Modelos probabilísticos de la distribución de la renta

Por

Doctor EUGENIO PRIETO PEREZ

Catedrático de la Universidad Autónoma de Madrid

Los estudios de naturaleza económico-estocástica, referentes a la distribución de la renta, se ocupan fundamentalmente de medir la desigualdad de ésta, por lo que respecta a las rentas personales.

Entre las medidas de la desigualdad de la renta cabe destacar las siguientes:

I. INDICE DE GINI. Sean $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n$ las rentas personales correspondientes a los individuos de un cierto colectivo. El índice de Gini es el cociente:

$$G=\frac{\Delta}{2a}$$

en donde:

$$a=\frac{1}{n}\sum_{i=1}^{n} x_i$$

esto es, la media aritmética de las rentas del colectivo, y para Δ es la media de las $\frac{n(n-1)}{2}$ diferencias positivas que puedan formarse con las rentas personales del colectivo. En consecuencia, tendremos:

$$\Delta=\frac{\sum\limits_{r\neq s}(x_r-x_s)}{\frac{n(n-1)}{2}}$$

y esto porque $1+2+3++\cdots/(n-1)=\frac{n(n-1)}{2}$.

Casos particulares

1. La concentración mínima tendrá lugar cuando sean

$$x_1=x_2=\cdots=x_n$$

En este supuesto se anula Δ y, naturalmente, es:

$$G=0$$

2. Por el contrario, la concentración será máxima cuando se verifique:

$$x_1=x_2=\cdots=x_{n-1}=0 \qquad x_n=\sum_{i=1}^{n} x_i$$

Tenemos:

$$\Delta=\frac{2}{n(n-1)}\cdot(n-1)\sum_{i=1}^{n} x_i=2\,\frac{\sum_{i}^{n} x_i}{n}=2a$$

y

$$G=\frac{2a}{2a}=1$$

Nótese, pues, que $0<G<1$.

Curva de Lorenz

La representación gráfica más usual de la medida de la distribución de la renta es la denominada *curva de Lorenz*. Se llega a ella del siguiente modo:

Sea el rectángulo OABC de la figura adjunta. La base del mismo es $\overline{OA}$, de medida n unidades, a altura $\overline{OC}$, tiene como medida:

$$\overline{OC}+x_1+x_2+\cdots+x_n.$$

Consideremos los siguientes puntos:

$$(1, u), (2, u_2), \cdots (n, u_n)$$

en donde las $u_i (i=1, 2, ..., n)$ vienen definidas por las igualdades:

$$\begin{aligned} u_1 &= x_1 \\ u_2 &= x_1 + x_2 \\ u_3 &= x_1 + x_2 + x_3 \\ &\cdots\cdots\cdots \\ u_n &= x_1 + x_2 + \cdots + x_n \end{aligned}$$

Estos puntos determinan la curva de Lorenz.

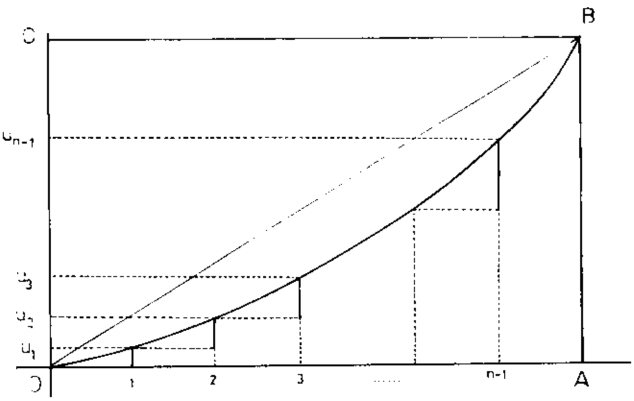


Figura 1

En el caso particular 1. de *mínima concentración*, la curva de Lorenz resulta ser la diagonal principal, OB, del rectángulo $\overline{\text{OABC}}$, pues, al ser en este caso

$$x_1 = x_2 = \cdots = x_n = x$$

tendremos, para $u_h (h=1, 2, ..., n)$:

$$u_h = hx$$

Por otro lado, dado que

$$\Delta u_i = u_{i+1} - u_i = x_{i+1}$$

y como

$$x_i \leqslant x_{i+1} \qquad \forall i = 1, 2, ...n-1$$

la curva de Lorenz es cóncava respecto de la parte positiva del eje de las YY, esto, claro está, en el supuesto de no-equidistribución.

