ÍNDICE

Estudios de Economía Aplicada
Nº 14, 2000. Págs. 153-172

Repercusiones de la ocultación de renta sobre la medición de la desigualdad

PRIETO ALAIZ, M.
Universidad de Valladolid
PENA TRAPERO, B.
Universidad de Alcalá de Henares

RESUMEN

Una de las características de los datos de renta procedentes de las Encuestas de Presupuestos Familiares es su escasa fiabilidad, ya que los hogares entrevistados ocultan parte de ella. Recientemente, Pena, Callealta, Casas, Merediz y Núñez (1996) han propuesto una metodología para corregir este problema de ocultación, que ha permitido generar una nueva serie de datos de renta.

En este trabajo, estudiamos las repercusiones que tiene la utilización de la serie de renta original y la corregida para el año 1990-91 sobre la desigualdad. En concreto, contrastamos el dominio en sentido de la curva de Lorenz entre las distribuciones generadas por ambas series, siguiendo la metodología de la inferencia estadística paramétrica.

Palabras clave: Distribución de la renta, ocultación de renta, modelización paramétrica, dominio en el sentido de la curva de Lorenz.

ABSTRACT

The lack of accurate is a relevant feature of income data from the "Encuesta de Presupuestos Familiares" because the interviewed households do not declare their true income. Recently, Pena, Callealta, Casas, Merediz y Núñez (1996) have generated a new set of data by mean of methodology that tries to correct this problem of hidden income .

Throughout this paper, we study the consequences on income inequality of using the original set of data and the correct one. Concretely, we test the Lorenz dominance between the income distribution generated by both sets of data follow a parametric statistical inference approach.

ÍNDICE

Key Words: Income distribution, hidden income, parametric estimation, Lorenz dominance.

Código UNESCO:1209.03, 5304.02, 5302.04,5307.19

Artículo recibido en noviembre de 1999. Aceptado en enero de 2000.

1. Introducción

Uno de los elementos básicos de cualquier análisis relacionado con la distribución personal de la renta[1] (DPR), como el de la desigualdad, es el conjunto de datos disponibles. En España, este conjunto de datos procede, fundamentalmente, de las Encuestas de Presupuestos Familiares (EPF).

La EPF es una investigación que viene realizando el Instituto Nacional de Estadística, aproximadamente, cada diez años desde 1954, correspondiendo la última al año 1990-91. El objetivo prioritario de la encuesta es obtener información sobre el consumo de los hogares para elaborar los pesos de los diferentes bienes y servicios en el índice de precios al consumo. Además, la EPF proporciona datos sobre otras características de los hogares lo que la convierten en una de las fuentes de información principales sobre variables económicas de forma desagregada.

En general, la información de la EPF está sujeta a dos tipos de errores: los de muestreo y los no de muestreo. Los errores de muestreo se producen al obtener conclusiones sobre la población a partir de una muestra. Los errores no de muestreo se cometen, substancialmente, porque los individuos seleccionados en la muestra no responden o lo hacen de forma incorrecta. Estos últimos son especialmente relevantes en los datos sobre la renta cuando ésta es aproximada mediante la renta disponible, ya que los individuos entrevistados ocultan parte de la misma. Además, el grado de subdeclaración de los hogares no es homogéneo lo que puede provocar cambios significativos en las conclusiones que se extraigan sobre la desigualdad. Precisamente, este artículo trata de estudiar la influencia de la ocultación de renta por parte de los hogares entrevistados en la EPF del año 1990-91 sobre la desigualdad.

Una de las formas de estudiar tal repercusión sería contrastar si existen diferencias significativas entre el nivel de desigualdad que se obtiene con los datos de renta reales de los hogares entrevistados y el que se obtiene con los que efectivamente declaran. Dado que la renta real de los hogares es desconocida, tal comparación resulta imposible. Sin embargo, se puede obtener una nueva serie de renta que trate de corregir el problema de la ocultación y que, por tanto, se aproxime más a la renta real. Pena et al. (1996) han propuesto una metodología para obtener esta nueva serie de renta aprovechando la información suministrada por la EPF y la de Contabilidad Nacional (CN). El resultado de aplicar esta metodología sobre los datos de

1. En este trabajo, utilizaremos de forma indistinta los términos "distribución personal de la renta" y "distribución de la renta".

renta originales es una nueva serie que nos permitirá valorar la repercusión de la escasa fiabilidad sobre la desigualdad.

Dicha repercusión se puede examinar comprobando si existen diferencias estadísticamente significativas entre los índices de desigualdad obtenidos con los dos conjuntos de datos, es decir, los originales y los corregidos. Consecuentemente, requeriría realizar tantos contrastes como índices considerados. No obstante, se puede tener en cuenta que el orden establecido por la curva de Lorenz es el mismo que el que determina una clase bien definida de índices desigualdad. Por tanto, otra forma de analizar el efecto de la ocultación sobre la desigualdad y que simplifica la anterior es contrastar si existen diferencias significativas entre la curva de Lorenz generada por los datos originales y la generada por los corregidos. En la sección 3, definiremos el dominio en sentido de la curva de Lorenz y su relación con la desigualdad.

Existen dos enfoques para realizar el contraste mencionado anteriormente: Uno indirecto y otro directo (ver, Maasumi 1994, p.214). El primero, a diferencia del segundo, precisa como paso previo la especificación de un modelo paramétrico para la DPR. Nosotros hemos adoptado este enfoque indirecto, puesto que, si el modelo está correctamente especificado, simplifica cualquier análisis basado en la DPR, al resumir toda la información sobre la misma en un número limitado de parámetros. En la sección 4, trataremos los aspectos más relevantes de la inferencia estadística paramétrica.

