Economía Española y Protección Social, VIII, Año 2016. 139-175

DISTRIBUCIONES SALARIALES: AJUSTE Y PROYECCIÓN. UNA APLICACIÓN A LA ESTIMACIÓN DE RETENCIONES A CUENTA DEL IRPF

WAGE DISTRIBUTIONS: ADJUSTMENT AND PROJECTION. AN APPLICATION TO THE ESTIMATED INCOME TAX WITHHOLDINGS

Pedro Valverde Caramés[1]

Jefe de Área. Servicio de Estudios Tributarios y Estadísticas. AEAT. España

Resumen

Este artículo parte de la carencia, en las fuentes estadísticas fiscales españolas, de información sobre el *tiempo de trabajo* en materia de retribuciones salariales. La información estadística sobre salarios anuales es muy precisa pero no tiene en cuenta el *tiempo efectivo* en que se obtiene esa renta.

Para poder abordar lo anterior se recurre a la Muestra Continua de Vidas Laborales (*MCVL*)**,** puesto que en dicha operación estadística se integran, conjuntamente, tanto información fiscal como laboral; salarios y tiempo de trabajo. El conocimiento detallado del tiempo de trabajo *efectivo* (en el que se obtiene la renta y que no tiene por qué coincidir con aquel al que se le imputa, generalmente el año natural) permite comprender de manera satisfactoria la forma que adopta la distribución de las rentas salariales anuales (inexplicable de otra manera) así como realizar una estratificación

[1] Correo electrónico / E-mail: pedro.valverdec@correo.aeat.es

muy fructífera en términos de análisis estadístico de la población de asalariados.

Todo lo anterior permite una adecuada modelización de las retribuciones salariales de la economía española. Como se verá, éstas se ajustan con mucha precisión a una distribución estadística de cuatro parámetros conocida como Beta Generalizada de Segunda Especie (GB2, en la literatura sobre el tema). Conocida la distribución de probabilidad a la que se ajustan los salarios anuales, se puede estudiar cómo evolucionan en el tiempo los parámetros que la definen, con lo cual se dispone de una herramienta provechosa para efectuar predicciones. Reemplazar los valores reales por sus equivalentes simulados, a través de la correspondiente GB2, permite el análisis de la evolución de los salarios y la simulación de acciones fiscales sobre ellos entre otras muchas posibilidades. A modo de ejemplo se presentará una propuesta para la estimación de las Retenciones a Cuenta del *IRPF*.

Palabras clave

Salarios; Tiempo de trabajo; Simulaciones numéricas; Funciones de distribución; Beta generalizada de segundo orden (GB2).

Abstract

Information about annual salaries is very accurate in Spanish tax statistics but does not consider the time period in which that income is obtained. To address the above, MCVL is used (MCVL is a statistical operation that integrates tax and employment information, wages and working time). The detailed working time (in which income is obtained and that does not have to match that to which is charged, usually a calendar year) allows us to understand satisfactorily the form of the distribution of incomes wages (otherwise inexplicable) and stratifying -very fruitful in terms of statistical analysis- the population of employees.

All this allows adequate modeling of the salaries of the Spanish economy. As it will be seen, they are adjusted very accurately to a statistical distribution known as Beta four parameters Generalized Second Kind (GB2, literature on the subject). Knowing the probability distribution which annual wages are adjusted, you can study how they evolve over time parameters that define it, showing a helpful tool for making predictions. Replace the actual values for their simulated counterparts, through the corresponding GB2, allows the analysis of the evolution of wages, the simulation of fiscal actions on them among many other possibilities. As an example, a proposal for the estimation of the personal income tax withholdings will be presented.

Keywords

Wages; Working time; Income distribution; Generalized beta of second kind (GB2); Numerical simulations.

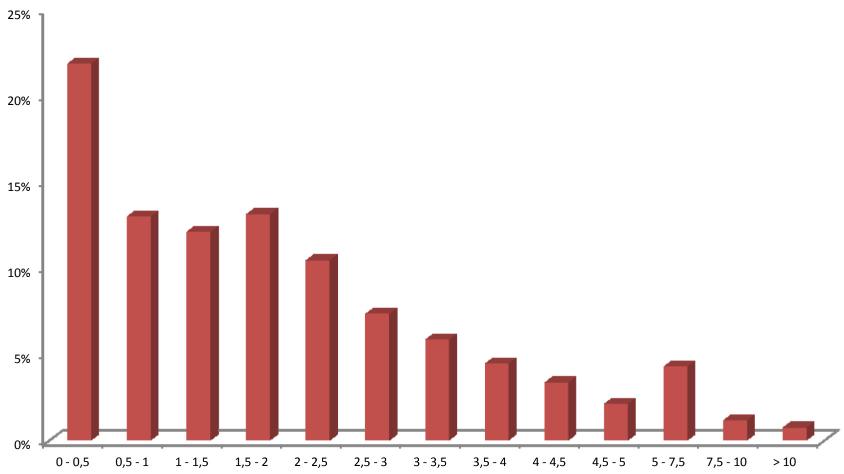
JEL: C13, C46, D63.

1. El problema de la forma

El presente artículo es una continuación de Valverde Caramés, P. (2011). Allí se discutía la necesidad de incorporar el tiempo de trabajo como variable fundamental **a la hora de estudiar la distribución de los salarios de la economía española y se proponía el uso de la Muestra Continua de Vidas Laborales (*MCVL*) como la herramienta adecuada para ese fin.** Partimos aquí del Gráfico 1 que muestra la distribución de las retribuciones salariales, en este caso como función del Salario Mínimo Interprofesional

(*SMI*), correspondiente al ejercicio de 2014. La información proviene de la estadística *Mercado de Trabajo y Pensiones en las Fuentes Tributarias* (*MTyPFT*) elaborada por la Agencia Estatal de Administración Tributaria (*AEAT*).

Gráfico 1. Distribución de las retribuciones salariales anuales 2014



Fuente: *Mercado de trabajo y pensiones en las fuentes tributarias* (*AEAT*).

La forma de este histograma es bien conocida y se repite de manera habitual al analizar la distribución de los salarios en la economía española. La existencia de *colas pesadas* en la parte izquierda de la distribución (salarios bajos) junto con un decrecimiento asintótico por el lado derecho (salarios altos) y una doble moda es una constante a lo largo del tiempo en los estudios sobre distribución de retribuciones salariales (véase, por todos, Melis Maynar, F. (1995)). Usando la información proporcionada por la explotación de la *MCVL* (la *MCVL-2014*, en este caso) se puede replicar la proporcionada por la estadística *MTyPFT* y así se obtiene la Tabla 1.

Como se puede observar, la *MCVL* replica de manera muy aproximada tanto la distribución como los salarios medios por tramo de *SMI* y además proporciona el tiempo de trabajo medio para cada intervalo. Esto último no es posible con la estadística *MTyPFT*.

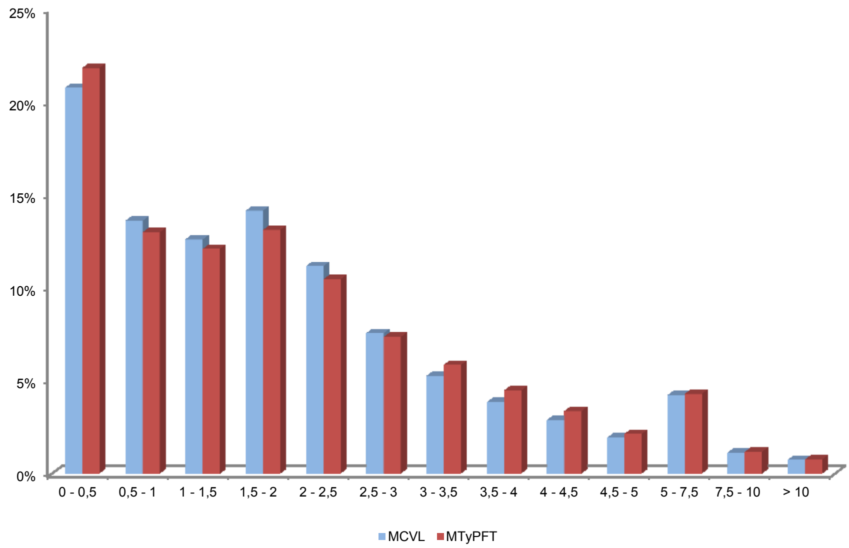
Tabla 1. Retribuciones salariales y tiempo de trabajo. Comparación entre *MCVL* y *MTyPFT*

TRAMOS SMI2014	**DISTRIBUCIÓN**		**SALARIOS MEDIOS**		**TIEMPO TRABAJO MEDIO**
	MCVL-2014	***MTyPFT***	***MCVL-2014***	***MTyPFT***	**MCVL-2014**
0 - 0,5	20,80%	21,86%	1.907	1.840	*101,05*
0,5 - 1	13,62%	13,00%	6.744	6.731	*243,85*
1 - 1,5	12,62%	12,11%	11.329	11.320	*311,11*
1,5 - 2	14,17%	13,14%	15.812	15.815	*346,43*
2 - 2,5	11,18%	10,47%	20.178	20.182	*356,30*
2,5 - 3	7,57%	7,36%	24.702	24.726	*358,96*
3 - 3,5	5,27%	5,86%	29.235	29.260	*360,13*
3,5 - 4	3,85%	4,48%	33.760	33.775	*361,29*
4 - 4,5	2,89%	3,37%	38.245	38.251	*361,41*
4,5 - 5	1,94%	2,13%	42.748	42.744	*361,25*
5 - 7,5	4,22%	4,29%	54.004	53.885	*361,53*
7,5 - 10	1,13%	1,17%	76.699	76.753	*361,87*
> 10	0,74%	0,76%	150.129	148.824	*356,35*
TOTAL	**100,00%**	**100,00%**	**18.138**	**18.420**	***281,91***

Fuente: elaboración propia.

