

**UNA MODELIZACION DE LA DISTRIBUCION PERSONAL DE LA RENTA:
EL CASO CASTELLANO-LEONES**

DEL ORO SAEZ, Carlos Pío
MOSQUERA CRIADO, M^a Jesús
TROI TINO COBAS, Anxela

Dpto. de Econometría e Métodos Cuantitativos
Universidad de Santiago de Compostela

1.-INTRODUCCION

De entre las diferentes formas funcionales que desde Pareto (1895) han sido propuestas para describir el fenómeno de la distribución de la renta, el modelo al que, por diferentes vías, han llegado Dagum, Lemmi y Fattorini, satisface un conjunto de propiedades (fundamentos económicos, convergencia a la ley de Pareto, existencia de un número pequeño de momentos finitos, significado económico de los parámetros, buen ajuste en todo el campo de variación de los ingresos,...), que actualmente se consideran esenciales a la hora de identificar una densidad como modelo de distribución de ingresos.

Sin embargo, tal y como sugiere Mandelbrot, tras considerar la observación de Miller acerca de la coexistencia en estos modelos de ingresos procedentes de muy diversas fuentes, es muy difícil que una única teoría pueda explicar todos los rasgos de la distribución de ingresos o que una simple expresión represente a todos los datos.

Este hecho es perfectamente contrastable cuando se pasa se un estudio de la distribución de ingresos familiares a otro concerniente a la distribución de ingresos personal. En general, la primera, es cero o unimodal, mientras que la segunda está habitualmente caracterizada por más de una moda relativa, debido, probablemente, a que el conjunto de ingresos personales es menos homogéneo que el de ingresos familiares.

Una solución en el segundo caso podría surgir de ajustar separadamente el modelo propuesto en partes homogéneas de la muestra en cuanto a distribución o también de representar, por ejemplo, con una parábola las zonas de distribución con aparente normalidad y con una recta los tramos de renta más altos aunque este tipo de solución no tenga ninguna justificación de tipo probabilístico.

Sin embargo, y pese a que este tipo de soluciones ofrece buenos resultados a nivel empírico, no es aconsejable adoptarlas en el campo económico puesto que se trata de un ajuste "a posteriori" de los datos. De hecho, se realiza después de la observación de los datos disponibles y no será a otros casos.

2.-PLANTEAMIENTO TEORICO

A la hora de resolver estos problemas, una propuesta que creemos más interesante y con mayor capacidad interpretativa, es la que resulta de unir la información que aportan las diferentes distribuciones unimodales que se observan en grupos de ingresos homogéneos.

Considerando que los ingresos son función de variables socioeconómicas se podrían obtener grupos homogéneos en relación, al menos, con alguna de estas variables. Una vez, determinados estos grupos, se podría realizar la estimación en cada grupo de la función de distribución de los ingresos y finalmente reconstruir la función de distribución total mediante una combinación lineal de las anteriores que podría venir dada por:

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + \dots + p_k F_k(x) \quad \sum_{i=1}^k p_i = 1$$

siendo F_i la función de distribución ajustada para el grupo homogéneo i -ésimo, k el número de grupos homogéneos de ingresos determinados y p_i la ponderación que se desee establecer. A efectos de este trabajo hemos utilizado como ponderación el porcentaje de individuos del total pertenecientes a cada grupo debido a su practicidad y fácil justificación intuitiva.

La correspondiente función de densidad podrá de esta forma, caso de presentarse en el histograma empírico del conjunto de la muestra, tener una forma con varias modas relativas. Además como los grupos se determinaron en base a consideraciones de tipo económico, los parámetros obtenidos en el ajuste de cada grupo y en la combinación conjunta conservarán su significado económico.

Dentro de otro ámbito de estudio, otra posibilidad de solución para el problema de la multimodalidad surge por medio de la estimación no paramétrica. El campo de la estimación no paramétrica permite abordar el estudio de los datos sin proponer a priori ningún modelo dependiente de parámetros. En general, se estimará directamente la función de densidad a partir de los datos exigiéndole propiedades muy generales como pueden ser continuidad y diferenciabilidad hasta órdenes bajos.

La técnica de estimación no paramétrica utilizada fue la estimación tipo Kernel o núcleo que se basa en la definición generalizada del histograma y propone como estimación de la densidad f de una variable aleatoria X en un punto x la siguiente expresión:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$$

siendo X_1, X_2, \dots, X_n realizaciones de la variable X , k la función Kernel y h el parámetro ventana.

El parámetro de suavización o ventana es el que controla el grado de suavidad o rugosidad de la estimación realizada. La función kernel suele ser en general una función de densidad que se elige en base a criterios de estimación pero que es de importancia pues la estimación realizada hereda las propiedades que tenga esta función. De hecho, en general se le exige que sea acotada, positiva e integre la unidad, propiedades casi universales para las funciones de densidad de una variable aleatoria real.