II. INDICE DE LORENZ. Se define a partir de la curva de Lorenz, así:

$$G_L=\frac{(a-u_1)+(2a-u_2)+\cdots+(\overline{n-1}a-u_{n-1})}{a+2a+\cdots+\overline{n-1}a}=$$

$$=1-\frac{\sum\limits_{1}^{n-r}}{a(1+2+\cdots+\overline{n-1})}=$$

$$=1-\frac{\sum\limits_{1}^{n-1} u_r}{a\dfrac{n(n-1)}{2}}=1-\frac{2}{n-1}\,\frac{\sum\limits_{1}^{n-1} u_r}{a\cdot n}=$$

$$=1-\frac{2}{n-1}\cdot\frac{\sum\limits_{1}^{n-1} u_r}{u_n},\quad \text{pues}\quad u_n=an=\sum_{1}^{n} z_i$$

En el supuesto de *concentración máxima* se obtiene $G_L=1$. En efecto, al ser

$$x_1=x_2=\cdots=x_{n-1}=0 \quad\Rightarrow\quad \sum_{1}^{n-1} u_r=0 \qquad G_L=1$$

La concentración mínima conduce a $G_L=0$. Pues,

$$\sum_{r=1}^{n-1} u_r=x+2x+\cdots+(n-1)\,x=\frac{n(n-1)}{2}\,x \qquad \text{y} \qquad u_n=nx$$

Por tanto,

$$G_L=1-\frac{2}{n-1}\,\frac{\sum\limits_{1}^{n-1} u_r}{u_n}=1-\frac{2\dfrac{n(n-1)}{2}x}{(n-1)\,nx}=0$$

III. DISTRIBUCIÓN DE PARETO (*). A finales del siglo XIX formuló Pareto la ley de la distribución de la renta que lleva su nombre. Sucesor

(*) Nótese que los índices de Gini y Lorentz son utilizables para estudiar la distribución de las rentas personales de cualquier colectivo, pues traducen, dando una medida, la concentración de las rentas en tal colectivo. La distribución de Pareto, sin embargo, al ser éste un modelo probabilístico, puede darse el caso que resulte inaplicable en ciertos colectivos, es decir, su ajuste puede no resultar adecuado.

de L. Walras en la Cátedra de Economía Política de Lausanne, tras la observación de la distribución de la renta en distintos países, formuló para éstas la siguiente ley:

$$P(x) = Pr(\xi \geqslant x) = Ax^{-\alpha} \qquad [1]$$

en donde ξ es una variable aleatoria asociada a la renta personal que puede corresponder a un individuo cualquiera perteneciente a un cierto colectivo. $P(x)$ representa, pues, la probabilidad de encontrar rentas superiores a x, que si el colectivo sigue la ley de Pareto es de la forma: $Ax^{-\alpha}$. A una variable ξ distribuida según el modelo de Pareto le corresponde una función de distribución:

$$F(x) = P(\xi \leqslant x) = 1 - Ax^{-\alpha} \qquad [2]$$

y una función de densidad

$$f(x) = F'(x) = A\alpha x^{-(\alpha+1)} \qquad [3]$$

Si el recorrido de x fuera $(x_0, +\infty)$, es decir, si x_0 es el valor inferior de las rentas que figuran en el colectivo que sigue la ley de Pareto, tendríamos:

$$P(\xi \geqslant x_0) = 1 = Ax_0^{-\alpha} \quad \Rightarrow \quad A = x_0^{\alpha}$$

Llevando este valor del parámetro A a [2] y [3], resulta:

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\alpha} \qquad \forall x \geqslant x_0$$

$$f(x) = \alpha x_0^{\alpha} \cdot x^{-(\alpha+1)} = \frac{\alpha}{x_0}\left(\frac{x}{x_0}\right)^{-(\alpha+1)}$$

Luego, en el supuesto de que sea fijado el extremo inferior, x_0, la distribución truncada de Pareto depende solamente del parámetro α.

