

Estimación de la función de distribución y cuantiles en la población de pobres*

ENCARNACIÓN ÁLVAREZ-VERDEJO

Departamento de Métodos Cuantitativos para la Economía y la Empresa, UNIVERSIDAD DE GRANADA, ESPAÑA. E-mail: encarniav@ugr.es

MARÍA DOLORES ESTUDILLO-MARTÍNEZ

Departamento de Estadística e Investigación Operativa, UNIVERSIDAD DE JAÉN, ESPAÑA. E-mail: mdestudi@ujaen.es

SONIA CASTILLO-GUTIÉRREZ

Departamento de Estadística e Investigación Operativa, UNIVERSIDAD DE JAÉN, ESPAÑA. E-mail: socasti@ujaen.es

RESUMEN

La línea o umbral de pobreza es un indicador económico que permite clasificar a los individuos de una población como pobres o no. Los estudios del fenómeno de la pobreza prestan especial atención a la población de pobres, midiendo a partir de distintos indicadores la distribución y la intensidad de los pobres. Tales indicadores están basados en su mayoría en la función de distribución y en cuantiles. En este trabajo se plantea la estimación de estos parámetros en la población de pobres, y los estimadores presentados se evalúan numéricamente mediante estudios de simulación Monte Carlo y con datos extraídos de la Encuesta de Presupuestos Familiares.

Palabras clave: Línea de pobreza, diseño muestral, coeficiente de asimetría, simulación Monte Carlo.

Estimation of the Distribution Function and Quantiles for the Population of Poor

ABSTRACT

The poverty line is an economic indicator that divides a population into poor and non poor. Studies about poverty are highly interested on the population of poor, hence they measure the distribution and intensity of the poor through different indicators. Such indicators are generally based upon the distribution function and quantiles. Assuming an arbitrary sampling design, the estimation of the previous parameters is discussed in this paper. Proposed estimators are numerically evaluated via Monte Carlo simulation studies and using data extracted from the Spanish Household Panel Survey.

Keywords: Poverty Line, Sampling Design, Skewness Coefficient, Monte Carlo Simulation.

Clasificación JEL: C13, I32, C15, I31

* Este trabajo ha sido subvencionado por el proyecto de Excelencia "Nuevas metodologías para la estimación de la función de distribución y proporciones con aplicación a las fuentes de crecimiento económico" (código SEJ-7039) de la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía.

Artículo recibido en abril de 2012 y aceptado en agosto de 2012

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. e-30307

1. INTRODUCCIÓN

El concepto de pobreza tiene amplias y numerosas interpretaciones, lo que implica que se trata de un fenómeno complejo en el que intervienen múltiples factores. Por esta razón existen diversas formas de enfocar el estudio de la pobreza. Por ejemplo, dependiendo de la información que se utilice podemos distinguir entre pobreza objetiva y subjetiva. Dependiendo del baremo o referencia se puede hablar de pobreza absoluta o relativa, o bien se puede hablar de estudios transversales o estáticos según se tenga en cuenta la dimensión temporal o no en el estudio del fenómeno de la pobreza.

Los análisis de pobreza objetiva utilizan información procedente de la observación directa del investigador, mientras que los de pobreza subjetiva se basan en la percepción que tienen los propios individuos de su situación.

Dentro del estudio de la pobreza objetiva podemos distinguir entre pobreza absoluta y pobreza relativa. El primer concepto está relacionado con la carencia de necesidades básicas, de forma que una persona incluida dentro de este grupo en un determinado lugar, lo está en cualquier país o sociedad. El concepto de pobreza relativa se centra en una determinada sociedad, de forma que una persona es pobre cuando se encuentra en una situación de clara desventaja, económica y socialmente, respecto al resto de personas de su entorno. El que una persona sea calificada como pobre dentro de una sociedad en concreto, no tiene por qué serlo también en una sociedad distinta.

En el estudio de la pobreza relativa interviene el concepto de línea o umbral de pobreza que permite clasificar a los individuos como pobres o como no pobres dependiendo de a qué lado de la línea de pobreza se encuentren. Normalmente se utilizan indicadores basados en el ingreso o el gasto y se fija un nivel mínimo de la variable por debajo del cual las personas serán clasificadas como pobres, y clasificadas como no pobres en caso contrario. Numerosos organismos obtienen el umbral o línea de pobreza a partir de un porcentaje de la mediana de ingresos familiares por unidad de consumo.

El objetivo de los estudios de pobreza relativa no se centra en la determinación del umbral de pobreza y en la clasificación de los individuos en pobres y no pobres. Al contrario, la distribución de los individuos dentro de la propia población de pobres y la intensidad de la pobreza es un tema de relevante interés que las agencias y organismos responsables de medir la pobreza relativa tienen en cuenta en sus respectivos estudios.

Por ejemplo, la brecha de pobreza es un indicador económico que mide la intensidad de la pobreza, es decir, se trata de una medida de la distancia de los individuos pobres al umbral o línea de pobreza. La brecha relativa de pobreza se calcula como el cociente entre la brecha de pobreza y el número de personas por el umbral de pobreza, es decir, como si todas estuvieran en el umbral de po-

breza. EUROSTAT (Eurostat, 2000) define la brecha de pobreza como la diferencia entre el umbral de pobreza y la mediana de los ingresos por unidad de consumo de las personas consideradas como pobres, expresado como un porcentaje del umbral de pobreza.

Otra técnica utilizada para analizar la distribución de los individuos dentro de la población de pobres es comparar distintos cuantiles dentro de este grupo de individuos. Por ejemplo, el Instituto Nacional de Estadística (INE) utiliza los deciles de la población de pobres para analizar la distribución de los mismos. El uso de cuantiles para la población de pobres no se limita a este estudio, sino que también son necesarios para el cálculo de otros indicadores económicos, como por ejemplo la propia brecha de pobreza.

