



**Universidad  
Zaragoza**

***TRABAJO DE FIN DE GRADO***

DESIGUALDAD MONETARIA Y DÓNDE ENCONTRARLA

Un análisis distributivo general y regional de los datos declarados en el IRPF

MONETARY INEQUALITY AND WHERE TO FIND IT

A general and regional distributive analysis of the data declared in the personal income tax of Spain.

AUTOR

Maximiliano Greco

DIRECTOR

Fernando Rodrigo Sauco

5 julio 2017

RESUMEN

El presente trabajo analiza, por medio de un paquete de programación ideado y desarrollado por el autor, el comportamiento reciente del IRPF español en términos de progresividad y redistribución (principio impositivo de equidad). En la primera parte del trabajo, se intenta poner de relieve la importancia relativa de cada tramo de renta, donde se presentan los principales agregados monetarios y cómo evolucionan en cada fase del IRPF. Este análisis se divide en: la descripción estadística de los agregados, una evaluación de la distribución en términos de equidad (usando los indicadores habituales de Gini, Theil y Atkinson) y un análisis de progresividad aplicando tipos impositivos medios y el índice de Kakwani. En la segunda parte del trabajo, se aproximan las desviaciones regionales de las comunidades autónomas respecto del promedio redistributivo de España, mostrando tanto el comportamiento de estadísticos descriptivos básicos, como las medidas estándar de desigualdad y progresividad.



ABSTRACT

The present work analyzes, through a programming package conceived and developed by the author, the recent behavior of the personal income tax (IRPF as known for its acronym in Spanish) in terms of progressivity and redistribution (tax equity principle). In the first part of the work, we try to highlight the relative importance of each income segment, where the main monetary aggregates are presented and its evolution in each phase of the IRPF is analyzed. This analysis is divided into: the statistical description of the aggregates, an assessment of the distribution in terms of equality (using the usual indicators of Gini, Theil and Atkinson) and a progressivity analysis applying average tax rates and the Kakwani index. In the second part of the work, the regional deviations of the autonomous communities are approximated with respect to the redistributive average of Spain, showing both the behavior of basic descriptive statistics and the standard measures of inequality and progressivity.

## Tabla de contenido

<b>1 INTRODUCCIÓN: MOTIVACIÓN, OBJETIVOS Y ESTRUCTURA GENERAL DEL TRABAJO .....</b>	<b>6</b>
1.1 Motivación.....	6
1.2 Objetivos.....	7
1.3 Estructura .....	8
<b>2 EI ESQUEMA LIQUIDATORIO DEL IRPF .....</b>	<b>9</b>
2.1 Descripción de la metodología y análisis descriptivo de la muestra .....	9
2.2 La distribución de la renta.....	10
2.3 Distribución de los tramos por agregado monetario.....	13
2.4 Un análisis complementario de desigualdad.....	14
2.5 Progresividad: ¿cómo evoluciona la equidad vertical en el IRPF?.....	17
2.5.1 Mas sobre progresividad, Kakwani.....	19
<b>3 EI COMPORTAMIENTO DEL IRPF DESDE EL ÁMBITO REGIONAL.....</b>	<b>20</b>
3.1 La distribución de la renta.....	20
3.1.1 El camino desde la renta declarada hasta la cuota líquida .....	21
3.2 Hablando de progresividad, los tipos medios.....	25
3.3 Medidas de desigualdad .....	27
3.4 Una medida de progresividad adicional, Kakwani.....	28
<b>4 CONCLUSIONES, LIMITACIONES Y EXTENSIONES .....</b>	<b>30</b>
<b>5 ANEXOS.....</b>	<b>32</b>
5.1 Representación resumida “minimal example” del código implementado en IneqPy: .	32
5.2 Esquemas de liquidación general del IRPF (ejercicio 2012).....	39
5.3 Notas metodológicas .....	42
5.4 Tablas y gráficos adicionales del capítulo 3.....	45

## Índice de tablas

Tabla 2.1: Distribución de declarantes por tramos de renta.....	10
Tabla 2.2: Resumen estadístico de los agregados monetarios .....	12
Tabla 2.3: Indicadores de desigualdad.....	14
Tabla 2.4: Cuantiles de los agregados monetarios.....	17
Tabla 2.5: Índice de Kakawani .....	19
Tabla 3.1: Media de los principales agregados por comunidad autónoma .....	21
Tabla 3.2: Ranking de la media de los principales agregados por comunidades.....	22
Tabla 3.3: Resumen de indicadores de desigualdad y progresividad .....	27
Tabla 5.1: Definición de los principales agregados monetarios .....	42
Tabla 5.2: Códigos y tramos de renta definidos .....	42
Tabla 5.3: Codificación de las comunidades autónomas .....	43
Tabla 5.4: Media de cada agregado monetario para cada tramo de renta.....	44
Tabla 5.5: Estadísticos descriptivos por comunidad autónoma.....	45
Tabla 5.6: Tipo impositivo medio por decilas de renta para el cociente entre ci y bl. ....	47

## Índice de figuras

Gráfico 2.1: Distribución de la renta por tramos .....	11
Gráfico 2.2: Importancia de cada tramo por agregado monetario .....	13
Gráfico 2.3: Indicadores de desigualdad.....	15
Gráfico 2.4: Tipo impositivo medio por decila.....	18
Gráfico 3.1: Distribución de la renta por CCAA .....	20
Gráfico 3.2: Diferencias regionales respecto de la media para España .....	24
Gráfico 3.3: Tipos impositivos medios por CC.AA .....	26
Gráfico 3.4: Ranking de contribución a la desigualdad.....	28
Gráfico 3.5: Desviación regional del índice de Kakwani .....	29
Gráfico 5.1: Distribución de tramos de renta por CCAA .....	46

# 1 INTRODUCCIÓN: MOTIVACIÓN, OBJETIVOS Y ESTRUCTURA GENERAL DEL TRABAJO

## 1.1 Motivación

En el presente trabajo, se ha llevado a cabo un análisis distributivo del impuesto sobre la renta de las personas físicas (IRPF) en España. Según se cuenta en el preámbulo de la propia ley del impuesto (Ley 35/2006), en vigor desde el ejercicio 2007, el IRPF está encaminado a promover una mayor igualdad (por su diseño progresivo) y financiar el gasto público. Si se elige este impuesto es porque, en la actualidad, es el instrumento fiscal más potente con el que cuenta el legislador para reequilibrar las desigualdades monetarias surgidas de los mercados laborales y de capitales.

Por lo tanto, cabe preguntarse ¿qué es y por qué debería importarnos la desigualdad? En primer lugar, en economía decimos que existe desigualdad cuando, principalmente, se hace referencia a la desigualdad monetaria, es decir, la diferencia de renta<sup>1</sup> o ingresos entre ciudadanos o declarantes en este caso<sup>2</sup>. En segundo lugar, desde un punto de vista teórico, corregir la desigualdad (asumiendo que la actual es excesiva, aunque exista un nivel óptimo, (Atkinson, 2016)), supone enfrentarnos a un problema de incentivos. Por ejemplo, con una medida como el IRPF se podrían conseguir dos efectos: por un lado, redistribuir riqueza, aumentando el bienestar social y, por tanto, promover así el desarrollo de la economía (Novales, 2011). Por otro lado, y como efecto colateral, desincentiva tanto a los más pobres cómo a los más ricos a seguir trabajando y produciendo para obtener esa renta (Atkinson, 2016)).

La teoría económica proporciona unos requisitos deseables que un tributo debería cumplir para contribuir a alcanzar los objetivos de carácter público. Uno de ellos, que el impuesto sea justo (que sea equitativo, tanto horizontal, como verticalmente), se puede traducir

---

<sup>1</sup> Asumiendo la definición de renta propia de la norma fiscal. que no haya recibido rentas o no este sujeto al impuesto.

<sup>2</sup> Según lo que dispone la legislación (Ley 35/2006) un ciudadano puede no ser declarante, por varias razones, una de ellas es que no haya recibido rentas o no este sujeto al impuesto.

en que las personas que cobren o tengan una situación similar sean tratadas de forma similar y, del mismo modo, personas con situaciones distintas, sean tratadas de forma distinta.

La redistribución se consigue, así, mediante un correcto diseño impositivo del IRPF que, por ejemplo, construya un impuesto progresivo ayudando a reducir las desigualdades (principio de equidad vertical) y a su vez, no las origine cuando los individuos obtengan la misma renta de mercado (principio de equidad horizontal).

No obstante, como veremos, la progresividad general o global del impuesto no sólo viene determinada por una estructura de tipos crecientes en su tarifa, sino que también viene influida por su carácter dual (tratamiento distinto de las rentas según su fuente de procedencia), o por el comportamiento más o menos regresivo de las reducciones en su base y de las bonificaciones en su cuota.

## **1.2 Objetivos**

En el presente trabajo, se evalúa el comportamiento del IRPF de España para el año 2012. ¿Es el IRPF globalmente un impuesto progresivo?, ¿contribuye a reducir la desigualdad de forma significativa?, ¿a través de qué elementos concretos del impuesto?, ¿existen diferencias entre CCAA?, ¿afecta a todos los perfiles de contribuyentes por igual? Estas son algunas de las preguntas a responder.

Junto con todo lo anterior, la desigualdad, resulta ser un tema muy ameno para un investigador social, con un intenso debate social: se puede percibir de forma frecuente en la ciudadanía y en los medios de comunicación, sobre todo, tras la gran recesión originada en España a partir de 2008.

En el anexo también se incluye las funciones más relevantes<sup>3</sup>. Por todo lo comentado anteriormente, la desigualdad, y la mejora en la toma de decisiones acerca de este concepto, parece ser una de las asignaturas pendientes dentro de la esfera política española. La habilidad de programar tiende a ser una habilidad propia de informáticos que resulta de increíble utilidad en muchas áreas de conocimiento, ignorar su existencia y tendencia, en opinión del autor, representa una desventaja. En cuanto a otros contenidos empíricos, lo novedoso respecto de

---

<sup>3</sup> Para ampliar ver anexo página 36.

otros trabajos es quizás también el análisis regional del IRPF por comunidades autónomas que se presenta en el mismo.

### **1.3 Estructura**

La parte central del trabajo se encuentra dividido en tres partes fundamentales. Tras una breve presentación de cuál es el esquema de liquidación que adoptaba el IRPF en el ejercicio 2012 (es la última muestra de declarantes disponible para el investigador al comenzar el presente trabajo, aunque actualmente ya está disponible la muestra de 2013) en una primera parte se realiza un análisis de forma agregada para el conjunto de España, se presentan estadísticos descriptivos de la *Muestra de Declarantes del IRPF de 2012*, e indicadores de desigualdad y progresividad. En la segunda parte, se realiza el mismo análisis distinguiendo por CCAA. Termina con un apartado de conclusiones sobre lo aprendido, las limitaciones con las que se ha contado para llevarlo a cabo, y las futuras extensiones (nuevos interrogantes surgidos) a las que autor querría dar respuesta en un posible futuro académico.



## 2 EL ESQUEMA LIQUIDATORIO DEL IRPF

El impuesto sobre la renta de las personas físicas (IRPF) es un impuesto personal, progresivo y directo, que grava la renta obtenida en un año por las personas físicas residentes en España y de “[...] importancia fundamental para hacer efectivo el mandato del artículo 31 de la Constitución Española, que exige la contribución de todos «... al sostenimiento de los gastos públicos de acuerdo con su capacidad económica mediante un sistema tributario justo inspirado en los principios de igualdad y progresividad que, en ningún caso, tendrá alcance confiscatorio»”<sup>4</sup>.

En los esquemas liquidatorios del IRPF, se puede ver a grandes rasgos los diferentes elementos que intervienen en su liquidación, y funcionamiento. A priori, los elementos más relevantes del IRPF son, por un lado, la separación de las bases en dos conceptos (general y del ahorro), siendo la tarifa de la primera más progresiva. También existen reducciones, favoreciendo algunas de ellas a las rentas más altas por estar vinculadas a la capacidad de ahorro de los individuos y cuyo ahorro es proporcional al tipo marginal. Sin embargo, los mínimos personales y familiares, por su especial configuración, provocarán, casi siempre, provocan un ahorro fiscal equivalente a la cuantía de los mismos multiplicada por el tipo más bajo de la tarifa general. Finalmente, existen deducciones que fundamentalmente son utilizadas por las rentas más altas (donativos, inversión en vivienda, inversiones empresariales, etc.), aunque por una cuestión de limitación de espacio, no se analizan específicamente en el presente trabajo. Todos estos elementos generan efectos que, en ocasiones, favorecen la redistribución y, en otras, la entorpezcan (algunas reducciones y deducciones, principalmente).

Por último, pero no menos importante, existe un efecto que este trabajo no puede medir: el efecto que genera el fraude fiscal en la equidad. Si éste está concentrado en las rentas más altas, la redistribución calculada por medio de los datos declarados estará claramente sobreestimada.

