_i = 0 \quad (2.6)$$

Lo cual sugiere la solución $(\partial t / \partial n_i)_{i \neq j} = 0$. Sin embargo, debemos considerar las restricciones sobre U y N, por lo que se debe cumplir:

$$\sum \delta n_i = \delta N = 0 \quad (2.7)$$

$$\sum \epsilon_i \delta n_i = \delta U = 0 \quad (2.8)$$

Por lo que no todas las variables n_i son independientes. Para resolver el problema multiplicamos la ecuación 2.7 por α y la ecuación 2.8 por β y sumamos a la ecuación 2.6 con lo que obtenemos:

$$\delta t(n) = \left(\frac{\partial t}{\partial n_1} + \alpha + \beta \epsilon_1 \right) \delta n_1 + \left(\frac{\partial t}{\partial n_2} + \alpha + \beta \epsilon_2 \right) \delta n_2 + \dots \left(\frac{\partial t}{\partial n_j} + \alpha + \beta \epsilon_j \right) \delta n_j + \dots \quad (2.9)$$

Como $\delta t(n) = 0$ el miembro derecho de la ecuación debe ser igual a cero para toda δn_i . Sin embargo, de las ecuaciones 2.7 y 2.8 podemos despejar δn_1 y δn_2 . Así que si escogemos α y β tales que:

$$\frac{\partial t}{\partial n_1} + \alpha + \beta \epsilon_1 = 0 \quad (2.10)$$

$$\frac{\partial t}{\partial n_2} + \alpha + \beta \epsilon_2 = 0 \quad (2.11)$$

Entonces $\delta t = 0$ en la ecuación 2.6 solo contiene del tercer termino en adelante y $\delta n_3, \delta n_4, \dots$, etc. son variaciones independientes luego todos los coeficientes de δn_i pueden hacerse cero, esto es:

$$\frac{\partial t}{\partial n_i} + \alpha + \beta \epsilon_i = 0 \quad (2.12)$$

Que constituyen la solución del problema. Ahora vamos a aplicar este resultado a la expresión de $\ln t(n)$, de donde obtenemos:

$$\ln t(n) = \ln N! - \sum \ln n_j! \quad (2.13)$$

Usando la aproximación de Stirling y suponiendo que n_i son grandes, podemos reescribir de la siguiente manera:

$$\ln t(n) = N \ln N - N - \sum (n_j \ln n_j - n_j) \quad (2.14)$$

$$= N \ln N - \sum n_j \ln n_j \quad (2.15)$$

Donde se usó la ecuación 2.2. Entonces obtenemos:

$$\frac{\partial \ln t(n)}{\partial n_i} = -\ln n_i - 1 \quad (2.16)$$

Si absorbemos la constante -1 en α , entonces sustituyendo en la ecuación 2.12 obtenemos:

$$\ln n_i^* = \alpha + \beta \epsilon_i \quad (2.17)$$

Donde n_i^* es la distribución que hace a $t(n)$ un extremal, de la ecuación anterior obtenemos:

$$n_i^* = \exp\{\alpha + \beta \epsilon_i\} \quad (2.18)$$

Si definimos $A = e^\alpha$ obtenemos:

$$n_i^* = A e^{\beta \epsilon_i} \quad (2.19)$$

2.3 Mecánica Estadística de la Distribución del Dinero

Uno de los factores principales para la derivación de la ley de Boltzmann Gibbs es la conservación de la energía. Es por ello que podemos suponer que si una cantidad se conserva en un sistema de muchos cuerpos, la distribución de probabilidad en equilibrio, debe seguir un patrón exponencial, similar al de Boltzmann Gibbs, descrito por la siguiente ecuación:

$$P(\epsilon) = Ce^{-\epsilon/T} \quad (2.20)$$

Dragulescu y Yakovenko [2] afirman que en un sistema económico cerrado, la cantidad de dinero total se conserva, por lo que en equilibrio la distribución de probabilidad del dinero debe seguir una distribución similar a la de Boltzmann-Gibbs, con la siguiente forma $P(m) = Ce^{-m/T}$ donde m es la cantidad de dinero y T es la temperatura efectiva, que en este caso representa la cantidad de dinero promedio por agente. El supuesto detrás de la conservación del dinero, es que los agentes promedio no pueden imprimir o producir papel moneda, acción exclusiva de los bancos centrales.

Para el modelo solo se consideran una cantidad fija de agentes, donde la única interacción entre agentes es el cambio en la tenencia de dinero: $[m_i, m_j] \rightarrow [m'_i, m'_j]$. Donde la cantidad de dinero se conserva en cada transacción, es decir:

$$m_i + m_j = m'_i + m'_j \quad (2.21)$$

Se asume que el sistema es cerrado, no hay flujo externo de dinero, por lo que la cantidad total de dinero M se conserva en el sistema. Además, se asume que no hay deuda por lo que la cantidad de dinero $m_i \geq 0$.

La distribución de probabilidad bajo estos supuestos tiene una forma de Boltzmann-Gibbs dada por la expresión 2.20. La función $P(m)$ debe cumplir:

$$\int P(m)dm = 1 \quad (2.22)$$

$$\int mP(m)dm = \frac{M}{N} \quad (2.23)$$

De las ecuaciones 2.22 y 2.23 obtenemos $C = 1/T$ y $T = M/N$; por lo que T es la cantidad promedio de dinero por agente.

2.4 Leyes de Potencias

Las distribuciones de leyes de potencia aparecen de forma recurrente en diversos fenómenos físicos y económicos. Las distribuciones de leyes de potencia se generan cuando los valores no se acumulan alrededor de un valor típico, sino que fluctúan en rango dinámico muy amplio. Además, se usan para describir fenómenos donde los eventos de gran magnitud son poco frecuentes, mientras que los eventos pequeños son muy frecuentes. Algunos ejemplos son los temblores, donde los grandes terremotos son poco frecuentes, pero los pequeños son muy comunes; El tamaño de las ciudades, pues son pocas las grandes metrópolis, sin embargo la cantidad de ciudades pequeñas es muy abundante. Otro ejemplo es el uso de las palabras, donde pocas palabras se usan de manera muy frecuente como es el caso de los artículos o preposiciones y por el otro lado tenemos muchas palabras que se usan muy poco.

En la figura 2.1 se muestra el histograma de las ciudades con más de 10,000 habitantes en los Estados Unidos. En la figura 2.1 del lado izquierdo se observa que la gran mayoría de ciudades tienen una población baja y una cantidad muy pequeña de ciudades tienen valores muy altos, lo que produce la cola alargada. Por lo que el histograma muestra una distribución de leyes de potencia.

Del lado derecho de la figura 2.1 se muestra el histograma de las ciudades, pero aplicando logaritmo en ambos ejes obtenemos una relación lineal; de tal forma que se cumple la siguiente relación: $\log p(x) = -\alpha \ln x + c$. Donde α y c son constantes y $p(x)$ representa la fracción de ciudades con valor de población x . De la relación anterior y aplicando exponencial en ambos lados de la ecuación, obtenemos que si un fenómeno sigue la ley de potencias, entonces la distribución que sigue es de la forma:

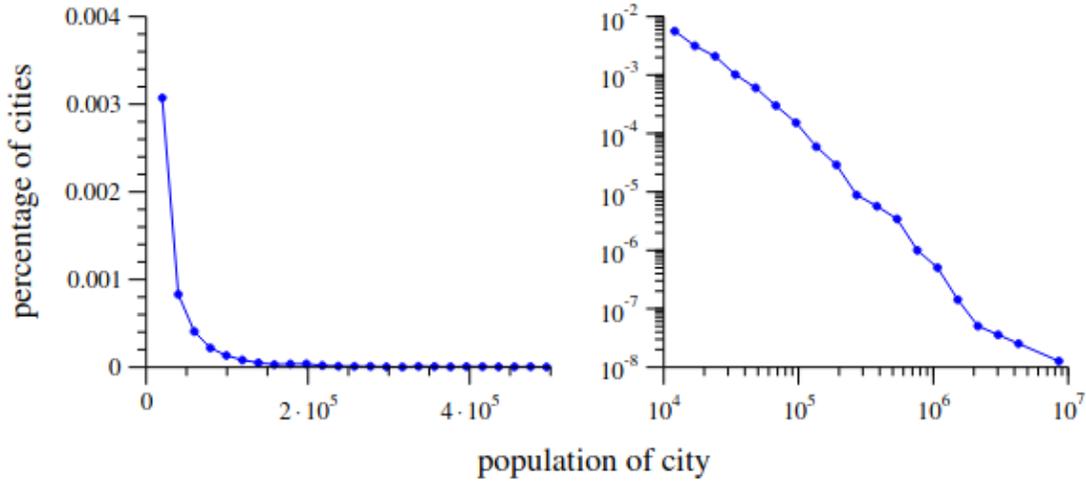


Figure 2.1: Distribución CCDF para GDP

$$p(x) = Cx^{-\alpha} \quad (2.24)$$

Donde $C = e^c$ y α es una constante de distribución conocida como exponente o parámetro de escalamiento.

Al trabajar con datos que simulan leyes de potencias, podemos encontrar ciertos desafíos, sobre todo en la cola de la distribución. En la cola de la distribución al tener pocos datos se puede obtener gráficas con demasiado ruido. Para solucionar este problema se hará uso de la función de distribución acumulada (CDF) en lugar de usar el histograma.

La función de distribución acumulada indica la probabilidad $P(x)$ de que x tenga un valor mayor o igual que x :

$$P(x) = \int_x^{\infty} p(x') dx' \quad (2.25)$$

Si la distribución sigue una ley de potencias, es decir $p(x) = Cx^{-\alpha}$. Entonces la función de distribución acumulada tiene la siguiente forma:

$$P(x) = \int_x^{\infty} p(x) = C(x')^{-\alpha} dx' = \frac{C}{\alpha - 1} [x^{-(\alpha-1)}] \quad (2.26)$$

Por lo que la función de distribución acumulada sigue una ley de potencias con un exponente $\alpha - 1$.

Dentro de las leyes de potencias, tenemos la ley de Zipf la cual relaciona el tamaño y de ocurrencia de un evento relativo a su rango r . Además, tenemos la ley de Pareto usada principalmente en el estudio de la distribución del ingreso. La ley de Pareto indica cuantas personas tienen un ingreso mayor a x .

2.5 Distribución Exponencial Estirada (SED)

Las leyes de potencia son comunes en muchos fenómenos físicos y económicos, sin embargo, en muchos de estos fenómenos el ajuste no es perfecto en todo el rango. Las distribuciones SED surgen como alternativa a las leyes de potencia para modelar la distribución en el rango completo.

J. Laherreérel y D. Sornette [17] proponen un modelo en término de exponenciales estiradas. Este modelo está basado en la demostración de que la cola de los PDF de productos de un número finito de variables aleatorias es generalmente una exponencial estirada [18], en el que el exponente c es el inverso del número de productos multiplicados. Este modelo está dado por la siguiente expresión:

$$P(x)dx = c(x^{c-1}/x_0^c)exp[-(x/x_0)^c]dx \quad (2.27)$$

Por lo que la distribución acumulada es:

$$P(x) = exp[-(x/x_0)^c] \quad (2.28)$$

Las exponenciales estiradas se caracterizan por un exponente c más pequeño que uno. En el límite $c = 1$ corresponde a la distribución exponencial habitual. Para c menor que uno, la distribución 2.28 presenta una curvatura al graficar en escala log-log, mientras que exhibe un comportamiento aparentemente lineal grande cuanto más pequeña es c .

3 Modelo

La función de distribución acumulada complementaria (CCDF), la cual se define como $CCDF = 1 - CDF$. Donde CDF es la función de distribución acumulada. CDF esta definida como

$$CDF(m) = \int PDF(m') dm'$$

donde PDF es la función de distribución de probabilidad y m es la variable a analizar.

La función CCDF tiene la propiedad de que si PDF sigue una ley de potencias entonces el CCDF sigue una ley de potencias.