Algunas referencias importantes que definen y comentan con detalle los indicadores económicos para la medición de la pobreza así como otros estudios relevantes relacionados con el fenómeno de la pobreza pueden consultarse en Kakwani (1980), Atkinson (1987), Pérez-Mayo y Fajardo (2003), Bárcena y Cowell (2006), Domínguez y Martín (2006), Foster (2007), Silber (2007), Grádin et al. (2008), Núñez (2009) y Pérez-Moreno (2009).

La determinación de un cuantil está directamente relacionada con la función de distribución, y por tanto el problema de la estimación de la función de distribución puede considerarse tan importante como el problema de la estimación de un cuantil.

Actualmente existen trabajos descriptivos donde se obtienen numerosas medidas de pobreza, entre ellas el cálculo de cuantiles para analizar la distribución de los ingresos de la población pobre. Por ejemplo, en INE (2005) pg. 15, y con datos de la Encuesta de Condiciones de Vida 2004, se analiza la distribución de los ingresos de los individuos que se encuentran por debajo del umbral de pobreza mediante el cálculo de deciles en la población de pobres. En este estudio descriptivo tan sólo se conoce el valor de la estimación de los distintos deciles, pero se desconoce la precisión que tienen estas medidas así como si esta precisión puede estar seriamente afectada por algún factor como puede ser el orden del cuantil. El objetivo de este trabajo es plantear métodos de estimación para la función de distribución y cuantiles para la población de pobres, y medir mediante estudios de simulación Montecarlo la precisión de los estimadores planteados. La precisión se medirá en términos de sesgo relativo y error cuadrático medio relativo, y se analizará tal precisión en función de distintos aspectos, como por ejemplo el grado de asimetría de la distribución en estudio, el orden del cuantil, el tamaño muestral, el tipo de muestreo utilizado, etc. Por tanto, el planteamiento del cálculo de la función de distribución y cuantiles en la población de pobres para distintos diseños muestrales, un estudio exhaustivo del comportamiento empírico de los estimadores presentados, así como la evolución de la precisión de los métodos planteados en función de distintos factores son

las principales aportaciones de este trabajo. Desde un punto de vista económico, este estudio permitirá identificar las situaciones en las que la precisión asociada a la estimación de las medidas de pobreza planteadas en este trabajo puede no ser apropiada.

En la Sección 2 se describe la notación que se utilizará y se define formalmente la línea de pobreza, la cual nos permitirá identificar la población de pobres. Dentro de esta población se definirán los parámetros objeto de este estudio: la función de distribución y el cuantil poblacional para los pobres. En la Sección 3 se plantea la estimación de tales parámetros con estimadores basados en el diseño, es decir, estimadores válidos para muestras seleccionadas mediante un diseño muestral arbitrario. No obstante, en la Sección 4 se definen los estimadores comentados anteriormente para muestreo aleatorio simple, el diseño muestral más conocido y utilizado en la práctica. Los distintos métodos de estimación se compararán numéricamente en la Sección 5 con datos reales extraídos de la Encuesta de Presupuestos Familiares. Utilizaremos el sesgo relativo y el error cuadrático medio relativo como criterios para la comparación empírica de los distintos estimadores. Este estudio se completará con un conjunto de poblaciones con distintas características y generadas a partir de distintas distribuciones. El objetivo de este nuevo estudio es analizar los métodos de estimación presentados en este trabajo en distintos escenarios y en función de distintos aspectos que pueden presentarse en la práctica. El uso de información auxiliar en la etapa de estimación es una técnica muy utilizada que puede mejorar considerablemente la estimación de un determinado parámetro. En la Sección 6 se plantea la estimación de los parámetros objeto de estudio en este trabajo con información auxiliar en la etapa de estimación y se comentan los inconvenientes que tiene esta técnica cuando se aplica a la población de pobres. Para finalizar, las conclusiones más relevantes de este trabajo se describen en la Sección 7.

2. DEFINICIÓN DE UMBRAL DE POBREZA Y POBLACIÓN DE POBRES

Consideramos una población finita U con N individuos en la cual se desea estudiar una variable de interés y . En este trabajo supondremos que la variable y son los ingresos familiares por unidad de consumo. Esta variable es utilizada comúnmente para la determinación del umbral de pobreza, el cual delimita la población que se considera pobre de la población que no es considerada como tal. En concreto, el umbral de pobreza suele obtenerse a partir de un determinado cuantil de los ingresos por unidad de consumo. El cuantil poblacional de orden β de la variable y se define como

$$Y_{\beta} = \inf\{t : F(t) \geq \beta\},$$

donde

$$F(t) = \frac{1}{N} \sum_{i=1}^N \Delta(t - y_i)$$

es la función de distribución de la variable y en un argumento específico t , $\Delta(a) = 1$ si $a \geq 0$ y $\Delta(a) = 0$ en caso contrario. Numerosos organismos oficiales de estadística definen la línea o umbral de pobreza como $L = \alpha Y_\beta$. Por ejemplo, el Instituto Nacional de Estadística o EUROSTAT consideran $\alpha = 0.6$ y $\beta = 0.5$ para la determinación del umbral de pobreza, es decir, consideran el 60% de la mediana de ingresos por unidad de consumo.

Una vez determinado el umbral de pobreza, se considera que un individuo es pobre o pertenece a la población de pobres si su ingreso por unidad de consumo es menor o igual que el mencionado umbral de pobreza, es decir, $y_i \leq L$. Esta clasificación nos permite definir los valores de la variable y para la población de pobres como

$$y_i^* = \{y_i : y_i \leq L\} \quad i = 1, \dots, N$$

Supongamos que el número de valores que verifican esta condición a nivel poblacional es N^* , es decir, N^* es el número total de pobres en la población.