Media y varianza de la distribución de Pareto

a) Media:

$$E(\xi) = \int_{x_0}^{\infty} xf(x)\,dx = \int_{x_0}^{\infty} x \cdot \alpha x_0^{\alpha} x^{-(\alpha+1)}\,dx = \alpha x_0^{\alpha} \int_{x_0}^{\infty} x^{-\alpha}\,dx$$

$$= -\frac{\alpha x_0^{\alpha}}{\alpha - 1}\left[x^{-(\alpha+1)}\right]_{x_0}^{\infty} = \frac{\alpha x_0}{\alpha - 1}$$

expresión esta que sólo tiene sentido económico para $\alpha > 1$.

b) Varianza.

$$D^2(\xi)=\int_{x_0}^{\infty} x^2 \cdot f(x)\,dx-\left(\frac{\alpha x_0}{\alpha-1}\right)^2=$$

$$=\int_{x_0}^{\infty} x^2 \cdot \alpha x_0^{\alpha} \cdot x^{-(\alpha+1)}\,dx-\left(\frac{\alpha x_0}{\alpha-1}\right)^2=$$

$$=\alpha x_0^{\alpha}\int_{x_0}^{\infty} x^{-(\alpha-1)}\,dx-\left(\frac{\alpha x_0}{\alpha-1}\right)^2=-\frac{\alpha x_0^{\alpha}}{\alpha-2}\Big[x^{-(\alpha-2)}\Big]_{x_0}^{\infty}-$$

$$-\left(\frac{\alpha x_0}{\alpha-1}\right)^2=\frac{\alpha x_0^{\alpha}}{\alpha-2}\cdot x_0^{-(\alpha-2)}-\left(\frac{\alpha x_0}{\alpha-1}\right)^2=$$

$$=\frac{\alpha}{\alpha-2}x_0^2-\left(\frac{\alpha x_0}{\alpha-1}\right)^2$$

Métodos de estimación del parámetro α

Dado que la distribución truncada de Pareto sólo depende del parámetro α, pasamos a su estimación.

a) Método de los momentos. Aplicando el método de los momentos de K. Pearson y denotando por $\bar{x}$ a la media muestral, resulta:

$$\bar{x}=\frac{\alpha x_0}{\alpha-1} \Leftrightarrow \bar{x}\alpha-\bar{x}=\alpha x_0 \Leftrightarrow \alpha=\frac{\bar{x}}{\bar{x}-x_0}.$$

b) Método de la máxima verosimilitud. La función de verosimilitud $L(x_1, x_2, \ldots, x_n, \alpha)$ viene dada por:

$$L(x_1,\ x_2,\ \ldots,\ x_n,\ \alpha)=\prod_{i=1}^{n}\alpha x_0^{\alpha}x_i^{-(\alpha+1)}$$

y tomando logaritmos, resulta:

$$\log L=\sum_{1}^{n}\Big[\log\alpha+\alpha\ \log x_0-(\alpha+1)\ \log x_i\Big]$$

El valor de α que hace máximo L, aparece al exigir que sea

$$\frac{d\ \log L}{d\alpha}=0.$$

$$\frac{d \log L}{d\alpha}=\sum_{1}^{n}\left[\frac{1}{\alpha}+\log x_0-\log x_i\right]=0 \quad \Leftrightarrow$$

$$\Leftrightarrow\Leftrightarrow \quad \frac{n}{\alpha}=-\sum_{1}^{n}\log\frac{x_0}{x_i}=\log\prod_{i=1}^{n}\left(\frac{x_0}{x_i}\right) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{1}{\alpha}=-\log\sqrt[n]{\prod_{1}^{n}\left(\frac{x_0}{x_i}\right)}=\log\sqrt[n]{\prod_{1}^{n}\left(\frac{x_i}{x_0}\right)}$$

y

$$\alpha=\frac{1}{\log\sqrt[n]{\prod_{1}^{n}\left(\frac{x_i}{x_0}\right)}}$$

Interpretación económica del parámetro α

A efectos de llevar a cabo la interpretación económica del parámetro α, hagamos:

$$\log P(x)=\log A-\alpha\ \log X \qquad [4]$$

Por tanto, la representación de [4] en escalas logarítmicas es una recta de coeficiente angular $(-\alpha)$.