A través de la *MCVL* se pueden conocer los días que el trabajador ha estado en alta a lo largo del año. Esto permite estudiar el período de tiempo -efectivo- en el que se han obtenido dichas retribuciones. El análisis así concebido tiene en cuenta una perspectiva temporal de la que carece si únicamente se dispone de la información de carácter fiscal. Representando las dos distribuciones de la tabla anterior se obtiene el Gráfico 2. Ambos histogramas tienen la misma forma, como era de prever, ya que ambas estadísticas comparten la misma información para determinar la masa salarial total. Es importante destacar que la *MCVL* combina en una única operación estadística tanto información fiscal, como serían los salarios en este caso, como información propia de seguridad social, el tiempo de trabajo. La explotación del módulo fiscal de la *MVCL* (el que contiene la información sobre salarios, retenciones etc. y ha sido utilizado en este trabajo) permite un análisis más completo que si se usan las bases de cotización, que al estar topadas aproximan mal los salarios más altos.

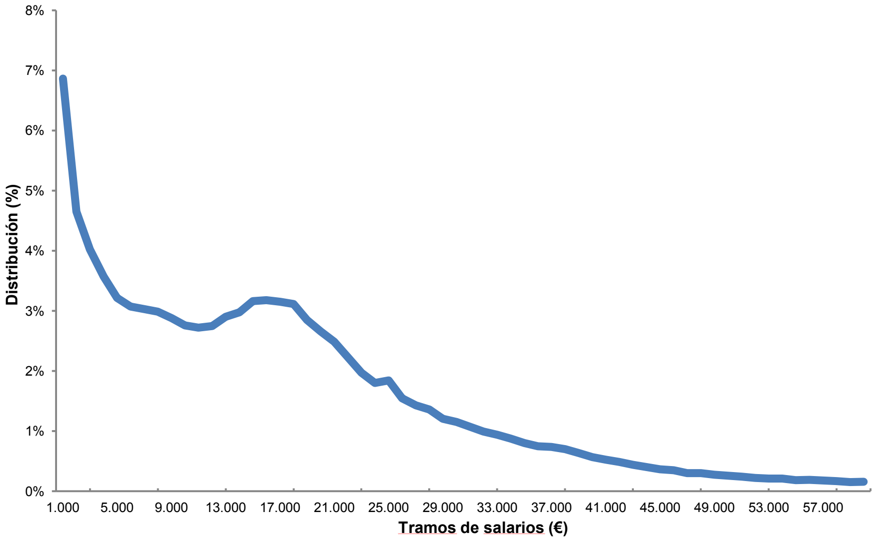
Gráfico 2. Distribución de las retribuciones salariales anuales 2014. Comparación entre *MCVL* y *MTyPFT*



Fuente: elaboración propia.

Usando la *MCVL-2014* se pueden construir las curvas de densidad (reduciendo la amplitud de los intervalos considerados) para las retribuciones salariales. El Gráfico 3 permite observar esto por tramos de salario bruto anual para el ejercicio de 2014. La curva obtenida es una envolvente al histograma inicial.

Gráfico 3. Curva de densidad para los salarios del ejercicio 2014



Fuente: elaboración propia. Miles de euros.

2. *Permanentes* y *Eventuales*

Al disponer, para cada uno de los trabajadores, no sólo de su salario anual sino además del número de días en alta laboral, se puede avanzar en el análisis descomponiendo la población total en dos colectivos claramente diferenciados. De forma natural el conjunto admite una desagregación en función de si se ha estado en alta laboral todo el año o sólo una fracción del mismo. A cada trabajador estudiado se le asigna, para cada ejercicio, un vector de 365 componentes, una para cada día del año. Cada componente tendrá el valor 1 si ha estado de alta ese día y 0 si no lo ha estado, de tal manera que quedará definido por un vector del tipo (1,1,1,…..,1,0,0,0,….).

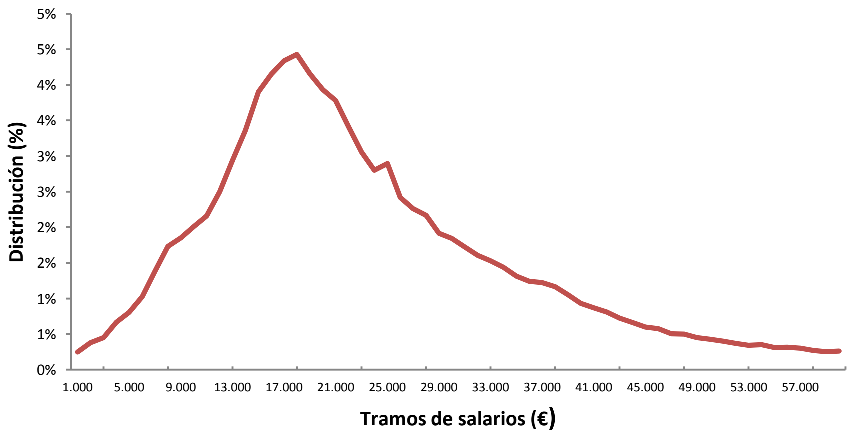
En base a esa estructura se puede clasificar al total de la muestra en dos conjuntos disjuntos:

1. Permanentes: será aquel subgrupo de trabajadores que están en alta laboral los 365 días del año, con lo cual tendrán un vector del tipo (1,1,….,1). No entran ni salen de mercado laboral en el año a estudio.

2. Eventuales: serán todos lo que no cumplen el ser Permanentes, tendrán una o varias entradas y/o salidas del Régimen General de la Seguridad Social. Han estado en alta menos de 365 días en el año. Tienen un vector asociado del tipo (1,0,1,0,…,0,0,1…1).

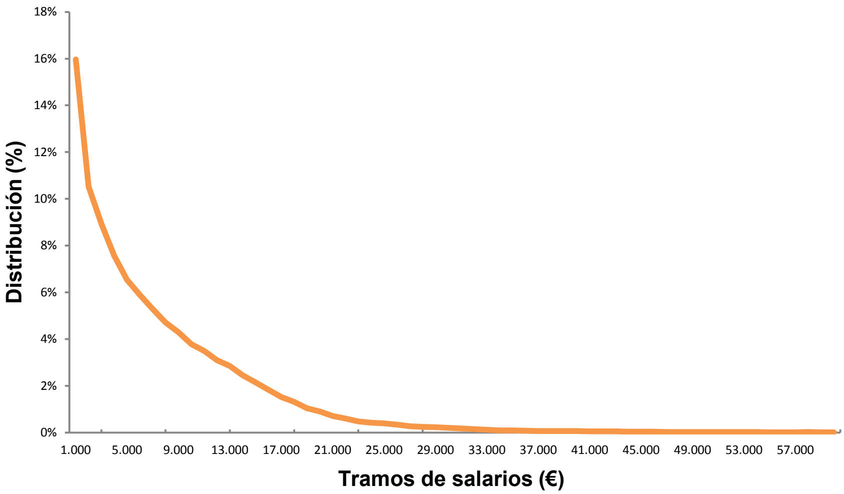
Como se verá, esta descomposición del colectivo total en función del tiempo de trabajo es especialmente fructífera a hora de explicar el problema de las *colas pesadas* en la distribución de salarios. La curva de densidad de los salarios para el colectivo de *Eventuales* es totalmente diferente a la que se obtiene para el de *Permanentes*. Ello se desprende de la observación directa de los Gráficos 4 y 5.

Gráfico 4. Curvas de densidad para los salarios. Permanentes 2014



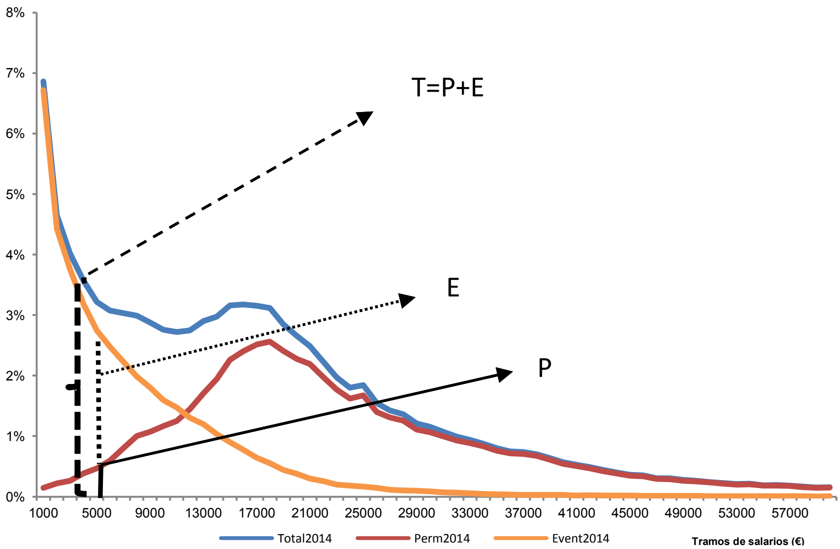
Fuente: elaboración propia.