El primer objetivo de este trabajo es plantear diferentes metodologías para la estimación de la función de distribución en la propia población de pobres, es decir, estimar

$$F^*(t) = \frac{1}{N^*} \sum_{i=1}^{N^*} \Delta(t - y_i^*).$$

En segundo lugar y , a partir de la función de distribución anterior, plantearemos la estimación del cuantil de orden γ en la población de pobres, es decir, se abordará el problema de la estimación de

$$Y_\gamma^* = \inf \{t : F^*(t) \geq \gamma\}.$$

En la siguiente sección se plantea la estimación de los parámetros poblacionales anteriormente descritos a partir de una muestra seleccionada en la población U objeto de estudio.

3. ESTIMACIÓN DE LA FUNCIÓN DE DISTRIBUCIÓN Y CUANTILES EN LA POBLACIÓN DE POBRES

En esta sección se definen estimadores para la función de distribución en la población de pobres, $F^*(t)$, y para el cuantil de orden γ de la mencionada población de pobres, Y_γ^* . Con el fin de realizar esta estimación asumiremos que

en la población U se extrae una muestra s , de tamaño n y mediante un diseño muestral arbitrario con probabilidades de inclusión de primer orden dadas por π_i . En primer lugar habría que estimar el parámetro desconocido L , lo que nos permitirá conocer el número de pobres en la muestra. El estimador estándar del umbral de pobreza L viene dado por $\hat{L} = \alpha \hat{Y}_\beta$, donde

$$\hat{Y}_\beta = \inf \{t : \hat{F}(t) \geq \beta\}$$

es el estimador estándar para Y_β y $\hat{F}(t)$ es un estimador para $F(t)$. Se pueden plantear dos estimadores basados en el diseño muestral para $F(t)$. El estimador de tipo Horvitz-Thompson (Horvitz y Thompson, 1952) viene dado por

$$\hat{F}_{HT}(t) = \frac{1}{N} \sum_{i=1}^n d_i \Delta(t - y_i),$$

donde $d_i = \pi_i^{-1}$ es el peso muestral basado en el diseño para el i -ésimo individuo. Otra opción para estimar $F(t)$ es utilizar un estimador de tipo Hájek (Hájek, 1964), que en el caso de la función de distribución dicho estimador vendrá dado por

$$\hat{F}_H(t) = \frac{1}{\hat{N}} \sum_{i=1}^n d_i \Delta(t - y_i),$$

donde $\hat{N} = \sum_{i=1}^n d_i$. Cuando el argumento t sea elevado puede resultar conveniente utilizar estimadores de tipo Hájek, puesto que se verifica

$$\lim_{t \rightarrow \infty} \hat{F}_H(t) = 1,$$

mientras que el estimador de tipo Horvitz-Thompson podría dar estimaciones por encima de 1 en esta situación.

Los valores de la variable de interés para los individuos que son considerados pobres a nivel muestral serán expresados como

$$\hat{y}_i^* = \{y_i : y_i \leq \hat{L}\} \quad i = 1, \dots, n,$$

mientras que n^* será el número total de pobres en la muestra. De esta forma, el estimador de tipo Horvitz-Thompson para $F^*(t)$ viene dado por

$$\hat{F}_{HT}^*(t) = \frac{1}{N^*} \sum_{i=1}^{n^*} d_i \Delta(t - \hat{y}_i^*).$$

Podemos observar que $\hat{F}_{HT}^*(t)$ no podría obtenerse en la práctica, puesto que el número de pobres en la población, N^* , es desconocido. En este caso, proponemos estimar N^* del siguiente modo. Puesto que la fracción de muestreo se define como $f = n/N$, el tamaño de la población puede obtenerse como $N = n/f$. De esta forma, el tamaño en la población de pobres podría estimarse como $\tilde{N}^* = n^*/f$, lo que nos daría el siguiente estimador alternativo para $F^*(t)$

$$\tilde{F}^*(t) = \frac{1}{\tilde{N}^*} \sum_{i=1}^{n^*} d_i \Delta(t - \hat{y}_i^*).$$

El estimador de tipo Hájek para $F^*(t)$ viene dado por

$$\hat{F}_H^*(t) = \frac{1}{\hat{N}^*} \sum_{i=1}^{n^*} d_i \Delta(t - \hat{y}_i^*),$$

donde $\hat{N}^* = \sum_{i=1}^{n^*} d_i$.

Al igual que en el caso de la función de distribución, no se podría utilizar en la práctica el estimador de tipo Horvitz-Thompson para estimar Y_γ^* . En su lugar, proponemos utilizar el estimador

$$\tilde{Y}_\gamma^* = \inf \{t : \tilde{F}^*(t) \geq \gamma\}.$$

Por otra parte, el estimador de tipo Hájek para Y_γ^* viene dado por

$$\hat{Y}_\gamma^* = \inf \{t : \hat{F}_H^*(t) \geq \gamma\}.$$

4. ESTIMACIÓN BAJO MUESTREO ALEATORIO SIMPLE

En la sección anterior hemos considerado que la muestra s es seleccionada mediante un diseño muestral arbitrario. Sin embargo, el muestreo más conocido y utilizado es el muestreo aleatorio simple. En esta sección se formularán los estimadores para este caso particular.