En base a minimizar el error cuadrático medio hemos tomado como función kernel, el de Epanechnikov, que viene dado por la función,

$$k(a) = \begin{cases} 3(1-a^2)/4 & |a| \leq 1 \\ 0 & |a| > 1 \end{cases}$$

que cumple las propiedades antes mencionadas para el kernel y que es óptimo para dicho criterio. El parámetro de suavización se eligió utilizando el criterio de Cross-Validation óptimo también en el sentido mencionado.

3.- APLICACION EMPIRICA A LOS DATOS DE LA ENCUESTA DE PRESUPUESTOS FAMILIARES: COMUNIDAD DE CASTILLA-LEON

Aunque los datos de la E.P.F. que habitualmente se utilizan en la literatura están referidos en general a los hogares, existe en dicha encuesta bastante información a nivel individual y, a la hora de aplicar los distintos métodos de ajuste para estimar la distribución de los ingresos, se observa un peor comportamiento de los modelos teóricos al tratar los datos individuales que los datos familiares.

De hecho, en el gráfico del polígono de frecuencias observado para los datos de renta familiar por individuo de la comunidad autónoma de Castilla-León del año 1980 se detecta la presencia de una especie de joroba en niveles de ingresos medio-altos que no se refleja en el ajuste de la función de densidad de Dagum realizado para dichos datos.

Considerando que esta joroba podría ser consecuencia de la coexistencia de grupos no homogéneos de ingresos hemos tratado de identificar dichos grupos con respecto a alguna característica socioeconómica eligiendo para ello la "condición profesional". Así, proponemos clasificar a los individuos en "perceptores de ingresos por cuenta ajena", "perceptores de ingresos de capital", "perceptores de ingresos por cuenta propia" y "perceptores de ingresos de transferencias".

Establecidos dichos grupos, se observó que el grupo de individuos que percibían ingresos de capital era en la comunidad de Castilla-León irrelevante en cuanto a su dimensión con respecto a los otros grupos por lo que se asimiló al grupo de ingresos por cuenta propia.

Por otra parte, al realizar los gráficos de los polígonos de frecuencias de los tres grupos constituidos (gráfico 1), resulta que en la comunidad castellano-leonesa el grupo de perceptores de ingresos por cuenta propia y capital presenta una distribución de frecuencias por intervalos muy similar a la del grupo de individuos que perciben ingresos por transferencias. Así que, se decidió formar un grupo mayor, más homogéneo en cuanto a la distribución de los ingresos de sus individuos, formado por los dos grupos citados anteriormente.

No se realizaron más agregaciones de los grupos iniciales puesto que descriptivamente se observa que el polígono de frecuencias correspondiente al grupo de perceptores por cuenta ajena se encuentra desplazado hacia la derecha frente al otro grupo formado indicando este hecho la existencia de dos grupos homogéneos claramente diferenciados.

Posteriormente se realizó el ajuste de la función de Dagum para los dos grupos de individuos (perceptores por cuenta ajena frente al resto) y a partir de estas distribuciones se determinó la conjunta para el total de los individuos tal y como se indicó en el punto anterior (gráfico 2). Dicha estimación se realizó tanto para la función de Dagum de tres parámetros como para la de cuatro, escogiéndose en cada grupo la que presentó un mejor ajuste global. Los cálculos se realizaron utilizando el programa EPID diseñado por C. Dagum y los resultados obtenidos pueden ser facilitados.

La función mistura resultante se adapta mejor a los datos empíricos y muestra mejores resultados que la obtenida realizando el ajuste sin subdividir el total de individuos en grupos homogéneos de ingresos (gráfico 3). Hay que comentar que los resultados obtenidos dependen totalmente de los criterios que se tomen para elaborar los grupos homogéneos de ingresos. Pudiera ser que no existiese una total homogeneidad entre los individuos de cada grupo establecido y ello llevaría a malos ajustes como se puede comprobar en el caso castellano-leonés al realizar el ajuste con tres grupos de individuos. En este sentido existen técnicas que permiten definir de manera objetiva los grupos y asignar a los individuos a los mismos, así como otras propuestas de clasificaciones que en determinados casos han obtenido buenos ajustes.

Para el estudio no paramétrico se partió del total de

los ingresos individuales y para ellos se realizó la estimación de la función de densidad. La ventana obtenida por el método de Cross-Validation fue de 0.05785E6 y el gráfico obtenido se presenta al final del trabajo (gráfico 4).

Es de destacar la buena adaptación de la estimación no paramétrica al gráfico de la densidad de frecuencia por intervalo así como el hecho de que claramente refleja la joroba citada anteriormente. También citar que este tipo de estimación es muy flexible en cuanto a estimación y permite obtener cuantas modas sean observadas descriptivamente en los datos.