Gráfico 5. Curvas de densidad para los salarios. Eventuales 2014



Fuente: elaboración propia.

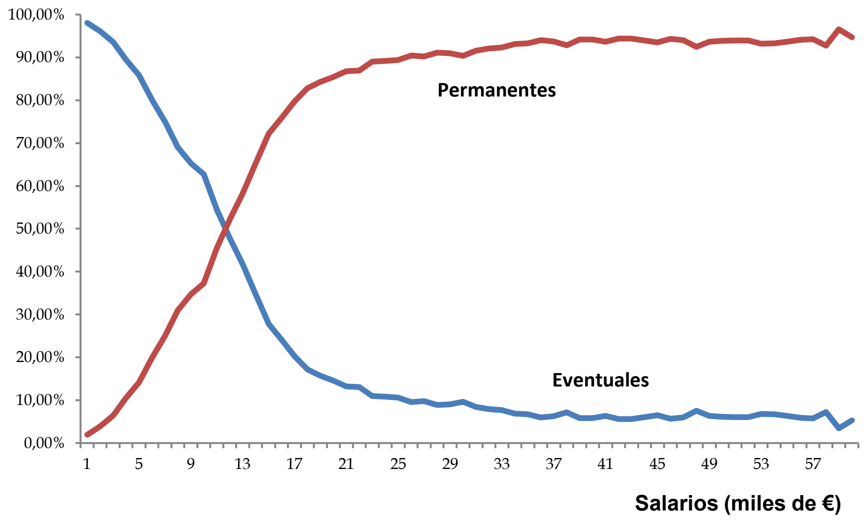
Gráfico 6. Descomposición de la curva de densidad para el total de la población 2014 en Permanentes y Eventuales



Fuente: elaboración propia.

Discriminar el colectivo total entre *Permanentes* y *Eventuales* permite explicar las *colas pesadas* de la distribución de salarios y la existencia de una importante bolsa de trabajadores con salarios muy bajos, por debajo del *SMI* anual. Como se deduce del análisis de estos colectivos, la mayoría de los trabajadores con sueldos muy bajos son *asalariados eventuales* que trabajan sólo una fracción del año. Por ello, más que hablar de sueldos bajos se debería hablar de poco tiempo trabajado. La composición de ambas distribuciones, *Eventuales* y *Permanentes*, promediadas sobre el total del colectivo, que se presenta en el Gráfico 7, así lo pone de manifiesto.

Gráfico 7. Relación entre *Permanentes* y *Eventuales* por nivel salarial



Fuente: elaboración propia.

La Tabla 2, resume la información obtenida, en materia de salarios y tiempo de trabajo, de la *MCVL* para el período 2004-2014.

Tabla 2. Resumen información obtenida de la *MCVL* (2004-2014)

POBLACION	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***	***2010***	***2011***	***2012***	***2013***	***2014***
Número de trabajadores	544.370	597.429	624.925	653.741	659.005	624.452	609.664	600.233	601.888	562.152	571.842
Salario anual medio	15.805,52	16.167,58	16.976,00	17.763,23	18.540,95	18.635,63	18.670,23	18.770,71	17.731,21	18.237,60	18.137,99
Días de trabajo medios	294,52	292,64	294,12	294,63	291,06	285,51	286,21	285,76	288,82	282,03	281,91
Salario día medio	*53,67*	*55,25*	*57,72*	*60,29*	*63,70*	*65,27*	*65,23*	*65,69*	*61,39*	*64,67*	*64,34*
PERMANENTES											
Número de trabajadores	314.363	335.239	350.125	367.899	374.591	366.748	361.282	354.794	362.294	332.217	331.157
Salario anual medio	21.670,38	22.400,42	23.360,74	24.334,72	25.556,89	25.785,41	25.828,66	26.032,23	24.363,96	25.693,24	25.835,38
Días de trabajo medios	365,00	365,00	365,00	365,00	365,00	365,00	365,00	365,00	365,00	365,00	365,00
Salario día medio	59,37	61,37	64,00	66,67	70,02	70,64	70,76	71,32	66,75	70,39	70,78
%Permanentes	*57,7%*	*56,1%*	*56,0%*	*56,3%*	*56,8%*	*58,7%*	*59,3%*	*59,1%*	*60,2%*	*59,1%*	*57,9%*
EVENTUALES											
Número de trabajadores	230.007	262.190	274.800	285.842	284.414	257.704	248.382	245.439	239.594	229.935	240.685
Salario anual medio	7.789,70	8.198,20	8.841,81	9.305,25	9.300,52	8.460,52	8.257,99	8.273,81	7.701,71	7.465,46	7.547,21
Días de trabajo medios	198,20	200,14	203,83	204,06	193,68	172,39	171,60	171,22	172,12	162,16	167,59
Salario día medio	39,30	40,96	43,38	45,60	48,02	49,08	48,12	48,32	44,75	46,04	45,03
%Eventuales	*42,25%*	*43,89%*	*43,97%*	*43,72%*	*43,16%*	*41,27%*	*40,74%*	*40,89%*	*39,81%*	*40,90%*	*42,09%*

Fuente: elaboración propia. Sobre el *salario día medio*, véase Valverde Caramés, P. (2011).

3. Modelización paramétrica de una distribución empírica

Se puede definir la modelización paramétrica como una técnica de estadística matemática cuyo objetivo fundamental es resumir los datos empíricos mediante una función matemática dependiente de un número pequeño de parámetros, y sin que ello suponga una pérdida de información importante. En el caso que nos ocupa la aplicaremos a la información obtenida sobre rentas salariales. El punto de partida de la modelización paramétrica siempre será un ***modelo de probabilidad*** definido por una familia de funciones de distribución. Dicho modelo se propone con el objeto de representar el conjunto de datos disponibles y su elección quedará determinada por las características específicas del fenómeno a estudio. Para mayor información véase Prieto Alaiz, M. y C. García Pérez (2006).

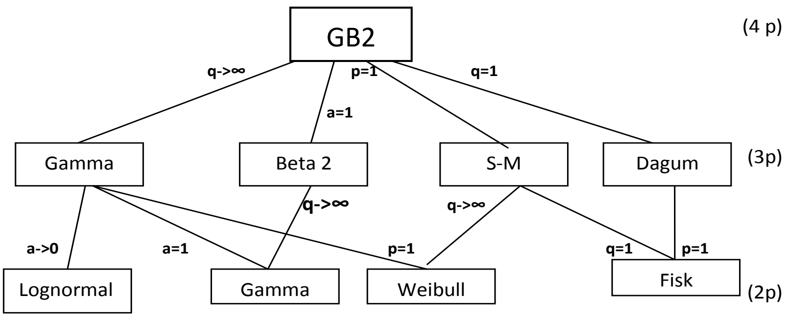
3.1. La distribución Beta Generalizada de segundo orden y derivadas

Se presenta en este apartado a la distribución Beta Generalizada de Segunda Especie (*GB2*). La expresión de la función de densidad de la distribución beta generalizada de segunda especie es la siguiente:

$$GB2(r;a,b,p,q) \;=\; \frac{ar^{ap-1}}{b^{ap}\;B(p,q)(\,1-(\frac{r}{b})^{a})^{(\,p+q)}} \qquad r \geq 0$$

donde se tiene que *a*, *b*, *p*, *q* >0 y $B(p,q)$ es la función Beta. El parámetro *b* lo es de escala y los restantes tres, *a*, *p* y *q* de forma. La *GB2* es una distribución que proporciona una adecuada descripción de la distribución de la renta, con un número razonable de parámetros (4).

Figura 1. La distribución GB2 y sus derivadas



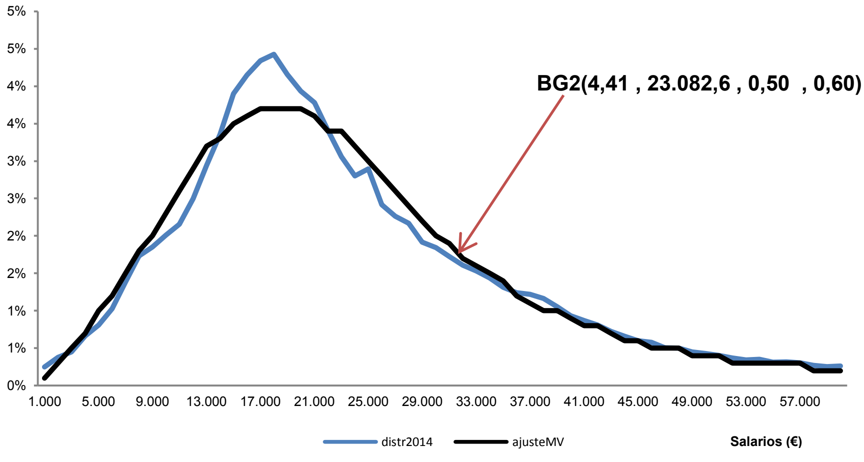
Fuente: Kleiber, C. y S. Kotz (2003).