Finalmente, en la sección 5, presentaremos los resultados y, en la sección 6, las conclusiones obtenidas.

2. Algunas aclaraciones conceptuales.

En esta sección concretaremos qué entendemos por renta y revelaremos ciertas características de los datos de renta procedentes de la EPF que ponen de manifiesto su falta de fiabilidad. Finalmente, presentaremos la metodología de Pena et al. (1996) para resolver este problema.

El término renta se ha asociado con diferentes variables. Dentro del análisis de la desigualdad, estrechamente relacionado con el del bienestar económico[2], la renta ha estado unida a variables que reflejaran la posición económica de los individuos o, como Ruiz-Castillo (1987) señala, el "conjunto de posibilidades de consumo al que los individuos vienen constreñidos". Una de estas variables ha sido la renta disponible[3].

---

2. En Atkinson (1970) y Dagum (1990), entre otros, se puede encontrar la conexión entre desigualdad y bienestar.
3. Otras variables que se han relacionado con la renta son la riqueza (ver, por ejemplo, Naredo, 1993) y el gasto (ver, por ejemplo, Ruiz-Castillo, 1986).

La renta disponible, en palabras de Pena (1996, p. xxii), es la cantidad de dinero «que les queda a los individuos[4] después de los procesos redistributivos que se desarrollan en las sociedades modernas para corregir o modificar las reglas del mercado» y, por tanto, «es la más ligada con los usos que los individuos pueden hacer de sus ingresos». La tabla 1 recoge sus componentes, que se identifican con aquellos de la cuenta de renta del sector hogares en CN. Básicamente, se asocian con los ingresos totales del hogar en la EPF (ver, Merediz y Pena 1996, p. 217).

Tabla 1. Componentes de la Cuenta de Renta del Sector Hogares

Recursos	*Empleos*
1. Remuneración de asalariados	1'. Cotizaciones Sociales
2. Intereses efectivos recibidos-Intereses efectivos pagados	2'. Impuesto de la renta y el patrimonio
3. Intereses imputados derivados de los contratos de seguros.	
4. Dividendos y otras rentas distribuidas por sociedades	
5. Rentas de la tierra y otros activos inmateriales.	
6. Excedente Bruto de Explotación.	
Renta primaria=1+2+3+4+5+6	
7. Prestaciones Sociales	
8. Transferencias netas.	
9. Indemnizaciones de seguros de accidentes netos de las primas pagadas.	
Renta Familiar Bruta=Renta primaria+7+8+9	
Renta Disponible=Renta Familiar Bruta-1'-2'	

A pesar de que la EPF constituye una de las fuentes de información fundamentales sobre variables económicas relevantes de forma desagregada, los datos sobre la renta no son fiables. Como Pena (1996, p. xxii) señala, esta información es «falsa en los niveles y en la distribución”.

Uno de los hechos que sirven para evidenciar la falta de fiabilidad en los niveles es que el 57% de los hogares entrevistados declaran tener mayores gastos que renta.

4. La EPF determina la unidad de análisis, ya que los datos de la EPF se refieren al hogar y no se dispone de información sobre las transferencias a unidades más pequeñas. Por tanto, hablaremos indistintamente de hogar y de individuos.

Sin duda, este rasgo resulta poco realista y, normalmente, se ha aceptado que es consecuencia de la falta de fiabilidad de los datos de renta, ya que, al estar la encuesta especialmente diseñada para conocer la estructura del gasto de los hogares, se considera que los datos del gasto son más correctos que los de la renta.

Sin embargo, algunos trabajos como Sanz (1992), Pazos (1994), Pena (1996) y Carrascal (1997) han cuestionado la fiabilidad de los datos de gasto debido a la manera de recoger esta información. Mientras al hogar se le pide información sobre los ingresos anuales, no ocurre lo mismo con todos los gastos. Así, se le pide información sobre el gasto semanal que realiza en una serie productos, por ejemplo, alimentos, bebida y tabaco; sobre el gasto mensual en productos como vestido y calzado; sobre el gasto trimestral en elementos tales como las consultas médicas y, finalmente, sobre el gasto anual en ciertos bienes duraderos como los electrodomésticos. Dado que los gastos no tienen el mismo periodo de referencia, al intentar expresarlos bajo un período común, el año, es necesario aplicar un factor de elevación temporal. Este factor es de 52 para los gastos semanales, de 12 en los gastos mensuales y de 4 en los trimestrales.

Por consiguiente, el dato del *gasto total de cada hogar* está influenciado, además de por el área geográfica al que pertenezca el hogar, por la semana del año en la que se realiza la encuesta, no pudiendo ser considerado como el verdadero gasto total del hogar. Hay que señalar que esta influencia estacional no afecta a la obtención del *gasto total de todos los hogares*, dado que la muestra se distribuye uniformemente a lo largo de las 52 semanas que dura la encuesta. Por otra parte, si pretendemos medir la desigualdad, Kakwani (1980, p.167 y ss.) demuestra que: “El gasto de consumo personal, domina en sentido de la curva de Lorenz al ingreso disponible”. Por lo que, la medida de desigualdad obtenida a partir del gasto enmascara el resultado real.

Merediz y Pena (1996, p. 201) señalan otro hecho que sirve para revelar la falsedad en los niveles. Al comparar la renta nacional disponible deducida de los datos de la EPF y la de los de la CN, se produce una infraestimación de la primera con respecto a la segunda en torno al 30%.