Usaremos tres distribuciones diferentes. La primera de ellas es la distribución de Pareto, la cual esta dada por la siguiente expresión

$$PDF(x) = b \frac{u^b}{x^{b+1}}, b > 0 \tag{3.1}$$

donde exponente b es un parámetro y u es el umbral más bajo. El CCDF de la distribución de Pareto es

$$CCDF \propto a \frac{u^a}{x^{a+1}} \tag{3.2}$$

Al graficar en escalas logarítmicas verticales y horizontales, se obtiene una línea recta.

$$\log CCDF \propto constant - (a + 1) \log x \tag{3.3}$$

La segunda distribución es la distribución exponencial o DBG que produce una línea recta en una escala semilogarítmica. La última es la distribución exponencial estirada (SED)

$$f(x) \propto x^{\beta-1} e^{-(x/x_0)^\beta} \quad (3.4)$$

De la expresión anterior obtenemos el siguiente CCDF:

$$CCDF \propto e^{-(x/x_0)^\beta} \quad (3.5)$$

Donde β es el exponente y x_0 es el parámetro de escala de la distribución. Si $\beta > 1$ SED decrece más rápido que una exponencial, para $\beta = 1$ se obtiene una exponencial y para $\beta < 1$, obtenemos una distribución similar a una ley de poder, esto ya que tenemos una cola larga en la región de valores grandes, aunque decrece más rápido que una ley de poder, pero más lento que una exponencial. Para observar si los datos se ajustan a una distribución SED aplicamos logaritmo a la Eq. (3.5), de donde obtenemos:

$$\log CCDF(x) \propto -(x/x_0)^\beta$$

$$-\log CCDF(x) \propto (x/x_0)^\beta$$

$$\log[-\log CCDF(x)] \propto \beta \log x - \beta \log x_0 \quad (3.6)$$

De esta forma una SED de un solo exponente debería mostrar una línea recta al graficar una escala de logaritmo triple con una pendiente igual a β

4 Base de Datos

4.1 Indicadores Económicos

Uno de los indicadores más utilizados para medir la riqueza de un país es el producto interno bruto (PIB), el cual es una medida estándar del valor generado por el país, que engloba los bienes finales y los servicios producidos en un periodo de tiempo dado. El PIB se puede medir de varias maneras, las cuales nos proveen de distinta información. El PIB nominal es el valor, a precios de mercado, de la producción de bienes y servicios finales producidos en un país durante un período de tiempo. El PIB real es el valor de esa producción a precios constantes. El PIB nominal refleja aumentos o disminuciones de los precios si hay inflación o deflación, respectivamente. Por otra parte, el PIB real se basa en precios fijos a un año y permite comparar la producción de un país determinado en diferentes períodos aislando las variaciones de precios. Para el PIB real, utilizaremos los precios de 2015 como base o referencia y encontraremos resultados más estables a lo largo de los años dentro de esta descripción.

El costo de vida es distinto en cada país, por lo que es necesario usar un indicador que nos permita comparar el poder adquisitivo entre los países, para ello se usara la paridad de poder adquisitivo (PPA), el cual es un deflactor de precios y un convertidor de divisas que tiende a eliminar los efectos de las diferencias en los niveles de precios entre países. La PPA nos permite comparar monedas de diferentes países a través de una canasta de cotización de productos. El PIB de paridad de poder adquisitivo (PIB-PPA), nos permitira comparar mejor la productividad económica y el nivel de vida entre los países. Para comparar el nivel de riqueza entre países, en el presente

trabajo, se empleará el PIB per capita, que es el PIB dividido por la población total de la región y se denotara por PCPIB y se usara el PCPIB de paridad de poder adquisitivo que denotaremos PCPIB-PPA.

4.2 OCDE

La Organización para la Cooperación y el Desarrollo Económicos es un organismo de cooperación internacional, compuesto por 37 países. Tiene como misión diseñar mejores políticas para mejorar la calidad de vida; buscando promover la prosperidad, oportunidades, igualdad y el bienestar de sus países miembros y no miembros. Los países miembros de la OCDE se conoce como el club de países ricos. El PIB nominal grupal de los países de la OCDE represento el 62.2% del PIB nominal total en el 2017.

La OCDE recaba datos en tres niveles territoriales distintos. El primer nivel es TL2 que son regiones territoriales extensas o grandes y se refiere a las divisiones políticas de cada país (estados, principados, comunidades autónomas, ciudades autónomas, etc.). El segundo nivel es TL3 que hace referencia a regiones pequeñas, estas regiones serán usadas para estimar la información de las áreas metropolitanas o ciudades que representan el tercer nivel¹.

En cada fuente se indica que años abarca la información tomada y el nivel de región. El nivel de región se denota con 2, 3 y áreas metropolitanas que hacen referencia a TL2, TL3 y áreas metropolitanas. En el caso de los datos por región se toman los datos correspondientes a TL2 (estados, principados, comunidades autónomas, ciudades autónomas, etc.), es decir, es la información de cada división política de cada país. Para todos los países se encuentran datos por región, con excepción de Israel e Islandia, para los cuales solo se reporta la información nacional.

Para determinar las ciudades se define una unidad denotada por FUA (funcional urban áreas). Para determinar las FUA se hace una división territorial y se buscan los centros urbanos. Los centros urbanos serán celdas contiguas con una densidad poblacional alta, en este caso con más de 50 mil habitantes. Al centro urbano se

¹Las fuentes para cada país se pueden encontrar en el siguiente libro [19]

le suman localidades vecinas menos pobladas, pero las cuales tienen una conexión con el centro urbano; esta conexión será que al menos el 15% de los empleados de la localidad trabaje en el centro urbano. Si en las fuentes de información existen datos de las ciudades se hacen ajustes en la información. Solo se da información de FUA's con más de 250 mil habitantes y se tienen 668 FUA's [20].

Para obtener los datos de las ciudades se usan tres métodos. El primero es obtenerlo directamente de las fuentes. El segundo es estimar la información de las ciudades usando información de regiones más grandes. Para esto se usa la distribución de la población y la división territorial. En la mayoría de los países se usa la información de las regiones TL3 que son datos de la OCDE, salvo en México, Australia y Colombia, donde se usa la información de TL2 de la OCDE. Para Canadá se usa la información del CMA que es el censo de áreas metropolitanas de Canadá. El tercer método es agregar información administrativa a nivel local. Este último método solo se usa para EUA.

La base de datos Metropolitanos de la OCDE recaba datos de tipo demográficos, económicos, laborales, sociales, de digitalización y de organización territorial. Para calcular los distintos indicadores se emplean cuatro métodos. El primer método es agregar los datos de las administraciones locales. El segundo método es procesar datos geográficos de datos de geolocalización. El tercer método es estimar datos usando la información de regiones más grandes. Por último, tenemos los indicadores derivados, los cuales se obtienen de otros indicadores como es el caso de la densidad.

Para las grandes regiones (TL2) y las pequeñas (TL3) se utilizó la base de datos de la OCDE Regional Economy [21]. Para las ciudades (FUA) se utilizó la base de datos metropolitana de la OCDE que proporciona indicadores socioeconómicos y ambientales de 688 zonas urbanas o metropolitanas funcionales de la OCDE en 36 países de la OCDE [22].

Table 4.1

País	Regiones	Ciudades	País	Regiones	Ciudades
AUS	9	6	ISL	1	0
AUT	10	6	ISR	1	0
BEL	4	5	ITA	22	19
BGR	7	0	JPN	11	19
BRA	28	0	KOR	8	8
CAN	14	16	LTU	3	2
CHE	8	4	LUX	2	1
CHL	16	5	LVA	2	1
CHN	32	0	MEX	33	7
COL	34	0	MLT	2	0
CZE	9	4	NLD	13	13
DEU	17	68	NOR	8	4
DNK	6	4	NZL	13	0
ESP	20	16	PER	26	0
EST	2	1	POL	18	14
FIN	6	4	PRT	8	3
FRA	19	30	ROU	9	0
GBR	13	40	RUS	84	0
GRC	14	2	SVK	5	1
HRV	3	0	SVN	3	2
HUN	9	4	SWE	9	4
IND	68	0	TUR	27	0
IRL	4	2	USA	52	166
			ZAF	10	0

5 Resultados

5.1 Ciudades y Regiones pequeñas

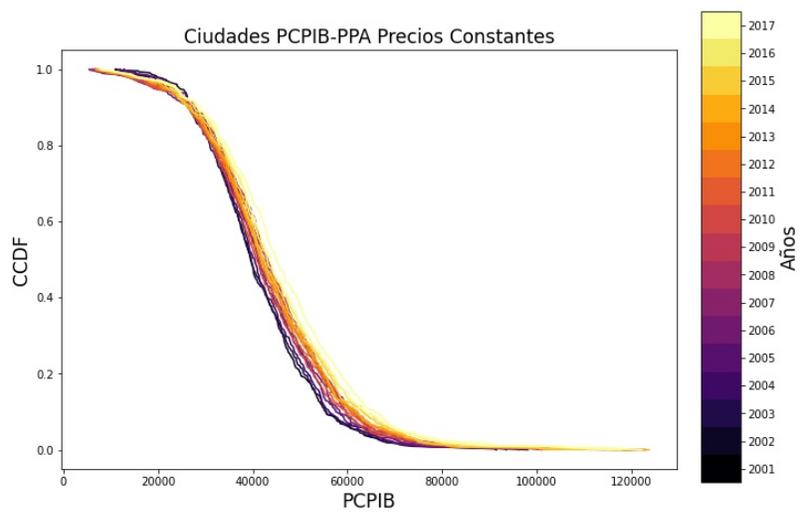


Figure 5.1: Distribución CCDF para PCPIB-PPA en Ciudades

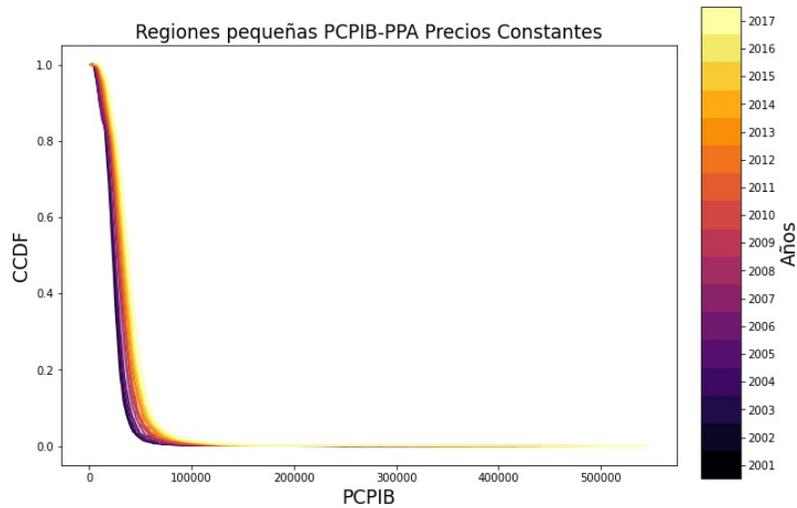


Figure 5.2: Distribución CCDF para PCPIB-PPA en Regiones pequeñas

En la Fig. 5.1 y Fig. 5.2 se muestran las distribuciones del PCPIB-PPA para ciudades y regiones pequeñas. Se observa que la gran mayoría de países se encuentran acumulados en los valores bajos, por lo que sugiere que la distribución del PCPIB-PPA siga una distribución de leyes de potencias.

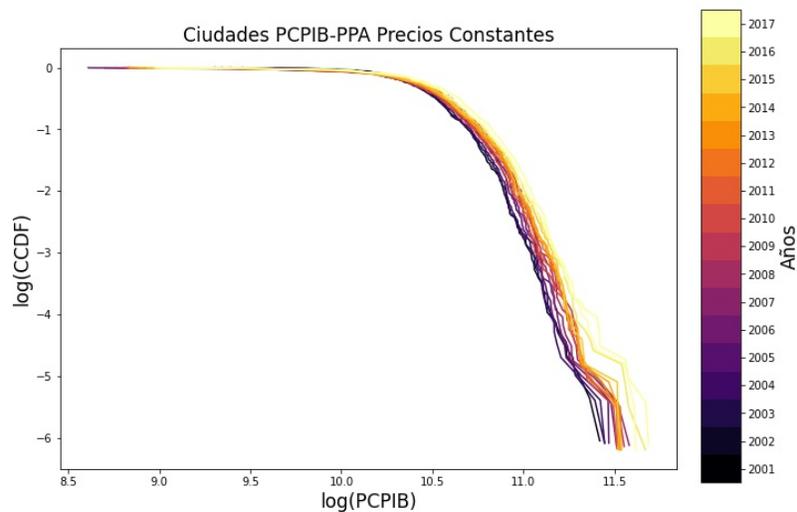


Figure 5.3: CCDF del PCPIB-PPA por ciudad en escala en log-log

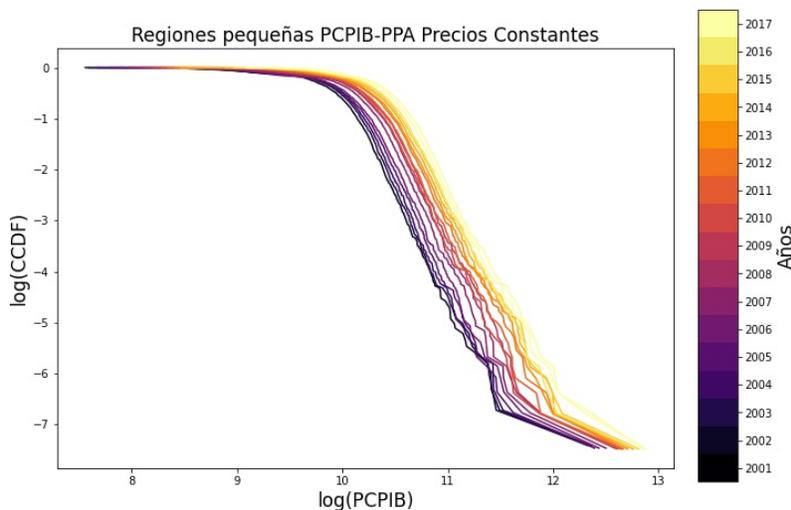


Figure 5.4: CCDF del PCPIB-PPA por Regiones pequeñas en escala log-log

Para analizar cuál es el mejor ajuste a la distribución del PCPIB mundial, se graficarán en dólares americanos el CCDF del PCPIB-PPA para las ciudades comprendiendo los años desde el 2001 al 2017 en escalas log-log, semi-log Y triple-log.