La comparación entre el ajuste "mistura" y la estimación no paramétrica (gráfico 5) revela una similar adecuación a los datos con la salvedad de que el ajuste no paramétrico refleja de modo claro la joroba que se aprecia en el histograma. Parece que la pérdida de fundamentos económicos que tiene el utilizar un ajuste no paramétrico permite adecuarse mejor a los datos observados.

4.-BIBLIOGRAFIA

DAGUM, C., LEMMI, A. (1987) A contribution to the analysis of income distribution and income inequality, and a case of study: Italy. Research paper 8706. Dept. Economics, Univ. of Ottawa.

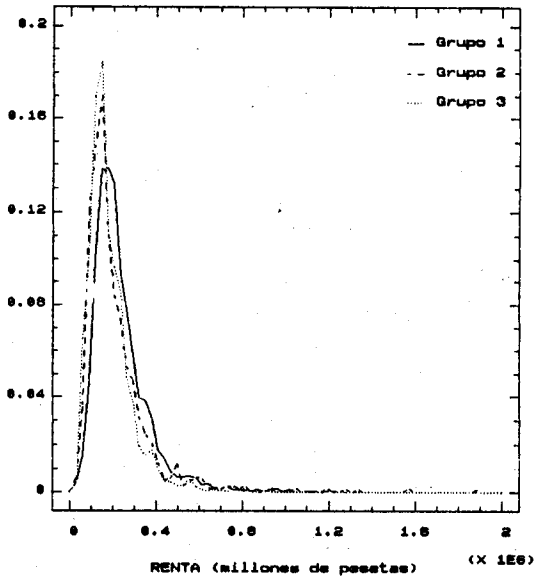
FERER, A. (1988) Comparazioni e costruzioni modellistiche per distribuzioni empiriche di redditi. Tesi di Laurea. Univ. degli studi di Siena.

SILVERMAN B.W. (1988) Density estimation for statistics and data analysis. Chapman and Hall. Londres.

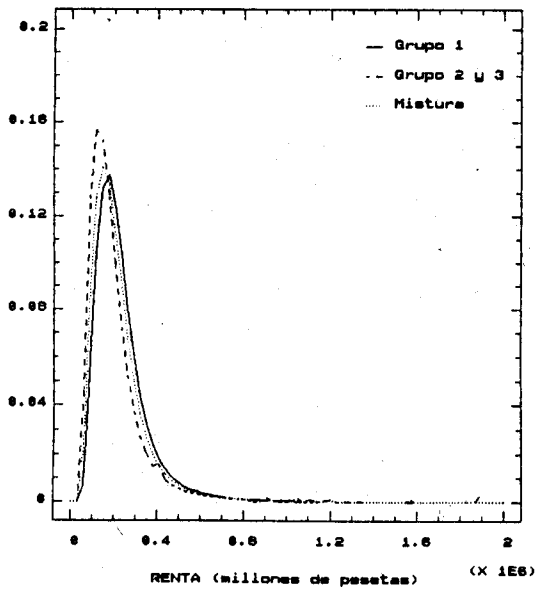
5.-GRAFICOS

Unicamente se presentan los gráficos citados en el texto para una mayor brevedad en la exposición del tema.

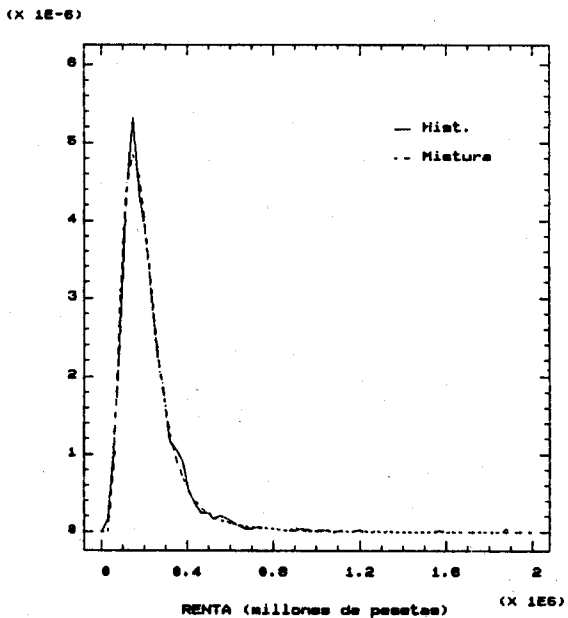
POLIGONOS DE FRECUENCIAS



AJUSTES POR GRUPOS Y MISTURA



AJUSTE MISTURA



AJUSTE NO PARAMETRICO