### 2.1 Descripción de la metodología y análisis descriptivo de la muestra

En este capítulo se analiza la *Muestra de declarantes del IRPF AEAT-IEF 2012*. Se presenta una descripción estadística de la muestra de declarantes del IRPF, el tamaño de la muestra, la

---

<sup>4</sup> Boe.es. (2017). BOE.es - Documento consolidado BOE-A-2006-20764.<http://www.boe.es/buscar/act.php?id=BOE-A-2006-20764>.

población que representa y los principales estadísticos descriptivos. Posteriormente, se procede a definir las principales variables monetarias que conforman el IRPF (la renta, base imponible, base liquidable, cuota íntegra, cuota líquida), el resultado para toda la población y la metodología usada<sup>5</sup>. Finalmente, ya en el anexo del TFG, se presentan una notas metodológicas<sup>6</sup> que hacen referencia a la definición de la variables usadas y su codificación.

## 2.2 La distribución de la renta

La **¡Error! No se encuentra el origen de la referencia.** muestra la distribución de declarantes por tramos de renta, por declarante se entiende a cada individuo que ha realizado una declaración conjunta o individual, más adelante se describen los detalles de la construcción.

A continuación se puede observar la distribución de los tramos de renta :

*Tabla 2.1: Distribución de declarantes por tramos de renta*

Tramo	Observaciones	Declaraciones	Porcentaje
1	445.254,00	452.923,51	2,34 %
2	800.792,00	3.823.019,99	19,73 %
3	125.915,00	3.697.083,81	19,08 %
4	49.548,00	3.634.723,57	18,76 %
5	25.676,00	2.677.408,78	13,82 %
6	15.815,00	1.772.186,21	9,14 %
7	9.849,00	1.235.906,58	6,38 %
8	7.347,00	646.528,63	3,34 %
9	5.427,00	383.326,75	1,98 %
10	4.188,00	259.028,82	1,34 %
11	134.501,00	181.538,08	0,94 %
12	449.913,00	615.706,76	3,18 %
TOTAL	2.074.225,00	19.379.381,50	100,00 %

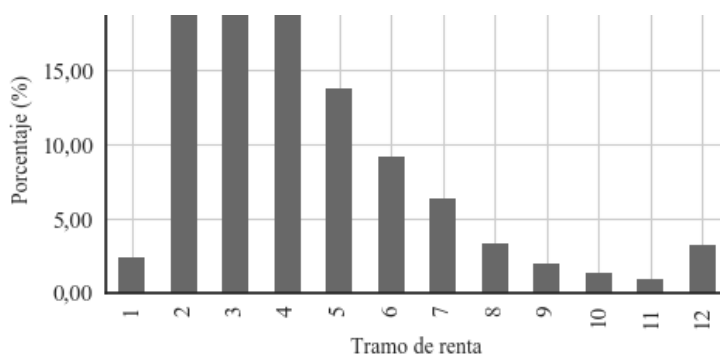
Veamos la distribución de frecuencias y la representatividad de los tramos en la muestra. Por tramo de renta<sup>7</sup>, de la muestra se obtiene el número de observaciones, es decir, el número de muestras que hay en dicho tramo y además cada observación ha de ponderarse por un peso para obtener la población que representan, es decir, la equivalencia de dicha muestra en la población real de declaraciones.

<sup>5</sup> Ver en el anexo para más información:

<sup>6</sup> Ver en el anexo para más información:

<sup>7</sup> Ver: **¡Error! No se encuentra el origen de la referencia.**

Gráfico 2.1: Distribución de la renta por tramos



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

Si se concibe el IRPF como un ‘proceso productivo’ en el que entran *inputs* (ingresos o rentas) y salen *outputs* (cuota líquida), en esta caja habría un primer filtro por el que pasan todas rentas percibidas por las personas físicas, procedentes de tres orígenes o fuentes de renta: rentas del trabajo, rentas del capital y rentas de actividades económicas.

Estas fuentes de renta reciben un tratamiento u otro en función de lo que disponga la ley, es decir pasan por distintas fases. Primero, se computan las rentas brutas (*renta\_periodo*) sujetas al impuesto en función de su origen, la primera fase consiste en convertir dichas rentas brutas en rendimientos netos (*renta*) quitándole los posibles gastos deducibles y reducciones. Se procede con una segunda fase posterior, que consiste en imputar renta y agregar ganancias/pérdidas patrimoniales que, como resultado, dan lugar a la base imponible (*bi*). La tercera fase, consiste en quitar a lo anterior las reducciones (por tributación conjunta, seguridad social, discapacidad, etc.) y se obtiene la base liquidable (*bl*). La cuarta fase devuelve la cuota íntegra (*ci*), que consiste en aplicar la escala de gravamen estatal y autonómico a la base imponible anterior y reducir según el cálculo de los mínimos personales y familiares. La quinta fase se calcula aplicando a lo anterior deducciones estatales y autonómicas, como resultado se tiene la cuota líquida (*cl*). La renta después de impuestos se ha calculado cómo la diferencia entre la renta y la cuota líquida<sup>8</sup>.

<sup>8</sup> La renta después de impuestos se ha seguido la metodología marcada por (Picos Sánchez, Lopez, & Marín, 2011).

Adicionalmente se han añadido dos variables: la “*renta\_periodo*” (léase cómo renta del periodo) es la suma de rentas e ingresos procedentes de las distintas fuentes de renta percibidas por los declarantes durante el periodo impositivo antes de aplicar gastos deducibles, y la “*renta\_ddi*” que es la renta del periodo menos la cuota líquida ( “*renta\_periodo – cl*” ).

Por lo tanto, para este apartado se analizan los anteriores agregados monetarios para el conjunto del territorio español, para luego ir al detalle por comunidades autónomas en el siguiente capítulo.

Para el conjunto de la economía española el comportamiento estadístico de los principales agregados monetarios queda reflejado en la siguiente tabla:

*Tabla 2.2: Resumen estadístico de los agregados monetarios*

Variable	Media	Percentil 10%	Mediana	Percentil 90%	Varianza	Cv	Asimetría	Curtosis
<b>renta_periodo</b>	21.997,44	5.531,50	17.218,48	39.657,32	2.192.090.401,53	0,47	399,84	374.184,90
<b>renta</b>	19.252,86	2.184,64	14.589,70	37.046,41	2.192.442.234,45	0,41	399,70	374.063,28
<b>bi</b>	19.287,32	2.226,44	14.648,14	37.056,86	2.194.142.533,28	0,41	399,30	373.490,78
<b>bl</b>	18.340,69	1.698,70	13.715,85	35.813,96	2.166.585.853,05	0,39	406,64	383.022,07
<b>ci</b>	3.740,52	0,00	1.692,56	8.217,39	301.432.853,93	0,22	313,97	214.296,68
<b>cl</b>	3.482,28	0,00	1.350,20	7.818,07	298.491.466,11	0,20	317,45	218.135,45
<b>renta_ddi</b>	15.798,37	2.181,74	13.246,82	29.595,63	966.710.939,00	0,51	453,06	491.704,70

Para el estadístico media: **la media de la renta de todos los declarantes** es de aproximadamente unos 20.000,00 euros anuales, que equivalen a 1.666 € al mes (suponiendo 12 pagas). El 10% más pobre (cerca de 2 millones de declarantes) cobran menos de 5.531,50 euros al año, que equivalen a aproximadamente 461 euros al mes, una cantidad bastante reducida si se compara con la media muestral: 1.833,12 € por mes. Es decir, el 27,66 % de lo que cobran en media los declarantes equivale a lo que cobra el 10% más pobre.

Analizando la renta antes y después de impuestos, en media pasamos de 21.997,44 € y 18514,86€ respectivamente, es decir, que en media se paga 3.482,58 € (cl) a la hacienda pública estatal y regional.

Los percentiles: El 10% más pobre según la *renta del periodo* (“*renta\_periodo*”, esto es antes de impuestos) tiene una renta inferior a los 5.531,5 €, tras la primera fase (variable *renta*), el 10% más pobre según la variable *renta* tiene una renta inferior a los 2.184,64 €. Esto supone una disminución del valor del percentil de más de la mitad: 3.346,86€ causados por los conceptos de gastos deducibles y reducciones. El 10% más rico tiene una renta del periodo superior a los 39.657,32 €, pero cuando esta distribución pasa la primera fase el 10% más rico

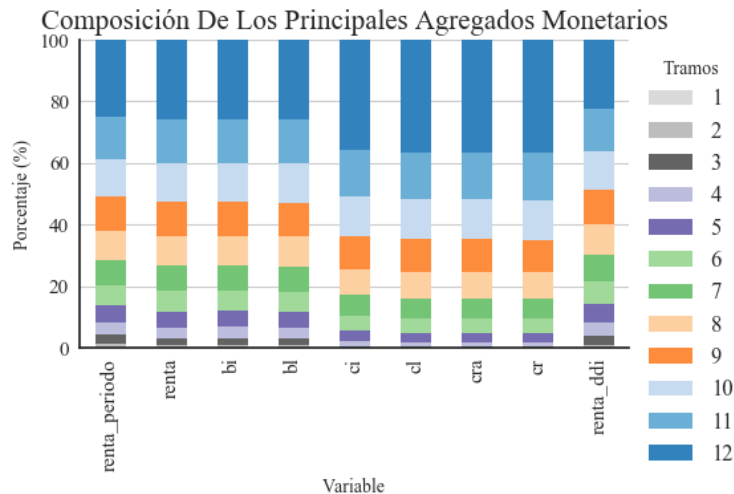
de la distribución nueva tiene una renta superior a los 37.046,41 €, lo que supone una reducción de 2.610,9 €. Esto parece indicar que los gastos deducibles y otras reducciones en el IRPF presentan una mayor incidencia en las rentas inferiores.

### 2.3 Distribución de los tramos por agregado monetario

El lector se puede preguntar ¿cuál es la composición de los principales agregados monetarios si se atiende a los distintos tramos de renta? La respuesta a esta pregunta daría una idea de la importancia que cada tramo tiene en cada magnitud del IRPF.

Para los valores promedio de las magnitudes fiscales mencionadas, se construye una figura que muestra la contribución al mismo de cada uno de los 12 tramos de renta.

Gráfico 2.2: Importancia de cada tramo por agregado monetario



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

Así, el Gráfico 3, informa que el tramo más importante en términos monetarios es, lógicamente, el tramo 12: su importancia es fundamental, especialmente, si se atiende a la cuota líquida, aunque presenta una baja frecuencia (poco denso). Si se compara la primera variable, renta del periodo (*renta\_periодо*) con la cuota líquida (*cl*), se ve que la composición ha cambiado notablemente: después de impuestos hay una mayor presencia de los tramos 8 a 12, y menor por tanto del resto. Esto indica una mejor distribución de la renta tras impuestos.

También resulta interesante notar que los tramos bajos de renta apenas son visibles en todas las variables, pero especialmente en las variables que representan cuotas, es decir, los pagos impositivos representan una parte muy pequeña entre los tramos 1 y 4 respecto del total de pagos.

## 2.4 Un análisis complementario de desigualdad.

Para medir la desigualdad, el indicador por excelencia es el coeficiente de **Gini** (Gini, 1936), que indica el nivel de desigualdad de una distribución en este caso para cada agregado monetario, dónde 0 indica máxima igualdad y 1 máxima desigualdad, pero no nos dice nada acerca de la distribución.

Para calcular la desigualdad, otro método usado es el coeficiente de **Theil** (Theil, 1989) que está basado en el concepto informativo de entropía<sup>9</sup>.

Y por último, se puede usar el coeficiente de **Atkinson** (Atkinson, 1970). Este índice fija de forma implícita un parámetro ( $\epsilon$ ) que indica el grado de aversión a la desigualdad y que puede utilizarse como medida normativa:  $\epsilon = 0$  implica que no hay aversión alguna, mientras que  $\epsilon = \infty$  implica una aversión infinita ante la desigualdad. En nuestro caso, se utiliza el parámetro  $\epsilon = 0,5$  que es el que toma por defecto *DASP* (Abdelkrim & Duclos, 2007). El coeficiente de Atkinson varía entre 0 y 1, y sirve para medir la ganancia que podría obtenerse con una redistribución determinada de la renta (cuanto más cercano a 1, peor será la distribución de la renta obtenida).