Por lo que se refiere a la falta de fiabilidad en la distribución, como Pena (1996, p. xxii) apunta, la subdeclaración es diferente según los distintos tramos de renta. Callealta, Casas y Nuñez (1996, pp. 429-430) señalan que la ocultación recae más en las clases de mayores ingresos. Éstas son las clases que pueden acceder con mayor facilidad a la propiedad de ciertos bienes generadores de renta y es, precisamente, en las rentas del capital y de la propiedad donde se produce un mayor nivel de ocultación. Estos autores indican que la participación de las rentas de la propiedad en el total de ingresos del hogar para los tramos altos de renta es ridícula si la comparamos con la participación que le asignan determinadas fuentes fiscales.

ÍNDICE

Los trabajos de Angel y Julio Alcaide (1974, 1977 y 1983) constituyen los primeros intentos en abordar el problema de la ocultación en los datos de renta procedentes de la EPF. Estos trabajos tienen como fundamento la relación entre los microdatos de la EPF y los datos agregados de la CN. Recientemente, basándose en la misma idea, Pena et al. (1996) han propuesto una metodología que ha permitido modificar los datos de renta originales de cada uno de los hogares entrevistados en la EPF para que fueran consistentes con los datos agregados de la cuenta de renta del sector hogares de la CN, cuyo saldo es la renta disponible de los hogares españoles.

La corrección se llevó a cabo en dos fases: en primer lugar, se hizo una corrección del nivel de renta que tienen los hogares españoles clasificados por la Comunidad Autónoma, la categoría socioprofesional y la clase de hábitat (ver, Merediz y Pena, 1996); y en segundo lugar, se hizo una corrección de la distribución (ver, Callealta, Casas y Nuñez, 1996).

3. La curva de Lorenz y el dominio en sentido de la curva de Lorenz

Consideremos la distribución de la renta representada por un elemento del conjunto de funciones de distribución continuas:

$$\Phi := \{F : \Re_+ \to [0,1]\}$$

donde $\Re_+ := [0,\infty)$. La función de densidad es $f(x) = \dfrac{d(F(x))}{dx}$, la media de *F* es $\mu_F = \int_0^\infty z\,dF(z)$ y la inversa de la función de distribución es $F^{-1}(t) = \inf_x \{x \mid F(x) \geq t\}$ para $t \in [0,1]$.

La *curva de Lorenz* relaciona la proporción acumulada de las unidades receptoras de renta con la proporción de renta acumulada por dichas unidades ordenadas de forma creciente, es decir,

$$L(p;F) = \frac{1}{\mu_F}\int_0^p zF^{-1}(t)dt$$

para $p \in [0,1]$ (ver Gastwirth, 1971); es una función creciente, continua y convexa en *p*, con *L(0;F) = 0* y *L(1;F) = 1*[5].

5. En Kakwani (1980, p.89) y Nygard y Sändstrom (181, pp. 150-157) se analizan detalladamente éstas y otras propiedades.

La curva de Lorenz tiene una interpretación directa en términos de desigualdad. Si todos los individuos tuvieran la misma renta (igualdad total), la distribución de la renta sería degenerada en el punto μ_F, es decir, $F^{-1}(p)=\mu_F$, para cualquier $p \in [0,1]$. Esto implicaría que la curva de Lorenz coincidiría con p; *L(p;F)=p* es llamada línea igualitaria. Consecuentemente, cuanto más cercana esté la curva de Lorenz a la línea igualitaria, menor desigualdad habrá. Así, dadas dos distribuciones, *F* y *G* $\in \Phi$, la distribución *F* presenta menos desigualdad que la *G* si la curva de Lorenz de *F* está por encima de la de *G*. Si las curvas de Lorenz se cortan, no se pueden obtener conclusiones definitivas sobre el nivel de desigualdad que presenta *F* con respecto a *G*.

La posición relativa de las curvas de Lorenz de dos distribuciones es lo que define el dominio en el sentido de Lorenz. Así, diremos que *la distribución F domina, en sentido de la curva de Lorenz, a G,* $F \geq_L G$*, si* $L(p;F) \geq L(p;G)$ *para* $\forall p \in [0,1]$ *con* $L(p;F) \neq L(p;G)$ *para algún* $p \in [0,1]$.

El tipo de desigualdad que mide la curva de Lorenz está caracterizado por dos rasgos. En primer lugar, la desigualdad es relativa, ya que es independiente de la renta media. En segundo lugar, el dominio en el sentido de la curva de Lorenz es equivalente al principio de transferencias de Pigou- Dalton (ver Nygard y Sändstrom, 1981, p.70)[6].

La ordenación del dominio en el sentido de la curva de Lorenz es el mismo que el que establece los índices de desigualdad que son independientes de la media y cumplen el principio de transferencias. En otros términos, si $F \geq_L G$ implica que $I(F) < I(G)$, para cualquier índice de desigualdad, *I(.)*, que cumple estas dos propiedades (ver Lambert, 1993, p.116). Esta relación es importante, ya que los índices más habituales, como el índice de Gini, el coeficiente de variación, la familia de índices de Atkinson y la familia de desigualdad de entropía generalizada, satisfacen las dos propiedades mencionadas.