En el caso de muestreo aleatorio simple, las probabilidades de inclusión de primer orden vienen dadas por $\pi_i = n/N$, y de ahí que los pesos básicos del diseño sean $d_i = N/n$. En esta situación, el estimador de tipo Horvitz-Thompson tiene la siguiente expresión

$$\hat{F}_{HT}(t) = \frac{1}{N} \sum_{i=1}^n d_i \Delta(t - y_i) = \frac{1}{n} \sum_{i=1}^n \Delta(t - y_i),$$

es decir, el estimador de tipo Horvitz-Thompson es simplemente la media aritmética de la variable $z = \Delta(t - y)$. Además se verifica $\hat{N} = \sum_{i=1}^n d_i = N$, lo que implica, tal como es conocido, que los estimadores de tipo Horvitz-Thompson y Hájek coinciden bajo muestreo aleatorio simple, y a su vez esto implica que el umbral de pobreza y la población de pobres se calcularán del mismo modo tanto usando un método como el otro. Para el caso de la población de pobres se verifica que

$$\hat{N}^* = \sum_{i=1}^{n^*} d_i = \frac{N}{n} n^*,$$

y el estimador de tipo Hájek para $F^*(t)$ también se calcularía a partir de una media aritmética, en concreto

$$\hat{F}_H^*(t) = \frac{1}{n^*} \sum_{i=1}^{n^*} \Delta(t - \hat{y}_i^*).$$

Bajo muestreo aleatorio simple, el estimador de tipo Horvitz-Thompson para $F^*(t)$ tampoco podría obtenerse al ser N^* desconocido, aunque podría calcularse el estimador $\tilde{F}^*(t)$ anteriormente definido, el cual coincide con $\hat{F}_H^*(t)$, tal como puede comprobarse a continuación

$$\tilde{F}^*(t) = \frac{1}{\tilde{N}^*} \sum_{i=1}^{n^*} d_i \Delta(t - \hat{y}_i^*) = \frac{f}{n^*} \frac{N}{n} \sum_{i=1}^{n^*} \Delta(t - \hat{y}_i^*) = \frac{1}{n^*} \sum_{i=1}^{n^*} \Delta(t - \hat{y}_i^*) = \hat{F}_H^*(t).$$

Esta última equivalencia implica que los estimadores \hat{Y}_γ^* y \tilde{Y}_γ^* para el cuantil poblacional Y_γ^* también coinciden.

5. ESTUDIOS DE SIMULACIÓN MONTE CARLO

En esta sección se utiliza el método de simulación de Monte Carlo para comparar numéricamente los distintos estimadores definidos en este trabajo, tanto para el cuantil de orden γ , como para la función de distribución correspondiente. Para ello se utilizan dos medidas de comparación: el Sesgo Relativo (*SR*) y el Error Cuadrático Medio Relativo (*ECMR*) definidos para un determinado estimador $\hat{\theta}$ del parámetro θ , de la siguiente manera:

$$SR = \frac{E[\hat{\theta}] - \theta}{\theta}; \quad ECMR = \frac{\sqrt{ECM[\hat{\theta}]}}{\theta},$$

donde $E[\hat{\theta}]$ y $ECM[\hat{\theta}]$ denotan, respectivamente el valor esperado y el error cuadrático medio del estimador $\hat{\theta}$, es decir,

$$E[\hat{\theta}] = \frac{1}{D} \sum_{d=1}^D \hat{\theta}_d; \quad ECM[\hat{\theta}] = \frac{1}{D} \sum_{d=1}^D (\hat{\theta}_d - \theta)^2,$$

donde $D = 10000$ es el número total de muestras seleccionadas en cada estudio de simulación y $\hat{\theta}_d$ denota el estimador $\hat{\theta}$ cuando es calculado en la d -ésima simulación.

Para el estudio de simulación se han utilizado varias poblaciones con distintas características. En primer lugar, parece razonable utilizar datos reales cuya variable de interés sean los ingresos, con el fin de analizar el comportamiento de los estimadores presentados en este trabajo en una aplicación real. Para ello, se ha considerado una muestra de 3114 datos procedentes de la Encuesta de Presupuestos Familiares (EPF) llevada a cabo por el INE, los cuales han sido utilizados como una población de tamaño $N = 3114$. La variable principal en esta población son los ingresos familiares por unidad de consumo y el objetivo será determinar, para cada muestra seleccionada, los individuos que se encuentran en pobreza y, a partir de esta información, estimar la función de distribución y cuantiles en la población de pobres. Un histograma de esta población puede consultarse en la Figura 1. Por la naturaleza de este tipo de poblaciones parece razonable asumir que tales poblaciones serán asimétricas hacia la derecha. De hecho, el coeficiente de asimetría de Fisher en la población EPF es $g_1 = 2.37$. Con el fin de analizar distintas características de este tipo de poblaciones, se han generado 3 poblaciones con tamaño $N = 1000$ cada una de ellas, con distribución Gamma con la misma media ($\mu = 10$) y distintas varianzas ($\sigma^2 = \{20, 40, 60\}$). En otras palabras, se han generado las siguientes distribuciones Gamma:

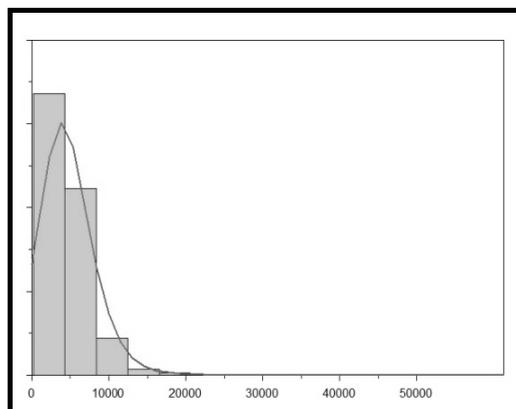
$$\gamma\left(\frac{10}{2}, 2\right); \quad \gamma\left(\frac{10}{4}, 4\right); \quad \gamma\left(\frac{10}{6}, 6\right).$$

Los coeficientes de asimetría de Fisher en estas poblaciones son, respectivamente, $g_1 = \{0.84, 1.35, 1.57\}$.