La gran ventaja de la distribución GB2 es la riqueza de modelos relacionados con ella, que incluye las distribuciones *triparamétricas Dagum* (*DAGUM*) y *Singh-Maddala* (*SM*), junto a los modelos del mismo orden beta de segunda especie y gamma generalizada, y los modelos de dos parámetros *Lognormal*, *Gamma*, *Weibull* y *Fisk.* Todas estas funciones con casos particulares de la GB2. Para más información sobre la GB2, véase Kleiber, C. y S. Kotz (2003).

4. Aplicación de la GB2 a los resultados de la MCVL-2014

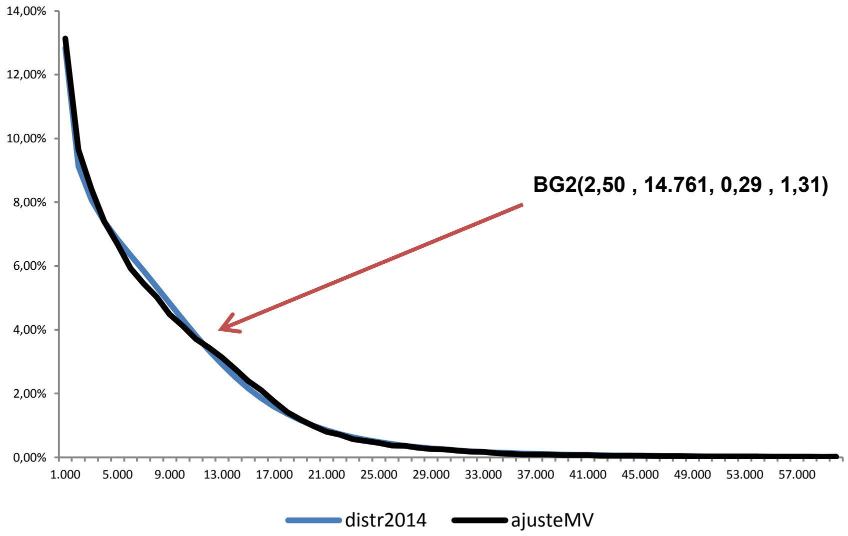
En el caso particular que nos ocupa, con la modelización paramétrica de las distribuciones empíricas para las retribuciones salariales anuales, obtenidas de la *MCVL-2014*, se busca estimar los parámetros que definen el ajuste óptimo de una distribución *GB2*. El *Método de máxima verosimilitud* proporciona este tipo de estimadores. Se ha aplicado el Principio de Máxima Verosimilitud sobre los datos de rentas salariales individuales de acuerdo a la información obtenida de la *MCVL-2014*. El procedimiento consiste en ajustar una distribución *GB2* para el conjunto de *Permanentes* y *Eventuales* de un ejercicio y reconstruir la distribución del total como agregación por intervalos de ambas distribuciones.

Gráfico 8. Ajuste del colectivo de Permanentes



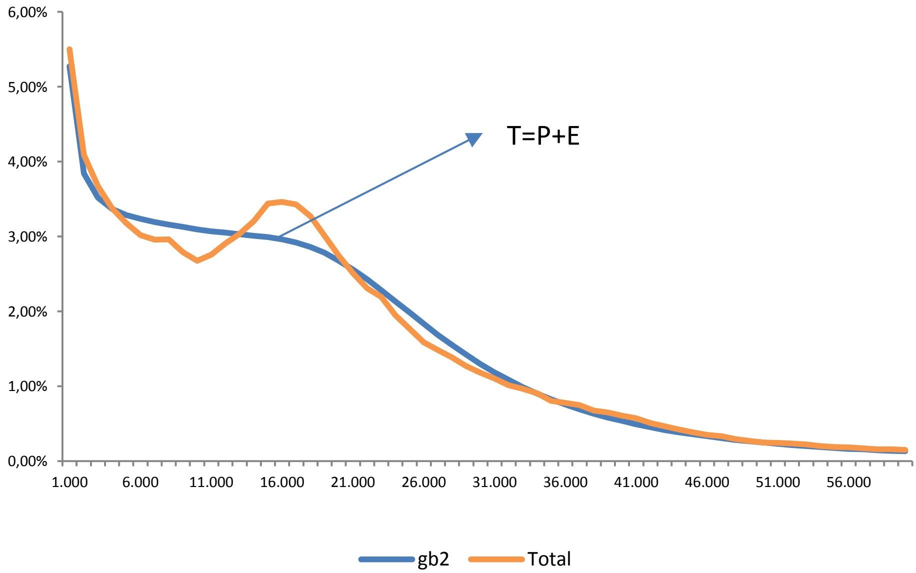
Fuente: elaboración propia.

Gráfico 9. Ajuste del colectivo de Eventuales



Fuente: elaboración propia.

Gráfico 10. Ajuste de la distribución del colectivo total (T=P+E)



Fuente: elaboración propia.

Tabla 3. Media y Mediana para los colectivos de Permanentes y Eventuales. Comparación entre empíricos y ajustados. Error cometido

Empíricos MCVL	***Permanentes***	***Eventuales***
Media	25.835,38	7.547,21
Mediana	20.954,30	5.077,89
Ajuste GB2		
Media	25.909,46	7.469,45
Mediana	21.358,31	5.169,36
Error		
Media	-0,29%	1,03%
Mediana	-1,93%	-1,80%

Fuente: elaboración propia.

5. Propuesta para la estimación de las Retenciones del IRPF sobre los Rendimientos del Trabajo. Ajuste numérico de la distribución salarial

5.1. Planteamiento de la cuestión

Como ejemplo de la aplicación de lo que se ha presentado hasta este momento se pretende en lo que sigue elaborar una metodología para la estimación de la recaudación por retenciones sobre *IRPF* en ejercicios futuros (*esto es,* partiendo de la información disponible en un ejercicio *n*, calcularla *en n+i, i=1 ó 2*) que sea consistente con las previsiones que informa el *Cuadro Macroeconómico* (información que se supone conocida en el ejercicio n y referida a *n+1 o n+2*). Las previsiones sobre la recaudación por retenciones deben ser consecuentes con lo anterior para que tengan utilidad, no se puede calcular de manera independiente al resto de las variables macroeconómicas. Se parte de que la estimación buscada ha de ser consistente con dos parámetros, *Salario medio y Número de trabajadores*, que son fijos y externos a cualquier modelo que se plantee.

Son por ello condiciones a las que se debe sujetar toda propuesta en este sentido. Como se vio anteriormente, la *MCVL* es una herramienta de la que se puede extraer un conocimiento de gran importancia a la hora de abordar este problema. De hecho, del análisis de la *MCVL* se pueden destacar una serie de evidencias empíricas, que se incorporarán como hipótesis de trabajo a partir de este momento. Así se tiene que:

1. La población de salarios por cuenta ajena puede dividirse en dos colectivos disjuntos: *Permanentes y Eventuales*. Estos conjuntos se muestran especialmente importantes a la hora de explicar la forma de la distribución total de los salarios.
2. Ambos colectivos se ajustan de manera óptima a una distribución Beta Generalizada de Segundo Orden (*GB2*). Esta distribución depende de cuatro parámetros (a, b, p, q) donde *a, p* y *q* son parámetros de forma y *b* lo es de escala. El parámetro *a* gobierna la forma general de la distribución, *p* la cola de la izquierda y *q* la de la derecha.
3. Se conocen los ajustes de la *MCVL* a las correspondientes *GB2* (*Permanentes* y *Eventuales*) para un período dado. En este caso usaremos el 2004-2013.
4. Sustituir los salarios reales por los obtenidos de una distribución $GB2(a, b, p, q)$ óptima en términos de ajuste máximo verosímil supone incurrir en errores, con respecto al salario medio real, por debajo del 2% para ambos colectivos (año 2013).

Partiendo de lo anterior y dado que se conoce el ajuste de los dos colectivos (*Permanentes* y *Eventuales*) para el período 2004-2013 (y subsecuentes con el paso de los ejercicios) se podría pensar en proyectar el vector de parámetros (a, b, p, q) al ejercicio a estudio, a partir de él reconstruir una población de salarios teórica (*que se supone replica a la población real futura y cuyo tamaño debe ser N*) y a ésta se aplicaría el algoritmo de retenciones correspondiente al ejercicio a estimar. De ahí se extraería el montante de la recaudación esperada. Esta primera aproximación al problema plantea una serie de inconvenientes a tener en cuenta:

1. La evolución de los parámetros de ajuste de las distribuciones $GB2(a,b,p,q)$ de los colectivos *Permanentes* y *Eventuales* se muestra *errática* en el tiempo y resulta complejo estimar los posibles valores que va a tener en un ejercicio posterior (*n+1 o n+2*). La última Muestra disponible ($MCVL_n$) será, como mínimo, la del ejercicio anterior al que se quiera estimar.