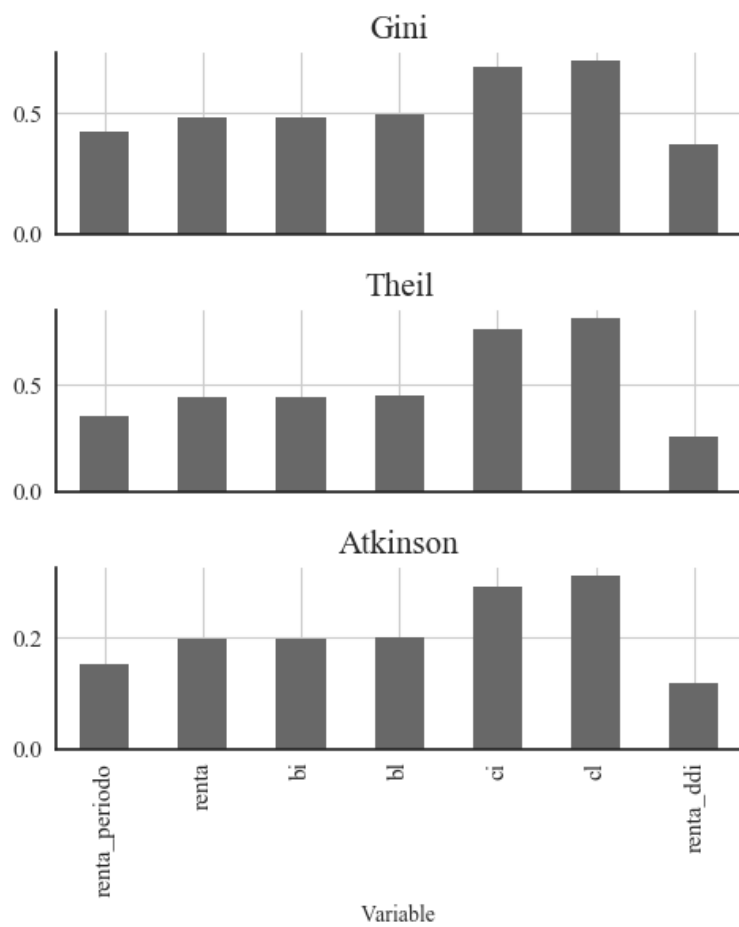
Dichos estadísticos se resumen en la siguiente tabla:

*Tabla 2.3: Indicadores de desigualdad*

Variable	Gini	Theil	Atkinson
renta_periodo	0,4215	0,3511	0,1519
renta	0,4835	0,4441	0,1985
bi	0,4825	0,4430	0,1979
bl	0,4948	0,4536	0,2013
ci	0,6926	0,7575	0,2908
cl	0,7174	0,8109	0,3114
renta_ddi	0,3693	0,2570	0,1180

<sup>9</sup> Entropía: Concepto adoptado de la física termodinámica y telecomunicaciones, Enri Theil utilizó este concepto para medir el grado de desigualdad. Se trata de una medida de desorden que puede ser usada como una medida del cambio necesario para eliminar dicho desorden o en nuestro caso la desigualdad.

Gráfico 2.3: Indicadores de desigualdad



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

Para el conjunto de la economía española, según estos indicadores, el IRPF se comporta reduciendo la desigualdad (el resultado de comparar la renta del periodo o la renta antes de impuestos con la renta después de impuestos).

Se puede ver cómo va evolucionando la desigualdad de la renta en cada estado dentro del IRPF. Así, dada la renta del periodo de los declarantes consignada en el IRPF se transforma en renta gravada. Dependiendo del origen, una es tratada como base imponible general, y otra como base imponible del ahorro. En este análisis no se distinguen ambos canales y se tratan en todo momento de forma agregada cómo base imponible (general y del ahorro) y base liquidable (general y del ahorro).

Por un lado la renta antes de impuestos medida por la variable “*renta\_periodo*”, presenta una distribución menos desigual que la variable renta, cuyo efecto se podría atribuir a la diferencia entre ambas que consiste en reducciones y gastos deducibles. Estas minoraciones hacen que la distribución asociada a la renta que computa el IRPF sea más desigual.

La renta se transforma en base imponible (magnitud en la que se agrega tanto la base imponible general como la del ahorro) mediante la aplicación de imputaciones de renta, o la compensación del saldo negativo de ganancias o pérdidas patrimoniales de años anteriores. Esta fase de cálculo no se traduce en un cambio reseñable en la desigualdad, de hecho, se reduce ligeramente.

El siguiente paso es la base liquidable: esta fase se puede asociar a la aplicación de una serie de reducciones (tributación conjunta, aportaciones a sistema de previsión social, pensiones compensatorias, etc.). En este caso, sí existe un cambio reseñable en la distribución ya que la hace más desigual, lo que confirma lo que antes presumíamos: que los ahorros fiscales, fundamentalmente por las reducciones por planes de pensiones son claramente regresivos.

El efecto de la tarifa se puede ver restando a la renta del período la cuota íntegra, la distribución asociada tiene un coeficiente de Gini de 0,3691, lo que supone una mejora de la redistribución notable debido a los tipos impositivos crecientes conforme aumenta la base liquidable declarada.

Por último encontramos con la renta después de impuestos que es el resultado de restar a la renta del período la cuota líquida del IRPF, para el caso en que se resta la cuota líquida el coeficiente de Gini toma el valor de 0,3693. Por tanto, la comparación entre el anterior indicador y el del presente párrafo supone un empeoramiento de la distribución, lo que lleva a pensar que las deducciones en España son regresivas.



## 2.5 Progresividad: ¿cómo evoluciona la equidad vertical en el IRPF?

A continuación se analiza un aspecto importante de los impuestos: la progresividad. Esto es importante porque es una característica deseable y necesaria para que un impuesto sea justo y contribuya a reducir la desigualdad. Un impuesto es progresivo si al aumentar la renta, aumenta la cuota impositiva. Con esta definición, si se analiza la media de los principales agregados monetarios para el conjunto de España poniendo el foco en las decilas, esto proporciona unas pistas sobre la progresividad del impuesto de la renta en cada una de las fases del impuesto y de forma agregada.

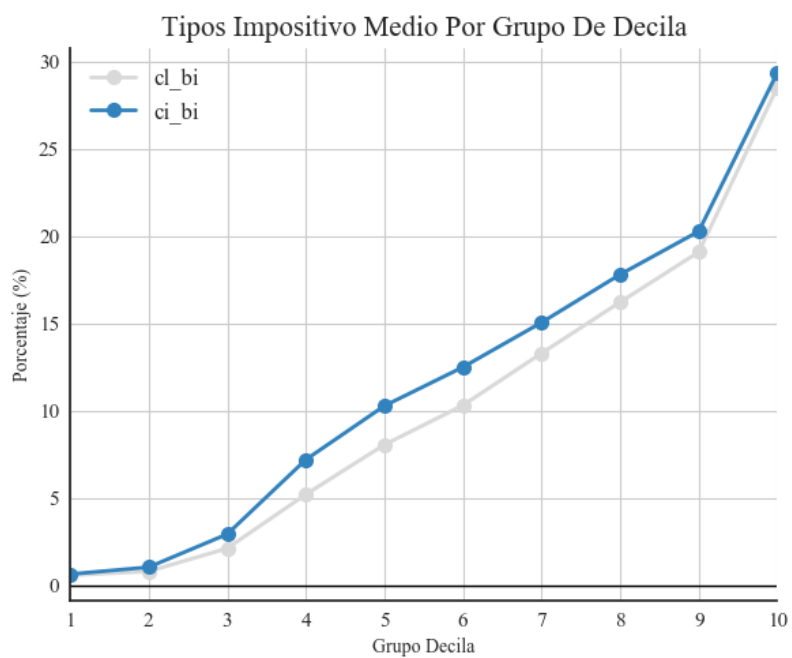
*Tabla 2.4: Cuantiles de los agregados monetarios*

Percentiles	renta_periodo	renta	bi	bl	ci	cl	renta_ddi
10%	5.531,50	2.184,64	2.226,44	1.698,70	0,00	0,00	2.181,74
25%	10.298,66	6.926,37	6.968,07	6.241,13	0,00	0,00	6.881,79
50%	17.218,48	14.589,70	14.648,14	13.715,85	1.692,56	1.350,20	13.246,82
75%	27.382,58	24.746,77	24.777,01	23.790,50	4.460,17	4.091,76	20.802,40
90%	39.657,32	37.046,41	37.056,86	35.813,96	8.217,39	7.818,07	29.595,63

Los cuantiles de los principales agregados monetarios, nos da algunas pistas de cómo se comporta el impuesto. Así se ve que la renta del periodo (gris más claro) es la variable que está por encima de la renta después de impuestos (azul). Esto implica que después de impuestos lógicamente los individuos en media tienen menos renta, pero la pendiente de la curva “renta\_ddi” tiene una menor pendiente en los últimos percentiles, es decir la progresividad es acusada para los declarantes de más renta.

Observando las líneas del Gráfico 2.4, el lector se puede ver la progresividad del impuesto: un impuesto es progresivo cuando aumenta el porcentaje pagado a medida que aumenta la renta (o lo que es lo mismo, cuando el denominado tipo medio del impuesto ( $ci/bi$  o  $cl/bi$ ) crece conforme lo hace su base imponible).

Gráfico 2.4: Tipo impositivo medio por decila



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

### 2.5.1 Mas sobre progresividad, Kakwani

El índice de Kakwani es un indicador de progresividad que se calcula a partir de una renta y una cuota la progresividad de dicho impuesto. Matemáticamente coincide con la diferencia entre la concentración de la cuota impositiva y el índice de Gini de la renta antes de impuestos (Jenkins, 1988).

Tabla 2.5: Índice de Kakawani

Cuota\Renta	bi	bl	renta_periodo	renta_ddi
ci	0,2101	0,1978	0,2711	0,3232
cl	0,2349	0,2226	0,2959	0,3480

La anterior Tabla 2.5, se ha calculado utilizando cómo variables de renta: las que aparecen nombradas en columnas y de impuestos: las que aparecen en las filas.

Así pues, el índice de Kakwani para la base imponible (bi) y la cuota íntegra (ci) es de 0,21 que es positivo y por tanto el efecto que tiene es de progresividad. En general, puede observarse que para todos los casos índice de Kakwani es positivo y además aumenta a medida que se pasa por las distintas rentas, y cuotas.

Los resultados relevantes aquí son  $Kakwani(ci, bi) = 0,21$  que hace referencia al efecto de las reducciones y  $Kakwani(cl, renta\_ddi) = 0,28$  que es el efecto total que tiene el IRPF. Así pues, se puede de afirmar que el IRPF es un impuesto globalmente progresivo.

La variable *renta\_ddi*, es sin lugar a dudas la variable que presenta más progresiva de las distribuciones, reforzando la idea de que el IRPF tiene un efecto final positivo sobre este indicador.

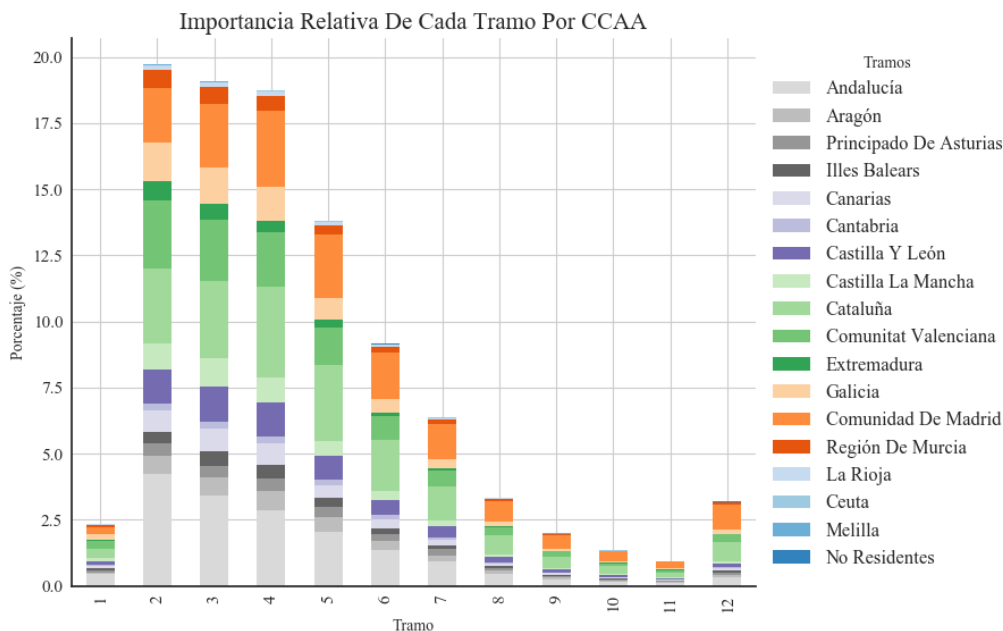
### 3 EL COMPORTAMIENTO DEL IRPF DESDE EL ÁMBITO REGIONAL

En este apartado, añadimos, como señalábamos antes, un enfoque espacial al análisis anterior. El lector encontrará la misma estructura que la desarrollada en el anterior apartado pero ahora desagregando por comunidad autónoma.

#### 3.1 La distribución de la renta

Recordará el lector la, en la que se presentaba la distribución por tramos de renta, ahora nos centramos en las comunidades autónomas. En el Gráfico 3.1 se representa la misma variable distinguiendo en cada tramo la proporción del tramo que pertenece a cada comunidad. Así pues todos los tramos presentan tantos colores como comunidades autónomas haya en la muestra de declarantes (se ha de recordar que País Vasco y Navarra no están incluidas en esta muestra, al tener modelos de declaración propios derivados de sus competencias particulares por ser territorios forales).

Gráfico 3.1: Distribución de la renta por CCAA



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

### 3.1.1 El camino desde la renta declarada hasta la cuota líquida

A continuación se analizan los principales agregados monetarios por comunidades autónomas, pasando nuevamente desde la renta declarada en el periodo impositivo 2012 hasta la renta después de impuestos (renta declarada – cuota líquida). En el anexo se han incluido las estadísticas descriptivas por comunidad autónoma: dado que sólo describe cantidades pero no aporta mucho sobre el análisis de impuesto se ha separado esta parte del texto.