En este estudio comprobaremos que el grado de asimetría de la distribución en estudio juega un papel muy importante en la precisión de los estimadores de cuantiles en la población de pobres, especialmente para órdenes pequeños. En concreto, la estimación de cuantiles en la población de pobres y para este tipo

de poblaciones (asimétricas a la derecha) hacen que las estimaciones no presenten sesgos muy elevados, incluso para cuantiles de órdenes muy pequeños. Sin embargo, parece razonable pensar que los estimadores de cuantiles con órdenes pequeños pueden tener sesgos más elevados en poblaciones más simétricas, puesto que los parámetros en estas situaciones (los cuantiles en la población de pobres con órdenes pequeños) tienen un mayor rango de posibles valores que pueden tomar. Aunque estas situaciones no son muy probables de presentarse en la práctica, en especial para el caso donde la variable principal son los ingresos, merece la pena analizar a partir de qué orden de cuantil las estimaciones se vuelven más inestables y presentan grandes sesgos y, por tanto, tenemos el riesgo de encontrarnos en la práctica con estimaciones menos precisas. Para corroborar este hecho y medir la precisión de los estimadores en estas situaciones extremas (en el sentido de que es menos común encontrarse distribuciones simétricas cuando la variable en estudio sean los ingresos), también se han generado tres distribuciones Normales con tamaño $N = 1000$ cada una de ellas, con los mismos parámetros media y varianza utilizados en la distribución Gamma, es decir, con media $\mu = 10$ y varianzas $\sigma^2 = \{20, 40, 60\}$. En la Figura 2 pueden consultarse los histogramas correspondientes a las distribuciones Normales y distribuciones Gamma comentadas anteriormente.

Figura 1
Histograma de la población EPF



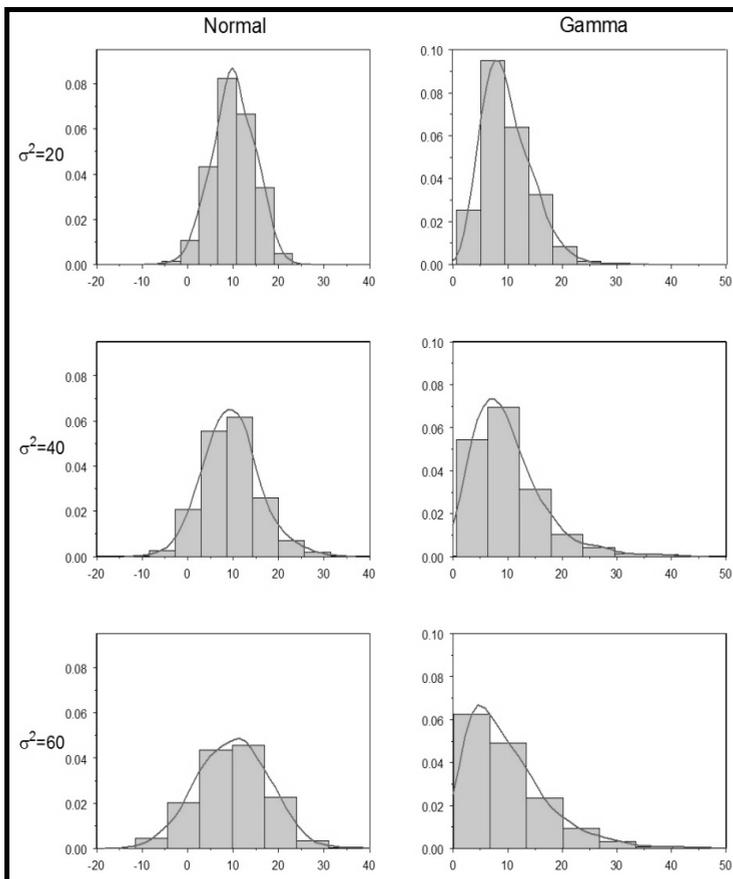
Fuente: Elaboración propia.

Se han utilizado dos esquemas de muestreo diferentes: muestreo aleatorio simple (Tablas 1, 3 y 5) y muestreo con probabilidades de selección desiguales utilizando el método de Midzuno (Tablas 2, 4 y 6). El método de Midzuno para la selección de muestras puede consultarse, por ejemplo, en Särndal et al.

(1992) y Singh (2003). Los estudios de simulación se han realizado calculando el umbral de pobreza del mismo modo que numerosos organismos oficiales de estadística, es decir, con $\alpha = 0.6$ y $\beta = 0.5$. Alternativamente se ha considerado el caso $\alpha = 1$ para la población EPF. Destacamos que cuando $\alpha = 1$, el cuantil de orden γ de la población de pobres también es un cuantil de la población general, en concreto es el cuantil de orden $\gamma/2$. Por tanto, el problema se simplifica y sólo sería necesario estudiar la población general. No obstante, consideramos este caso con el fin de analizar el efecto de los estimadores con respecto a α . Del mismo modo, se analiza el comportamiento de los estimadores respecto a distintos valores de g_1 , γ y la fracción de muestreo $f = n/N$.

Figura 2

Histogramas de las poblaciones generadas mediante distribución Normal y distribución Gamma. En todos los casos $\mu = 10$



Fuente: Elaboración propia.