Por otro lado, si denominamos m a la renta media del colectivo, resulta:

$$m=\frac{\alpha x_0}{\alpha-1} \quad \Leftrightarrow \quad \alpha=\frac{m}{m-x_0} \qquad [5]$$

Entonces, se observa en [5] que al aumentar la diferencia $m-x_0$, es decir, la diferencia entre la renta media (m) y la renta mínima (x_0), el parámetro α disminuye y viceversa. Por ello, se puede interpretar a α como *una medida de la distribución de la renta*, la que será tanto más justa cuanto menor sea la diferencia $m-x_0$ y, por tanto, cuanto mayor sea el valor del parámetro α.

Obsérvese, asimismo, que α *es la elasticidad de la función de distribución de la renta*. En efecto:

$$E=\frac{x}{y}\cdot\frac{dy}{dx}=-\alpha \qquad [6]$$

en donde,

$$y = P(x) = Ax^{-\alpha}$$

La [6] puede escribirse así:

$$\frac{dy}{y} = -\alpha \frac{dx}{x} \qquad [7]$$

La [7] indica que *el decrecimiento relativo del número de personas* (y) *con ingresos iguales o superiores a* x_0 *es proporcional al incremento relativo de la renta. Además, si para rentas* $x_1 < x_2$ *tomamos variaciones iguales* $dx_1 = dx_2$, las variaciones relativas serán:

$$\frac{dx_1}{x_1} > \frac{dx_2}{x_2}$$

Entonces, la ley de Pareto puede formularse así:

El decrecimiento relativo del número de personas (*) *a medida que la renta aumenta es cada vez más pequeño y la disminución de aquél es proporcional al nivel absoluto de la renta, o sea,*

$$\frac{dy}{y} = \alpha \frac{dx}{x}$$

o también, *a medida que aumenta el nivel de la renta personal es relativamente más fácil pasar a un nivel superior de renta.*

Resulta, en consecuencia, que si fuera:

$$\frac{dx}{x} = c \quad \Rightarrow \quad \frac{dy}{y} = -\alpha c$$

o sea, *a variaciones relativas de la renta, constantes, le corresponden variaciones relativas del número de rentistas proporcionales a* α. Por tanto, si el valor de α es *grande*, a un porcentaje de incremento relativo de la renta corresponde un decrecimiento porcentual grande de rentistas. Luego

(*) Oscar Lauge utiliza para este fenómeno la palabra *tamizado* y compara oportunamente la situación con la de los exámenes en los varios años de estudio. «El relativamente mayor tamizado —dice Lauge— durante los exámenes ocurre durante el primer año de estudio, es menor durante el segundo y así sucesivamente.» (Véase Oscar Lauge: «Introducción a la Econometría». Fondo de Cultura Económica).

cuanto mayor es α más elevado es el porcentaje de rentistas que quedan eliminados al aumentar la renta.

La ley de Pareto puede decirse que fue la primera investigación econométrica, hasta el extremo que la curva de Pareto puede ser considerada como el valor medio del número de individuos que van adquiriendo rentas personales más elevadas y no representando un número exacto de personas con ingresos no inferiores a x.

Interpretación sociológica de la ley de Pareto

Varias son las interpretaciones sociológicas que, sobre todo en la literatura económica, fueron dadas a la ley de Pareto. Destacaremos, siguiendo a Oscar Lauge, las siguientes:

I. La primera interpretación que parece lógica exponer es la dada por el propio Wilfredo Pareto (*), quien *creyó encontrar una lev sociológica general que consideró como la ley natural que tendría validez en todo tiempo y en toda sociedad.* Como consecuencia de esta interpretación, las diferentes reformas sociales encaminadas a eliminar la desigualdad en la distribución de la renta estarían destinadas al fracaso de antemano, *ya que la ley de la naturaleza sobre la distribución de la renta actuaría en cualesquiera condiciones , y dicha distribución de la renta tomaría siempre la forma indicada por la fórmula que él estableció»*.

La justificación de esta postura rígida de Pareto está, sin duda, en que de las muchas observaciones para diferentes países y épocas hechas por él, siempre había encontrado valores para α muy próximos a 1,5. Veremos, sin embargo, que la ley de Pareto no es una ley natural y, en consecuencia, sólo puede ser aplicable adecuadamente en ciertos grupos y situaciones históricas.