Tabla 3.1: Media de los principales agregados por comunidad autónoma

	renta_periodo	renta	bi	bl	ci	cl	renta_ddi
ANDALUCÍA	18.918,20	16.119,53	16.148,07	15.098,08	2.768,59	2.563,71	13.555,82
ARAGÓN	21.462,19	18.710,76	18.767,01	17.845,75	3.457,89	3.230,97	15.479,79
PRINCIPADO DE ASTURIAS	21.876,42	19.129,45	19.212,26	18.209,78	3.567,86	3.354,91	15.774,55
ILLES BALEARS	21.643,51	18.954,72	18.982,97	18.217,38	3.691,61	3.413,53	15.541,19
CANARIAS	19.954,75	17.178,20	17.215,29	16.372,00	3.142,13	2.840,24	14.337,96
CANTABRIA	21.592,38	18.890,39	18.937,55	17.960,47	3.491,58	3.230,30	15.660,09
CASTILLA Y LEÓN	20.257,58	17.550,56	17.594,05	16.628,51	3.086,01	2.835,97	14.714,59
CASTILLA LA MANCHA	18.751,10	15.975,34	16.023,52	14.925,37	2.623,68	2.408,79	13.566,55
CATALUÑA	24.575,90	21.886,97	21.914,05	21.050,66	4.538,30	4.253,65	17.633,32
COMUNITAT VALENCIANA	19.792,22	17.005,40	17.032,55	16.145,11	3.067,51	2.844,28	14.161,12
EXTREMADURA	16.831,66	13.957,23	14.021,94	12.993,27	2.207,26	2.056,40	11.900,83
GALICIA	19.400,99	16.628,10	16.668,68	15.815,38	2.951,32	2.787,69	13.840,40
COMUNIDAD DE MADRID	28.534,71	25.845,60	25.871,72	24.879,36	5.915,53	5.567,73	20.277,87
REGIÓN DE MURCIA	19.292,51	16.465,24	16.487,50	15.525,37	2.787,40	2.574,47	13.890,77
LA RIOJA	20.639,85	17.970,75	18.031,57	17.147,24	3.225,78	2.963,03	15.007,72
CEUTA	27.614,44	24.950,61	24.979,64	23.799,46	5.095,75	2.411,14	22.539,47
MELILLA	26.344,91	23.579,04	23.635,21	22.532,61	4.680,39	2.192,56	21.386,48
NO RESIDENTES	20.783,07	18.820,02	18.831,71	17.836,81	4.106,63	3.959,08	14.860,94

En la Tabla 3.1, se presenta la media de cada agregado monetario por comunidad autónoma (cuyas unidades son euros). Para facilitar la interpretación de los datos, se ha añadido la Tabla 3.2, que representa el ranking de cada variable, lo que permite hacer la comparación entre comunidades autónomas de forma rápida.

Tabla 3.2: Ranking de la media de los principales agregados por comunidades

	renta_periodo	renta	bi	bl	ci	cl	renta_ddi
CA							
ANDALUCÍA	16	16	16	16	16	11	17
ARAGÓN	8	9	9	8	9	7	8
PRINCIPADO DE ASTURIAS	5	5	5	6	7	15	5
ILLES BALEARS	6	6	6	5	6	2	7
CANARIAS	12	12	12	12	11	14	12
CANTABRIA	7	7	7	7	8	16	6
CASTILLA Y LEÓN	11	11	11	11	12	10	11
CASTILLA LA MANCHA	17	17	17	17	17	13	16
CATALUÑA	4	4	4	4	4	3	4
COMUNITAT VALENCIANA	13	13	13	13	13	9	13
EXTREMADURA	18	18	18	18	18	6	18
GALICIA	14	14	14	14	14	4	15
COMUNIDAD DE MADRID	1	1	1	1	1	5	3
REGIÓN DE MURCIA	15	15	15	15	15	12	14
LA RIOJA	10	10	10	10	10	8	9
CEUTA	2	2	2	2	2	18	1
MELILLA	3	3	3	3	3	17	2
NO RESIDENTES	9	8	8	9	5	1	10

Lo primero que se puede extraer de la Tabla 3.1: Son las comunidades que presenta una mayor media en la renta del periodo, en primer lugar, se encuentra Madrid, le siguen Ceuta, Melilla y Cataluña. En último lugar se encuentra Extremadura, le preceden Castilla la Mancha, Andalucía y Murcia.

Si se analiza la renta después de impuestos, se comprueba cierto efecto de reordenación debido a factores como la distribución de la renta de la población y las diferencias normativas que introduce cada comunidad autónoma (por ejemplo, cada comunidad autónoma puede legislar sobre los tipos impositivos que aplica en la parte autonómica de la declaración, o puede incluir deducciones propias sobre los contribuyentes residentes en la región).

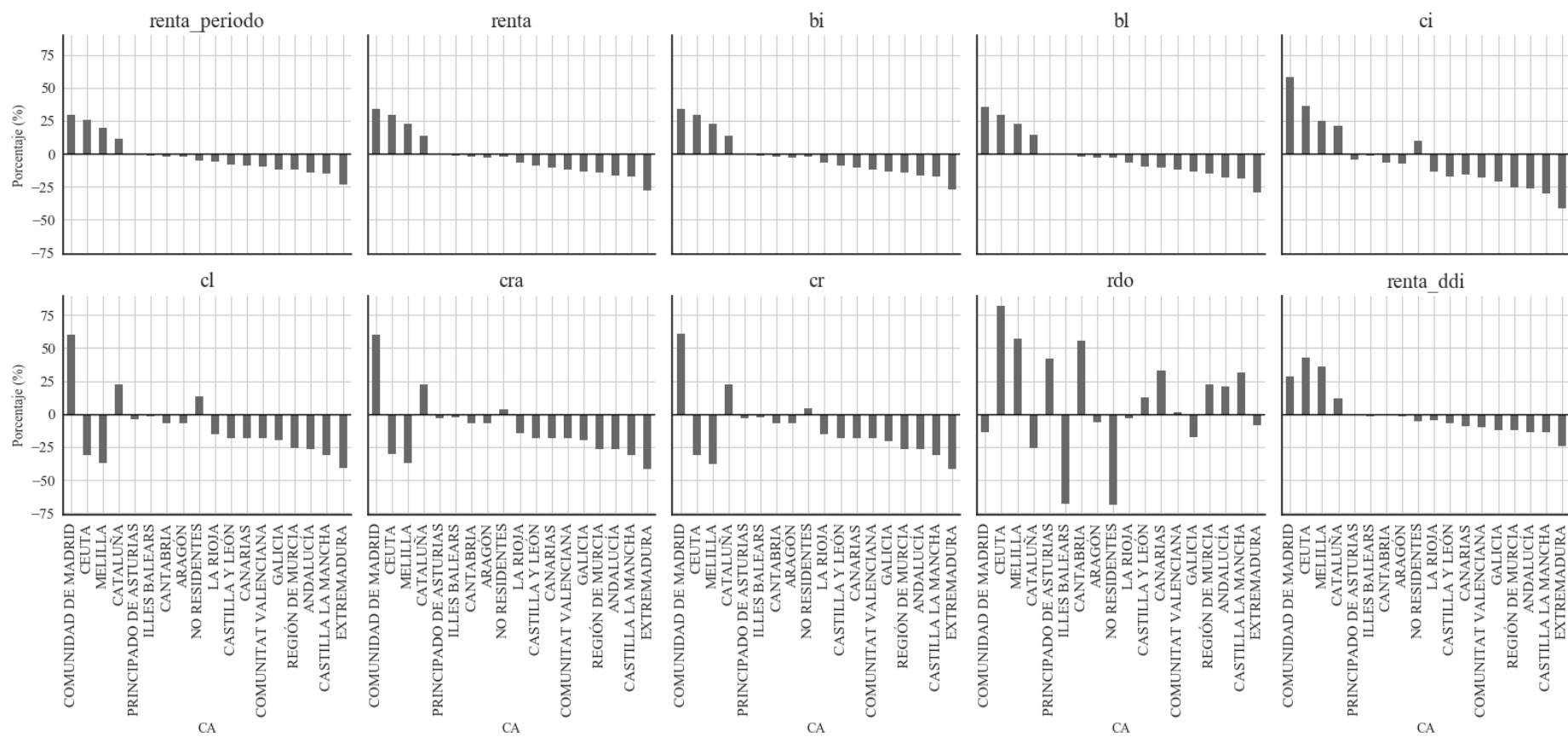
En gráfico 3.2, se representa la diferencia relativa de cada variable respecto de la media española. Las comunidades autónomas están ordenadas según su renta del periodo. Si las

cantidades representadas son positivas, implica que la variable en cuestión es un X% superior a la media de toda España; si es negativa, un X% inferior.

Así, puede observarse cómo va difiriendo el orden en cada fase del IRPF. Básicamente el orden se mantiene hasta que entra en juego la escala de gravamen: esto es, al llegar a la cuota íntegra (ci), las diferencias de progresividad en las escalas a nivel autonómico provoca que algunas comunidades autónomas tengan mayores cuotas que otras comunidades. En la cuota líquida (cl), se producen más cambios que principalmente afectan a Ceuta y Melilla, cuyo origen son deducciones estatales (de forma particular, el art. 68.4 establece que los residentes en Ceuta y Melilla se deducirán el 50% de las cuotas íntegras que correspondan a rentas obtenidas en esas ciudades autónomas) y autonómicas. Estos cambios son los que se provocan finalmente la reordenación de la distribución de la renta después de impuestos reflejada en la última columna

Gráfico 3.2: Diferencias regionales respecto de la media para España

Diferencia Relativa De Cada Agregado Respecto De La Media De España



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012



### 3.2 Hablando de progresividad, los tipos medios.

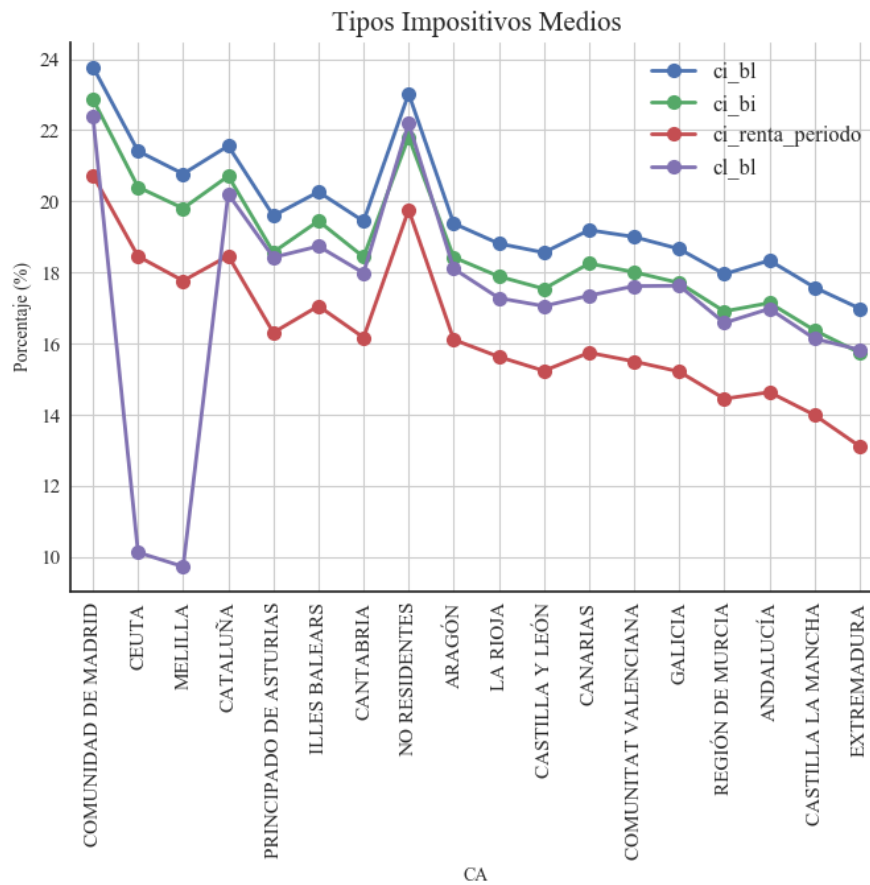
Como ya se ha señalado, los tipos impositivos medios informan sobre la progresividad proporcionada por cada fase del IRPF posterior a la aplicación de las tarifas, representándose en el **¡Error! No se encuentra el origen de la referencia.**. La evolución de estos tipos medios se atribuyen, por lo tanto, a distintas deducciones (estatales y autonómicas), tipos de gravamen (contenidos en la tarifa estatal y autonómica), mínimos personales y familiares (estatales y autonómicos), etc.,

Por lo tanto, en esta evolución de los tipos medios regionales no solo influye la desigual distribución de las rentas en el territorio nacional, sino también la acción normativa conjunta del sector público central y el autonómico.

A continuación, se describen dicho tipos impositivos medios.