Al analizar el PCPIB-PPA por ciudad vemos, en la Fig. 5.4, que el ajuste lineal en la escala log-log no es bueno en el rango completo de ciudades. Por lo que la distribución no sigue una ley de potencias de Pareto. En la Fig. 5.6 el caso semi log, el ajuste mejora y a partir del año 2008 en adelante obtenemos una $R^2 > 0.95$. En el caso semi-log al eliminar las ciudades con menor GDP, la gráfica parece lineal. Por lo tanto, la distribución presenta un comportamiento exponencial sobre todo en los últimos años. En las Fig. 5.3 y Fig. 5.5 observamos las distribuciones del PCPIB-PPA para regiones pequeñas en escalas log-log y semi-log respectivamente. Sin embargo, no tiene un comportamiento semi-log o log-log, ya que los ajustes del modelo lineal no son buenos, obteniendo una $R^2 < 0.90$.

En las Fig. 5.4 y Fig. 5.5 se realizaron tres análisis diferentes. El primero evalúa el ajuste eliminando las diez regiones más ricas. En este caso el ajuste mejora, teniendo para las regiones pequeñas, una $R^2 > 0.94$ a partir del año 2008. En las ciudades, el ajuste no es tan preciso, ya que se obtiene $0.85 < R^2 < 0.90$. El segundo, evalúa el ajuste únicamente tomando las regiones con un nivel medio de PIB per cápita, en este caso se eliminan las diez ciudades más ricas y las 130 y 375 ciudades y regiones

más pobres respectivamente, en ambos casos se obtiene una $R^2 > 0.97$. Finalmente, se analizan solo las ciudades y regiones más pobres, con el modelo semi-log se obtiene una $R^2 > 0.97$ para todos los años.

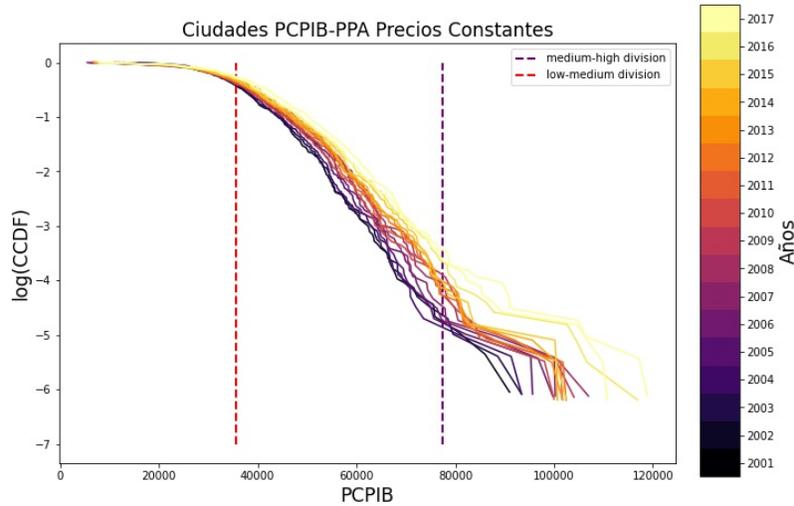


Figure 5.5: CCDF del PCPIB-PPA por ciudades en escala semi log, las líneas verticales discontinuas representan los límites entre ingresos pobres, medios y altos en 2014.

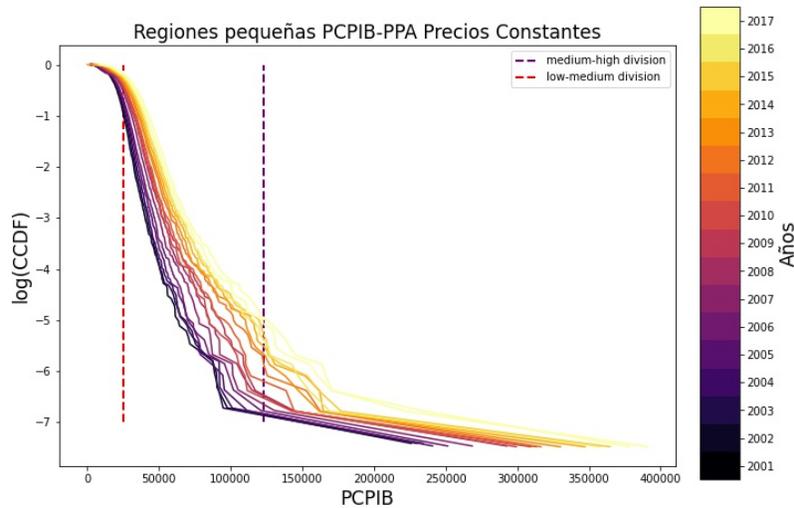


Figure 5.6: CCDF del PCPIB-PPA por Regiones pequeñas en escala semi log, las líneas verticales discontinuas representan los límites entre ingresos pobres, medios y altos en 2014.

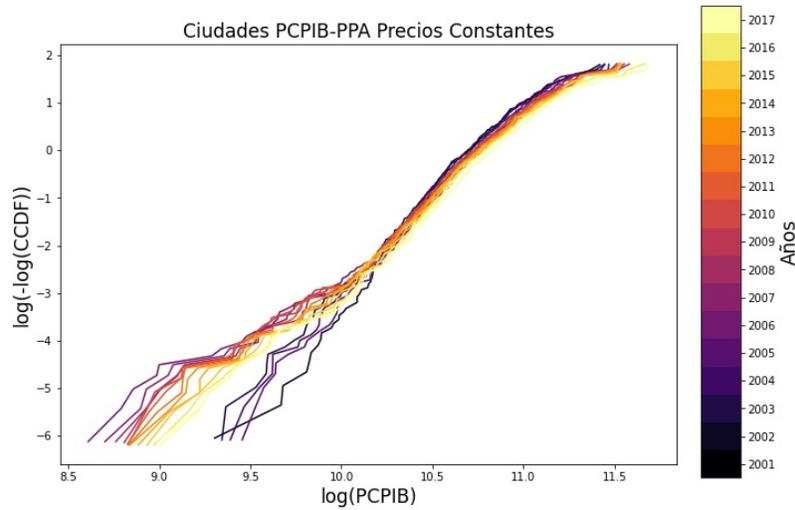


Figure 5.7: CCDF del PCPIB-PPA por ciudades en escala triple log

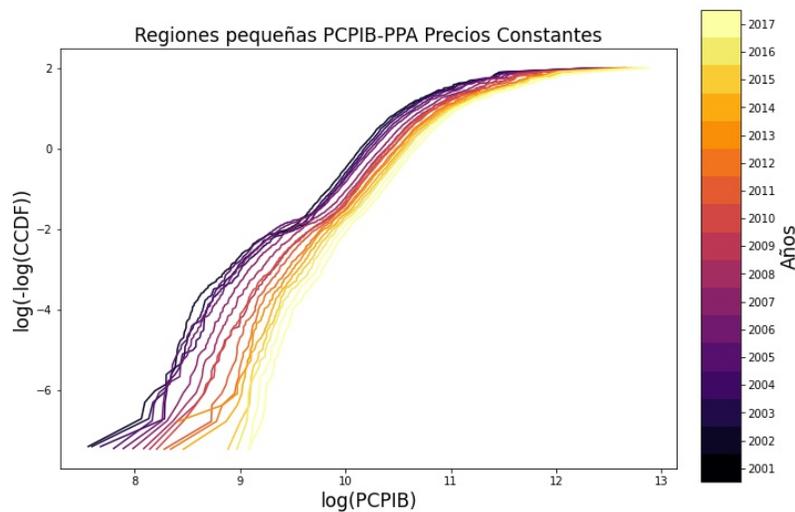


Figure 5.8: CCDF del PCPIB-PPA por Regiones pequeñas en escala triple log

En la Fig.5.7 se grafica el PCPIB-PPA en escala triple-log para las ciudades. Podemos observar que la gráfica muestra un comportamiento lineal, obteniendo para todos los años valores de R^2 superiores a 0.95; por lo que el PCPIB-PPA se comporta como una distribución SED con coeficiente $\beta > 1$. En la Fig.5.8 observamos el PCPIB-PPA para regiones pequeñas en escala triple log. Obtuvimos un buen ajuste lineal, ya que para todos los años obtuvimos un $R^2 > 0.95$. Por lo tanto, la distribución del PCPIB-PPA utilizando regiones pequeñas sigue o se ajusta a un modelo SED.

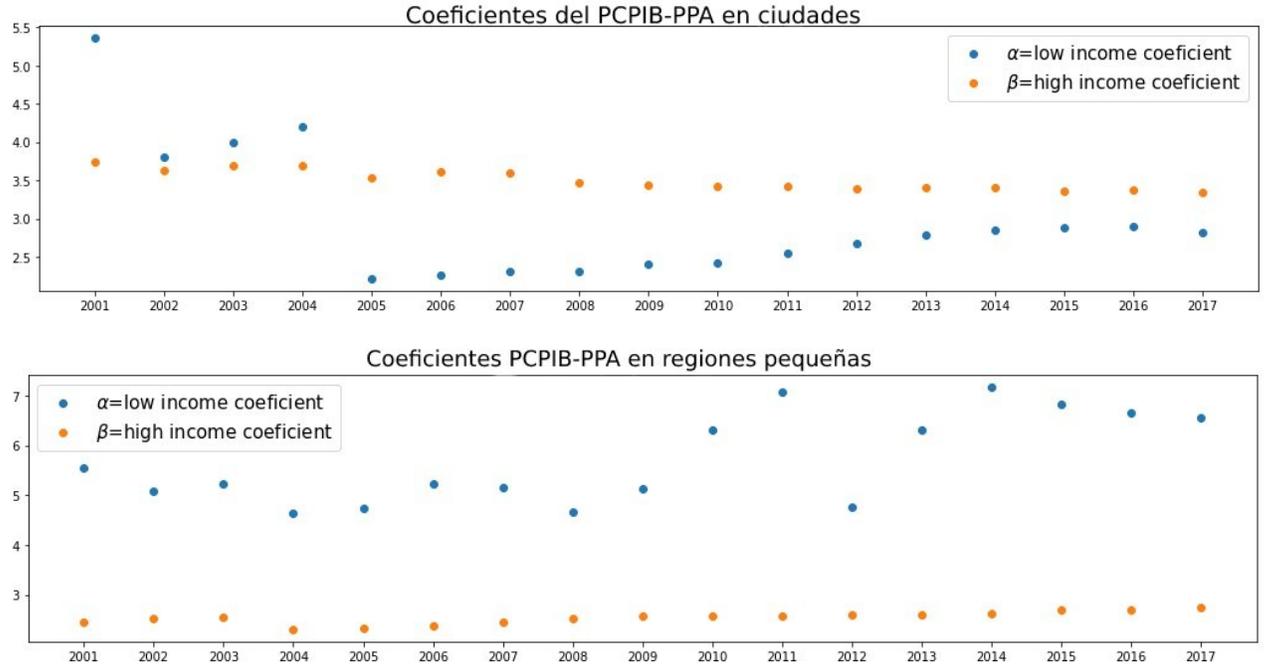


Figure 5.9: Evolución de α y β en la distribución SED para PCPIB-PPA en ciudades y regiones pequeñas

Si dividimos los datos en valores de ingresos bajos y altos, podemos realizar un análisis dividiendo los países en dos grupos. Para las regiones pequeñas obtenemos en ambos casos un buen ajuste con la distribución SED, ya que en todos los casos $R^2 > 0.95$. Fig. 5.9 muestra los parámetros α y β obtenidos en cada caso, ambos son mayores que 1 dentro de los modelos SED, por lo que tenemos distribuciones que disminuyen más rápido que una exponencial.

5.2 Regiones

Para analizar la distribución de riqueza en las regiones seguiremos el mismo procedimiento empleado en el caso de las ciudades y regiones pequeñas. Al observar el CCDF en la Fig.5.10 y Fig.5.11 observamos en ambos casos que la gran mayoría de los países están acumulados en los niveles de PCPIB-PPA más bajos, lo que sugiere una distribución de leyes de potencias.