Asumiendo muestreo aleatorio simple, la Tabla 1 muestra los valores, expresados en porcentaje, de SR y $ECMR$ para los estimadores de tipo Hájek de los parámetros Y_γ^* y $F^*(t)$ cuando la población objeto de estudio es la obtenida a través de la Encuesta de Presupuestos Familiares. Observando los valores obtenidos, podemos decir que en lo que respecta al sesgo del estimador \hat{Y}_γ^* se han obtenido valores muy pequeños, prácticamente insignificantes, con sesgos menores al 1% y en valor absoluto, siendo algo mayores estos valores cuando $\alpha = 0.6$ y $\gamma = 0.1$, y en todas las fracciones de muestreo. Con $\alpha = 1$ los valores de SR son muy pequeños, obteniéndose el valor mayor (1.8% en valor absoluto) para $\gamma = 0.1$ y $f = 5\%$.

En cuanto al sesgo del estimador $\hat{F}^*(t)$, se observan también sesgos muy pequeños, con valores nunca mayores del 2.5%. Los valores de SR más elevados suelen observarse cuando la fracción de muestreo es pequeña ($f = 5\%$). En general, los valores del sesgo son menores cuando $\alpha = 1$ que cuando $\alpha = 0.6$.

En lo que respecta al error cuadrático medio de ambos estimadores, podemos observar, como resulta razonable, que los estimadores presentan un menor valor de $ECMR$ a medida que la fracción de muestreo aumenta, siendo esta conclusión también válida para los valores de γ . Este patrón también se repite tanto para $\alpha = 0.6$ como para $\alpha = 1$. También podemos concluir que los estimadores tienen menor $ECMR$ cuando $\alpha = 1$ que cuando $\alpha = 0.6$.

Asumiendo un diseño muestral con probabilidades de selección desiguales, la Tabla 2 muestra los valores, expresados en porcentaje, de SR y $ECMR$ para los diferentes estimadores de los parámetros Y_γ^* y $F^*(t)$ estudiados en este trabajo, cuando la población objeto de estudio es la obtenida a través de la Encuesta de Presupuestos Familiares.

En lo que respecta al sesgo de los estimadores de Y_γ^* podemos comprobar que se han obtenido valores muy pequeños y prácticamente similares en ambos estimadores. Los valores mayores (en torno al 2% en valores absolutos) se obtienen cuando $\alpha = 0.6$, $\beta = 0.5$ y $\gamma = 0.1$. En cuanto al sesgo de los estimadores de $F^*(t)$, se observan sesgos relativos en torno al 2% cuando $\alpha = 0.6$, $\beta = 0.5$ y $f = 5\%$, excepto cuando $\gamma = 0.9$. En el resto de los casos los sesgos son insignificantes para ambos estimadores.

Tabla 1

Valores de SR y ECMR de los estimadores de tipo Hájek para Y_γ^* y $\hat{F}^*(t)$, con

$$t = Y_\gamma^* \text{ en la población EPF}$$

Muestras seleccionadas mediante muestreo aleatorio simple

α	β	γ	f	SR (%)		ECMR (%)	
				\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.6	0.5	0.1	0.05	-1.4	1.8	17.5	55.5
			0.10	-1.3	0.1	12.6	37.4
			0.20	-1.7	0.3	9.3	24.9
	0.3	0.05	0.05	0.0	2.5	8.7	30.5
			0.10	0.4	0.3	5.8	20.5
			0.20	0.4	0.2	3.7	13.4
	0.5	0.05	0.05	-0.1	2.4	7.8	21.7
			0.10	0.1	0.6	5.5	14.5
			0.20	-0.2	0.4	3.6	9.6
	0.7	0.05	0.05	-0.7	2.2	7.0	16.9
			0.10	-0.1	0.8	4.7	11.4
			0.20	0.1	0.4	3.1	7.5
0.9	0.05	0.05	-1.0	0.2	6.1	10.8	
		0.10	-0.5	0.2	4.2	8.1	
		0.20	-0.3	0.3	2.7	5.7	
1.0	0.5	0.1	0.05	-1.8	0.8	10.1	34.1
			0.10	-0.7	-0.5	6.9	23.6
			0.20	-0.3	-0.1	4.5	15.7
	0.3	0.05	0.05	0.3	0.2	8.0	18.6
			0.10	0.0	-0.5	5.5	12.9
			0.20	-0.1	0.0	3.7	8.6
	0.5	0.05	0.05	-0.7	0.2	6.6	13.5
			0.10	-0.2	-0.4	4.5	9.4
			0.20	-0.1	-0.1	2.9	6.3
	0.7	0.05	0.05	0.1	0.2	5.6	10.6
			0.10	0.5	-0.4	3.9	7.3
			0.20	0.3	-0.2	2.5	4.9
0.9	0.05	0.05	0.6	-0.3	5.5	7.8	
		0.10	0.5	-0.4	3.8	5.8	
		0.20	0.2	-0.2	2.5	4.0	

Fuente: Elaboración propia.

En cuanto al ECMR de los estimadores de Y_γ^* , podemos observar que se obtienen valores prácticamente iguales para ambos estimadores. Se verifica que los estimadores tienen menor ECMR cuanto mayor es la fracción de muestreo y

cuanto mayor es γ . También se obtienen estimadores con menor ECMR cuando $\alpha = 1$ que cuando $\alpha = 0.6$. Conclusiones similares son válidas para los estimadores de $F^*(t)$.