II. En esta misma línea, Harold Davis (**), economista norteamericano, señala que un valor $\alpha = 1{,}5$ parece indicar un estado de equilibrio deseable, de tal modo que, si el valor de α difiere mucho de 1,5, el equilibrio se ha trastocado. «La sociedad —en opinión de Davis— está algo acostumbrada a la distribución de la renta según la curva de ecuación $y = Ax^{\alpha}$, donde es $\alpha = 1{,}5$. De forma que, si las desigualdades en la distribución del ingreso se incrementan ($\alpha > 1{,}5$), el pueblo empieza a sublevarse y a promover la revolución. Por otra parte, cuando la distribución de la renta se vuelve más equitativa ($\alpha < 1{,}5$), las clases altas luchan por sus propios intereses, por afianzar su dominio, y esto conduce al fascismo o alguna otra forma antidemocrática de gobierno de las clases poseedoras (***).»

(*) Véase V. Pareto: «Cours d'économie politique». Lausanne, 1897.
(**) H. T. Davis: «Theory of Econometrics». Bloomington, 1941.
(***) Véase Oscar Lange, obra citada pág. 155.

III. En no pocas ocasiones, la ley de Pareto se interpretó como significativa de la distribución de la habilidad humana. Así, se emplea frecuentemente como modelo de ajuste para predicciones de resultados de exámenes y pruebas de habilidad en general. En este sentido, quiere decir *que el progreso ulterior en el desarrollo de la eficiencia educativa y física es más fácil para aquellas personas que ya han alcanzado niveles altos.*

Las más modernas investigaciones suelen ser claras al respecto de que el ajuste de la curva de Pareto es deficiente en los ingresos más bajos y superiores. ¿Dónde está, pues, el error de observación de Wilfredo Pareto? Todo parece indicar que el tipo de sociedad estudiado por él se caracterizaba porque sus rentas procedían principalmente de la tierra y del capital.

Los estudios posteriores indican, sin ninguna duda, que el ajuste del modelo paretiano *no es adecuado* cuando se consideran grupos homogéneos y con rentas del trabajo. En estos casos existe una *renta modal* y desviaciones respecto de ésta que obedecen a múltiples causas, casi siempre sin apenas significación. De aquí que tenga sentido el ajuste del modelo de distribución normal o el logarítmico normal que pasamos a estudiar.

IV. DISTRIBUCIÓN LOGARÍTMICO-NORMAL. El modelo de distribución logarítmico-normal surge a partir de la distribución normal, cuando se considera como variable aleatoria el logaritmo de la variable aleatoria normal, es decir:

Sea ξ una variable aleatoria $N(\mu, \sigma)$, al hacer el cambio

$$\eta = \log \xi$$

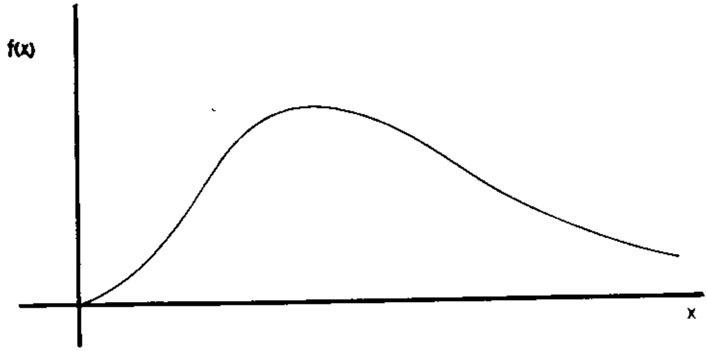
encontramos que las funciones de distribución y densidad de η son:

$$F(y) = \int_0^y f(y)\,dy = \int_0^x \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(\log x - \mu)}{\sigma^2}} \frac{1}{x} dx$$

Luego

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(\log x - \mu)^2}{\sigma^2}} \cdot \frac{1}{x}$$

La representación gráfica es:



Momentos

a) Esperanza matemática

$$E(\eta) = e^{\mu + \frac{1}{2}\sigma^2}$$

b) Varianza.

$$D^2(\eta) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = m^2(e^{\sigma^2} - 1)$$

En efecto, el momento del orden n se calcula así:

$$\alpha_n = \int_0^\infty x^n \frac{1}{x\sqrt{2\pi}\,\sigma} e^{-\frac{(\log x - M)^2}{2\sigma^2}} dx$$