Tabla 4. Evolución de los parámetros (a,b,p,q) en el período 2004-2013

PARÁMETROS	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Permantes										
a	4,85	5,18	5,07	5,18	5,19	4,66	4,52	4,71	4,91	4,37
b	17.056,16	17.280,60	18.091,28	18.861,97	20.204,48	21.088,97	21.493,20	22.082,06	21.693,97	22.727,20
p	0,56	0,53	0,54	0,52	0,50	0,53	0,54	0,50	0,48	0,52
q	0,51	0,46	0,48	0,47	0,47	0,54	0,57	0,55	0,52	0,61
Eventuales										
a	7,70	7,80	8,28	7,67	6,07	2,93	2,92	3,06	2,53	1,24
b	12.291,65	12.684,11	13.458,26	14.015,95	14.100,22	14.343,74	14.477,32	13.959,52	15.063,62	15.262,00
p	0,10	0,10	0,10	0,11	0,13	0,28	0,27	0,25	0,30	0,33
q	0,40	0,40	0,37	0,39	0,49	1,05	1,08	0,98	1,30	1,57

Fuente: elaboración propia.

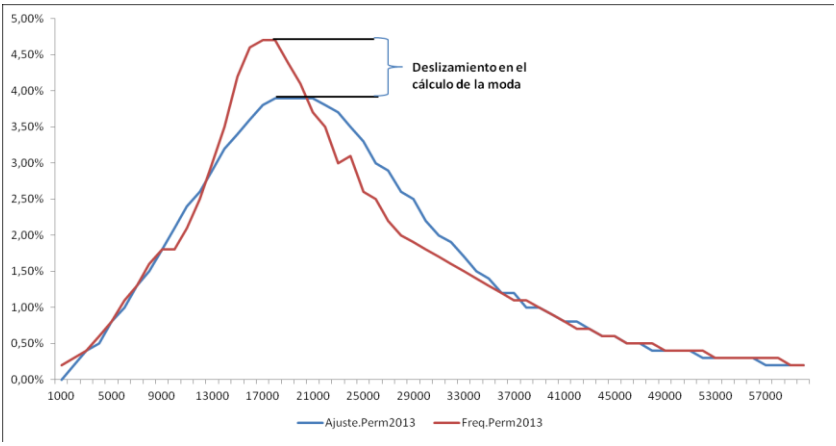
2. La distribución *GB2* es muy sensible a las variaciones de los parámetros que la definen. Por tanto, posibles errores, aunque sean pequeños, en la estimación del vector de parámetros $(\hat{a},\hat{b},\hat{p},\hat{q})$ con respecto al que sería óptimo $(a,b,p,q)_{opt}$ se transmitirán a la distribución de salarios teórica y de ahí a la de retenciones estimadas.
3. Aun cuando se consiguiera una correcta estimación de dichos parámetros para el ejercicio a estudio (*n+1* o *n+2*), cosa que sólo se podría comprobar en el ejercicio siguiente (*n+2* o *n+3*), nada

garantiza que esa distribución sea consecuente con el salario medio propuesto en el *Cuadro Macroeconómico* y que, por definición, es un parámetro dado al que se deben sujetar las estimaciones.

4. El cálculo de las retenciones es una función del salario (y de otras variables de menor relevancia, que generalmente no se tendrán en cuenta) estructurada en forma de árbol de decisión lo que da lugar a una función no lineal fuertemente dependiente de la *forma* de la distribución de los salarios. Véase el Apéndice 2.

5. El ajuste directo de la distribución *GB2* sobre el colectivo de *Permanentes* de la *MCVL*, usando estimación por Máxima Verosimilitud, da un ajuste que, si bien es óptimo en términos estadísticos, no recoge bien la amplitud del intervalo modal. Éste siempre aparece suavizado con respecto a los datos reales. Parece deseable conseguir un ajuste que se aproximase mejor en este tramo de la distribución aunque se perdiese el *óptimo* estadístico. Ahí se concentran la mayoría de los salarios que contribuyen a la recaudación por retenciones.

Gráfico 11. Ajuste por Máxima Verosimilitud de una distribución $GB2(a,b,p,q)$ a Permanentes 2013



Fuente: elaboración propia.

En cualquier caso, el tener que atenerse obligatoriamente a las indicaciones del *Cuadro Macroeconómico,* sobre el salario medio del ejercicio a estimar, es razón suficiente para tener que plantearse otra estrategia de ajuste diferente.

5.2. Planteamiento de un problema de minimización: aproximación por métodos numéricos

Una primera alternativa para abordar el objetivo propuesto (estimar la recaudación por retenciones ajustándose a un parámetro externo) sería plantearse su solución a través de un problema de minimización (o similar). Se trataría, *grosso modo*, de encontrar para cada uno de los dos colectivos en que se descompone la población total aquella combinación de parámetros a, b, p, q que den lugar, a través de la distribución *GB2*, a una población de salarios teórica cuya media sea consistente con la propuesta en el *Cuadro Macroeconómico*. Es decir, se buscaría solucionar el siguiente planteamiento general:

Sea *S* el salario medio dado por el cuadro macroeconómico y *N* el número de asalariados. Si definimos $\tilde{S} = E(GB2(a, b, p, q))$ esto es, $\tilde{S}$ es la esperanza matemática de una población (de salarios en este caso) generada por la distribución *GB2* de tamaño *N*. Se supone que $(a, b, p, q) \in C$ donde C es aquí el campo de definición adecuado para el vector de parámetros. Recuérdese que a, b, p, q son todos números reales estrictamente positivos. Por tanto tiene sentido plantearse el siguiente problema:

$$\mathrm{Min}\big(\tilde{\mathrm{S}} - \mathrm{S}\big)_{\mathrm{a,b,p,q}} \ \ / \ \ (\mathrm{a, b, p, q}) \in \mathrm{C}$$

Se trataría por tanto de buscar en C aquel vector de parámetros $(a,\widetilde{b,p},q)$ que asegure que la esperanza matemática de la distribución *GB2* es precisamente el valor buscado *S*. Obsérvese que $\tilde{S}$ es una variable aleatoria y que, por tanto, una extracción particular de $GB2((a,\widetilde{b,p},q);N)$ no tiene por qué tener una media que coincida, exactamente, con el valor *S*, pero sí se tiene la seguridad que es el mejor proceso aleatorio consistente con ese valor. En caso de que el problema anterior tuviese solución, que no tendría por qué ser única (*no hay garantías de la existencia de un óptimo global en* C), la *forma* de la función de densidad empírica puede no ser la adecuada. (Véase inconveniente 4).

Sería interesante por todo ello que no sólo se buscase una función *GB2* que minimizase la diferencia con respecto a *S* sino que a la vez su *forma* se aproximase lo más posible a los datos empíricos conocidos (*dados por la MCVL de ejercicios pasados*). Se puede pensar entonces en ampliar las condiciones del problema anterior para buscar no únicamente un vector $(a,\widetilde{b,p},q)$ que minimice la distancia con respecto a *S* sino que además genere una función de densidad que acote lo máximo posible el inconveniente 4. Para abordar esto último se necesitaría disponer de indicadores de la forma de la distribución que puedan ser calculados en función de los parámetros de la *GB2*. En realidad de (*a,p,q*), que son los que gobiernan la estructura de la función.

Por simplicidad, en lo que sigue se hará mención únicamente al colectivo de *Permanentes*, siendo todo extensible al de *Eventuales* sin más que aplicarlo sobre la *GB2* correspondiente, tal y como hemos comprobado. La razón para esto, más allá de no extendernos aquí repitiendo el mismo análisis y sí respetar las normas de publicación, es también de orden práctico ya que más del 90% del total de las retenciones provienen del colectivo de *Permanentes*.

Tabla 5. Magnitudes relevantes Permanentes 2007-2013

Medidas	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Mediana (M)	19.283,53	20.346,78	20.625,39	20.756,74	20.991,05	20.578,33	20.768,55
Media (m)	24.334,72	25.556,89	25.785,41	25.828,66	26.032,24	25.512,63	25.693,24
qsr (80%,20%)	5,376	5,439	5,578	5,619	5,677	5,683	5,903
Gini	0,339	0,339	0,341	0,341	0,341	0,340	0,344
Relaciones							
Ratio M/m	79,24%	79,61%	79,99%	80,36%	80,63%	80,66%	80,83%
Tasa variación anual (M)		5,51%	1,37%	0,64%	1,13%	-1,97%	0,92%
Tasa variación anual (qsr)		5,02%	0,89%	0,17%	0,79%	-2,00%	0,71%
Tasa variación anual (Gini)		1,17%	2,56%	0,74%	1,03%	0,11%	3,87%

Fuente: elaboración propia.

La tabla anterior muestra los valores reales de algunos indicadores relacionados con la forma de la distribución de densidad de los datos reales. Como se puede ver:

1. La mediana y su relación con la media evolucionan de una manera muy estable en el tiempo. El ratio M/m se mantiene en el entorno del 80%. Bajo el supuesto (*razonable*) de mantenimiento de dicho valor en el ejercicio a estudio, conocer la media (*dato*) implica tener un valor (*o un intervalo*) de referencia para la mediana.
2. La relación entre el percentil 80 y el 20, $\left(\frac{P_{80}}{P_{20}}\right)$, dada por el indicador *qsr*, vinculado de manera directa a las colas de la distribución, también mantiene un comportamiento estable y permitiría estimar un rango de variación para el ejercicio a estudio.
3. Todo lo dicho antes es extensible a la evolución del Índice de Gini.