- La variable “ci\_bl”, resulta del cociente entre la cuota íntegra y la base liquidable: indica, por lo tanto, el efecto conjunto de la doble escala de gravamen a la que se somete cualquier renta en España en territorio común (estatal y autonómica).
- *La variable ‘cl\_bl’*, resulta del cociente entre la cuota líquida y la base liquidable: Indica el efecto neto de las deducciones (autonómicas y estatales), al compararlo con la serie anterior.
- La variable “ci\_bi”, resulta del cociente entre la cuota íntegra y la base imponible: añade al anterior el efecto de las reducciones aplicadas en la base imponible.
- La variable “ci\_renta\_periódico”: añade a lo anterior el efecto de la reducción por rendimientos del trabajo y mínimo personal y familiar.

Gráfico 3.3: Tipos impositivos medios por CC.AA



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

El Gráfico 3.3 muestra los tipos impositivos medios, se han ordenado las comunidades desde las más ricas a la izquierda a las más pobres a la derecha. Se puede apreciar bastantes diferencias entre comunidades autónomas, pero, lógicamente, se observa el perfil creciente esperable en un impuesto como el IRPF que se comporta, globalmente, de forma progresiva: los tipos mayores corresponden a las CCAA más ricas, por lo que, fundamentalmente, podemos concluir que es el efecto de la tarifa es el que acaba determinando la progresividad del IRPF (aunque algunas reducciones, y deducciones mitiguen esa progresividad).

### 3.3 Medidas de desigualdad

En este apartado se analiza la desigualdad por CC.AA. haciendo uso de los indicadores usados anteriormente. En la *Tabla 3.3*, se presenta un resumen de indicadores y entre paréntesis, se menciona la o las variables usadas para su cálculo y el hipotético de cada variable, ordenando de forma decreciente cada valor del índice correspondiente.

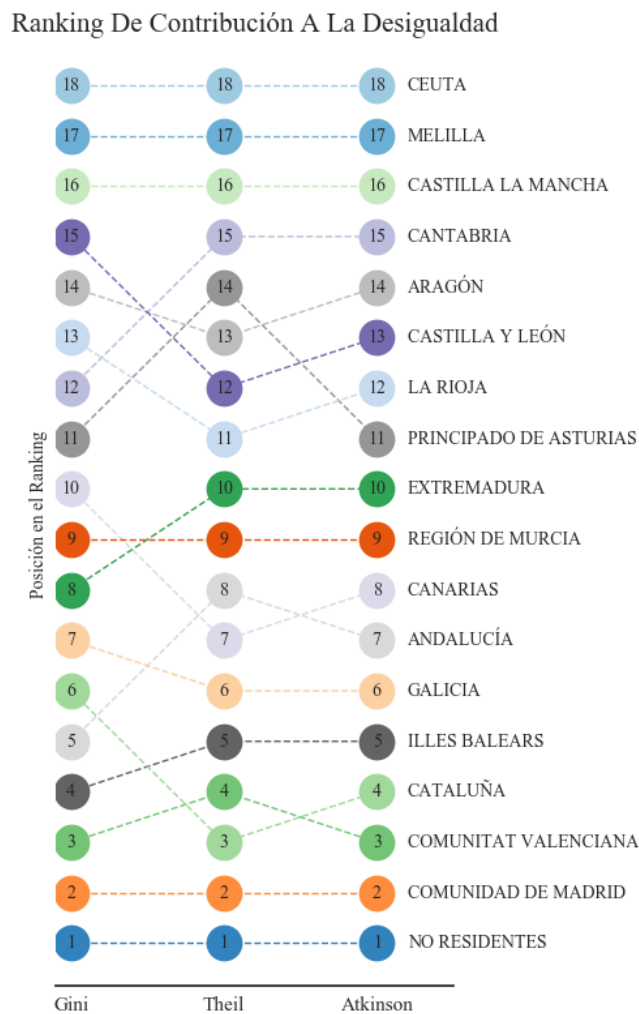
*Tabla 3.3: Resumen de indicadores de desigualdad y progresividad*

	Gini	Ran k	Concen- tración	Ran k	RS	Ran k	Theil	Ran k	Atkinson	Ran k	Ka- kwani	Ran k
	(renta_p erido)	Gini	(renta_dd i)	Con- cen- tración	(renta_p erido, renta_dd i)	RS	(renta_p erido)	Thei l	(renta_p erido)	At- kin- son	(ci, renta_p erido)	Ka- kwa ni
ANDALUCÍA	0,4144	14	0,4472	16	-0,0328	2	0,2978	11	0,1413	12	0,2972	15
ARAGÓN	0,3911	5	0,4080	3	-0,0169	12	0,2716	6	0,1297	5	0,2597	8
PRINCIPADO DE ASTU- RIAS	0,3953	8	0,4100	7	-0,0147	13	0,2681	5	0,1323	8	0,2445	5
ILLES BALEARS	0,4156	15	0,4279	10	-0,0123	15	0,3320	14	0,1439	14	0,2681	10
CANARIAS	0,4075	9	0,4334	12	-0,0258	5	0,3024	12	0,1387	11	0,2837	12
CANTABRIA	0,3943	7	0,4084	4	-0,0141	14	0,2679	4	0,1295	4	0,2570	7
CASTILLA Y LEÓN	0,3909	4	0,4100	6	-0,0191	9	0,2742	7	0,1297	6	0,2707	11
CASTILLA LA MANCHA	0,3907	3	0,4178	9	-0,0271	4	0,2608	3	0,1273	3	0,3013	16
CATALUÑA	0,4140	13	0,4160	8	-0,0019	16	0,3541	16	0,1485	15	0,2501	6
COMUNITAT VALEN- CIANA	0,4206	16	0,4464	15	-0,0258	6	0,3416	15	0,1505	16	0,2902	14
EXTREMADURA	0,4080	11	0,4552	17	-0,0472	1	0,2835	9	0,1365	9	0,3265	18
GALICIA	0,4115	12	0,4370	14	-0,0256	7	0,3213	13	0,1421	13	0,2865	13
COMUNIDAD DE MA- DRID	0,4400	17	0,4310	11	0,0090	18	0,4370	17	0,1717	17	0,2364	3
REGIÓN DE MURCIA	0,4076	10	0,4360	13	-0,0285	3	0,2912	10	0,1385	10	0,3056	17
LA RIOJA	0,3919	6	0,4095	5	-0,0176	10	0,2778	8	0,1307	7	0,2653	9
CEUTA	0,3623	1	0,3799	1	-0,0176	11	0,2514	1	0,1168	1	0,2277	2
MELILLA	0,3757	2	0,3994	2	-0,0237	8	0,2563	2	0,1229	2	0,2410	4
NO RESIDENTES	0,5881	18	0,5845	18	0,0037	17	0,6396	18	0,2894	18	0,1952	1
ESPAÑA	0,4215		0,435		-0,0135		0,3511		0,1519		0,2711	

A continuación, se muestra también en el Gráfico 3.4 un ranking de desigualdad regional atendiendo a diversas mediciones. Las comunidades autónomas más desiguales son: Murcia y Extremadura. Por el contrario, Comunidad de Madrid, Comunidad Valenciana. Las más igualitarias-son: Ceuta, Melilla, Castilla y León y Cantabria.

Se debe reseñar que algunos indicadores no ofrecen la misma ordenación en la clasificación de desigualdad de las comunidades autónomas. Por ejemplo, la Comunidad Valenciana está en el puesto 4 según el coeficiente de Atkinson, el 3 según Theil y el 6 según Gini.

Gráfico 3.4: Ranking de contribución a la desigualdad



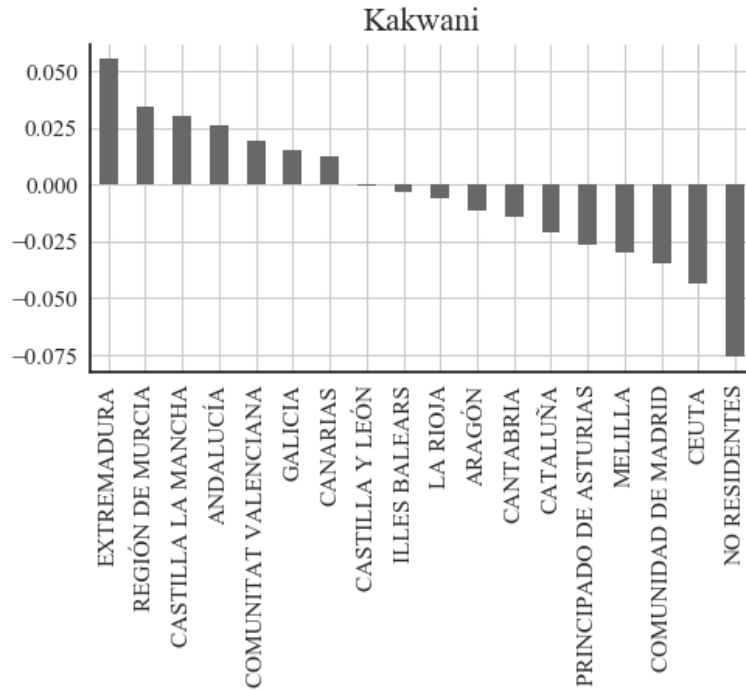
Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

### 3.4 Una medida de progresividad adicional, Kakwani

A continuación en el Gráfico, se informa, atendiendo al índice de Kakwani (se ha usado para este indicador las variables “ci” y “renta\_periodo”), acerca de qué comunidades contribuyen de forma más intensa a la progresividad y cuáles menos.

El lector podrá observar en dicho gráfico 3.5 que Extremadura, Murcia y Castilla la Mancha son las que contribuyen más a la progresividad global del IRPF. En cambio, Ceuta y Madrid y Melilla son las que lo hacen menos.

Gráfico 3.5: Desviación regional del índice de Kakwani



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

## 4 CONCLUSIONES, LIMITACIONES Y EXTENSIONES

A lo largo de este trabajo se ha intentado ir analizando, con distintas medidas y enfoques, los aspectos distributivos más relevantes del actual IRPF. Por una parte, se ha intentado poner de relieve la importancia relativa de cada tramo de renta dentro de la población de declarantes. Así, se ve que los tramos centrales de renta (como era esperable) son los más densos (dicho de otra forma: las clases medias son los tramos modales en España).

Luego, se pone en perspectiva los estadísticos descriptivos de los principales agregados monetarios del impuesto, describiendo a grandes rasgos cómo es la distribución de cada uno de ellos y cómo evolucionan en cada fase de la liquidación del IRPF. Se ha visto que en cada fase los agregados monetarios cambian, y por lo tanto también su distribución.

Para poder evaluar si la distribución mejora en términos de equidad se han utilizado los índices de desigualdad habituales de Gini, Theil y Atkinson, y se puede observar, como cabía esperar, que el efecto final según los mismos es un aumento de la igualdad en las rentas de los ciudadanos españoles.

También se ha analizado la progresividad del impuesto, mirando los tipos impositivos medios y el índice de Kakwani, mostrando de esta forma la intensidad que la citada progresividad presenta en nuestro IRPF actual (lo que permitiría su comparación con otros períodos anteriores y posteriores, si se repitiera este mismo ejercicio con muestras de otros años).

En la segunda parte, se pone el foco en las distintas comunidades autónomas y en cómo modulan su progresividad y reducen por consiguiente la desigualdad existente de forma previa. Se han puesto en contexto las diferencias intrarregionales de los principales agregados, comparando la media de cada comunidad autónoma con la media del total de España, así como los índices de desigualdad y progresividad. Individualmente, cada comunidad autónoma presenta un tratamiento global progresivo sobre los declarantes, aunque la intensidad de la acción redistribuidora es muy desigual entre comunidades.

Durante el desarrollo del presente trabajo, se han encontrado algunas limitaciones como, por ejemplo, la dificultad para encontrar un trabajo que sea reproducible, con la intención de aprender y seguir el mismo hilo que ha seguido el investigador.

Otra dificultad ha sido encontrar trabajos que traten de forma detallada la metodología usada en el análisis distributivo de un impuesto, por ejemplo, a partir de la información de la *Muestra de declarantes del IRPF AEAT-IEF 2012* proporcionada por el Instituto de Estudios Fiscales.

Es por ello que el presente análisis se ha realizado con herramientas *open-source* y todo su contenido está disponible en *GitHub*<sup>10</sup> para así facilitar la reproducibilidad del trabajo y su posible ampliación en el futuro.

Entre las herramientas *open-source* utilizadas, se encuentra el lenguaje de programación Python<sup>11</sup> (apoyándonos en el conocimiento autodidacta conseguido para el uso de esta herramienta), que está muy extendida entre científicos de datos de áreas como Finanzas, Física, Bioquímica, Estadística, y que puede llegar a convertirse en sustitutivo de *Matlab* o *R*.

Por motivos de tiempo y espacio, no se ha podido realizar todo lo que se había proyectado en un principio en el presente trabajo. Entre las posibles extensiones, se podría realizar un análisis del caso específico de Aragón: cuál es el perfil del contribuyente del IRPF en ésta comunidad; cuál es el comportamiento redistributivo de la actual legislación aragonesa y su evolución en el tiempo; cómo se comportan desde el punto de vista distributivo las últimas deducciones introducidas por el Parlamento aragonés; etc..

Otra extensión muy enriquecedora hubiera sido el planteamiento de simulaciones de posibles reformas de elementos centrales del actual IRPF (eliminación de determinadas deducciones; introducción de subsidios salariales para rentas bajas; extensión o reducción de los tramos del impuesto, etc.).