Al hacer $y = lx \Rightarrow x = e^4 \Rightarrow dx = e^y dy$, resulta:

$$\alpha_n = \int_{-\infty}^\infty e^{ny} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-y} \cdot e^{-\frac{(y-k)^2}{2\sigma^2}} e^y dy$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{ny - \frac{(y-k)^2}{2\sigma^2}} \cdot dy =$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{1}{2}\left[\frac{y-k}{\sigma} - n\sigma\right]^2 + n\mu + \frac{n^2\sigma^2}{2}} \cdot dy$$

Haciendo el cambio:

$$z=\frac{y-\mu}{\sigma}-n\sigma \quad \Rightarrow \quad d_z=\frac{1}{\sigma}dy$$

y, en consecuencia:

$$\alpha_n=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2}z^2+n\mu+\frac{n^2\sigma^2}{2}}\sigma d_z=$$

$$=e^{n\mu+\frac{n^2\sigma^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}d_z=e^{n\mu+\frac{n^2\sigma^2}{2}}$$

Para $n=1,\ 2$, tenemos:

$$\alpha_1=E(\eta)=E^{\mu+\frac{\sigma^2}{e}}$$

$$\alpha_2=E(\eta^2)=e^{2\mu+2\sigma^2}$$

Así que:

$$D^2(\xi)=\alpha_2-(\alpha_1)^2=e^{2\mu+2\sigma^2}-e2\left(k+\frac{\sigma^2}{2}\right)=$$

$$=2\mu+\sigma^2\cdot[e^{\sigma^2}-1]=\alpha_1^2[e^{\sigma^2}-1]$$

Análisis de la distribución de la renta en España

A continuacion procedamos a ajustar y discutir los diferentes modelos analizados a los datos empíricos que ofrecía la distribución de la renta en España en 1964, recogido en la tabla n.° I.

El profesor Bernardo Pena (*), hace notar que «las declaraciones de ingresos seleccionados en la muestra presentan un sesgo por defecto, que se pone de manifiesto al relacionar los ingresos declarados con los gastos de consumo correspondientes». Así —sigue diciendo B. Pena— mientras los gastos anuales de consumo medio por hogar han sido de 79.055 pesetas, los ingresos anuales medios declarados resultan ser: 75.139 pesetas.

(*) Véase B. Pena Tampero: «Modelos Econométricos para el estudio de la distribución de la renta en España». Estadística Española, n.° 27. 1965.

Tabla I

Renta mensual Intervalos	Marcas de clase x_i	Porcentaje de familias y_i
0 – 1.800	1.800	9
1.800 – 2.000	1.900	2
2.000 – 2.500	2.250	6
2.500 – 3.000	2.750	7
3.000 – 3.500	3.250	7
3.500 – 4.000	3.750	8
4.000 – 4.500	4.250	7
4.500 – 5.000	4.750	7
5.000 – 6.000	5.500	11
6.000 – 8.000	7.000	15
8.000 – 10.000	9.000	8
10.000 – 12.000	11.000	5
12.000 – 15.000	13.500	4
15.000 – 20.000	17.500	2
20.000 y más	30.000	2

Fuente: Encuesta de Presupuestos familiares, 1964. I. N. E.

Si por otra parte, se tiene en cuenta que, según los datos de la Contabilidad Nacional, el ahorro familiar representaba en 1964 el 10 por 100 del consumo de las economías domésticas, tendremos que los ingresos medios deberían ser 86.900 pesetas. Con ésta, podemos calcular un coeficiente de ocultación en los ingresos declarados y que ha sido del orden del 13,60 por 100.

Cuanto acabamos de decir, indica, que el grado de ocultación —que naturalmente, será mayor para las familias de rentas mayores—, no estan tan exagerados como para invalidar la distribución empírica recogida en la Tabla I. En todo caso, queda el valor didáctico de cuanto realicemos.

Ajuste del modelo de Pareto

Partiendo de $y = Ax^{-\alpha}$, al aplicar logarítmos obtenemos:

$$\log y = \log A - \alpha \log x \qquad [8]$$

Para simplificar haremos:

$$z = \log y \qquad t = \log x \qquad a = \log A$$

La [8], toma la forma:

$$z = a - \alpha t \qquad [9]$$

Los valores de a y α, los estimaremos por mínimos cuadrados, tomando los logaritmos de los valores de la Tabla I. Las operaciones a realizar aparecen en la Tabla II.