A la vista de los anterior se puede concluir que se dispone de tres indicadores de la forma de la distribución salarial —Mediana, *qsr*, Gini— con un comportamiento conocido y estable. La relación entre estas tres magnitudes y los parámetros (a, p, q) es conocida analíticamente (aunque no por ello sea sencilla) de manera que se pueden estimar para una $GB2\,(a,b,p,q)$ los indicadores teóricos. El problema a resolver sería encontrar uno o varios (*pocos*) vectores $(a,b,p,q) \in C$ cuya distribución *GB2* teórica tuviese una esperanza igual a la media (*S)* y que a su vez la Mediana, *qsr* y Gini teóricos fuesen los dados por hipótesis. Se busca por tanto un $(a,b,p,q)_{opt} \in C$ tal que verifique simultáneamente las siguientes condiciones:

a) $S = E(GB2(a,b,p,q)^{opt})$

b) $\widehat{Med} = Med(GB2(a,b,p,q)^{opt})$

c) $\widehat{qsr} = qsr(GB2(a,b,p,q)^{opt})$

d) $\widehat{Gini} = Gini(GB2(a,b,p,q)^{opt})$

donde $\widehat{Med}$, $\widehat{qsr}$, $\widehat{Gini}$ serían valores dados por hipótesis.

La resolución analítica de semejante problema se adivina de una complejidad importante en el supuesto de que fuese abordable por procedimientos analíticos. Como alternativa se propone una resolución por métodos numéricos y en base a un algoritmo secuencial. El algoritmo propuesto es el siguiente:

1. Se resuelve un problema similar al de minimización presentado anteriormente (pero ahora de manera menos restrictiva). Se busca la colección de puntos Q ⊂ C que verifican:

$$\mathrm{Q} = \{(\mathrm{a,b,p,q})/\mathrm{Abs}(\tilde{\mathrm{S}} - \mathrm{S}) \leq \mathrm{K \ tal\ que\ K} > 0\ \mathrm{a,b,p,q} \in \mathrm{C}\}$$

   Q es el subconjunto de C que garantiza distribuciones *BG2* cuya esperanza matemática dista de S (dato) un valor menor que *K* (arbitrario).

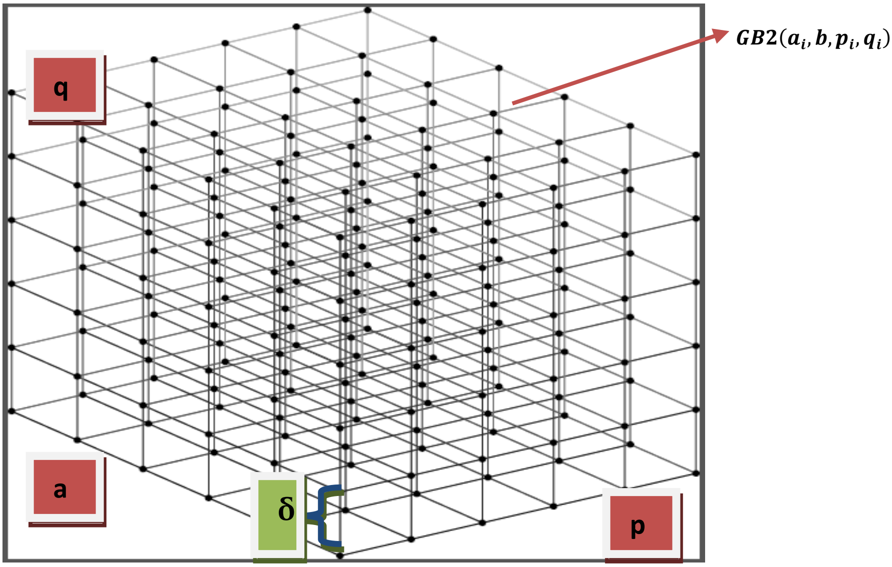
2. Sobre $Q \subset C$ se van a buscar aquellos puntos que proporcionen valores deseados para Mediana, *qsr* y Gini. El procedimiento será secuencial y en el orden anterior:

   a. Se extrae sobre Q el siguiente colectivo:
   $Q1 = \{(a, b, p, q) \subset Q / Abs\left(Med(a, b, p, q) - \widehat{Med}\right) \leq M$
   $tal\ que\ M > 0\}$
   Donde $Med(a, b, p, q)$ es el valor teórico de la Mediana
   y $\widehat{Med}$ el valor hipotético.
   b. Análogamente se extrae sobre $Q1$ el siguiente colectivo:
   $Q2 = \{(a, b, p, q) \subset Q1 / Abs(qsr(a, b, p, q) - \widehat{qsr}) \leq R$
   $tal\ que\ R > 0\}$
   c. Finalmente sobre $Q2$ se extrae $Q3$
   $Q3 = \{(a, b, p, q) \subset Q2 / Abs\left(Gini(a, b, p, q) - \widehat{Gini}\right) \leq G$
   $tal\ que\ G > 0\}$

5.3. Aplicación del algoritmo y comparación con datos reales. Colectivo de Permanentes 2013

El esquema anterior se ha programado en lenguaje *R*, utilizando como apoyo fundamentalmente el paquete *GB2* en su versión 2.1. La estrategia de programación consiste en partir de una malla de puntos (a, b, p, q) definida *ad hoc* y en cada punto computar $GB2(a, b, p, q)$. A partir de ahí se van realizando de forma secuencial los pasos del algoritmo. El paquete *GB2* permite el cálculo de todos los parámetros teóricos necesarios en cada paso.

Figura 2. Representación idealizada del espacio C para un valor *b*=cte



Fuente: elaboración propia.

En base a la información proporcionada por la muestra se pueden estimar los intervalos de variación de las variables (a, b, p, q) esto es $[a_{min}, a_{max}]$, $[b_{min}, b_{max}]$, $[p_{min}, p_{max}]$, $[q_{min}, q_{max}]$ y que define el espacio $\boldsymbol{C}$ de búsqueda. Una vez definidos los límites se debe establecer el granulado de la malla, esto es, la distancia δ que para cada una de ellas existe entre dos valores consecutivos. Por simplicidad tomaremos para *a, p, q* un mismo δ, δ=0,01 y para b, un paso ε=10. Así se estudiaran todas las combinaciones de valores de (a, b, p, q) entre sus límites y con los pasos, δ y ε anteriores. Por ejemplo, una malla construida con los siguientes intervalos y pasos:

a∈[4, 6, by = 0,01]

b∈ [20.000, 24.000, by=10]

p∈ [0,3, 0,5, by =0,01]

q∈[0,3, 0,5, by =0,01]

da lugar a un retículo (el espacio C) con 3.713.661 puntos a estudio, todos ellos posibles candidatos a resolver el problema.

Obsérvese que, si se reduce el paso, la malla sería mucho más tupida con lo que el número de candidatos a estudio se multiplicaría y la complejidad computacional también. Si se toma un paso demasiado grande el problema se resuelve más rápido, al reducirse el número de puntos a evaluar y por ello se corre el riesgo de no encontrar solución. En cualquier caso el *óptimo* del problema planteado va a depender del grado de precisión que se desee (o se pueda) emplear. Los valores extremos de los intervalos se toman a partir de la información de la Tabla 5. Sobre este conjunto de puntos se va a simular una posible solución, situados en 2012 para el ejercicio de 2013. Los únicos datos que se suponen conocidos serían la media, *S*=25.693,24 y el número, *N*=332.217*.* Resolviendo el problema para el ejercicio de 2013, con las siguientes hipótesis:

a) S=25.693,24 (dato conocido de la *MCVL*)
b) K=2,6 euros
c) Mediana $\in$ [20.500,20.800]
d) *qsr* $\in$ [5,60 , 6,0]
e) Gini $\in$ [0,340 , 0,350]

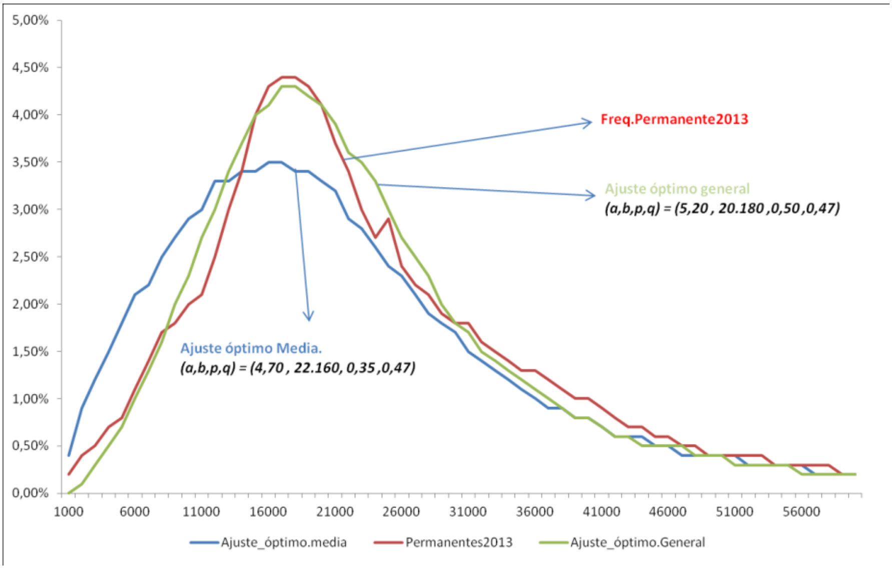
se obtienen, a pesar del valor tan restrictivo de *K*, once posibles candidatos a solución, tal y como aparece en la Tabla 6, cuyos resultados finales tras ser aplicados hemos comprobado que no presentan sensibles diferencias entre ellos. Si, por ejemplo, se toma un valor K=26 euros, el número de candidatos sería, en lugar de once, ochenta y ocho, en principio también todos ellos igual de válidos. Este resultado ilustra el principal problema de este método, la posible sobreabundancia de soluciones, todas factibles y la dificultad para elegir la “mejor”.