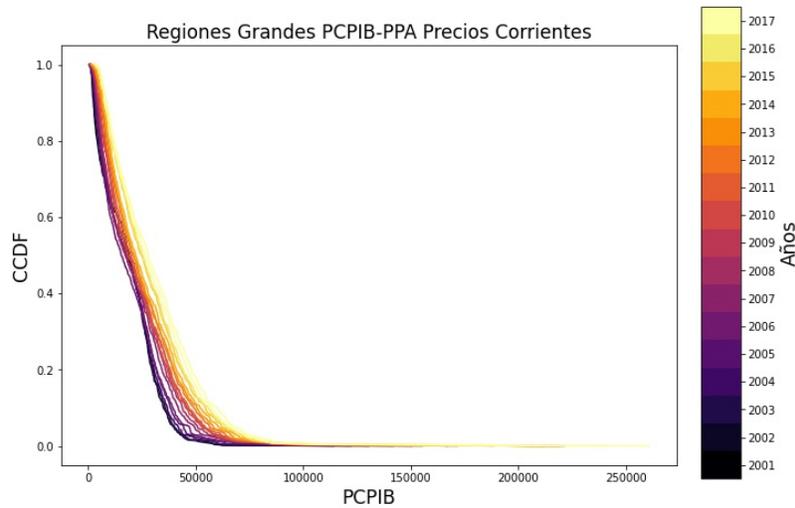


Figure 5.10: Distribución CCDF para PCPIB-PPA con precios corrientes

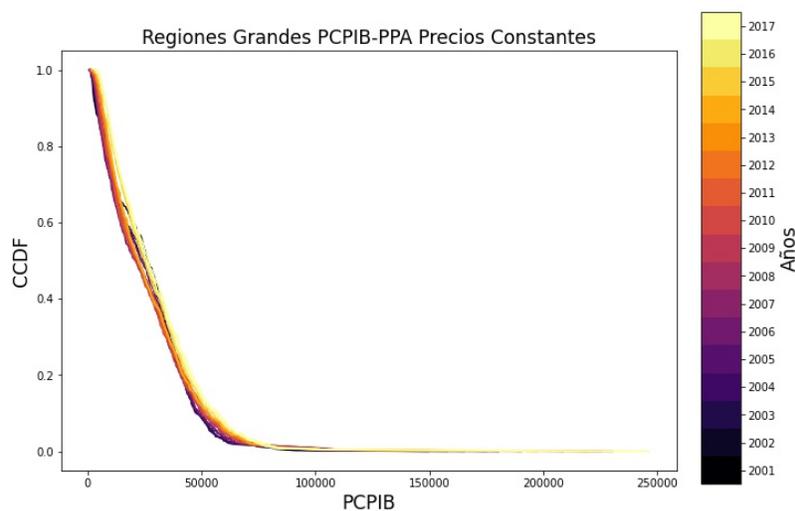


Figure 5.11: Distribución CCDF para PCPIB-PPA con precios constantes

Podemos observar que las gráficas en precios constantes en la Fig. 5.13 son mucho más compactas o constantes en el tiempo que en la Fig. 5.12 porque no hay efectos de la inflación. A precios corrientes, las distribuciones evolucionan en el tiempo hacia la derecha, porque generalmente, los precios tienden a aumentar (inflación), causando que el PIB sea más alto. Este efecto estará presente en todas las gráficas.

Al observar la gráfica de PCPIB-PPA en escala log-log en las Fig. 5.12 y Fig. 5.13, obtenemos que el ajuste no es lineal, por lo que el PCPIB-PPA no sigue una distribución de Pareto para ninguno de los dos casos, ya sean precios corrientes o

precios constantes con año base 2015.

En la Fig.5.14 y Fig.5.15 se muestra la distribución en escala semi-log para el PCPIB-PPA por regiones en precios constantes y corrientes. Para el PCPIB-PPA a precios corrientes tenemos un resultado combinado, ya que del año 2005 al 2015 tenemos una $R^2 > 0.95$. Mientras que en los años restantes tenemos $0.90 < R^2 < 0.95$. En el caso del PCPIB-PPA con precios constantes volvemos a observar el mismo caso, ya que del año 2005 al 2015 tenemos un $R^2 > 0.95$, mientras que en los años restantes tenemos $0.90 < R^2 < 0.95$. Por lo que el PCPIB-PPA a precios constantes y corrientes sigue una distribución exponencial (DBG) para la gran mayoría de los años analizados.

En las Fig.5.14 y Fig.5.15, se realizaron tres análisis diferentes. El primero evalúa el ajuste eliminando las diez regiones más ricas. En este caso el ajuste mejora, teniendo $R^2 > 0.90$ en todos los años posteriores a 2001 y $R^2 > 0.95$ de 2005 a 2017. El segundo, evalúa el ajuste únicamente tomando las regiones con un nivel medio de PIB per cápita, en este caso se eliminan las diez ciudades más ricas y las 400 y 370 regiones más pobres, para las regiones utilizando precios constantes y corrientes respectivamente. En ambos casos $R^2 > 0.97$. Finalmente, analizamos las regiones pobres, el modelo semi-log obtiene un $R^2 > 0.97$ para todos los años.

En la Fig.5.16 y Fig. 5.17 se muestra la distribución del PCPIB-PPA en una escala de logaritmo triple correspondiente al modelo SED. En el caso PCPIB-PPA a precios corrientes se obtiene un ajuste lineal con una $R^2 > 0.95$ para los años comprendidos entre 2003 y 2013. En el caso de PCPIB-PPA a precios constantes, el ajuste lineal mejora, ya que del año 2001 al 2013 se obtiene una $R^2 > 0.95$ y todos los años reportan una $R^2 > 0.94$. Por lo que el PCPIB-PPA por regiones sigue una distribución SED con parámetro $1 < \beta < 2$.

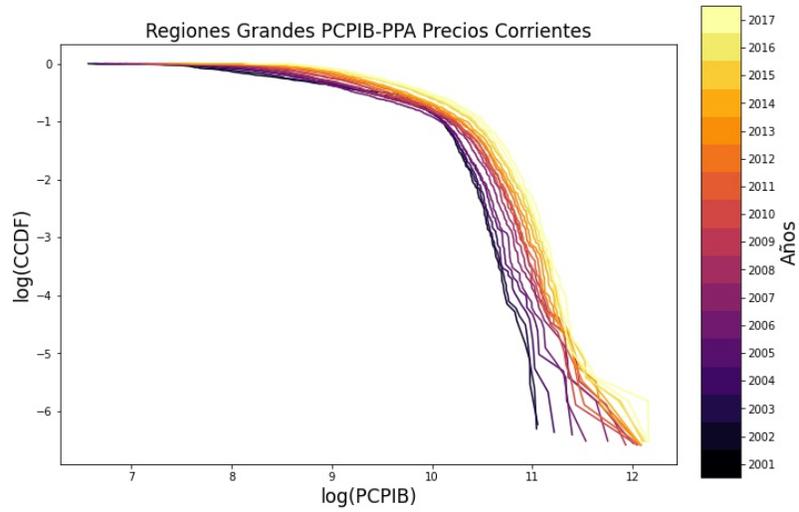


Figure 5.12: CCDF del PCPIB-PPA corriente para regiones grandes en escala log-log

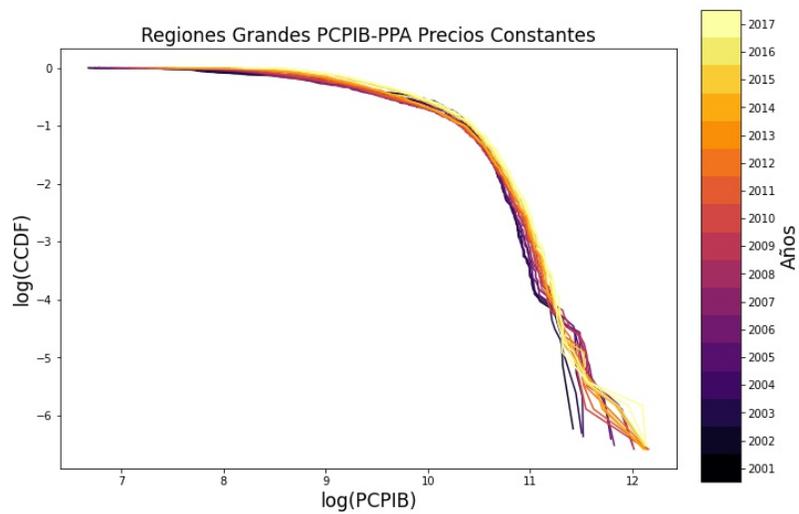


Figure 5.13: CCDF del PCPIB-PPA constante para regiones grandes en escala log-log

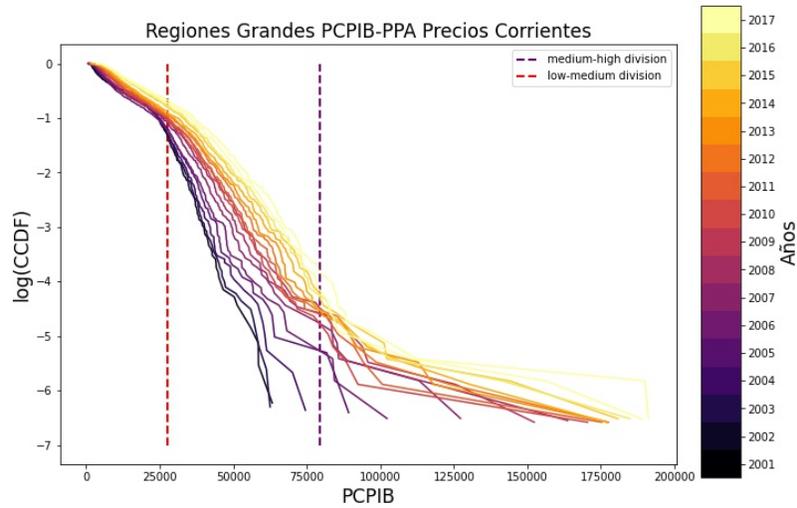


Figure 5.14: CCDF del PCPIB-PPA corriente para regiones grandes en escala semi log, las líneas verticales discontinuas representan los límites entre ingresos pobres, medios y altos en 2014.

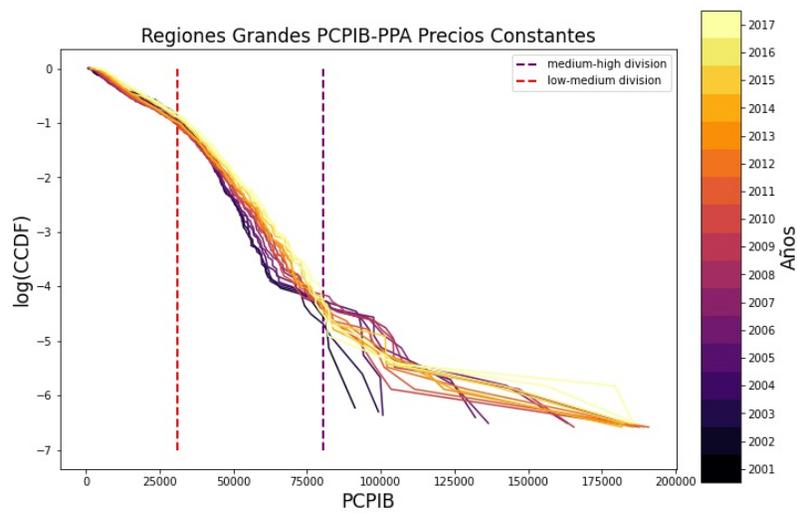


Figure 5.15: CCDF del PCPIB-PPA constante para regiones grandes en escala semi log, las líneas verticales discontinuas representan los límites entre ingresos pobres, medios y altos en 2014.