4. Inferencia estadística: un enfoque paramétrico

Como ya hemos comentado, las conclusiones que se establecen sobre la desigualdad en España, generalmente, están basadas en la información obtenida de la EPF. Esta información consiste en un conjunto de observaciones $(x_{1,\ldots,}x_n)$, que se supone generadas por *F* (*distribución original*). Además, fruto del proceso de corrección aludido, hay otro conjunto de observaciones $(y_{1,\ldots,}y_n)$ que se supone generadas por *G*

6. El tratamiento continuo de la distribución de la renta asume implícitamente la propiedad de simetría, que es necesaria, en el caso discreto, para que el principio de transferencias sea equivalente al dominio en sentido de la curva de Lorenz

(*distribución corregida*). Nuestro objetivo es contrastar si la curva de Lorenz de *F* es la misma que la de *G*, frente a la hipótesis alternativa de que *F* domina, en sentido de la curva de Lorenz, a *G*, es decir,

$$H_o: \quad L(p;F) = L(p;G) \qquad \text{para} \quad \forall\, p \in [0,1]$$
$$H_1: \quad L(p;F) \geq L(p;G) \qquad \text{para} \quad \forall\, p \in [0,1]$$

con desigualdad estricta para algún p

La inferencia estadística paramétrica es el marco en el que vamos a realizar este contraste. En términos generales, este enfoque, que Maasumi (1994) califica como indirecto, precisa, como paso previo, la estimación de una serie de parámetros determinantes de la función de distribución de la renta.

Si los estimadores de dichos parámetros tienen “buenas“ propiedades y la función estimada se ajusta bien a los datos, este enfoque permite simplificar cualquier análisis basado en la DPR, ya que toda la información sobre la misma se sintetiza en un número limitado de parámetros. En nuestro caso, permitirá expresar la hipótesis de interés como una función de los parámetros de la distribución elegida. Sin embargo, la desventaja, como Maasumi (1994) señala, es que las conclusiones que se obtienen son vulnerables a posibles errores de especificación, especialmente, si la distribución paramétrica elegida se ajusta mal a los datos en los extremos inferior y superior.

Consecuentemente, la modelización paramétrica de la DPR entraña la elección no sólo de formas funcionales que sean, a priori, apropiadas para la distribución de renta, sino también métodos de estimación que garanticen buenas propiedades para los estimadores, técnicas de bondad del ajuste que cuantifiquen la diferencia entre la función ajustada y la distribución empírica y contrastes de hipótesis que permitan la elección entre diferentes formas funcionales. A continuación, expondremos algunas consideraciones sobre estos aspectos de la modelización paramétrica que es preciso tener en cuenta.

El enfoque paramétrico asume que la distribución de la renta puede ser representada por un elemento del conjunto de distribuciones continuas, perfectamente especificada, salvo un vector de orden (kx1) de parámetros desconocidos,, es decir, por un miembro de

$$\Psi := \{F_\theta, \theta \in \Theta\}$$

donde $\Psi \subset \Phi$ y $\Theta \subset \Re^k$ es el espacio paramétrico. $f(x;\theta)$ es la función de densidad de F_θ.

ÍNDICE

El primer paso en la modelización paramétrica consiste en elegir, entre todas las funciones de distribución continuas con representación paramétrica, aquellas distribuciones que reúnan una serie de propiedades deseables para el caso de la renta. Algunas de estas propiedades son la asimetría a la derecha, la convergencia a la ley de Pareto, la parquedad en la especificación, el origen económico de la formulación y la generalidad[7]. Esta última propiedad se ha destacado en los trabajos más recientes que persiguen la especificación de funciones cada vez más flexibles que comprendan el mayor número posible de distribuciones como casos especiales.

Una de las formas funcionales más generales es la función beta generalizada de segunda especie (BG2) propuesta por McDonald (1984). La función de densidad de la BG2 viene dada:

$$f^{BG2}(x;\theta)=\frac{ax^{ap-1}}{b^{ap}B(p,q)\left(1+\left(x/b\right)^{a}\right)^{p+q}} \quad \text{para } x \geq o \qquad \textbf{(2)}$$

donde *B(.,.)* es la función beta y $\theta=(a,b,p,q)^{T}$ [8].

Entre los casos especiales de la BG2, se encuentran, para *q=1*, la distribución Dagum, (1977) y, para *p=1*, la Singh y Maddala, (1976). La gran ventaja de estas dos distribuciones es que su función de distribución tiene forma cerrada. McDonald (1985) detalla las distribuciones que comprende la BG2, que, además de las mencionadas, abarca distribuciones biparamétricas ampliamente utilizadas en el caso de la modelización paramétrica de la renta, como la lognormal, la gamma y la log-logísitica o distribución de Fisk. En el apéndice 1 se encuentra las formas de las funciones de densidad que hemos considerado en este artículo y su relación la BG2.

Una vez elegida una función, el siguiente paso es la estimación del vector de parámetros desconocidos. Uno de los procedimientos más utilizados por las buenas propiedades de los estimadores es el método de máxima verosimilitud (ver, por ejemplo, Zacks, 1981).

Una propiedad de los estimadores máximo verosímiles que nos interesa resaltar es la propiedad de la invarianza en el sentido de Zehna (Zacks, 1981). Así, sea *H (q)* un vector de orden (rx1), formado por r funciones continuas del vector paramétrico, es decir, $H(q)=[h_1(q),\ldots,h_r(q)]^{T}$. Si el estimador máximo verosímil de *q* es $\hat{\theta}$, el esti-

7. Una revisión de estas propiedades se puede encontrar en Dagum, (1977) y Pena et al. (1996).

8. Con "T" denotaremos la transpuesta del vector.

mador máximo verosímil de *H(q)* es $H(\hat{\theta})$. Por tanto, $H(\hat{\theta})$ es un estimador consistente y asintóticamente eficiente, con distribución asintótica determinada por

$$\sqrt{n}\,(H(\hat{\theta}) - H(\theta)) \xrightarrow{a} N(0, \Omega) \qquad \textbf{(5)}$$

donde la matriz de varianzas y covarianzas asintótica de $H(\hat{\theta})$, W, se puede expresar como

$$\Omega = D^T \Sigma D$$

donde *D* es una matriz de orden (rxk) cuyos elementos son de la forma

$$D_{ij} = \left[\frac{\partial h_i(\theta)}{\partial \theta_j} \right] \begin{matrix} i = 1,\ldots r \\ j = 1,\ldots, k \end{matrix}$$

y Σ la matriz de varianzas y covarianzas asintótica de $\hat{\theta}$. Un estimador consistente de W es $\hat{\Omega} = \hat{D}^T \hat{\Sigma} \hat{D}$ con $\hat{D}$ obtenida al sustituir q por $\hat{\theta}$.