Y, por último, y como se ha mencionado, se podría realizar un análisis intertemporal con otras muestras de declarantes que permitiera preguntarse sobre la evolución de la capacidad redistributiva del IRPF durante los últimos años (que ha presentado cambios legislativos relevantes en casi todos los períodos impositivos).

---

<sup>10</sup> Es una plataforma de desarrollo colaborativo, para alojar proyectos usando control de versiones. <http://www.github.com>

<sup>11</sup> <https://www.python.org>

## 5 ANEXOS

### 5.1 Representación resumida “minimal example” del código implementado en IneqPy:

- Ejemplo de código usado para calcular las curvas de concentración, a partir de las cuales se determinan los índices de Gini, Kakwani y Reynolds-Smolensky:

```
def concentration(data=None, income=None, weights=None, sort=True):
    """This function calculate the concentration index, according to the
    notation used in [Jenkins1988]_ you can calculate the:
     $C_x = 2 / x \cdot \text{cov}(x, F_x)$ 
    if  $x = g(x)$  then  $C_x$  becomes  $C_y$ 
    when there are taxes:

     $y = g(x) = x - t(x)$ 

    Parameters
    -----
    income : array-like
    weights : array-like
    data : pandas.DataFrame
    sort : bool

    Returns
    -----
    concentration : array-like

    References
    -----
    Jenkins, S. (1988). Calculating income distribution indices
    from micro-data. National Tax Journal. http://doi.org/10.2307/41788716
    """
    # TODO complete docstring

    # check if DataFrame is passed, if yes then extract variables else make a copy
    if data is not None:
        data = data.copy()
        income = data[income].values
        if weights is not None:
            weights = data[weights].values
    else:
        income = income.copy()
        weights = weights.copy() if weights is not None else np.ones(len(income))

    # if sort is true then sort the variables.
    if sort:
        idx_sort = np.argsort(income)
        income = income[idx_sort]
        weights = weights[idx_sort]
    # main calc
    f_x = weights / weights.sum()
    F_x = f_x.cumsum()
    mu = np.sum(income * f_x)
    cov = np.cov(income, F_x, rowvar=False, aweights=f_x)[0,1]
    return 2 * cov / mu
```



- Código implementado para el cálculo del índice de Atkinson agregado:

```
def atkinson(data=None, income=None, weights=None, e=0.5):
    """More precisely labelled a family of income inequality measures, the
    theoretical range of Atkinson values is 0 to 1, with 0 being a state of
    equal distribution, with 1 being a state of equal distribution.
    An intuitive interpretation of this index is possible: Atkinson values can
    be used to calculate the proportion of total income that would be required
    to achieve an equal level of social welfare as at present if incomes were
    perfectly distributed.

    For example, an Atkinson index value of 0.20 suggests
    that we could achieve the same level of social welfare with only
    1 - 0.20 = 80% of income.

    Parameters
    -----
    income : array or str
        If `data` is none `income` must be an 1D-array, when `data` is a
        pd.DataFrame, you must pass the name of income variable as string.
    weights : array or str, optional
        If `data` is none `weights` must be an 1D-array, when `data` is a
        pd.DataFrame, you must pass the name of weights variable as string.
    e : int, optional
        Epsilon parameter interpreted by atkinson index as inequality aversion,
        must be between 0 and 1.
    data : pd.DataFrame, optional
        data is a pd.DataFrame that contains the variables.

    Returns
    -----
    atkinson : float

    Reference
    -----
    Atkinson index. (2017, March 12). In Wikipedia, The Free Encyclopedia.
    Retrieved 14:35, May 15, 2017, from
    https://en.wikipedia.org/w/index.php?title=Atkinson_index&oldid=769991852
    """
    if (income is None) and (data is None):
        raise ValueError('Must pass at least one of both `income` or `df`')

    if data is not None:
        data = data.copy()
        income = data[income].values
        weights = data[weights].values
    else:
        income = income.copy()
        weights = weights.copy() if weights is not None else None

    # not-null condition
    if np.any(income <= 0):
        mask = income > 0
        income = income[mask]
        if weights is not None:
            weights = weights[mask]

    # not-empty condition
    if len(income) == 0:
        return 0

    N = len(income) # observations

    # not-empty wights
    if weights is None:
        weights = np.ones(N)

    # auxiliar variables: mean and distribution
```

```
mu = mean(variable=income, weights=weights)
f_i = weights / sum(weights) # density function

# main calc
if e == 1:
    atkinson = 1 - np.power(np.e, np.sum(f_i * np.log(income) - np.log(mu)))
elif (0 <= e) or (e < 1):
    atkinson = 1 - np.power(np.sum(f_i * np.power(income / mu, 1 - e)),
                            1 / (1 - e))
else:
    assert (e < 0) or (e > 1), "Not valid e value, 0 ≤ e ≤ 1"
    atkinson = None
return atkinson
```

- Código implementado para el calculo del índice de Atkinson desagregado:

```
def atkinson_group(data=None, income=None, weights=None, group=None, e=0.5):
    """The Atkinson index (also known as the Atkinson measure or Atkinson
    inequality measure) is a measure of income inequality developed by
    British economist Anthony Barnes Atkinson. The measure is useful in
    determining which end of the distribution contributed most to the observed
    inequality. The index is subgroup decomposable. This means that overall
    inequality in the population can be computed as the sum of the corresponding
    Atkinson indices within each group, and the Atkinson index of the group mean
    incomes.

    Parameters
    -----
    income : str or np.array
        Income variable, you can pass name of variable in `df` or array-like
    weights : str or np.array
        probability or weights, you can pass name of variable in `df` or
        array-like
    groups : str or np.array
        stratum, name of stratum in `df` or array-like
    e : int, optional
        Value of epsilon parameter
    data : pd.DataFrame, optional
        DataFrame that's contains the previous data.

    Returns
    -----
    atkinson_by_group : float

    Reference
    -----
    Atkinson index. (2017, March 12). In Wikipedia, The Free Encyclopedia.
    Retrieved 14:52, May 15, 2017, from
    https://en.wikipedia.org/w/index.php?title=Atkinson_index&oldid=769991852

    TODO
    ----
    - Review function, has different results with stata.
    """
    if weights is None:
        if data is None:
            weights = np.repeat([1], len(income))
        else:
            weights = np.repeat([1], len(data))

    if data is None:
        data = _to_df(income=income, weights=weights, group=group)
        income = 'income'
        weights = 'weights'
        group = 'group'

    N = len(data)

    def a_h(df):
        """
        Funtion alias to calculate atkinson from a DataFrame
        """
        if df is None:
            raise ValueError

        res = atkinson(income=df[income].values,
                       weights=df[weights].values,
                       e=e) * len(df) / N

        return res

    # main calc:
    if data is not None:
```

```
data = data.copy()
atk_by_group = data.groupby(group).apply(a_h)
mu_by_group = data.groupby(group).apply(lambda dw: mean(dw[income],
                                                    dw[weights]))

return atk_by_group.sum() + atkinson(income=mu_by_group.values)
else:
raise NotImplementedError
```

- Código implementado para el cálculo del índice de Kakwani:

```
def kakwani(data=None, tax=None, income_pre_tax=None, weights=None):
    """The Kakwani (1977) index of tax progressivity is defined as twice the
    area between the concentration curves for taxes and pre-tax income,
    or equivalently, the concentration index for  $t(x)$  minus the Gini index for
     $x$ , i.e.


$$K = C(t) - G(x)$$


$$= (2/t) \text{cov}[t(x), F(x)] - (2/x) \text{cov}[x, F(x)].$$


Parameters
-----
data : pandas.DataFrame
    This variable is a DataFrame that contains all data required in
    columns.
tax_variable : array-like or str
    This variable represent tax payment of person, if pass array-like
    then data must be None, else you pass str-name column in `data`.
income_pre_tax : array-like or str
    This variable represent income of person, if pass array-like
    then data must be None, else you pass str-name column in `data`.
weights : array-like or str
    This variable represent weights of each person, if pass array-like
    then data must be None, else you pass str-name column in `data`.

Returns
-----
kakwani : float

References
-----
Jenkins, S. (1988). Calculating income distribution indices from micro-data.
National Tax Journal. http://doi.org/10.2307/41788716
"""
    if weights is None:
        if data is None:
            weights = np.repeat([1], len(tax))
        else:
            weights = np.repeat([1], len(data))

    if data is None:
        data = _to_df(income_pre_tax=income_pre_tax,
                     tax=tax,
                     weights=weights)
        income_pre_tax = 'income_pre_tax'
        tax = 'tax'
        weights = 'weights'

    # main calc
    c_t = concentration(data=data, income=tax, weights=weights, sort=True)
    g_y = concentration(data=data, income=income_pre_tax, weights=weights,
                       sort=True)
    return c_t - g_y
```

## Código usado para

```
def theil(data=None, income=None, weights=None):
    """The Theil index is a statistic primarily used to measure economic
    inequality and other economic phenomena. It is a special case of the
    generalized entropy index. It can be viewed as a measure of redundancy,
    lack of diversity, isolation, segregation, inequality, non-randomness, and
    compressibility. It was proposed by econometrician Henri Theil.

    Parameters
    -----
    data : pandas.DataFrame
        This variable is a DataFrame that contains all data required in it's
        columns.
    income : array-like or str
        This variable represent tax payment of person, if pass array-like
        then data must be None, else you pass str-name column in `data`.
    weights : array-like or str
        This variable represent weights of each person, if pass array-like
        then data must be None, else you pass str-name column in `data`.

    Returns
    -----
    theil : float

    References
    -----
    Theil index. (2016, December 17). In Wikipedia, The Free Encyclopedia.
    Retrieved 14:17, May 15, 2017, from
    https://en.wikipedia.org/w/index.php?title=Theil_index&oldid=755407818
    """

    if weights is None:
        if data is None:
            weights = np.repeat([1], len(income))
        else:
            weights = np.repeat([1], len(data))

    if data is not None:
        data = data.copy()
        income = data[income].values
        weights = data[weights].values
    else:
        income = income.copy()
        weights = weights.copy()

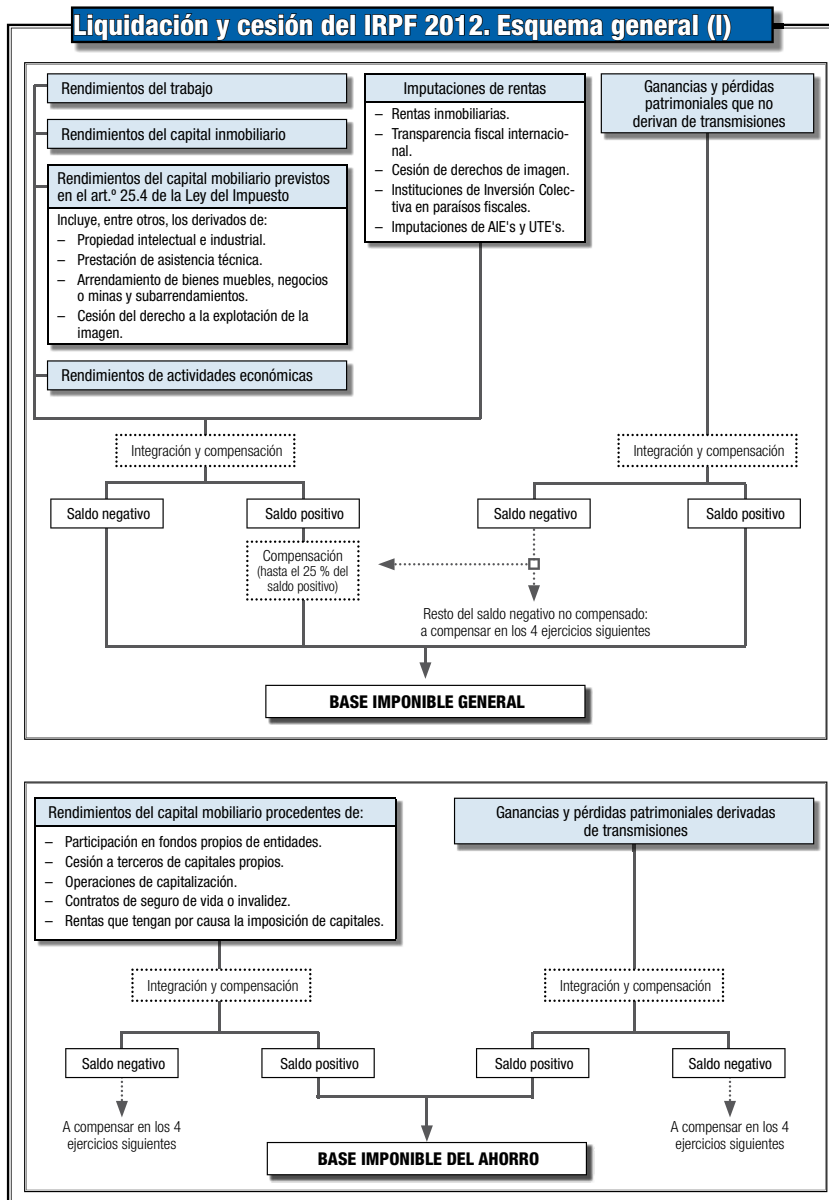
    if np.any(income <= 0):
        mask = income > 0
        income = income[mask]
        weights = weights[mask]

    # variables needed
    mu = mean(variable=income, weights=weights)
    f_i = weights / np.sum(weights)
    # main calc
    theil = np.sum((f_i * income / mu) * np.log(income / mu))
    return theil
```