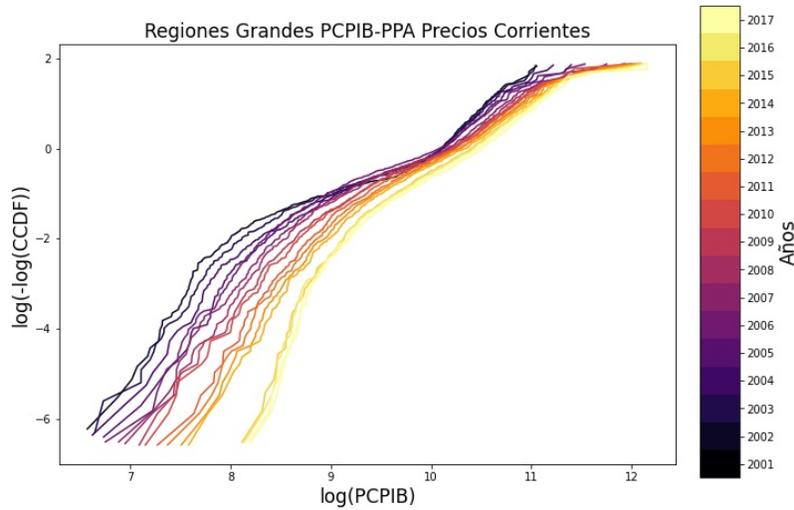


Figure 5.16: CCDF del PCPIB-PPA corriente para regiones grandes en escala triple log

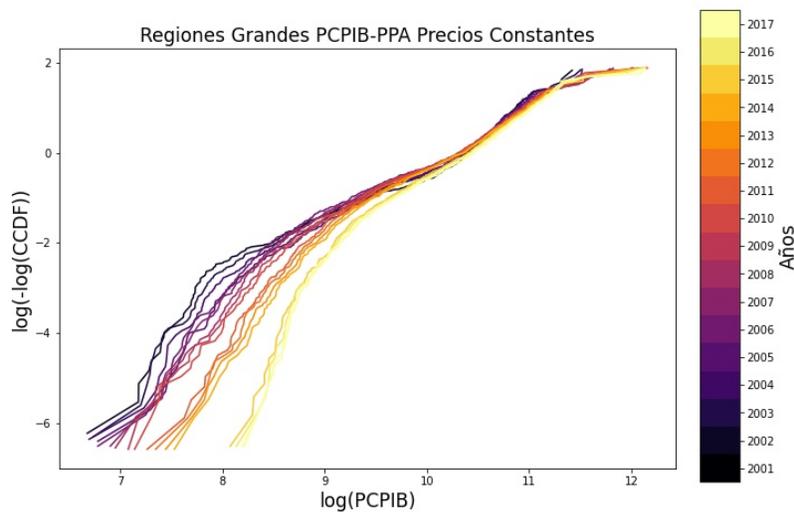


Figure 5.17: CCDF del PCPIB-PPA constante para regiones grandes en escala triple log

Cuando ajustamos los modelos SED, encontramos un buen ajuste, dado que en la mayoría de los casos, en ambos modelos se obtiene una $R^2 > 0.95$. En la Fig. 5.18 se muestra el valor de los parámetros α y β de los modelos SED. Los valores de α son mayores a 1, por lo que tenemos un SED que decrece más rápido que una exponencial. En el caso de β , tenemos un valor que ronda 0.25, por lo que tenemos una distribución similar a una ley de potencias, por lo que tenemos un buen ajuste usando la distribución DBG.

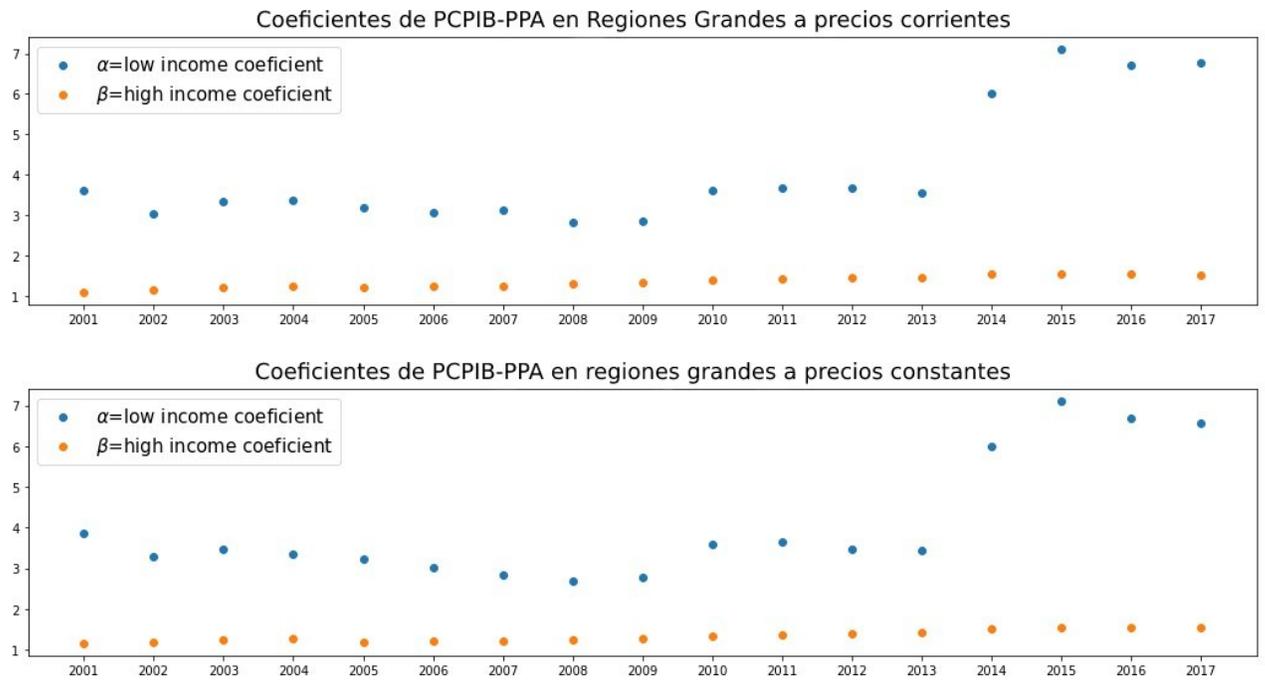


Figure 5.18: Evolución de α y β en la distribución SED para PCPIB-PPA

6 Conclusiones

Empleando datos de países de la OCDE e incluyendo datos de otros países en el período 2001-2017, estudiamos en tres diferentes niveles geográficos (ciudades, pequeñas y grandes regiones) la distribución del PCPIB per cápita ajustado por el poder de paridad de compra. El PCPIB-PPA se estudió a precios corrientes (nominales) y ajustados a la inflación (reales o constantes). Para el PIB real, utilizamos los precios de 2015 como base.

Comparando los resultados de grandes regiones obtenidos durante el período de tiempo analizado entre valores de corriente y constantes (Figs. 9-16), encontramos que las gráficas en valores constantes son mucho más compactas (colapsando a curvas simples) porque no hay efectos de inflación. Además, a precios corrientes, observamos que la distribución se desplaza hacia la derecha porque, en general, los precios tienden a aumentar (inflación). Este efecto está presente en todas las parcelas.

En las ciudades, ninguna distribución se ajusta perfectamente a los datos en todos los años. Sin embargo, el ajuste mejora con el paso de los años. En los últimos años, el PCPIB-PPA de la ciudad sigue un DBG exponencial. En el caso de las ciudades, el PCPIB-PPA se ajusta bien a una distribución SED con un coeficiente $\beta > 1$. Al analizar regiones pequeñas, encontramos que después de 2008 la distribución está bien ajustada a una distribución DBG. Sin embargo, el ajuste mejora a lo largo de los años al utilizar una distribución SED.

Al analizar las regiones grandes, se obtuvo un buen ajuste para el PCPIB-PPA utilizando la escala semi-log, especialmente para los años entre 2005-2015 con $R^2 > 0.95$. La distribución SED es adecuada para todos los años, especialmente de 2001 a 2013 ($R^2 > 0.95$), para el resto de los años se tiene $R^2 > 0.94$. El modelo SED

tiene un coeficiente $1 < \beta < 2$ para todos los años. Por lo tanto, el PCPIB-PPA de las regiones en general se pueden ajustar adecuadamente a un DBG de un solo parámetro.

En general, encontramos un ajuste excelente cuando usamos el modelo BGD en ciudades y regiones pobres y de nivel medio al analizarlas por separado, obteniendo un $R^2 > 0.97$. Al retirar las 10 ciudades y regiones más ricas, el ajuste mejora para las regiones en general, ya que a partir del año 2005 obtenemos un $R^2 > 0.94$ para las regiones largas y para las regiones pequeñas tenemos un $R^2 > 0.90$ para todos los años.

Al ajustar dos modelos SED a los datos (uno a las regiones pobres y otro a las ricas), conseguimos un ajuste muy bueno para ambos casos ($R^2 > 0.95$). Para las regiones con un PIB per cápita bajo, obtenemos un SED con $\beta > 2$ y para regiones con un PIB alto un coeficiente de $1 < \beta < 2$

En todos los casos, el modelo SED de dos parámetros es el que mejor se ajusta, pero para las regiones pequeñas y grandes el DBG de un parámetro también produce excelentes ajustes, que no es el caso con las ciudades. En el caso de las ciudades, solo se toman en cuenta los países miembros de la OCDE, para las regiones se incluyen más países. Dado que los países de la OCDE son países desarrollados, en su mayoría, las ciudades analizadas puede representar una muestra menos representativa.

El hecho de que el DBG exponencial simple rinda tan buen ajuste para regiones pequeñas y grandes indica que el modelo de econofísica, que emplea la conservación del dinero en las transacciones económicas, es una buena descripción del comercio económico, siempre que el sistema esté cerrado.

Referencias

- [1] E. Hernández-Ramírez, M. del Castillo-Mussot, and J. Hernández-Casildo. World per capita gross domestic product measured nominally and across countries with purchasing power parity: Stretched exponential or boltzmann–gibbs distribution? *Physica A: Statistical Mechanics and its Applications*, 568:125690, 2021.
- [2] V. Yakovenko A. Dragulescu. Statistical mechanics of money. *ur. Phys. J. B*, 17(4):723–729, 2000.
- [3] A. Christian Silva and Victor M Yakovenko. Temporal evolution of the “thermal” and “superthermal” income classes in the USA during 1983–2001. *Europhysics Letters (EPL)*, 69(2):304–310, jan 2005.
- [4] M. Gallegat C. Di Guilmi, E. Gaffeo. Power law scaling in world income distribution. *Econ. Bull*, 15(6):1–7, 2003.
- [5] R Bustos-Guajardo and Cristian F Moukarzel. World distribution of gross domestic product per-capita. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(5):P05023, may 2015.
- [6] Michael A. Williams, Grace Baek, Yiyang Li, Leslie Y. Park, and Wei Zhao. Global evidence on the distribution of gdp growth rates. *Physica A: Statistical Mechanics and its Applications*, 468:750–758, 2017.
- [7] Tomson Ogwang. Power laws in top wealth distributions: evidence from Canada. *Empirical Economics*, 41(2):473–486, October 2011.

- [8] P. Soriano-Hernández, M. del Castillo-Mussot, O. Córdoba-Rodríguez, and R. Mansilla-Corona. Non-stationary individual and household income of poor, rich and middle classes in Mexico. *Physica A: Statistical Mechanics and its Applications*, 465:403–413, 2017.
- [9] F. Clementi and M. Gallegati. Power law tails in the Italian personal income distribution. *Physica A: Statistical Mechanics and its Applications*, 350(2):427–438, 2005.
- [10] Adrian Drăgulescu and Victor 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):213–221, 2001. Application of Physics in Economic Modelling.
- [11] Maciej Jagielski and Ryszard Kutner. Modelling of income distribution in the European Union with the Fokker–Planck equation. *Physica A: Statistical Mechanics and its Applications*, 392(9):2130–2138, 2013.
- [12] Maciej Jagielski and Ryszard Kutner. Modelling of income distribution in the European Union with the Fokker–Planck equation. *Physica A: Statistical Mechanics and its Applications*, 392(9):2130–2138, 2013.
- [13] Sitabhra Sinha. Evidence for power-law tail of the wealth distribution in India. *Physica A: Statistical Mechanics and its Applications*, 359:555–562, 2006.
- [14] Eder Johnson de Area Leão Pereira, Marcus Fernandes da Silva, and H.B.B. Pereira. Econophysics: Past and present. *Physica A: Statistical Mechanics and its Applications*, 473:251–261, 2017.
- [15] O. Miramontes and K. Volke. *Fronteras de la física en el siglo XXI*. CopIt arXives, Mexico, 2013.
- [16] García-Colín L. Termodinámica estadística. *UAM Unidad Iztapalapa, México*, 1995.

- [17] Jean Laherrere and Didier Sornette. Stretched exponential distributions in nature and economy: “fat tails” with characteristic scales. *Physics of Condensed Matter*, 2:525–539, 04 1998.
- [18] Uriel Frisch and Didier Sornette. Extreme deviations and applications. *Journal de Physique I*, 7, 05 1997.
- [19] OECD. *OECD Regions and Cities at a Glance 2018*. 2018.
- [20] OECD. *Territorial grids and regional typology*. 2011.
- [21] Organisation for Economic Cooperation and Development (OCDE). *Regional economy*, 2020.
- [22] Organisation for Economic Cooperation and Development (OCDE). *Metropolitan areas*, 2021.