Esta propiedad resulta especialmente útil para deducir las propiedades de los estimadores de las herramientas utilizadas en la medición de la desigualdad, ya que la mayoría se pueden expresar como funciones de los parámetros.

La validez de la modelización paramétrica de la DPR depende de la bondad del ajuste. En otras palabras dependerá de si el modelo estimado se ajusta bien a los datos, porque, si lo hace y los estimadores tienen buenas propiedades, entonces el modelo se puede considerar adecuado para describir la DPR. Los contrastes de bondad del ajuste establecen bajo la hipótesis nula que las observaciones han sido generadas por una distribución que puede estar completamente especificada (hipótesis nula simple) o no (hipótesis nula compuesta). Este último es el caso relevante, ya que, generalmente, , es desconocido. Los estadísticos para realizar estos contrastes se basan en la cuantificación de la distancia entre la función de distribución que se supone bajo la hipótesis nula y la distribución empírica, $F_n(x)$, definida como

$$F_n(x) = n^{-1} \sum_{i=1}^{n} \Delta_{x_i}(x)$$

tal que

$$\Delta_{x_i}(x) = \begin{cases} 0 & x < x_i \\ 1 & x \geq x_i \end{cases}$$

Desgraciadamente, cuando la hipótesis nula es compuesta y para la mayoría de los modelos empleados en la DPR, las distribuciones de los estadísticos de contraste de bondad del ajuste, como el de Kolmogorov-Smirnov o los de la familia de Cramer-Mises, no se encuentran tabuladas (ver Stephens, 1986 y Gibbons and Chakraborti, 1992)[9]. Sin embargo, estos estadísticos pueden ser usados para estudiar las diferencias entre la distribución empírica y la estimada. Entre los estadísticos de bondad del ajuste, Stephens (1986) recomienda el estadístico de Anderson-Darling (A^2), porque es más poderoso para detectar posibles desviaciones en las colas. La forma del estadístico cuando q está estimado es

$$A^2 = n\int_0^\infty \left(F_n(x) - F_{\hat{\theta}}(x)\right)^2 \left[F_{\hat{\theta}}(x)\left(1 - F_{\hat{\theta}}(x)\right)\right]^{-1} f(x;\hat{\theta})dx$$

Finalmente, relacionados con la modelización de la renta se encuentran los contrastes de selección de modelos. La hipótesis nula de este tipo de contraste es que la distribución de la renta es de una determinada clase y la hipótesis alternativa es que dicha distribución es de otra clase. Los estadísticos de contraste a utilizar dependerán de si la hipótesis nula es abarcada por la hipótesis alternativa o no. En el caso de que sí lo sea, se puede utilizar cualquiera de los estadísticos de la terna sobradamente conocida compuesta por el de la razón de verosimilitud, el de Wald o el de los multiplicadores de Lagrange con distribución igual a una Chi cuadrado con tantos grados de libertad como restricciones imponemos bajo la hipótesis nula. Sin embargo, cuando la hipótesis nula no queda comprendida en la alternativa, los estadísticos a utilizar son los del tipo de Cox (ver Cox, 1961 y 1962 Atkinson ,1970, White, 1982 y, más recientemente, Victoria-Feser, 1997) con distribuciones más complejas que en el caso anterior. La gran ventaja de utilizar distribuciones como la BG2 es que permite realizar contrastes utilizando el trío mencionado.

Además, existen determinados criterios para seleccionar modelos con diferente número de parámetros. Entre los más empleados se encuentran el criterio de Akaike (*AIC*) y el de Schwarz (*SBIC*). Estos se definen como:

$$AIC = -2\ell/n + 2k/n$$

y

$$SBIC = -2\ell/n + (k\log n)/n$$

9. Una forma de aproximar la distribución de estos estadísticos es por medio de las técnicas de bootstrap.

donde $\ell = \ell(\hat{\theta})$ es el valor del logaritmo natural de la función de verosimilitud evaluada en los estimadores máximo verosímiles de los parámetros.

Concluido el proceso de modelización, la hipótesis de interés (1) se puede expresar como una función de los parámetros. Sea F_θ la distribución original y $G_{\theta'}$ la distribución corregida, tales que F_θ es un caso particular de $G_{\theta'}$ o $G_{\theta'}$ es un caso particular de F_θ. La hipótesis (1) se puede formular como

$$\begin{aligned} &H_o : H(\theta_F) = H(\theta_G) \\ &H_1 : H(\theta_F) > H(\theta_G) \text{ o } H(\theta_F) < H(\theta_G) \end{aligned} \qquad \textbf{(6)}$$

donde *H(.)* el vector formado por las $r \leq k$ funciones de los parámetros que determinan el dominio en sentido de la Curva de Lorenz. El estadístico de contraste sería,

$$\frac{\sqrt{n}\,\iota^T\left(H_F(\hat{\theta}_F) - H_G(\hat{\theta}_G)\right)}{\sqrt{\iota^T \hat{\Omega}_{F+G}\,\iota}} \qquad \textbf{(7)}$$

donde ι es un vector de orden (rx1) formado por unos y $\hat{\Omega}_{F+G} = \left(\hat{D}^T_{H_F}\hat{\Sigma}_F\hat{D}_{H_F}\right) + \left(\hat{D}^T_{H_G}\hat{\Sigma}_G\hat{D}_{H_G}\right)$. La distribución asintótica de dicho estadístico bajo la hipótesis nula es una distribución normal estándar.