Tabla 2

Valores de SR y ECMR de los diferentes estimadores de Y_γ^* y $\hat{F}^*(t)$, con $t = Y_\gamma^*$ en la población EPF
Muestras seleccionadas con probabilidades desiguales mediante el Método de Midzuno

α	β	γ	f	SR (%)				ECMR (%)			
				\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$
0.6	0.5	0.1	0.05	-2.4	-2.6	2.0	2.3	18.0	18.1	55.9	56.2
			0.10	-1.7	-1.8	0.4	0.6	12.6	12.6	37.2	37.3
			0.20	-1.8	-1.8	0.5	0.5	9.4	9.4	25.0	25.1
	0.3	0.05	0.05	-0.2	-0.2	1.8	2.2	8.8	8.8	30.5	30.7
			0.10	0.2	0.2	0.6	0.8	5.8	5.8	20.4	20.4
			0.20	0.3	0.3	0.3	0.4	3.8	3.8	13.5	13.5
	0.5	0.05	0.05	-0.6	-0.7	2.0	2.3	7.7	7.7	21.6	21.7
			0.10	-0.2	-0.2	0.4	0.6	5.4	5.4	14.5	14.5
			0.20	-0.2	-0.2	0.4	0.5	3.6	3.6	9.6	9.6
	0.7	0.05	0.05	-0.7	-0.7	2.1	2.4	7.0	7.0	16.8	16.9
			0.10	0.0	0.0	0.5	0.7	4.7	4.7	11.3	11.4
			0.20	0.0	0.0	0.5	0.5	3.1	3.1	7.5	7.5
0.9	0.05	0.05	-1.1	-1.2	0.2	0.5	6.1	6.2	10.8	10.8	
		0.10	-0.4	-0.4	0.0	0.2	4.1	4.1	8.1	8.1	
		0.20	-0.3	-0.3	0.3	0.4	2.7	2.7	5.7	5.7	
1.0	0.5	0.1	0.05	-1.6	-1.6	0.5	0.8	9.9	9.9	33.9	34.0
			0.10	-0.6	-0.6	-0.4	-0.2	6.8	6.8	23.1	23.1
			0.20	-0.2	-0.2	-0.3	-0.2	4.5	4.5	15.7	15.7
	0.3	0.05	0.05	0.5	0.5	0.0	0.3	8.1	8.1	18.7	18.7
			0.10	-0.1	-0.1	-0.2	-0.1	5.5	5.5	12.7	12.7
			0.20	-0.1	-0.1	-0.1	0.0	3.7	3.7	8.7	8.7
	0.5	0.05	0.05	-0.4	-0.4	-0.2	0.0	6.7	6.7	13.6	13.6
			0.10	-0.1	-0.1	-0.4	-0.3	4.5	4.5	9.3	9.3
			0.20	-0.1	-0.1	-0.2	-0.2	3.0	3.0	6.3	6.3
	0.7	0.05	0.05	0.3	0.3	-0.2	0.0	5.8	5.8	10.7	10.8
			0.10	0.6	0.6	-0.5	-0.3	3.8	3.8	7.3	7.3
			0.20	0.3	0.3	-0.2	-0.2	2.6	2.5	4.9	4.9
0.9	0.05	0.05	0.8	0.6	-0.6	-0.4	5.6	5.7	8.0	8.0	
		0.10	0.6	0.6	-0.5	-0.4	3.9	3.9	5.8	5.8	
		0.20	0.2	0.2	-0.2	-0.1	2.6	2.6	4.0	4.0	

Fuente: Elaboración propia.

La Tabla 3 muestra, nuevamente, los valores expresados en porcentaje de SR y $ECMR$ para los estimadores de tipo Hájek de los parámetros Y_γ^* y $F^*(t)$ bajo muestreo aleatorio simple, cuando la población bajo estudio tiene distribución Normal. Podemos observar que en estas poblaciones, donde la distribución es simétrica, los sesgos en el problema de estimación de cuantiles son, en valor absoluto, extremadamente elevados para valores de γ menores que 0.3, incluso para tamaños muestrales más elevados. La justificación de este hecho, en comparación con la población EPF estudiada en tablas anteriores, es que al tener la cola de la izquierda de la distribución una mayor dispersión, las estimaciones del cuantil con órdenes pequeños se vuelven más inestables, lo que provoca importantes sesgos. A partir de la Tabla 3 podemos observar que se resuelve considerablemente este problema a medida que se aumenta el tamaño muestral. También se puede observar que los sesgos toman valores más razonables (en general menores del 5%) para valores de γ mayores o iguales que 0.5. Conclusiones similares a las comentadas sobre el sesgo pueden derivarse para el ECMR del estimador de cuantiles, es decir, el hecho de trabajar con distribuciones simétricas puede provocar estimaciones muy poco precisas con un error cuadrático medio extremadamente elevado.

Los problemas observados para la estimación de cuantiles en la población de pobres, especialmente para valores pequeños de γ , no se observan en el problema de la estimación de la función de distribución. En este caso, todo los sesgos relativos están siempre por debajo, en valor absoluto, del 4%, mientras que los valores de ECMR son razonables.

Tabla 3

Valores de SR y $ECMR$ de los estimadores de tipo Hájek para Y_γ^* y $\hat{F}_H^*(t)$, con

$$t = Y_\gamma^* \text{ en la población Normal, } \alpha = 0.6, \beta = 0.5$$

Muestras seleccionadas mediante muestreo aleatorio simple

σ^2	γ	f	SR (%)		$ECMR$ (%)	
			\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
20	0.1	0.10	-229.60	3.34	479.50	69.60
		0.20	6.00	1.67	160.50	45.80
	0.3	0.10	-3.60	2.77	27.70	35.60
		0.20	-2.60	2.02	17.30	23.70
	0.5	0.10	-4.50	2.51	16.90	24.70
		0.20	-3.20	2.25	11.90	16.60
0.7	0.10	-2.70	2.62	11.80	17.60	
	0.20	-1.90	2.11	8.00	12.00	
0.9	0.10	-3.10	1.59	8.40	10.70	
	0.20	-2.00	1.55	5.80	8.10	

Tabla 3 (continuación)