A Tablas

A.1 Ciudades y Regiones más ricas

Table A.1: Las 10 ciudades más ricas, regiones largas y regiones pequeñas

Large Regions (Current)	Large Regions (Constant)	Small Regions	Cities
Nenets Autonomous Okrug, Rusia	Nenets Autonomous Okrug, Rusia	Camden City of London, UK	Durham USA
District of Columbia, USA	District of Columbia, USA	Westminster, UK	Luxembourg
Yamalo-Nenets Autonomous Okrug, Rusia	Yamalo-Nenets Autonomous Okrug, Rusia	Wolfsburg Kreisfreie Stadt Germany	Washington USA
Campeche, Mexico	Luxembourg	Ingolstadt Kreisfreie Stadt Germany	San Francisco, USA
Luxembourg	Campeche, Mexico	Tower Hamlets, UK	Boston USA
Northwest Territories, Canada	Northwest Territories, Canada	Basel-Stadt Switzerland	Seattle, USA
Khanty-Mansi Autonomous Okrug – Yugra, Rusia	Brussels Capital Region, Belgium	München Landkreis, Germany	Basel Switzerland
Brussels Capital Region, Belgium	Zurich, Switzerland	Zug, Switzerland	Ottawa, Canada
North Dakota, USA	Khanty-Mansi Autonomous Okrug – Yugra, Rusia	Midland-Odessa, USA	New York, USA
Zurich, Switzerland	North Dakota, USA	Minot, USA	Munich, Germany

A.2 Ciudades y Regiones pequeñas

A.2.1 PCPIB-PPA Regiones pequeñas

Table A.2: Regresión lineal de PCPIB-PPA precios constantes para Regiones pequeñas, escala log-log

Intercep	Coef	R^2	Year
14.775749	-1.5854891	0.64745616	2001
15.197943	-1.6211037	0.65003821	2002
15.574122	-1.6535808	0.65784149	2003
14.131343	-1.5096314	0.64886954	2004
14.60894	-1.5521253	0.66491298	2005
15.232233	-1.6014061	0.6738655	2006
15.916892	-1.6595966	0.69003701	2007
16.936749	-1.7505689	0.71395646	2008
17.296584	-1.789806	0.7164813	2009
17.934022	-1.8444237	0.73989609	2010
18.288933	-1.8700562	0.75518799	2011
18.93947	-1.9286963	0.77421995	2012
19.03267	-1.9310663	0.77264582	2013
19.517744	-1.9714896	0.78477694	2014
19.963031	-2.0095874	0.78487214	2015
20.432203	-2.0455306	0.79141054	2016
20.916687	-2.0825948	0.7980164	2017

Table A.3: Regresión lineal de PCPIB-PPA precios constantes para Regiones pequeñas, escala semi-log

Intercep	Coef	R^2	Year
0.85531423	-7.906E-05	0.84951257	2001
0.89196581	-7.765E-05	0.85004863	2002
0.91104783	-7.621E-05	0.85336948	2003
0.83065254	-7.179E-05	0.86475867	2004
0.82316196	-6.922E-05	0.869591	2005
0.84747811	-6.511E-05	0.87051535	2006
0.83636104	-6.138E-05	0.86788591	2007
0.84934667	-5.895E-05	0.87309639	2008
0.87378246	-6.134E-05	0.87050161	2009
0.85674532	-0.0000583	0.87640599	2010
0.85060527	-5.531E-05	0.88825085	2011
0.83895743	-5.381E-05	0.88808502	2012
0.833976	-5.176E-05	0.8841624	2013
0.82885622	-5.001E-05	0.88202007	2014
0.88795319	-5.061E-05	0.88683856	2015
0.84562055	-0.0000473	0.87274505	2016
0.87573517	-4.601E-05	0.87870176	2017

Table A.4: Regresión lineal de PCPIB-PPA precios constantes para Regiones pequeñas, escala triple-log

Intercep	Coef	R^2	Year
-25.533273	2.5078096	0.97200936	2001
-26.274003	2.571292	0.97373976	2002
-26.758614	2.6116284	0.97426163	2003
-24.658634	2.4022947	0.97457236	2004
-25.0499	2.4330546	0.97526275	2005
-25.879589	2.4958782	0.97488696	2006
-26.666745	2.5593575	0.9750116	2007
-27.721815	2.6491393	0.97340944	2008
-27.979589	2.6805481	0.97268156	2009
-28.558994	2.725599	0.96666899	2010
-28.807839	2.7368104	0.96311236	2011
-28.906151	2.7406288	0.96150599	2012
-29.426481	2.7806642	0.95754154	2013
-29.77969	2.8057424	0.95268851	2014
-30.507994	2.8690688	0.95382791	2015
-30.893158	2.8931131	0.95054718	2016
-31.426523	2.931361	0.94604327	2017

A.2.2 PCPIB-PPA Ciudades

Table A.5: Regresión lineal de PCPIB-PPA precios constantes para ciudades, escala log-log

Intercep	Coef	R^2	Year
26.562294	-2.605675	0.737141	2002.0
27.415802	-2.682620	0.741294	2003.0
27.911948	-2.723629	0.746603	2004.0
20.434691	-2.023032	0.622640	2005.0
20.957999	-2.067102	0.622729	2006.0
21.550463	-2.119152	0.634849	2007.0
20.899208	-2.059966	0.641230	2008.0
21.507119	-2.125042	0.661087	2009.0
21.638815	-2.134362	0.666219	2010.0
21.983243	-2.163971	0.672179	2011.0
22.299282	-2.193177	0.686599	2012.0
22.554072	-2.215957	0.689752	2013.0
22.893558	-2.244720	0.696964	2014.0
23.387992	-2.287068	0.713625	2015.0
23.795918	-2.322924	0.720468	2016.0
23.111863	-2.253837	0.706201	2017.0

Table A.6: Regresión lineal de PCPIB-PPA precios constantes para ciudades, escala semi-log

Intercep	Coef	R^2	Year
2.27441554975229	-7.95237001635495E-05	0.917061395156369	2001
2.06004812644436	-7.47507058741204E-05	0.900273079337753	2002
2.12216365710296	-7.52940614737728E-05	0.896435734504753	2003
2.14531239235762	-7.41909678554252E-05	0.898052402738021	2004
1.78613106985487	-6.61180529576724E-05	0.861287084749373	2005
1.8304206419543	-6.55143380736809E-05	0.856035418146581	2006
1.85103142568045	-6.48142940238351E-05	0.866459599669563	2007
1.77528866686304	-6.34840969897247E-05	0.87424714791132	2008
1.78965224712839	-6.6432261382939E-05	0.887553245684898	2009
1.78206412983362	-6.52055050236289E-05	0.888783133726396	2010
1.79334262073523	-6.46332836809057E-05	0.886909337199369	2011
1.78264986719147	-6.41933226431689E-05	0.894911800565765	2012
1.79467084406467	-6.41290296375759E-05	0.892923795826399	2013
1.80150681319774	-6.33636239667143E-05	0.895543939039664	2014
1.78120066460577	-6.16923505383124E-05	0.912433927298486	2015
1.80499479943552	-6.16053535351042E-05	0.914435502645745	2016
1.76113059463578	-5.91450096918054E-05	0.907663608079474	2017

Table A.7: Regresión lineal de PCPIB-PPA precios constantes para ciudades, escala triple-log

Intercep	Coef	R^2	Year
-46.001127725427	4.29199130689124	0.9741417040929	2001
-42.0454653762112	3.92326276104784	0.986980204084878	2002
-43.2121192787975	4.02782817351354	0.986345391199935	2003
-43.7888328367547	4.073504965651	0.985164364250153	2004
-35.0780570095769	3.25971137164351	0.978249593941124	2005
-35.9542524664615	3.33392462281413	0.978383311526584	2006
-36.5416601254899	3.38307236623812	0.978396214100801	2007
-35.5519261644467	3.29295768345283	0.982085174436527	2008
-35.9782327928323	3.34559187346513	0.983000639663499	2009
-36.0910435988912	3.35106955849482	0.983695907531045	2010
-36.5096585612353	3.38639660114484	0.987084817878005	2011
-36.6226003626478	3.39602252931988	0.987411333989938	2012
-36.98632089791	3.42844169371955	0.988378421771326	2013
-37.3088019501693	3.45376636710181	0.98782182644128	2014
-37.5120524196816	3.46664522412653	0.983518235546634	2015
-37.9647887466882	3.50550846055775	0.982336015158814	2016
-37.2833783103292	3.43417512908903	0.984268979252454	2017

A.3 Regiones Grandes

A.3.1 PCPIB-PPA precios corrientes

Table A.8: Regresión lineal de PCPIB-PPA precios corrientes, escala log-log

Intercep	Coef	R^2	Year
5.9348216048189	-0.731695012888569	0.577227240350642	2001
6.32337588381091	-0.771849433061208	0.605129383632099	2002
6.91363047067546	-0.830474074736781	0.631727915660936	2003
7.41896493773601	-0.879400906469555	0.662715782951844	2004
7.44538765330907	-0.886471194219179	0.69781640958137	2005
7.70425042153739	-0.904382884226739	0.698012325368864	2006
7.91768391833238	-0.922693226498087	0.711691380315251	2007
8.22529565339029	-0.946869089100722	0.708864745471722	2008
8.56174756865365	-0.981430031581562	0.713083880266066	2009
9.12509974484225	-1.03244956444485	0.724708046590867	2010
9.27707917151742	-1.04256935023713	0.726801267657154	2011
9.57033495804834	-1.06852955531559	0.734943743819573	2012
9.76742742081257	-1.0839542118387	0.736540706590249	2013
10.8105513539487	-1.17661257293747	0.748603850489609	2014
10.6306973017851	-1.15712564098038	0.739844124199653	2015
10.6336556829122	-1.15278871471755	0.732731609595204	2016
10.5280400143568	-1.13792072252179	0.729350363227786	2017

Table A.9: Regresión lineal de PCPIB-PPA precios corrientes, escala semi-log

Intercep	Coef	R^2	Year
0.328358283125425	-7.04921764232112E-05	0.874086919162476	2001
0.313564496043885	-6.95980931102101E-05	0.88727645824307	2002
0.327592316784937	-6.90501470696173E-05	0.91720563509366	2003
0.311692092163971	-6.58422121939645E-05	0.9392899152749	2004
0.239728332478255	-6.34073632690492E-05	0.957845468722947	2005
0.241155144455201	-5.82213069908792E-05	0.958638416252368	2006
0.214582131236565	-5.48515892863577E-05	0.95372415414875	2007
0.257482783766124	-5.35139736960012E-05	0.963324776788016	2008
0.275894806298833	-5.52304729755443E-05	0.953519122221354	2009
0.300306886523265	-5.38963047051903E-05	0.956973197582997	2010
0.31090352196044	-5.19078091081075E-05	0.959536341499613	2011
0.325050158724541	-5.11920654422272E-05	0.964123147170833	2012
0.341999614911506	-4.99069651295414E-05	0.965971397481847	2013
0.40947486544471	-4.87984823514357E-05	0.962592579949959	2014
0.411563863377789	-4.81040482061926E-05	0.957100563509485	2015
0.416643189929602	-4.6480644622155E-05	0.948582707546472	2016
0.399339307070329	-4.39939312976173E-05	0.944329705487893	2017

Table A.10: Regresión lineal de PCPIB-PPA precios corrientes, escala triple-log

Intercep	Coef	R^2	Year
-12.0835430057574	1.21631018504062	0.937957890485423	2001
-12.6192095861387	1.27081659528563	0.94888473307973	2002
-13.3509906207242	1.34209064699363	0.95940750915848	2003
-13.8267945830276	1.38550154902614	0.959363612845173	2004
-13.4497838372631	1.35224119494285	0.948631754609842	2005
-13.8803855086892	1.38301272781866	0.95522067516182	2006
-14.0598033389526	1.39586371913369	0.951355311708721	2007
-14.58700971056	1.43895451117517	0.960753718511762	2008
-15.0503622221476	1.48647651977667	0.960456316228808	2009
-15.6830228336741	1.54183039551023	0.956612873986215	2010
-15.9782258347872	1.56353257018143	0.959160153229113	2011
-16.3452091800094	1.59546921342203	0.958376349680992	2012
-16.6021408081169	1.61411970493789	0.958935028752241	2013
-17.9332696077154	1.73011228509707	0.94876144478276	2014
-17.8165276359219	1.71608734009404	0.948341568137579	2015
-17.8673392785325	1.71458465128853	0.946808347400945	2016
-17.7886940363499	1.70013736157731	0.948916105872894	2017

A.3.2 PCPIB-PPA precios constantes

Table A.11: Regresión lineal de PCPIB-PPA precios constantes, escala log-log

Intercep	Coef	R^2	Year
6.09521461646473	-0.721929324883454	0.567707481400567	2001
6.47299490155099	-0.762142228668709	0.596327252443385	2002
6.9192341378685	-0.805997074612411	0.616248826334066	2003
7.43054571016332	-0.855827585153896	0.641427861052707	2004
7.2091046079822	-0.8414946137916	0.67417193627116	2005
7.35337731276324	-0.852503825173203	0.674920804503045	2006
7.50528690811236	-0.867304505197219	0.68821144644377	2007
7.75705814527312	-0.890079366798825	0.692662814835276	2008
8.11408309935789	-0.92804530178646	0.701325208508546	2009
8.72717875466928	-0.985416922069956	0.71808479143459	2010
8.87463439822815	-0.997595436833457	0.718397991691367	2011
9.21525754970572	-1.03042413035914	0.727411167156212	2012
9.47418469564326	-1.05437153914918	0.729503865196983	2013
10.5429772534179	-1.15068221745831	0.7441652456719	2014
10.6306973017851	-1.15712564098038	0.739844124199653	2015
10.7348324938403	-1.16531885838168	0.73814134813996	2016
10.6152413206656	-1.15161923897225	0.733945633360519	2017