Tabla 6. Candidatos óptimos a la solución del problema.

a	b	p	q	media	mediana	qsr	Gini
5,20	**20.180**	**0,50**	**0,47**	**25.693,01**	**20.631,90**	**5,78**	**0,349**
5,90	20.180	0,44	0,41	25.694,11	20.668,12	5,67	0,346
5,90	20.510	0,42	0,41	25.692,36	20.680,12	5,83	0,350
6,00	20.280	0,42	0,40	25.691,75	20.618,65	5,80	0,350
5,30	20.030	0,50	0,46	25.694,75	20.628,58	5,70	0,347
5,40	20.160	0,48	0,45	25.694,79	20.623,60	5,77	0,349
5,80	20.730	0,42	0,42	25.691,45	20.730,00	5,87	0,350
5,90	20.340	0,43	0,41	25.691,32	20.672,62	5,75	0,348
5,80	20.560	0,43	0,42	25.695,45	20.727,15	5,79	0,348
5,40	20.010	0,49	0,45	25.690,99	20.616,13	5,70	0,347
5,60	20.100	0,46	0,43	25.690,67	20.576,96	5,79	0,350

Fuente: elaboración propia.

Gráfico 12. Comparación gráfica de los datos reales con dos modelos de ajuste numérico



Fuente: elaboración propia.

A fin de comparar las ganancias del método aquí presentado con respecto a la minimización pura de la diferencia con la media, presentado como punto de partida de este apartado, se presenta el Gráfico 12. Se recogen en él la distribución real del colectivo de Permanentes (2013) y las frecuencias empíricas obtenidas de los dos métodos. Como se puede ver, el ajuste únicamente condicionado a la media (*Ajuste_óptimo.media*) no recoge la forma de la distribución real, la curva empírica es mucho menos apuntada y el error en el intervalo modal es grande. Si se aplica el segundo método, a modo de ejemplo y sin pérdida de generalidad, sobre el primer candidato obtenido (el sombreado en amarillo en la Tabla 6), el resultado es muy distinto. Al haber forzado un ajuste condicionado a la media y a la vez a otras medidas dependientes directamente de los parámetros de forma (Mediana, *qsr*, Gini) el resultado gráfico es mucho mejor. Si sobre esos modelos teóricos se extraen dos realizaciones aleatorias con *N*=332.217 elementos se pueden calcular y comparar con los valores reales, tanto la masa salarial total del colectivo como el total de retenciones. Téngase en cuenta que estos dos resultados (para *Ajuste_óptimo.media y Ajuste_óptimo.general)* son variables aleatorias al ser una realización concreta de sus respectivas $GB2(a, b, p, q)$ teóricas. El resultado se muestra en la tabla siguiente.

Tabla 7. Retribuciones y Retenciones totales para Permanentes2013 y los dos métodos numéricos

	Permanentes 2013	Ajuste óptimo_media	Ajuste óptimo_general
Retribuciones	8.535.731.807,10	8.597.629.868,00	8.530.269.999,00
Retenciones	1.534.298.544,20	1.747.969.570,00	1.593.624.129,00
Variación Ajuste sobre Permantes 2013. Retribuciones		0,7252%	-0,0640%
Variación Ajuste sobre Permantes 2013. Retenciones		13,9263%	3,8666%

Fuente: elaboración propia.

Como se puede constatar, ambos modelos ajustan casi a la perfección la masa salarial total, lo que era de esperar dado que ambos ajustes buscan minimizar la distancia $Abs(\tilde{S} - S)$. De hecho, el *Ajuste_óptimo.general* es casi perfecto y mejor que el proporcionado por el *Ajuste_óptimo.media.* Sin embargo al aplicarle a ambas el esquema de Retenciones las diferencias son muy notables; el que *Ajuste_óptimo.general* recoja mucho mejor la forma general de la distribución empírica (*Permanentes-2013*) sobre todo en los tramos centrales de la distribución, donde se acumulan la moda, hace que la estimación de la recaudación por Retenciones sea mucho más ajustada.

6. Conclusiones

El presente documento aborda la falta de información sobre el tiempo de trabajo existente en las fuentes estadísticas fiscales en materia de retribuciones salariales. El problema que se plantea es que sin esta información no es posible abordar un correcto análisis de la distribución personal de las rentas salariales. Se plantea entonces la utilización de la Muestra Continua de Vidas Laborales (*MCVL*) como marco adecuado donde abordar la solución al problema anterior, puesto que en ella se integran conjuntamente tanto información fiscal como laboral relativa a los individuos que la forman. El conocimiento detallado del tiempo de trabajo permite explicar de manera satisfactoria la forma que adopta la distribución de las rentas salariales anuales. Abundando en la misma línea, el tiempo de trabajo permite una estratificación muy fructífera de la población de asalariados. Conseguir una adecuada modelización de las retribuciones salariales es la siguiente cuestión que se aborda en este documento. Se plantea aquí un ajuste de estas retribuciones en términos de la distribución Beta Generalizada de Segunda Especie (*GB2*), distribución de cuatro parámetros que se muestra especialmente útil este fin.

Conocida la distribución de probabilidad se ajustan los salarios anuales y se dispone de una herramienta útil que permite abordar el análisis de las Retenciones a cuenta del IRPF, su simulación y su predicción. Poder estimar correctamente la recaudación por Retenciones de IRPF es de suma importancia para presupuestar correctamente las partidas de gasto del presupuesto, dada la importancia cuantitativa que el Impuesto sobre la Renta tiene para conjunto del sistema impositivo. Por otra parte, este trabajo es un ejemplo de la riqueza de análisis existente en la combinación de fuentes estadísticas distintas (MCVL en su combinación de datos de seguridad social y de datos fiscales), pero complementarias entre sí.

Fecha de recepción del artículo: 24 de abril de 2016
Fecha de aceptación definitiva: 21 de junio de 2016

7. Bibliografía

[1] García, C.; F.J. Callialta y J.J. Núñez (2006): “La Evolución de la Distribución Personal de la Renta en España (1973-2000) a través de los Parámetros del Modelo de Dagum”. *El Trimestre Económico*, 292. Págs. 783-808.

[2] Graf, M. y D. Nedyalkova (2014): “Modeling of Income and Indicators of Poverty and Social Exclusion Using the Generalized Beta Distribution of the Second Kind”. *Review of Income and Wealth*. Vol. 60, Issue 4. Págs. 821-842.

[3] Graf, M. y D. Nedyalkova (2011): “Parametric Estimation of Income Distributions and Indicators of Poverty and Social Exclusion”. https://www.unitrier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP2-D2.1-20110409.pdf

[4] Graf, M. y D. Nedyalkova (2015): “GB2: Generalized Beta Distribution of the Second Kind: Properties, Likelihood, Estimation.” https://cran.r-project.org/web/packages/GB2/index.html

[5] Kleiber, C. y S. Kotz (2003): *Statistical size distributions in economics and actuarial sciences*. Wiley series in Probability and Statistics.

[6] McDonald, J.B. y A. Mantrala. (1995): “The distribution of income: Revisited”. *Journal of Applied Econometrics*, 10. Págs 201-204.

[7] Melis Maynar, F. (1995): *La distribución personal del salario anual en 1992*. Papeles de trabajo del Instituto de Estudios Fiscales, nº 16/95.

[8] Prieto Alaiz, M. y C. García Pérez (2006): *Modelización paramétrica de la distribución personal de la renta en España. Una aproximación a partir de la distribución beta generalizada de segunda especie*. IEF. P.T. N 21/07.

[9] Valverde Caramés, P. (2011): “La distribución personal de los salarios y su relación con el tiempo de trabajo”. *Economía Española y protección Social*, nº 3. Págs. 5-35.

Apéndice 1: Estimación del error cometido en el cálculo de Retenciones usando la *MCVL*. Fuentes de error

Un hecho a tener en cuenta son los errores que en la estimación del cálculo de retenciones genera la aplicación de la *Función de retención (Fr)* sobre los resultados teóricos obtenidos en base al análisis anterior. Hay que tener en cuenta que el árbol de decisión utilizado para su cálculo es muy complejo y depende del salario total del trabajador, de cómo se estructura éste (si lo cobra de un único pagador o de varios), de la situación personal (casado o no, hijos, ascendiente a cargo etc.) y más. Por otra parte, no todas estas variables son fáciles de manejar usando la *MCVL* y por ello no han sido tenidas todas en cuenta, y en lo referente a la estructura del salario total anual del trabajador, a efectos de todo lo anterior, se ha supuesto un único retenedor. Es decir, se aplica *Fr* al total del salario anual percibido. Todo lo anterior hace necesario estimar qué error se comete por el uso de una *Fr* simplificada respecto a la realmente aplicada. El esquema de la *Fr* usado, es el siguiente (para el ejercicio 2013):

MINORACIONES

1. Minoración por Seguridad Social

2. Minoración por reducción (Artículo 18.2)

MINIMO PERSONAL

Mínimo personal=6000

ESCALA DE RETENCION

Escala del ejercicio 2013

CUOTAS

Dado que la MCVL ofrece los datos salariales y las retenciones **realmente** aplicadas a cada trabajador, se puede proceder como sigue:

Se definen las siguientes magnitudes:

R=Retenciones reales del colectivo a estudio.