5. Resultados

Las tablas 2 y 3 proporcionan la información sobre los resultados de las estimaciones de la distribuciones log-normal (LN), gamma (G), Fisk (F), Dagum (D), Singh Maddala (SM) y beta generalizada de segunda especie (BG2) para los datos de renta originales de la EPF normalizados por el tamaño familiar y los datos corregidos por subdeclaración, también en términos per capita. En dichas tablas se muestra el valor de los estimadores máximo verosímiles[10], el error estándar de los estimadores entre

---

10. Los problemas de optimización que se plantean para obtener los estimadores máximo verosímiles de los parámetros de las distribuciones analizadas son no lineales. Por tanto, fue necesario la aplicación de algún algoritmo de optimización numérica que en nuestro caso fue el de Newton- Raphson. Los criterios de convergencia establecidos fueron dos: que la diferencia entre los valores de los estimadores en dos iteraciones consecutivas fuera de 10-6 y que las condiciones de primer orden fueran menores que 10-2. El programa fue diseñado en lenguaje C.

ÍNDICE

paréntesis[11], el valor del logaritmo de la función de verosimilitud (γ) evaluado en el estimador máximo verosímil, el criterio de Akaike (*AIC*), el criterio de Schwarz (*SBIC*) y el estadístico de Anderson-Darling (A^2).

Tabla 2: Modelización paramétrica de la distribución de la renta original

	LN	*G*	*F*	*D*	*SM*	*BG2*
a	μ=-0.5214 (0.0040)	1	3.1347 (0.0180)	3.0976 (0.0323)	3.1802 (0.0336)	3.2043 (0.149)
b (*b*)	s^2=0.3367 (0.0033)	0.2344 (0.0023)	0.5926 (0.0023)	0.5828 (0.0075)	0.5814 (0.0071)	0.5816 (0.0073)
p	NA (0.0279)	3.061	1 (0.0263)	1.0356	1 (0.0665)	0.9890
q	NA	NA (0.0237)	1 (0.063)	1	0.9605	0.9506
ℓ	-7473.2226	-83339.20	-7114.38	-7113.37	-7113.067	-7113.06
AIC	0.7067	0.7886	0.6728	0.6728	0.6728	0.6729
SBIC	0.7075	0.7893	0.6735	0.6739	0.6739	0.6744
A^2	30.0473	∞	1.6077	1.0554	1.0033	1.0250

NA: No es aplicable

11. El error estandar de un elemento $\hat{\theta}_i$ de $\hat{\theta}$ se ha calculado $s.e.(\hat{\theta}_i) = \sqrt{\frac{\hat{\Sigma}_{ii}}{n}}$, donde $\hat{\Sigma}_{ii}$ es el valor del elemento i de la diagonal principal del estimador de la matriz de varianzas y covarianzas asintótica $\hat{\Sigma}$.

Tabla 3: Modelización paramétrica de la distribución de la renta corregida

	LN	*G*	*F*	*D*	*SM*	*BG2*
a	μ=-0.2554 (0.0042)	1	2.9647 (0.0170)	2.8981 (0.030)	3.0512 (0.0325)	3.1144 (0.1446)
b (*b*)	s^2=0.3790 (0.0037)	(0.3765) (0.0038)	0.7717 (0.0031)	0.7456 (0.0104)	0.7418 (0.0094)	0.7428 (0.0097)
p	NA	2.5377 (0.0232)	1	1.070 (0.0276)	1	0.9703 (0.0650)
q	NA	NA	1	1	0.9240 (0.0225)	0.8985 (0.0586)
ℓ	-14334.20	-15378.73	-13912.90	-13909.17	-13907.86	-13907.77
AIC	1.3554	1.4541	1.3155	1.3152	1.3151	1.3152
SBIC	1.3554	1.4549	1.3163	1.3164	1.3163	1.3167
A^2	∞	∞	2.5682	1.2339	1.1138	1.1942

NA: No es aplicable

Comprobamos que la lognormal y la gamma son las distribuciones que peor se ajustan a las dos series de datos según todos los criterios de bondad. Además, tanto la diferencia entre BG2 y la lognormal y la que existe entre la BG2 y la gamma son estadísticamente significativas usando el contraste de la razón de verosimilitud.

La distribuciones Fisk, Singh-Maddala, Dagum y BG2 son las que proporcionan mejores ajustes para las dos series. En el caso de los datos originales, no existen diferencias estadísticamente significativas entre la distribución BG2 y las distribuciones Fisk, Dagum y Singh-Maddala. Merece la pena destacar el comportamiento de la distribución Fisk que, pese a ser una distribución biparamética, se ajusta bien a los datos: el estadístico de Anderson-Darling apenas varía entre la distribución Fisk, Dagum y Singh-Maddala. Además, los criterios de Akaike y Swartz para la distribución Fisk son los menores. En el caso de la serie corregida, los resultados apuntan hacia la elección de la distribución Singh-Maddala. El valor del criterio de Anderson-Darling en el mencionado modelo es el menor de los obtenidos. No existen diferencias significativas entre la distribución Singh-Maddala y la BG2. Por el contrario, estas diferencias se producen entre la BG2 y el resto de los modelos que abarca dicha distribución. Además, la distribución Singh-Maddala presenta los mejores resultados en el resto de los criterios de selección de modelos.