Table A.12: Regresión lineal de PCPIB-PPA precios constantes, escala semi-log

Intercep	Coef	R^2	Year
0.356720577816469	-5.13415227441864E-05	0.89264953776263	2001
0.33177275591241	-5.16826318738267E-05	0.907135692589635	2002
0.334780622061996	-5.16174121104215E-05	0.926771861946211	2003
0.320968710154027	-5.04385504599658E-05	0.944104687318255	2004
0.231734861180942	-4.9533637811966E-05	0.957308209885511	2005
0.228250888547461	-4.76221873906316E-05	0.9542519218184	2006
0.208752114809632	-4.65535093183673E-05	0.953383326465275	2007
0.239087923668723	-4.69194656847549E-05	0.962575375832296	2008
0.25165368549417	-4.91114271890581E-05	0.953364654632665	2009
0.273467834302075	-4.84925460417449E-05	0.956968543935565	2010
0.287117560569898	-4.80855931297466E-05	0.958773195129427	2011
0.305213378105458	-4.85520796670937E-05	0.96295843013518	2012
0.326353364823232	-4.88727091858105E-05	0.963906145242896	2013
0.397721237126197	-4.84254364238884E-05	0.961074861837958	2014
0.411563863377789	-4.81040482061926E-05	0.957100563509485	2015
0.416081025009542	-4.75847363455356E-05	0.949217839624069	2016
0.402142718791839	-4.62040300093117E-05	0.945908803130681	2017

Table A.13: Regresión lineal de PCPIB-PPA precios constantes, escala triple-log

Intercep	Coef	R^2	Year
-12.6950019757601	1.23496477300802	0.957478658336777	2001
-13.0725749701126	1.27607579302133	0.965179024611014	2002
-13.680662800232	1.33499126267733	0.973012245738512	2003
-14.2181435175853	1.3861000709403	0.975361833806472	2004
-13.3902669100847	1.31454825049291	0.960806346384952	2005
-13.6192869855076	1.33195689048756	0.963850952810152	2006
-13.7180946974574	1.34108640871645	0.959510302812792	2007
-14.1037263230536	1.37602124241001	0.964814520926436	2008
-14.5234895231208	1.42137912701818	0.963160798528515	2009
-15.1887331092568	1.48145945406519	0.956565403203823	2010
-15.4972911343792	1.50842571147956	0.958855133490177	2011
-15.8451269798519	1.54129785875292	0.959653600422347	2012
-16.2443285579798	1.57834173364319	0.958575407708224	2013
-17.6229840227816	1.70051112772655	0.948073736825343	2014
-17.8165276359219	1.71608734009404	0.948341568137579	2015
-17.9860199462264	1.72989925176791	0.947896605711306	2016
-17.8631162987825	1.71490786681219	0.949620664207488	2017

B Código Implementado

```
hyperref titling longtable booktabs [inline]enumitem [normalem]ulem mathrsfs
    verbose,tmargin=1in,bmargin=1in,lmargin=1in,rmargin=1in
    [breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
import pandas as pd
import numpy as np

# Gráficos
#_
↳ =====

import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns

# Preprocesado y modelado
#_
↳ =====

from scipy.stats import pearsonr
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
import statsmodels.api as sm
```

```

import statsmodels.formula.api as smf

# Configuración matplotlib
#_
↳ =====

plt.rcParams['image.cmap'] = "bwr"
#plt.rcParams['figure.dpi'] = "100"
plt.rcParams['savefig.bbox'] = "tight"
style.use('ggplot') or plt.style.use('ggplot')

# Configuración warnings
#_
↳ =====

import warnings
warnings.filterwarnings('ignore')

Importamos los datos deseados
[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
missing_values = [".."]
GDP=pd.read_excel("/home/diegobalam/ecof/eco/data/Datos2/Region/
↳precios_corrientes.xlsx",na_values = missing_values)

Organizamos los datos por año y eliminamos los missing values
[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
datos1=list(GDP[2001])
datos1= [x for x in datos1 if np.isnan(x) == False]
datos2=list(GDP[2002])
datos2= [x for x in datos2 if np.isnan(x) == False]
datos3=list(GDP[2003])
datos3= [x for x in datos3 if np.isnan(x) == False]

```

```
datos4=list(GDP[2004])
datos4= [x for x in datos4 if np.isnan(x) == False]
datos5=list(GDP[2005])
datos5= [x for x in datos5 if np.isnan(x) == False]
datos6=list(GDP[2006])
datos6= [x for x in datos6 if np.isnan(x) == False]
datos7=list(GDP[2007])
datos7= [x for x in datos7 if np.isnan(x) == False]
datos8=list(GDP[2008])
datos8= [x for x in datos8 if np.isnan(x) == False]
datos9=list(GDP[2009])
datos9= [x for x in datos9 if np.isnan(x) == False]
datos10=list(GDP[2010])
datos10= [x for x in datos10 if np.isnan(x) == False]
datos11=list(GDP[2011])
datos11= [x for x in datos11 if np.isnan(x) == False]
datos12=list(GDP[2012])
datos12= [x for x in datos12 if np.isnan(x) == False]
datos13=list(GDP[2013])
datos13= [x for x in datos13 if np.isnan(x) == False]
datos14=list(GDP[2014])
datos14= [x for x in datos14 if np.isnan(x) == False]
datos15=list(GDP[2015])
datos15= [x for x in datos15 if np.isnan(x) == False]
datos16=list(GDP[2016])
datos16= [x for x in datos16 if np.isnan(x) == False]
datos17=list(GDP[2017])
datos17= [x for x in datos17 if np.isnan(x) == False]
```

Se definen las funciones para obtener los valores de CCDF

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
def ecdf(a):
    x, counts = np.unique(a, return_counts=True)
    cusum = np.cumsum(counts)
    return x, cusum / cusum[-1]

def plot_ecdf(a):
    x, y = ecdf(a)
    x = np.insert(x, 0, x[0])
    y = np.insert(y, 0, 0.)
    y=1-y
    plt.plot(x, y, drawstyle='steps-post')
    plt.grid(True)
    plt.savefig('ecdf.png')
    return y,x

```

Generamos los valores a graficar

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
#x representa el GDP importado en los datos
#y tiene los valores de CCDF

y1,x1=plot_ecdf(datos1)
y2,x2=plot_ecdf(datos2)
y3,x3=plot_ecdf(datos3)
y4,x4=plot_ecdf(datos4)
y5,x5=plot_ecdf(datos5)
y6,x6=plot_ecdf(datos6)
y7,x7=plot_ecdf(datos7)
y8,x8=plot_ecdf(datos8)
y9,x9=plot_ecdf(datos9)

```

```
y10,x10=plot_ecdf(datos10)
y11,x11=plot_ecdf(datos11)
y12,x12=plot_ecdf(datos12)
y13,x13=plot_ecdf(datos13)
y14,x14=plot_ecdf(datos14)
y15,x15=plot_ecdf(datos15)
y16,x16=plot_ecdf(datos16)
y17,x17=plot_ecdf(datos17)
```

```
#z contiene los valores de log(GDP)
#w contiene los valores de log(CCDF)
```

```
z1,w1=np.log(x1),np.log(y1)
z2,w2=np.log(x2),np.log(y2)
z3,w3=np.log(x3),np.log(y3)
z4,w4=np.log(x4),np.log(y4)
z5,w5=np.log(x5),np.log(y5)
z6,w6=np.log(x6),np.log(y6)
z7,w7=np.log(x7),np.log(y7)
z8,w8=np.log(x8),np.log(y8)
z9,w9=np.log(x9),np.log(y9)
z10,w10=np.log(x10),np.log(y10)
z11,w11=np.log(x11),np.log(y11)
z12,w12=np.log(x12),np.log(y12)
z13,w13=np.log(x13),np.log(y13)
z14,w14=np.log(x14),np.log(y14)
z15,w15=np.log(x15),np.log(y15)
z16,w16=np.log(x16),np.log(y16)
z17,w17=np.log(x17),np.log(y17)
```

```
#t contiene los valores de log(-log(CCDF))
#El cual se usara en el triple log.
```

```
t1=np.log(-w1)
t2=np.log(-w2)
t3=np.log(-w3)
t4=np.log(-w4)
t5=np.log(-w5)
t6=np.log(-w6)
t7=np.log(-w7)
t8=np.log(-w8)
t9=np.log(-w9)
t10=np.log(-w10)
t11=np.log(-w11)
t12=np.log(-w12)
t13=np.log(-w13)
t14=np.log(-w14)
t15=np.log(-w15)
t16=np.log(-w16)
t17=np.log(-w17)
```

Se generan los data frames A,B,C y D, los cuales contienen los valores del eje x y del eje y. En el caso de B se tienen en el eje x=GDP y en y=CCDF. En A x=log(GDP) y y=log(CCDF). En C x=GDP y en y=log(CCDF). Finalmente, en D x=log(GDP) y y=log(-log(CCDF)).

```
[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
X1= pd.DataFrame (x1,columns=['x1'])
Y1= pd.DataFrame (y1,columns=['y1'])
X2= pd.DataFrame (x2,columns=['x2'])
Y2= pd.DataFrame (y2,columns=['y2'])
```

```
X3= pd.DataFrame (x3,columns=['x3'])
Y3= pd.DataFrame (y3,columns=['y3'])
X4= pd.DataFrame (x4,columns=['x4'])
Y4= pd.DataFrame (y4,columns=['y4'])
X5= pd.DataFrame (x5,columns=['x5'])
Y5= pd.DataFrame (y5,columns=['y5'])
X6= pd.DataFrame (x6,columns=['x6'])
Y6= pd.DataFrame (y6,columns=['y6'])
X7= pd.DataFrame (x7,columns=['x7'])
Y7= pd.DataFrame (y7,columns=['y7'])
X8= pd.DataFrame (x8,columns=['x8'])
Y8= pd.DataFrame (y8,columns=['y8'])
X9= pd.DataFrame (x9,columns=['x9'])
Y9= pd.DataFrame (y9,columns=['y9'])
X10= pd.DataFrame (x10,columns=['x10'])
Y10= pd.DataFrame (y10,columns=['y10'])
X11= pd.DataFrame (x11,columns=['x11'])
Y11= pd.DataFrame (y11,columns=['y11'])
X12= pd.DataFrame (x12,columns=['x12'])
Y12= pd.DataFrame (y12,columns=['y12'])
X13= pd.DataFrame (x13,columns=['x13'])
Y13= pd.DataFrame (y13,columns=['y13'])
X14= pd.DataFrame (x14,columns=['x14'])
Y14= pd.DataFrame (y14,columns=['y14'])
X15= pd.DataFrame (x15,columns=['x15'])
Y15= pd.DataFrame (y15,columns=['y15'])
X16= pd.DataFrame (x16,columns=['x16'])
Y16= pd.DataFrame (y16,columns=['y16'])
X17= pd.DataFrame (x17,columns=['x17'])
Y17= pd.DataFrame (y17,columns=['y17'])
```

```
B1=pd.DataFrame()  
B2=pd.DataFrame()  
B3=pd.DataFrame()  
B4=pd.DataFrame()  
B5=pd.DataFrame()  
B6=pd.DataFrame()  
B7=pd.DataFrame()  
B8=pd.DataFrame()  
B9=pd.DataFrame()  
B10=pd.DataFrame()  
B11=pd.DataFrame()  
B12=pd.DataFrame()  
B13=pd.DataFrame()  
B14=pd.DataFrame()  
B15=pd.DataFrame()  
B16=pd.DataFrame()  
B17=pd.DataFrame()  
B1['x']=X1.iloc[:,0]  
B1['y']=Y1.iloc[:,0]  
B2['x']=X2.iloc[:,0]  
B2['y']=Y2.iloc[:,0]  
B3['x']=X3.iloc[:,0]  
B3['y']=Y3.iloc[:,0]  
B4['x']=X4.iloc[:,0]  
B4['y']=Y4.iloc[:,0]  
B5['x']=X5.iloc[:,0]  
B5['y']=Y5.iloc[:,0]  
B6['x']=X6.iloc[:,0]  
B6['y']=Y6.iloc[:,0]  
B7['x']=X7.iloc[:,0]
```

```

B7['y']=Y7.iloc[:,0]
B8['x']=X8.iloc[:,0]
B8['y']=Y8.iloc[:,0]
B9['x']=X9.iloc[:,0]
B9['y']=Y9.iloc[:,0]
B10['x']=X10.iloc[:,0]
B10['y']=Y10.iloc[:,0]
B11['x']=X11.iloc[:,0]
B11['y']=Y11.iloc[:,0]
B12['x']=X12.iloc[:,0]
B12['y']=Y12.iloc[:,0]
B13['x']=X13.iloc[:,0]
B13['y']=Y13.iloc[:,0]
B14['x']=X14.iloc[:,0]
B14['y']=Y14.iloc[:,0]
B15['x']=X15.iloc[:,0]
B15['y']=Y15.iloc[:,0]
B16['x']=X16.iloc[:,0]
B16['y']=Y16.iloc[:,0]
B17['x']=X17.iloc[:,0]
B17['y']=Y17.iloc[:,0]