## 5.2 Esquemas de liquidación general del IRPF (ejercicio 2012)

Figura 1: Esquema de liquidación del IRPF (1)

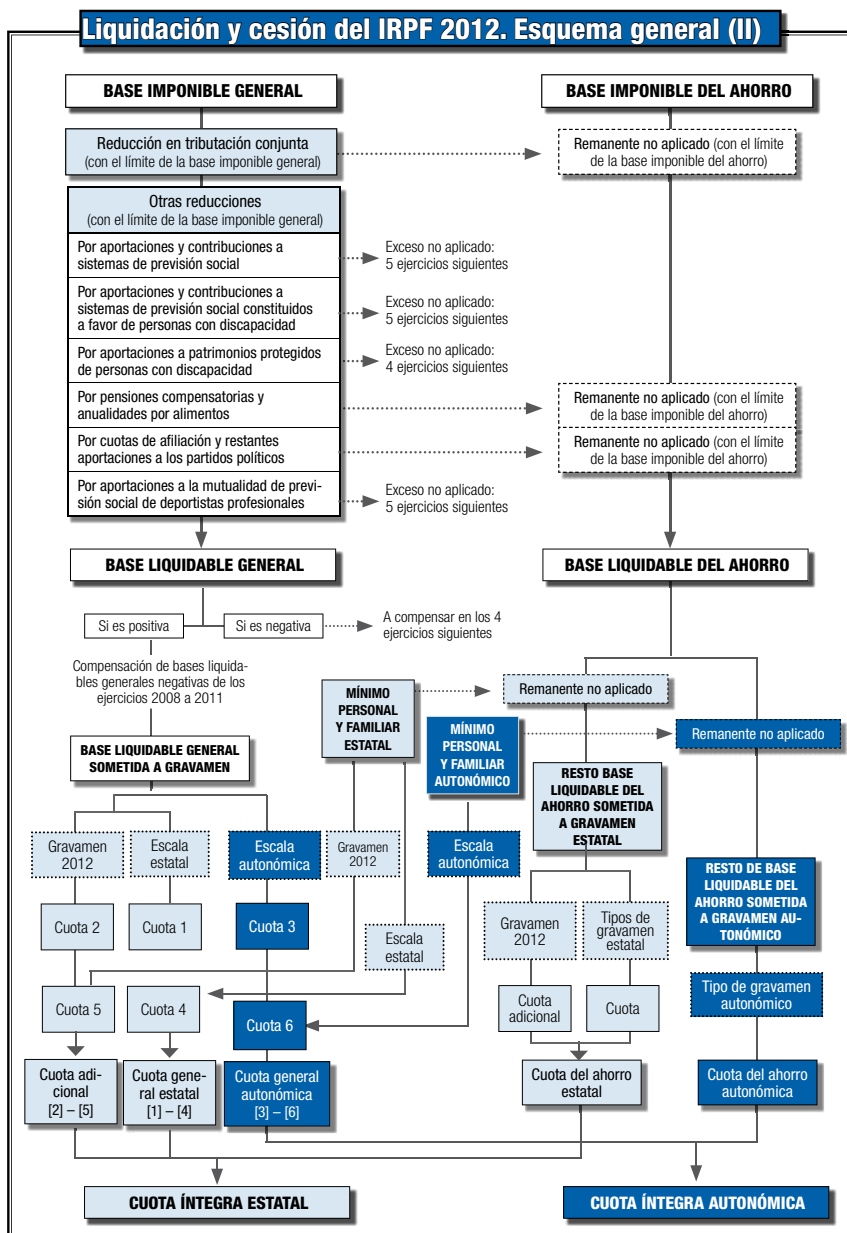
Liquidación del IRPF. Esquema general (I)



Fuente: (Agencia Tributaria, n.d.), pagina 65.

Figura 2: Esquema de liquidación del IRPF (2)

Liquidación del IRPF. Esquema general (II)

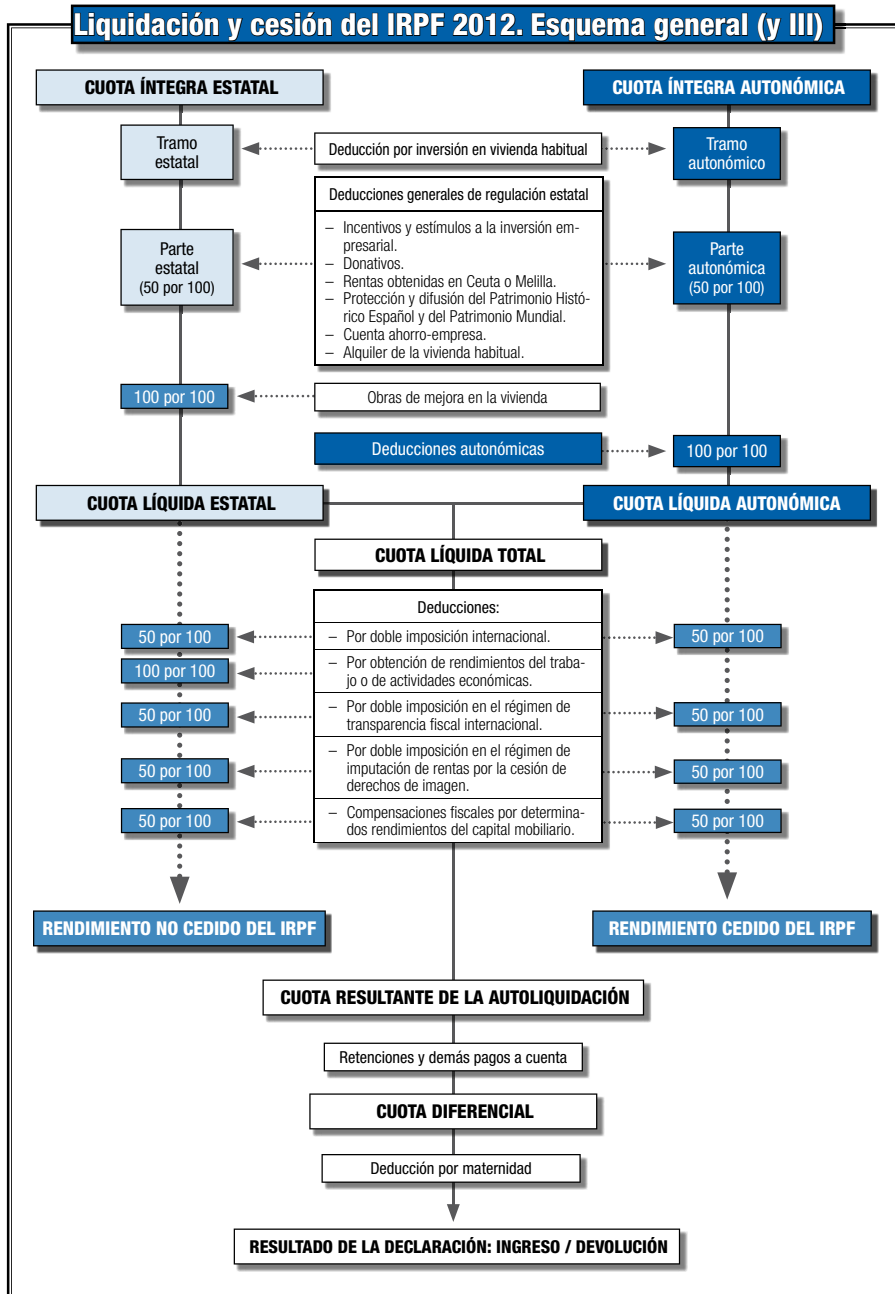


Fuente: (Agencia Tributaria, n.d.), página 66.



Figura 3: Esquema de liquidación del IRPF (3)

Liquidación del IRPF. Esquema general (III)



### 5.3 Notas metodológicas

Para realizar el análisis previamente se han definido las siguientes variables, que se muestran en la siguiente **¡Error! No se encuentra el origen de la referencia.:**

*Tabla 5.1: Definición de los principales agregados monetarios*

AGREGADO MONETARIO	VARIABLE	CONSTRUCCIÓN
Renta del Período	renta_periodo	PAR455 + PAR465 + PAR20 + PAR19 + PAR18 + PAR17 + PAR29
Renta	renta	PAR452 + PAR465
Base Imponible	bi	PAR455 + PAR465
Base Liquidable	bl	PAR620 + PAR630
Cuota Íntegra	ci	PAR698 + PAR699
Cuota Líquida	cl	PAR720+ PAR721
Cuota Resultante de Autoliquidación	cra	PAR741
Cuota Real	cr	PAR741 - PAR756
Resultado	rdo	PAR760
Renta después de impuestos	renta_ddi	renta_periodo - cl

Esta tabla informa que, por ejemplo, cuando se habla de “renta” se ha utilizado para el cálculo la variable “renta” construida cómo la suma de las variables PAR452 + PAR465 presente en los datos de la Muestra de declarantes del IRPF AEAT-IEF 2012, y de forma equivalente, el resto.

Siguiendo la metodología usada en (Lopez, Garcia, Prieto, & Kammel, 2015) se han definido 12 tramos de la variable renta de acuerdo a la siguiente tabla:

*Tabla 5.2: Códigos y tramos de renta definidos*

CÓDIGO	TRAMO
1	Negativas y 0.
2	Desde 0 hasta 6000 euros.
3	Desde 6000 hasta 12000 euros.
4	Desde 12000 hasta 18000 euros.
5	Desde 18000 hasta 24000 euros.
6	Desde 24000 hasta 30000 euros.
7	Desde 30000 hasta 36000 euros.
8	Desde 36000 hasta 42000 euros.
9	Desde 42000 hasta 48000 euros.
10	Desde 48000 hasta 54000 euros.
11	Desde 54000 hasta 60000 euros.
12	De más de 60.000 euros.

Las decilas de renta se han creado a partir de la variable “renta” y la propia definición matemática, es decir, la decila 1, equivale al 10% de las declaraciones de la cola inferior de la

distribución (en este caso de la variable ‘renta’) por lo tanto, equivale a la suma de la renta de cada individuo hasta alcanzar el 10% de la población. Es importante aclarar que se han usado decilas que no incluyen a las anteriores (por lo que estrictamente no son decilas, sino grupos).

*Nota metodológica importante:*

*En cuanto a las CCAA, la codificación que aparece en la “Guía Renta 2012: Libro electrónico” proporcionado por la AEAT en su página web, menciona una codificación que se corresponde con la siguiente:*

*Tabla 5.3: Codificación de las comunidades autónomas*

Clave	Nombre
1	ANDALUCÍA
2	ARAGÓN
3	PRINCIPADO DE ASTURIAS
4	ILLES BALEARS
5	CANARIAS
6	CANTABRIA
7	CASTILLA Y LEÓN
8	CASTILLA LA MANCHA
9	CATALUÑA
10	COMUNITAT VALENCIANA
11	EXTREMADURA
12	GALICIA
13	COMUNIDAD DE MADRID
14	REGIÓN DE MURCIA
17	LA RIOJA
18	CEUTA
19	MELILLA
20	NO RESIDENTES