Γ=Retenciones calculadas

S=Salarios totales reales (anuales)

Š=Salarios anuales simulados (obtenidos de Ajuste_óptimo.general)

donde:

$R = Y(S)$ donde Y es el *algoritmo completo* para el cálculo de retenciones, $\Gamma = \breve{Y}(\check{S})$ donde $\breve{Y}$ es la *versión reducida* del algoritmo Y ($Y \subset \breve{Y}$) la *Fr* simplificada. Las retenciones calculadas se obtienen al aplicar un algoritmo de retenciones reducido a los salarios prevenientes de un ajuste dado por una *GB2*.

Sea *Abs(R—Γ)*≡error cometido en la estimación de las retenciones reales, esto es:

$$\varepsilon = |R - \Gamma| = |Y(S) - \breve{Y}(\check{S})|$$

Sea $\breve{Y}(S)$ el error que se comete cuando se usa el algoritmo de retenciones *reducido* sobre el vector los salarios reales (*conocido*). Sumando y restando en la expresión anterior:

$$\varepsilon = |Y(S) - \breve{Y}(\check{S}) + \breve{Y}(S) - \breve{Y}(S)|$$

Reordenando los factores:

$$\varepsilon = |\left(Y(S) - \breve{Y}(S)\right) + \left(\breve{Y}(S) - \breve{Y}(\check{S})\right)| = |\varepsilon_d + \delta| \leq |\varepsilon_d| + |\delta|$$

El error total cometido se puede por tanto separar en dos componentes:

a) Error determinista (ε_d). Debido al uso del algoritmo reducido sobre los salarios reales.

b) Error aleatorio (ε). Debido al uso de salarios teóricos sobre el algoritmo reducido.

$$\boldsymbol{\varepsilon} \leq \left|\boldsymbol{\varepsilon}_{\mathbf{d}}\right| + \left|\boldsymbol{\delta}\right|$$

Tabla 8. Estimación del peso de los errores cometidos en el cálculo de la Retención

	Retribuciones	**Retenciones**
Permanentes 2013	8.535.731.807,10	1.534.298.544,20
Ajuste Óptimo General	8.530.269.999	1.593.624.129
E	59.325.584,80	
ε_d	41.859.709,80 70,56%	
Δ	17.465.875 29,44%	

Por tanto, se puede concluir que la mayor parte del error cometido en la estimación del montante total de las retenciones del colectivo Permanentes.2013 se debe a que no se aplica el algoritmo de retenciones exacto, sino una aproximación. De hecho el 70% del error cometido es imputable a esa causa y menos de un 30% al hecho de reemplazar los salarios reales por los obtenidos de *Ajuste_óptimo.general.*

Apéndice 2: Orden y recaudación

En los apartados anteriores se hizo especial énfasis en que la “forma” de la distribución de los salarios (o de las rentas en general), cuando se le aplica un impuesto progresivo, era algo determinante. De hecho, esa búsqueda de la “forma” adecuada era la clave de todo el análisis. Veamos por qué esto es así. Sea Y=$\{y_1, y_{2,} y_{3,} \dots y_N\}$ una colección de N rentas, $y_{i,} > 0$, $i_{=1,2,\dots,N}$ verificándose que $S = \sum_{i=1}^{N}(y_i)$ es la renta total de ese colectivo. Por simplicidad se supone que y_i es un número real positivo. Sea t, por hipótesis, una función de la renta, real, continua, derivable y creciente.

t:R$^+$ $\longrightarrow$ [0,1]/ t es la función de tipos marginales sobre la renta Y.

Para cada nivel de renta de Y, y_i, se verifica que la recaudación obtenida al aplicar t viene dada por:

$$T(y_i) = \int_0^{y_i} t(x)dx$$

La recaudación total obtenida de la aplicación de t al conjunto Y, T, viene dada por:

$$T(y_1 \dots y_N) = \sum_{i=1}^{N} T(y_i) = \sum_{i=1}^{N} \int_0^{y_i} t(x)dx$$

Llegados a este punto se puede plantear si existe algún conjunto de N rentas que sumando S, es decir, *conservando la renta total*, proporcione un valor óptimo (máximo y/o mínimo) para T. Se plantea así un problema de optimización sujeto a una restricción lineal. Este problema admite solución en los términos de un problema de optimización de Lagrange.

$$\left.\begin{array}{l} \text{Max } T = \sum_{i=1}^{N} \int_0^{y_i} t(x)dx \\ \text{Sujeto a} \\ S = \sum_{i=1}^{N} (y_i) \end{array}\right\} \quad \text{Problema de optimización de Lagrange}$$

Formando el lagrangiano L, se tiene:

$$L(y_{1,} \dots y_N; \lambda) = \left[\sum_{i=1}^{N} \int_0^{y_i} t(x)dx\right] - \lambda\left(\left(\sum_{i=1}^{N} (y_i)\right) - S\right)$$

Expresión de la función lagrangiana.

$\boxed{\frac{\delta L}{\delta y_i} = 0, i_{=1,2,\dots,N}}$ Ecuaciones de Lagrange.

$$\frac{\delta L}{\delta y_i} = \frac{\delta}{\delta y_i}\left[\sum_{i=1}^{N} \int_0^{y_i} t(x)dx\right] - \frac{\delta}{\delta y_i} \lambda\left(\left(\sum_{i=1}^{N} (y_i)\right) - S\right)$$

Aplicando el *Teorema Fundamental del Cálculo* se tiene:

$$\frac{\delta L}{\delta y_i} = \sum_{i=1}^{N} \frac{\delta}{\delta y_i} \int_0^{y_i} t(x)dx - \lambda = \frac{\delta}{\delta y_i} \int_0^{y_i} t(x)dx - \lambda = t(y_i) - \lambda = 0$$

$$\lambda = t(y_i) \;\; i = 1,2 \dots, N$$

De lo anterior se deduce que:

$$\boldsymbol{\lambda} = t(y_1) = t(y_2) \ = \cdots = t(y_N)$$

Para la función t, en los términos de las hipótesis del problema, la relación anterior se verifica si se está evaluando a t en el mismo punto, esto es:

$$y_1 = y_2 = \cdots = y_N = y.$$

Aplicado lo anterior sobre la restricción lineal:

$$S = \sum_{i=1}^{N}(y_i) = Ny \rightarrow \boldsymbol{y} = \frac{\boldsymbol{S}}{\boldsymbol{N}} = \overline{\boldsymbol{y}}$$

Así pues, existe un candidato a óptimo en el punto $\{\overline{y} \ldots . \overline{y}\}$. Aplicando lo anterior.

$$T_{op} = \sum_{i=1}^{N} t(y_i) = \sum_{i=1}^{N} t\left(\frac{\boldsymbol{S}}{\boldsymbol{N}}\right) = Nt\left(\frac{\boldsymbol{S}}{\boldsymbol{N}}\right) = Nt(\overline{\boldsymbol{y}})$$

$$T_{op} = Nt(\overline{y})$$

Pasando a los criterios de segunda derivada para el lagrangiano, L, se tiene:

$$\left.\begin{array}{l} \frac{\partial}{\partial y_j}\left(\frac{\delta L}{\delta y_i}\right) = \frac{\partial}{\partial y_j}(t(y_i) - \lambda) = 0, \text{ para todo } j \neq i \\ \frac{\partial}{\partial y_i}\left(\frac{\delta L}{\delta y_i}\right) = \frac{\partial}{\partial y_i}(t(y_i) - \lambda)) = \frac{d}{dy_i}t(y_i) = t'(y_i) \end{array}\right\}$$

De aquí se concluye que la matriz Hessiana tendrá la siguiente forma:

$$H=\begin{bmatrix} t'(y_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t'(y_N) \end{bmatrix}$$

El carácter de la matriz H, aplicada en el punto óptimo, dependerá del valor de $t'(\overline{y})$. Como por hipótesis t es una función creciente se tiene que $t'(\overline{y}) >$0 y con lo que todos los menores principales serán positivos. Ello implica que H es definida positiva y que $Y = \{\overline{\mathbf{y}} \ldots . \overline{\mathbf{y}}\}$ sería un mínimo para la recaudación total. *La recaudación mínima se da cuando la distribución de rentas es absolutamente equitativa*. Obsérvese que no existe una única combinación Y=$\{y_1, y_{2,} y_{3,} \ldots y_N\}$ que proporcione un máximo de recaudación, ya que cualquiera de ellas en las que N-1 unidades tengan la mínima renta posible (por hipótesis, no hay rentas nulas $y_{i,} > 0,\ i_{=1,2,\ldots,N}$) y la restante acumule el resto sería un máximo de recaudación. El caso ideal sería que toda la renta S se acumulase en un único individuo, lo que daría lugar al máximo de recaudación. *La máxima recaudación se consigue con la distribución de renta menos equitativa*. En conclusión, la forma de la distribución de las rentas es determinante de la recaudación obtenida cuando se aplica sobre ella un impuesto progresivo.