```

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
Z1= pd.DataFrame (z1,columns=['x1'])
W1= pd.DataFrame (w1,columns=['y1'])
Z2= pd.DataFrame (z2,columns=['x2'])
W2= pd.DataFrame (w2,columns=['y2'])
Z3= pd.DataFrame (z3,columns=['x3'])
W3= pd.DataFrame (w3,columns=['y3'])
Z4= pd.DataFrame (z4,columns=['x4'])

```

```
W4= pd.DataFrame (w4,columns=['y4'])
Z5= pd.DataFrame (z5,columns=['x5'])
W5= pd.DataFrame (w5,columns=['y5'])
Z6= pd.DataFrame (z6,columns=['x6'])
W6= pd.DataFrame (w6,columns=['y6'])
Z7= pd.DataFrame (z7,columns=['x7'])
W7= pd.DataFrame (w7,columns=['y7'])
Z8= pd.DataFrame (z8,columns=['x8'])
W8= pd.DataFrame (w8,columns=['y8'])
Z9= pd.DataFrame (z9,columns=['x9'])
W9= pd.DataFrame (w9,columns=['y9'])
Z10= pd.DataFrame (z10,columns=['x10'])
W10= pd.DataFrame (w10,columns=['y10'])
Z11= pd.DataFrame (z11,columns=['x11'])
W11= pd.DataFrame (w11,columns=['y11'])
Z12= pd.DataFrame (z12,columns=['x12'])
W12= pd.DataFrame (w12,columns=['y12'])
Z13= pd.DataFrame (z13,columns=['x13'])
W13= pd.DataFrame (w13,columns=['y13'])
Z14= pd.DataFrame (z14,columns=['x14'])
W14= pd.DataFrame (w14,columns=['y14'])
Z15= pd.DataFrame (z15,columns=['x15'])
W15= pd.DataFrame (w15,columns=['y15'])
Z16= pd.DataFrame (z16,columns=['x16'])
W16= pd.DataFrame (w16,columns=['y16'])
Z17= pd.DataFrame (z17,columns=['x17'])
W17= pd.DataFrame (w17,columns=['y17'])
A1=pd.DataFrame()
A2=pd.DataFrame()
A3=pd.DataFrame()
```

```
A4=pd.DataFrame()
A5=pd.DataFrame()
A6=pd.DataFrame()
A7=pd.DataFrame()
A8=pd.DataFrame()
A9=pd.DataFrame()
A10=pd.DataFrame()
A11=pd.DataFrame()
A12=pd.DataFrame()
A13=pd.DataFrame()
A14=pd.DataFrame()
A15=pd.DataFrame()
A16=pd.DataFrame()
A17=pd.DataFrame()
A1['x']=Z1.iloc[:,0]
A1['y']=W1.iloc[:,0]
A2['x']=Z2.iloc[:,0]
A2['y']=W2.iloc[:,0]
A3['x']=Z3.iloc[:,0]
A3['y']=W3.iloc[:,0]
A4['x']=Z4.iloc[:,0]
A4['y']=W4.iloc[:,0]
A5['x']=Z5.iloc[:,0]
A5['y']=W5.iloc[:,0]
A6['x']=Z6.iloc[:,0]
A6['y']=W6.iloc[:,0]
A7['x']=Z7.iloc[:,0]
A7['y']=W7.iloc[:,0]
A8['x']=Z8.iloc[:,0]
A8['y']=W8.iloc[:,0]
```

```
A9['x']=Z9.iloc[:,0]
A9['y']=W9.iloc[:,0]
A10['x']=Z10.iloc[:,0]
A10['y']=W10.iloc[:,0]
A11['x']=Z11.iloc[:,0]
A11['y']=W11.iloc[:,0]
A12['x']=Z12.iloc[:,0]
A12['y']=W12.iloc[:,0]
A13['x']=Z13.iloc[:,0]
A13['y']=W13.iloc[:,0]
A14['x']=Z14.iloc[:,0]
A14['y']=W14.iloc[:,0]
A15['x']=Z15.iloc[:,0]
A15['y']=W15.iloc[:,0]
A16['x']=Z16.iloc[:,0]
A16['y']=W16.iloc[:,0]
A17['x']=Z17.iloc[:,0]
A17['y']=W17.iloc[:,0]
```

```
[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
C1=pd.DataFrame()
C2=pd.DataFrame()
C3=pd.DataFrame()
C4=pd.DataFrame()
C5=pd.DataFrame()
C6=pd.DataFrame()
C7=pd.DataFrame()
C8=pd.DataFrame()
C9=pd.DataFrame()
C10=pd.DataFrame()
```

```
C11=pd.DataFrame()
C12=pd.DataFrame()
C13=pd.DataFrame()
C14=pd.DataFrame()
C15=pd.DataFrame()
C16=pd.DataFrame()
C17=pd.DataFrame()
C1['x']=X1.iloc[:,0]
C1['y']=W1.iloc[:,0]
C2['x']=X2.iloc[:,0]
C2['y']=W2.iloc[:,0]
C3['x']=X3.iloc[:,0]
C3['y']=W3.iloc[:,0]
C4['x']=X4.iloc[:,0]
C4['y']=W4.iloc[:,0]
C5['x']=X5.iloc[:,0]
C5['y']=W5.iloc[:,0]
C6['x']=X6.iloc[:,0]
C6['y']=W6.iloc[:,0]
C7['x']=X7.iloc[:,0]
C7['y']=W7.iloc[:,0]
C8['x']=X8.iloc[:,0]
C8['y']=W8.iloc[:,0]
C9['x']=X9.iloc[:,0]
C9['y']=W9.iloc[:,0]
C10['x']=X10.iloc[:,0]
C10['y']=W10.iloc[:,0]
C11['x']=X11.iloc[:,0]
C11['y']=W11.iloc[:,0]
C12['x']=X12.iloc[:,0]
```

```
C12['y']=W12.iloc[:,0]
C13['x']=X13.iloc[:,0]
C13['y']=W13.iloc[:,0]
C14['x']=X14.iloc[:,0]
C14['y']=W14.iloc[:,0]
C15['x']=X15.iloc[:,0]
C15['y']=W15.iloc[:,0]
C16['x']=X16.iloc[:,0]
C16['y']=W16.iloc[:,0]
C17['x']=X17.iloc[:,0]
C17['y']=W17.iloc[:,0]
```

```
[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
T1= pd.DataFrame (t1,columns=['y1'])
T2= pd.DataFrame (t2,columns=['y2'])
T3= pd.DataFrame (t3,columns=['y3'])
T4= pd.DataFrame (t4,columns=['y4'])
T5= pd.DataFrame (t5,columns=['y5'])
T6= pd.DataFrame (t6,columns=['y6'])
T7= pd.DataFrame (t7,columns=['y7'])
T8= pd.DataFrame (t8,columns=['y8'])
T9= pd.DataFrame (t9,columns=['y9'])
T10= pd.DataFrame (t10,columns=['y10'])
T11= pd.DataFrame (t11,columns=['y11'])
T12= pd.DataFrame (t12,columns=['y12'])
T13= pd.DataFrame (t13,columns=['y13'])
T14= pd.DataFrame (t14,columns=['y14'])
T15= pd.DataFrame (t15,columns=['y15'])
T16= pd.DataFrame (t16,columns=['y16'])
T17= pd.DataFrame (t17,columns=['y17'])
```

```
D1=pd.DataFrame()
D2=pd.DataFrame()
D3=pd.DataFrame()
D4=pd.DataFrame()
D5=pd.DataFrame()
D6=pd.DataFrame()
D7=pd.DataFrame()
D8=pd.DataFrame()
D9=pd.DataFrame()
D10=pd.DataFrame()
D11=pd.DataFrame()
D12=pd.DataFrame()
D13=pd.DataFrame()
D14=pd.DataFrame()
D15=pd.DataFrame()
D16=pd.DataFrame()
D17=pd.DataFrame()
D1['x']=Z1.iloc[:,0]
D1['y']=T1.iloc[:,0]
D2['x']=Z2.iloc[:,0]
D2['y']=T2.iloc[:,0]
D3['x']=Z3.iloc[:,0]
D3['y']=T3.iloc[:,0]
D4['x']=Z4.iloc[:,0]
D4['y']=T4.iloc[:,0]
D5['x']=Z5.iloc[:,0]
D5['y']=T5.iloc[:,0]
D6['x']=Z6.iloc[:,0]
D6['y']=T6.iloc[:,0]
D7['x']=Z7.iloc[:,0]
```

```

D7['y']=T7.iloc[:,0]
D8['x']=Z8.iloc[:,0]
D8['y']=T8.iloc[:,0]
D9['x']=Z9.iloc[:,0]
D9['y']=T9.iloc[:,0]
D10['x']=Z10.iloc[:,0]
D10['y']=T10.iloc[:,0]
D11['x']=Z11.iloc[:,0]
D11['y']=T11.iloc[:,0]
D12['x']=Z12.iloc[:,0]
D12['y']=T12.iloc[:,0]
D13['x']=Z13.iloc[:,0]
D13['y']=T13.iloc[:,0]
D14['x']=Z14.iloc[:,0]
D14['y']=T14.iloc[:,0]
D15['x']=Z15.iloc[:,0]
D15['y']=T15.iloc[:,0]
D16['x']=Z16.iloc[:,0]
D16['y']=T16.iloc[:,0]
D17['x']=Z17.iloc[:,0]
D17['y']=T17.iloc[:,0]

```

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
clf =LinearRegression()
clf2=LinearRegression()

```

Se muestra un ejemplo para generar una de las gráficas. En este caso se grafica en x el GDP y en y CCDF. Para los demas casos se cambían las variables x_i, y_i por z_i, w_i, t_i dependiendo la grafica que se desee.

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:

```

```

plt.figure(figsize=(10,7))
plt.scatter(x1,y1,marker='.',label="2001")
plt.scatter(x2,y2,marker='.',label='2002')
plt.scatter(x3,y3,marker='.',label='2003')
plt.scatter(x4,y4,marker='.',label='2004')
plt.scatter(x5,y5,marker='.',label='2005')
plt.scatter(x6,y6,marker='.',label='2006')
plt.scatter(x7,y7,marker='.',label='2007')
plt.scatter(x8,y8,marker='.',label='2008')
plt.scatter(x9,y9,marker='.',label='2009')
plt.scatter(x10,y10,marker='.',label='2010')
plt.scatter(x11,y11,marker='.',label='2011')
plt.scatter(x12,y12,marker='.',label='2012')
plt.scatter(x13,y13,marker='.',label='2013')
plt.scatter(x14,y14,marker='.',label='2014')
plt.scatter(x15,y15,marker='.',label='2015')
plt.scatter(x16,y16,marker='.',label='2016')
plt.scatter(x17,y17,marker='.',label='2017')
plt.title('Region GDP Current Price ')
plt.xlabel('GDP')
plt.ylabel('CCDF')
plt.legend(numpoints=1)
#plt.savefig("/home/diegobalam/ecof/eco/data/Datos2/Region/
↳gdp_region_current.jpg")

```

En el código siguiente, se realizan las regresiones lineales y se produce una tabla con los datos obtenidos de estas regresiones.

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder]:
coefB=pd.DataFrame(columns = ['Intercept' , 'Coef' , 'R2' , 'year'])

```

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,

```

```

colframe=cellborder[1]:
year=2001
for i in range(0,17):
    B=[B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14,B15,B16,B17]
    B[i]=B[i].dropna()
    #B[i]=B[i][:-1]
    x=np.array(B[i]['x'])
    y=np.array(B[i]['y'])
    #x=np.delete(x,-1)
    #y=np.delete(y,-1)
    x=x.reshape(-1, 1)
    y=y.reshape(-1, 1)
    clf.fit(x, y)
    a=list([float(clf.intercept_),float(clf.coef_.flatten()),clf.
↪score(x, y),year])
    year+=1
    coefB.loc[i]=a

#coefB.to_excel('/home/diegobalam/ecof/eco/data/Datos2/Region/coefB.
↪xlsx')

```

```

[breakable, size=fbox, boxrule=1pt, pad at break*=1mm,colback=cellbackground,
colframe=cellborder[1]:

```