Tabla 5.4: Media de cada agregado monetario para cada tramo de renta

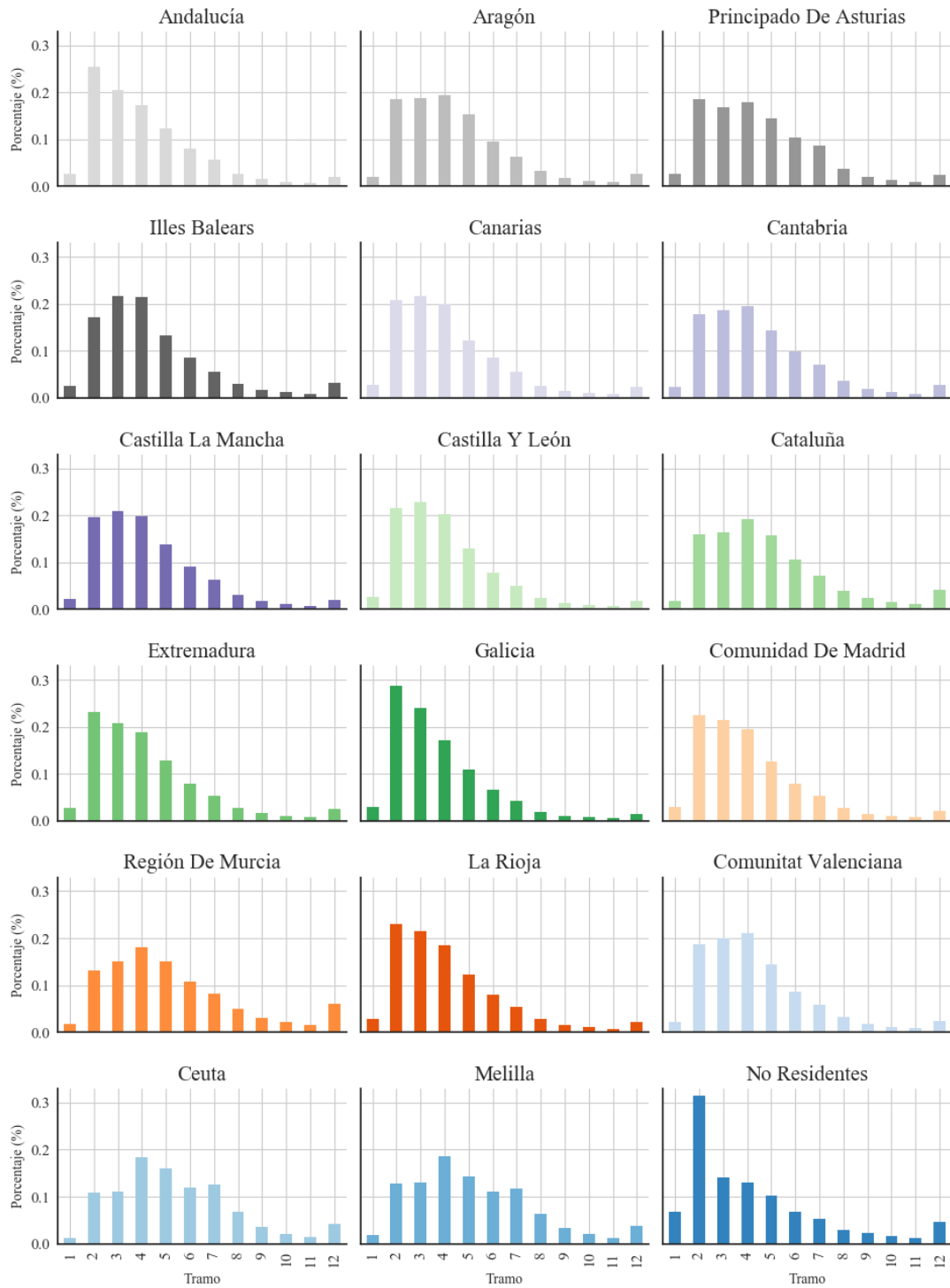
Tramo	1	2	3	4	5	6	7	8	9	10	11	12
<b>renta_periodo</b>	-1.345,76	6.097,80	11.960,73	17.424,04	23.435,27	29.526,49	35.459,21	41.333,60	47.390,09	53.348,92	59.423,74	109.874,81
<b>renta</b>	-2.912,10	2.950,47	9.039,64	14.845,93	20.825,11	26.904,62	32.839,17	38.761,24	44.742,86	50.755,16	56.827,11	107.305,16
<b>BI</b>	-2.834,69	2.990,67	9.085,01	14.891,35	20.860,80	26.916,63	32.837,77	38.765,27	44.753,52	50.764,40	56.837,24	107.337,60
<b>BL</b>	-2.875,62	2.670,34	8.328,25	13.917,31	19.826,60	25.855,42	31.637,71	37.303,17	43.044,98	48.982,87	54.635,99	103.813,95
<b>CI</b>	10,84	7,82	549,01	1.725,13	3.252,89	5.016,20	6.722,72	8.661,60	10.829,49	13.116,43	15.370,91	36.230,26
<b>CL</b>	9,98	5,26	397,28	1.395,55	2.892,40	4.627,48	6.345,97	8.212,75	10.347,26	12.580,51	14.832,19	35.573,92
<b>CRA</b>	11,73	3,30	301,63	1.394,70	2.894,25	4.622,52	6.338,22	8.198,97	10.331,58	12.567,65	14.799,75	35.350,69
<b>CR</b>	-10,03	-20,00	257,19	1.351,30	2.843,87	4.567,13	6.307,10	8.144,40	10.278,44	12.525,11	14.762,89	35.322,96
<b>RDO</b>	-229,69	-228,48	-409,94	-408,83	-422,37	-455,07	-413,77	-239,64	-39,77	-43,45	150,13	4.102,37

## capítulo 3

Tabla 5.5: Estadísticos descriptivos por comunidad autónoma.

CA	Media	Mediana	Percentil 1%	Percentil 10%	Percentil 25%	Percentil 75%	Percentil 90%	Percentil 99%	Asimetría	CV	Curtosis	Varianza
ANDALUCÍA	16.119,53	12.378,68	-258,72	1.363,26	5.219,26	22.115,62	33.445,45	74.673,66	61,81	0,74	13.914,62	473.223.993,29
ARAGÓN	18.710,76	15.137,55	0,00	2.737,38	7.404,79	24.450,90	36.044,17	83.110,76	16,02	0,88	855,32	447.863.885,91
PRINCIPADO DE ASTURIAS	19.129,45	15.711,52	-409,87	2.146,91	7.290,54	26.044,36	36.642,87	79.394,08	61,00	0,81	22.580,43	563.575.587,22
ILLES BALEARS	18.954,72	14.226,80	-1.208,43	2.804,05	7.558,37	23.213,32	36.044,78	98.010,47	50,21	0,62	6.005,41	948.749.953,42
CANARIAS	17.178,20	13.193,62	0,00	1.762,96	6.404,49	22.379,49	33.737,21	81.742,64	30,99	0,75	3.718,46	530.987.125,90
CANTABRIA	18.890,39	15.168,04	-593,67	2.559,87	7.604,83	24.991,35	36.341,73	83.936,44	11,37	0,93	444,40	412.177.261,57
CASTILLA Y LEÓN	17.550,56	14.038,57	0,00	2.483,75	6.921,59	23.428,54	34.294,85	75.441,96	83,28	0,71	15.938,44	614.782.675,35
CASTILLA LA MANCHA	15.975,34	12.793,57	-138,15	1.853,82	6.246,54	21.218,51	32.219,85	72.140,20	19,82	0,92	1.995,77	304.229.736,83
CATALUÑA	21.886,97	16.929,62	0,00	3.228,44	8.867,07	27.376,81	40.781,57	107.541,44	438,26	0,34	336.412,90	4.253.007.113,37
COMUNITAT VALENCIANA	17.005,40	12.927,42	-32,93	1.579,88	5.763,12	22.174,93	33.715,97	83.712,86	255,62	0,43	113.084,47	1.600.715.069,76
EXTREMADURA	13.957,23	10.477,86	-206,34	1.079,52	4.782,76	18.983,18	29.921,83	67.229,71	165,97	0,71	68.655,13	383.900.876,52
GALICIA	16.628,10	12.840,58	-1.018,38	1.882,77	5.874,12	21.854,31	33.401,18	77.353,23	770,72	0,29	745.354,74	3.239.011.172,77
COMUNIDAD DE MADRID	25.845,60	18.790,25	0,00	3.876,91	10.453,96	30.575,31	47.418,16	138.795,32	140,79	0,37	41.069,84	4.893.911.710,44
REGIÓN DE MURCIA	16.465,24	12.709,92	0,00	1.625,47	5.756,76	21.764,38	33.456,33	77.437,23	21,88	0,83	1.929,68	396.619.479,42
LA RIOJA	17.970,75	14.350,71	0,00	2.544,56	7.276,01	23.219,29	34.885,57	79.621,44	30,57	0,80	2.842,02	501.534.699,08
CEUTA	24.950,61	21.023,98	0,00	4.764,91	12.793,79	32.152,03	43.702,83	98.434,64	29,06	0,87	2.107,05	819.780.755,96
MELILLA	23.579,04	19.210,52	0,00	3.762,47	10.830,15	32.082,50	42.500,55	96.207,41	10,27	0,98	315,48	581.224.454,16
NO RESIDENTES	18.820,02	11.040,56	-1.625,94	84,82	2.185,38	23.782,71	41.492,72	136.318,15	12,35	0,50	252,16	1.416.587.866,15

*Gráfico 5.1: Distribución de tramos de renta por CCAA*  
 Distribucion De Tramos De Renta Relativos Por CCAA



Fuente: Elaboración propia a partir de la Muestra de declarantes del IRPF AEAT-IEF 2012

Tabla 5.6: Tipo impositivo medio por decilas de renta para el cociente entre  $ci$  y  $bl$ .

decila	1	2	3	4	5	6	7	8	9	10
ANDALUCÍA	0,0075	0,0103	0,0302	0,0726	0,1038	0,1261	0,1533	0,1821	0,2094	0,2924
ARAGÓN	0,0099	0,0109	0,0303	0,0790	0,1151	0,1368	0,1614	0,1862	0,2117	0,2890
PRINCIPADO DE ASTURIAS	0,0156	0,0144	0,0337	0,0823	0,1138	0,1339	0,1568	0,1807	0,2104	0,2888
ILLES BALEARS	-0,0215	0,0133	0,0370	0,0847	0,1182	0,1395	0,1648	0,1917	0,2098	0,3033
CANARIAS	0,0094	0,0118	0,0380	0,0826	0,1141	0,1365	0,1596	0,1874	0,2124	0,2987
CANTABRIA	0,0237	0,0150	0,0336	0,0814	0,1132	0,1334	0,1564	0,1851	0,2112	0,2891
CASTILLA Y LEÓN	0,0085	0,0141	0,0306	0,0788	0,1128	0,1353	0,1591	0,1853	0,2100	0,2843
CASTILLA LA MANCHA	0,0098	0,0111	0,0289	0,0701	0,1048	0,1288	0,1549	0,1822	0,2084	0,2854
CATALUÑA	-0,1198	0,0135	0,0329	0,0796	0,1120	0,1361	0,1602	0,1873	0,2127	0,3075
COMUNITAT VALENCIANA	0,0055	0,0099	0,0299	0,0772	0,1075	0,1293	0,1586	0,1852	0,2096	0,2962
EXTREMADURA	0,0056	0,0121	0,0285	0,0723	0,1054	0,1286	0,1564	0,1824	0,2088	0,2907
GALICIA	0,0115	0,0117	0,0307	0,0809	0,1121	0,1336	0,1589	0,1852	0,2102	0,2938
COMUNIDAD DE MADRID	0,0355	0,0110	0,0365	0,0850	0,1145	0,1395	0,1626	0,1884	0,2105	0,3198
REGIÓN DE MURCIA	0,0055	0,0098	0,0322	0,0702	0,0974	0,1221	0,1479	0,1795	0,2047	0,2877
LA RIOJA	0,0107	0,0118	0,0319	0,0803	0,1167	0,1365	0,1592	0,1871	0,2099	0,2854
CEUTA	0,0039	0,0116	0,0405	0,0758	0,1029	0,1289	0,1497	0,1786	0,2106	0,2871
MELILLA	0,0098	0,0154	0,0368	0,0704	0,0980	0,1249	0,1443	0,1737	0,2071	0,2851
NO RESIDENTES	0,0009	0,0267	0,0441	0,0881	0,1215	0,1384	0,1614	0,1867	0,2104	0,3101

## BIBLIOGRAFIA

Abdelkrim, A., & Duclos, J.-Y. (2007). DASP: Stata modules for distributive analysis. *Statistical Software Components*.

Agencia Tributaria. (n.d.). Manual práctico RENTA 2012. Retrieved from [http://www.agenciatributaria.es/static\\_files/AEAT/DIT/Contenidos\\_Publicos/CAT/AYUWEB/Biblioteca\\_Virtual/Manuales\\_practicos/Renta/Manual\\_renta\\_patrimonio\\_2012\\_es\\_es.pdf](http://www.agenciatributaria.es/static_files/AEAT/DIT/Contenidos_Publicos/CAT/AYUWEB/Biblioteca_Virtual/Manuales_practicos/Renta/Manual_renta_patrimonio_2012_es_es.pdf)

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3), 244–263. [http://doi.org/10.1016/0022-0531\(70\)90039-6](http://doi.org/10.1016/0022-0531(70)90039-6)

Atkinson, A. B. (2016). Desigualdad: ¿Qué podemos hacer? Fondo de Cultura Económica.

Gini, C. (1936). On the measure of concentration with special reference to income and statistics. Colorado College Publication.

Jenkins, S. (1988). Calculating income distribution indices from micro-data. *National Tax Journal*. <http://doi.org/10.2307/41788716>

Lopez, C. P., Garcia, J. V., Prieto, M. J. B., & Kammel, El, E. B. R. L. K. C. (2015). LA MUESTRA DE IRPF DE 2012. *Documentos De Trabajo*, 1–142.

Novales, A. (2011). Crecimiento económico, desigualdad y pobreza (pp. 419–435). Presented at the Anales de la Real Academia de Ciencias Morales y Políticas.

Picos Sánchez, F., Lopez, C. P., & Marín, C. G. (2011). La muestra de declarantes de IRPF de 2007. *Documentos De Trabajo*, 1–53. Retrieved from [http://www.ief.es/documentos/recursos/publicaciones/documentos\\_trabajo/2011\\_01.pdf](http://www.ief.es/documentos/recursos/publicaciones/documentos_trabajo/2011_01.pdf)

Theil, H. (1989). The development of international inequality 1960–1985. *Journal of Econometrics*, 42(1), 145–155. [http://doi.org/10.1016/0304-4076\(89\)90082-1](http://doi.org/10.1016/0304-4076(89)90082-1)