Tabla II

x_i	y_i	$\log x_i$	$\log y_i$	$\log x_i \cdot \log y_i$	$(\log x_i)^2$
1.800	100	3,2552	2,0000	6,5105	10,5968
1.900	91	3,2787	1,9590	6,4232	10,7502
2.250	89	3,3522	1,9494	6,5347	11,2371
2.750	83	3,4393	1,9191	6,6003	11,8290
3.250	76	3,5119	1,8808	6,6052	12,3333
3.750	69	3,5740	1,8388	6,5721	12,7737
4.250	61	3,6284	1,7853	6,4779	13,1652
4.750	54	3,6770	1,7324	6,3695	13,5181
5.500	47	3,7404	1,6721	6,2543	13,9903
7.000	36	3,8454	1,5563	5,9841	14,7848
9.000	21	3,9542	1,3222	5,2284	15,6360
11.000	13	4,0414	1,1139	4,4685	16,3329
13.500	8	4,1303	0,9031	3,7301	17,0597
17.500	4	4,2430	0,6020	2,5546	18,0034
30.000	2	4,4771	0,3010	1,3477	20,0446
	1.000	56,1481	22,5356	81,6611	212,0551

Hemos de encontrar a y α para que sea mínimo

$$Q(a, \alpha) = \sum_{i=1}^{n} [Z_i - (a - \alpha t_i)]^2$$

Tenemos:

$$\left.\begin{aligned} \frac{\partial Q(a, \alpha)}{\partial a} &= -2\sum_{1}^{n} [z_i - (a - \alpha t_i)] = 0 \\ \frac{\partial Q(a, \alpha)}{\partial \alpha} &= 2\sum_{1}^{n} [z_i - (a - \alpha t_i)]\, t_i = 0 \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left\{\begin{aligned} \sum_{1}^{n} z_i &= a \cdot n - \alpha \sum_{1}^{n} t_i \\ \sum_{1}^{n} z_i \cdot t_i &= a \sum_{1}^{n} t_i - \alpha \sum_{1}^{n} t_i^2 \end{aligned}\right.$$

En nuestro caso:

$$\left.\begin{array}{l} 22{,}5356 = a \cdot 15 - \alpha \cdot 56{,}1481 \\ 81{,}6611 = a \cdot 56{,}1481 - \alpha \cdot 212{,}0551 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \left\{ \begin{array}{l} a = \dfrac{\begin{vmatrix} 22{,}5356; & -\ 56{,}1481 \\ 81{,}6611; & -212{,}0551 \end{vmatrix}}{\begin{vmatrix} 15 \quad ; & -\cdot 56{,}1481 \\ 56{,}1481; & -212{,}0551 \end{vmatrix}} = 6{,}8657 \\ \\ \alpha = \dfrac{\begin{vmatrix} 15 \quad ; & 22{,}5356 \\ 56{,}1481; & 81{,}6611 \end{vmatrix}}{\begin{vmatrix} 15 \quad ; & -\ 56{,}1481 \\ 56{,}1481; & -212{,}0551 \end{vmatrix}} = 1{,}43 \end{array} \right.$$

La recta ajustada es:

$$z = 6{,}8657 - 1{,}43 \cdot t \Leftrightarrow y = 7341 \cdot x^{-1{,}43}$$

La bondad del ajuste se puede estudiar empleando el contraste de la χ^2; esto es:

$$\chi^2 = \sum_1^n \frac{(f_i - p_i)^2}{p_i} \qquad [10]$$

f_i = porcentaje empírico.
p_i = porcentaje teórico o ajustado.

Fijado un nivel de significación $\alpha = 0{,}05$, las tablas de la χ^2, permiten obtener A, tal que

$$P(\chi^2 > A) = 0{,}05$$

y si el valor obtenido para [10], utilizando los datos contenidos en la tabla III, pertenece al intervalo (A, ∞) rechazamos el ajuste del modelo por deficiente. En otro caso rechazamos la hipótesis.

Como basándonos en la muestra estimamos dos parámetros A y α, χ^2 tiene 12 grados de libertad.