Por tanto, el proceso de modelización de la distribución de la renta para las dos series concluye eligiendo la distribución Fisk, para el caso de los datos originales, y la

ÍNDICE

distribución Singh-Maddala, en el caso de los datos corregidos. Hay que señalar que la distribución Fisk es un caso restringido de la Singh-Maddala, cuando $q=1$.

La hipótesis (6) se puede concretar ahora, expresándola como una función de los parámetros de la distribución Fisk y de los de la Singh-Maddala aprovechando los resultados de Wilfling y Krämer (1993). Estos autores proporcionan las condiciones necesarias y suficientes para el dominio en sentido Lorenz entre dos distribuciones Sing-Maddala. Así, dos distribuciones Singh-Maddala F_q y $G_{q'}$ con parámetros $q=(a,b,p)^T$ y $q'=(a',b',p')^T$, F_q ³ $_L$ G_q si y sólo si $a>a'$ y $ap>a'p'$.

Dado que la distribución Fisk es un caso particular de la Singh-Maddala, las condiciones establecidas por Wilfling y Krämer (1993) se pueden reformular para las distribuciones Fisk y Singh-Maddala. Si los parámetros de la distribución Fisk, $F_è$, son $q=(a,b)^T$ y los de la distribución Singh-Maddala, $G_{q',}$ son $q'=(a',b',p')^T$, entonces $F_\theta \geq_L G_{\theta'}$ si y sólo si $a>a'$ y $a>a'p'$.

Por tanto, hipótesis establecidas en (6) se pueden expresar como

$$H_o : \begin{pmatrix} a \\ a \end{pmatrix} = \begin{pmatrix} a' \\ a'p' \end{pmatrix}$$

$$H_1 : \begin{pmatrix} a \\ a \end{pmatrix} > \begin{pmatrix} a' \\ a'p' \end{pmatrix}$$

El valor de estadístico (7) para este caso es 7.66 lo que nos lleva a rechazar la hipótesis nula a favor de la alternativa. En otras palabras la evidencia empírica nos muestra que la función de distribución original domina, en sentido de la curva de Lorenz, a la corregida. Si aceptamos que la distribución corregida es más cercana a la original, entonces lo que estamos diciendo es que el nivel de desigualdad que se obtiene con los datos originales es menor que el que realmente se obtendría en realidad.

6. Conclusiones

Nuestro objetivo es analizar las consecuencias de la falta de fiabilidad de los datos de renta procedentes de la EPF aproximada por la renta disponible. Para ello, contrastamos, desde una perspectiva paramétrica, el dominio en sentido de la curva de Lorenz de la distribución generada por los datos originales y por los datos corregidos mediante la metodología propuesta por Pena et al. (1996).

La ocultación tiene efectos sobre la forma funcional seleccionada, puesto que la distribución que mejor se ajusta a los datos originales es la Fisk y a los datos corregidos es la Singh-Maddala. Sin embargo, los resultados más significativos son los que se establecen en términos de desigualdad. Así, la evidencia empírica apunta a que la distribución generada por los datos originales domina, en sentido de la curva de Lorenz, a la generada por los datos corregidos. Por tanto, existen diferencias significativas entre el nivel de desigualdad de la distribución que se deriva de la EPF, de aquella otra que se supone más cercana a la realidad, siendo menor en la primera que en la segunda. Este resultado coincide con lo que se esperaba que ocurriera, ya que la ocultación se cree que es superior en los tramos de renta más elevados.

Bibliografía

ALCAIDE INCHAUSTI, A. y J. ALCAIDE INCHAUSTI (1974): Metodología para la estimación de la Distribución Personal de la Renta en España en 1970. *Hacienda Pública Española*, 26, 55-63.

ALCAIDE INCHAUSTI, A. y J. ALCAIDE INCHAUSTI (1977): Distribución Personal de la Renta en España y en los Países de la OCDE. *Hacienda Pública Española*, 47, 17-57.

ALCAIDE INCHAUSTI, A. y J. ALCAIDE INCHAUSTI (1984): Distribución Personal de la Renta Española en 1980. *Hacienda Pública Española*, 85, 485-509.

ATKINSON, A.B.: On the measurement of inequality. *Journal of Economic Theory* 2, 244-63.Traducido en Hacienda Pública Española,61, 217-233.

ATKINSON, A.C. (1970): A method for discriminting between models (with discussion). *Journal of Royal Statistic Society*. B, 32, 323-353

CALLEALTA,J., J.M. CASAS y J. NÚÑEZ (1996): Distribución de la renta per capita disponible en España: descripción, desigualdad y modelización en Pena B., J. Callealta, J.M. Casas, A. Merediz y J. Núñez *Distribución Personal de la Renta en España*. Ed. Pirámide. Madrid. Cap.5.

CARRASCAL, U. (1997): Consumo Familiar en España. Análisis y Obtención de Escalas de Equivalencia. Universidad de Valladolid.

COX, D. R. (1961): Test of separate families of hypothesis en *Proccedings 4th Berkeley Sympsium of Mathematicals Statistics and Pobability,* vol I, pp.105-123. Berkeley: University of California Press.

DAGUM, C. (1977): A new model of personal income distribution: specification and estimation, *Economiee Appliqée* 30, 413-437.