Tabla III

Bondad del Ajuste

Rentas mensuales Marcas de clase	Porcentaje observado f_i	Porcentaje ajustado p_i	$f_i - p_i$	$\frac{(f_i - p_i)^2}{p_i}$
1.800	100	158	58	24,29
1.900	91	146	55	20,72
2.250	89	113	23	4,72
2.750	83	86	3	0,10
3.250	76	68	8	0,94
3.750	69	55	14	3,56
4.250	61	46	15	4,89
4.750	54	39	15	5,77
5.500	47	32	15	7,08
7.000	36	22	13	7,34
9.000	21	16	5	1,56
11.000	13	13	0	0,00
13.500	8	9	1	0,11
17.500	4	6	2	0,67
30.000	2	3	1	0,33
		Total		79,08

Las tablas de χ^2 proporcionan para $n=12$ y $\alpha=0{,}05$, el valor 21,30. Entonces

$$79{,}08 \leqslant (21{,}30, \ \infty)$$

asi que, debemos rechazar la bondad del ajuste.

Ajuste del modelo logarítmico-normal

Tomamos como estimadores de μ y σ^2, los correspondientes momentos muestrales, c y d. Para estos, encontramos:

$$c = 3{,}70$$
$$d = 0{,}39$$

Por lo que respecta a la bondad del ajuste, podemos discutirlo, también empleando la χ^2 como hicimos con el *modelo de Pareto*. Las operaciones encaminadas a obtener el valor de [10], cuando el modelo ajustado es logarítmico normal, aparecen en las tablas IV y V, siguientes.

Tabla IV

x_i	$\log x_i$	f_i	$f_i \cdot \log x_i$	$f_i \cdot (\log x_i)^2$
1.800	3,25	9	29,25	93,04
1.900	3,28	2	6,56	21,52
2.250	3,35	6	20,10	67,32
2.750	3,44	7	24,08	82,81
3.250	3,51	7	24,57	86,24
3.750	3,57	8	28,56	101,92
4.250	3,63	7	25,41	92,26
4.750	3,68	7	25,76	94,78
5.500	3,74	11	41,14	153,89
7.000	3,84	15	57,60	221,25
9.000	3,95	8	31,60	124,80
11.000	4,05	5	20,20	81,60
13.500	4,12	4	16,52	68,24
17.500	4,24	2	8,48	35,96
30.000	4,48	2	8,96	40,14
	Total	100	368,79	1367,97

$$c=\frac{368,79}{100}\sim 3,70$$

$$d^2=\frac{1367,97}{100}3,7^2=13,6797-3,7^2\Leftrightarrow d=0,37$$

Tabla V

Intervalos $\log x_i$	f_i = frecuencias observados	p_i = porcentajes ajustados	$f_i - p_i$	$\frac{(f_i-p_i)^2}{p_i}$
0 – 3,25	12	9	3	0,7
3,25 – 3,30	2	2	0	0,0
3,30 – 3,40	7	6	1	0,1
3,40 – 3,48	6	7	1	0,1
3,48 – 3,54	6	7	1	0,1
3,54 – 3,60	5	8	3	1,8
3,60 – 3,65	5	7	2	0,8
3,65 – 3,70	5	7	2	0,8
3,70 – 3,78	8	11	3	1,1
3,78 – 3,90	16	15	1	0,0
3,90 – 4,00	8	8	0	0,0
4,00 – 4,08	5	5	0	0,0
4,08 – 4,18	5	4	1	0,2
4,18 – 4,30	4	2	2	1,0
4,3 y siguientes	6	2	4	2,6
		Total.........................		8,13

El valor de $\chi^2 = 8{,}13 \in (0,\ A = 21{,}30)$, así que aceptamos la hipótesis.

En resumen, la distribución personal de la renta en España, podemos suponer que sigue la ley logarítmica normal de parámetros

$$\mu = 3{,}7 \qquad \text{y} \qquad \sigma = 0{,}39$$

o sea

$$\log y = -\log 0{,}39\sqrt{2\pi}\,x - \frac{1}{2\times 0{,}39^2}(\log x - 3{,}7) \Leftrightarrow$$

$$\Leftrightarrow y = \frac{1}{x\sqrt{2\pi}\,0{,}39}\, e - \frac{(\log\ y - 3{,}7)^2}{2\times 0{,}39^2}$$