DAGUM, C. (1990): On the relationship between income inequality measures and social welfare functions, *Journal of Econometrics*, Annals Issue, 43, 91-102.

FISK, P.R. (1961): The Graduation of Income Distribution. *Econometrica*, 29, 171-185.

ÍNDICE

GASTWIRTH, J.L. (1971): A General Definition of the Lorenz Curve. *Econometrica* 39, 1037-1039.

GIBBONS, J. y CHAKRABORTI (1992): *Nonparametric Statistical Inference*. Marcel Dekker.

LAMBERT, P.J. (1993): *The Distribution and Redistribution of Income* Segunda Edición. Manchester University Press.

KAKWANI, WC (1980): *Income Inequality and Poverty. Methods of Estimation and Policy Applications.* World Bank Research Publication

MAASUMI, E. (1994): Empirical analysis of inequality and welfare en P. Schmidt y H. Pesaran (eds.), *Handbook of Applied Microeconomics*.

MCDONALD, J. B. (1984): Some generalized functions for the size distribution of income, *Econometrica* 52, 647-663.

MEREDIZ, A. (1996): Instrumentos Básicos en la Elaboración de las Cuentas de Renta de los Hogares en 1973,1980 y 1990 según Distintas Desagregaciones en Pena, B., J. Callealta, J.M. Casas, A. Merediz y J. Núñez, J *Distribución Personal de la Renta en España*. Ed. Pirámide. Madrid. Cap. II.

MEREDIZ, A. y B. PENA (1996): La Cuenta de los Hogares en 1973,1980 y 1990, por Autonomías, Categorías Socioprofesionales y Clases de Hábitat en

PENA, B., J. CALLEALTA, J.M. CASAS, A. MEREDIZ y J. NÚÑEZ,J: *Distribución Personal de la Renta en España*. Ed. Pirámide. Madrid. Cap.III.

NAREDO, J.M.(1993): Composición y Distribución de la Riqueza de los Hogares Españoles. *Actas del I Symposio sobre la Igualdad y Distribución de la Renta y la Riqueza*. Argentaria. Madrid

NYGARD, F. y A. SANDSTROM (1981): *Measuring Income Inequality*. University of Stockholm.

PAZOS, M. (1994): Variabilidad Semanal de los Gastos en la EPF. *Estadística Española*, 137, 431-444.

PENA, J.B (1996): Prólogo de la Distribución Personal de la Renta en España en Pena,J.B., Callealta,J., Casas,J.M., Merediz,A. y Núñez,J *Distribución Personal de la Renta en España*. Ed. Pirámide. Madrid.

PENA,J.B., J. CALLEALTA, J.M. CASAS, A. MEREDIZ y J. Núñez (1996): *Distribución Personal de la Renta en España*. Ed. Pirámide. Madrid.

RUIZ-CASTILLO, J. (1987): La Medición de la Pobreza y de la Desigualdad en España. *Estudios Económicos* no. 42. Banco de España. Madrid.

SANZ, B (1992): La Encuesta de Presupuestos Familiares 1990-1991. *Situación*, 2-3, 151-166.

SINGH, S. K. and G. S. MADDALA (1976): A Function for the size distribution of income *Econometrica* 44, 963-970.

ÍNDICE

STEPHENS, M.A. (1986): Test Based on EDF Statistics en D'Agostino, R.B. y M.A. Stephens *Goodness of Fit Techniques* . Marcel Dekker. Cap.4.

VICTORIA-FESER, M.P. (1997): A Robust test for non-nested Hypothesis*, Journal of Royal Statistical Society,* no 3, 715-727.

WILFLING, B. y W. KRAMER (1993): The Lorenz-Ordering of Singh-Maddala Income Distributions. *Economics Letters*, 43, 53-57.

WHITE, H. (1982): Discrimination between regression models to determine the pattern of enzyn synthesis in synchronous cell cultures. *Journal of Biometrics*, 28, 13-32.

ZACKS,S. (1981). *Parametric Statistical Inference: Basic Theory and Modern Approaches*. Pergemon Press

APÉNDICE
Funciones de densidad utilizadas y su relación con la beta generalizada de segunda especie.

- Función de densidad lognormal

$$f^{LN}(x;\mu,\sigma^2)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{\frac{-1}{2\sigma^2}(\ln(x)-\mu)^2\right\}$$

con x *>0 y* σ^2 *>0*

- Función de densidad gamma

$$f^{G}(x;\beta,p)=\frac{x^{p-1}\beta^{-p}\exp\left\{-x/\beta\right\}}{\Gamma(p)}$$

con x >0 y β y *p>0*

- Función de densidad log-logistica o Fisk

$$f^{F}(x;a,b)=\frac{ab^{-a}x^{a-1}}{\left(1+\left(x/b\right)^{a}\right)^{2}}$$

con x>0 y *a* y *b>0*

- Función de densidad Dagum

$$f^{D}(x;a,b,p)=\frac{ab^{-ap}px^{ap-1}}{\left(1+\left(x/b\right)^{a}\right)^{1+p}}$$

con $x>0$ y *a, b* y *p>0*

- Función de densidad Singh-Maddala

$$f^{SM}(x;a,b,q)=\frac{ab^{-a}qx^{a-1}}{\left(1+\left(x/b\right)^{a}\right)^{1+q}}$$

con $x>0$ y *a, b* y *q>0*

Relaciones con la beta generalizada d4 segunda especie.

Funciones de 4 parámetros:

Funciones de 3 parámetros:

Funciones de 2 parámetros