Valores de SR y ECMR de los estimadores de tipo Hájek para Y_γ^* y $\hat{F}^*(t)$, con

$$t = Y_\gamma^* \text{ en la población Normal, } \alpha = 0.6, \beta = 0.5$$

Muestras seleccionadas mediante muestreo aleatorio simple

σ^2	γ	f	SR (%)		ECMR (%)		
			\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$	
40	0.1	0.10	1.90	2.51	-91.60	58.60	
		0.20	-3.60	1.42	-57.10	39.00	
	0.3	0.10	-7.00	2.17	71.40	30.90	
		0.20	-6.40	1.63	48.50	20.50	
	0.5	0.10	-3.70	2.36	27.70	21.30	
		0.20	-0.90	1.44	19.30	14.30	
	0.7	0.10	-2.60	2.40	14.30	15.70	
		0.20	-1.30	1.41	8.90	10.70	
	0.9	0.10	-3.40	1.29	10.70	10.00	
		0.20	-2.30	1.30	7.50	7.50	
	60	0.1	0.10	-4.50	0.99	-34.40	52.20
			0.20	-2.40	0.36	-22.10	34.70
0.3		0.10	-127.40	1.35	478.30	27.50	
		0.20	-84.10	0.35	309.60	18.30	
0.5		0.10	-0.70	1.26	45.60	18.90	
		0.20	2.20	0.38	31.30	12.80	
0.7		0.10	-5.10	1.45	20.50	14.10	
		0.20	-3.10	0.60	14.20	9.50	
0.9		0.10	-2.30	0.59	12.10	9.20	
		0.20	-1.00	0.48	8.40	6.90	

Fuente: Elaboración propia.

En la Tabla 4 se repite el estudio en las poblaciones con distribución Normal pero se toman muestras con probabilidades de selección desiguales. Al igual que en el caso de muestreo aleatorio simple, las estimaciones de cuantiles son poco precisas, tanto en términos de SR como en términos de ECMR, para los casos donde γ es menor o igual que 0.3, y tales valores son razonables en el resto de situaciones. Los sesgos asociados a los estimadores de la función de distribución son siempre menores de 4%. En este caso se puede apreciar que el estimador \hat{Y}_γ^* es ligeramente mejor, en algunas ocasiones, que el estimador \tilde{Y}_γ^* , aunque la diferencia no es significativa.

Tabla 4

Valores de SR y ECMR de los diferentes estimadores de Y_γ^* y $\hat{F}^*(t)$, con $t = Y_\gamma^*$ en la población Normal, $\alpha = 0.6$, $\beta = 0.5$

Muestras seleccionadas con probabilidades desiguales mediante el Método de Midzuno.

σ^2	γ	f	SR (%)				ECMR (%)				
			\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	
20	0.1	0.10	-270.70	-290.80	1.78	2.35	497.20	516.80	68.60	69.10	
		0.20	2.60	0.70	1.59	1.84	158.60	161.90	45.00	45.10	
	0.3	0.10	-5.20	-5.30	2.67	3.25	28.30	28.30	35.80	36.10	
		0.20	-2.90	-2.90	1.51	1.76	17.70	17.70	23.60	23.70	
	0.5	0.10	-4.30	-4.40	2.41	2.98	16.90	16.90	24.70	24.90	
		0.20	-2.70	-2.70	1.56	1.81	11.80	11.80	16.60	16.60	
	0.7	0.10	-2.40	-2.50	2.52	3.09	12.00	12.10	17.60	17.80	
		0.20	-1.70	-1.70	1.69	1.94	8.00	8.00	12.10	12.10	
	0.9	0.10	-3.30	-3.50	1.19	1.76	8.70	8.90	10.90	11.10	
		0.20	-2.00	-2.00	1.42	1.67	5.80	5.80	8.10	8.20	
	40	0.1	0.10	2.60	7.90	1.24	1.83	-93.00	-98.80	58.40	58.80
			0.20	-5.90	-3.90	0.77	1.03	-56.40	-56.80	38.10	38.20
0.3		0.10	-8.00	-8.40	1.89	2.48	71.60	72.00	30.90	31.20	
		0.20	-7.20	-7.30	1.26	1.52	47.80	47.80	20.10	20.20	
0.5		0.10	-2.80	-3.00	1.79	2.38	27.80	27.90	21.30	21.50	
		0.20	-0.70	-0.70	1.37	1.63	19.10	19.10	14.20	14.30	
0.7		0.10	-2.20	-2.60	1.76	2.36	14.40	14.60	15.60	15.80	
		0.20	-1.40	-1.50	1.37	1.64	8.90	8.90	10.60	10.70	
0.9		0.10	-3.00	-3.40	0.86	1.45	10.60	10.80	10.00	10.20	
		0.20	-2.30	-2.50	1.23	1.49	7.50	7.60	7.50	7.60	
60		0.1	0.10	-6.70	-4.60	1.31	1.90	-34.60	-34.60	51.90	52.30
			0.20	-2.90	-1.90	0.90	1.16	-22.00	-21.90	34.10	34.20
	0.3	0.10	-126.80	-140.20	0.96	1.56	477.50	490.80	27.70	27.90	
		0.20	-87.80	-93.30	0.32	0.58	306.90	311.80	18.10	18.10	
	0.5	0.10	1.00	0.00	0.87	1.47	44.90	45.90	19.00	19.20	
		0.20	2.30	2.10	0.38	0.64	30.70	30.90	12.70	12.80	
	0.7	0.10	-5.10	-5.80	0.86	1.46	20.70	21.10	14.00	14.20	
		0.20	-3.10	-3.40	0.37	0.63	14.20	14.30	9.50	9.50	
	0.9	0.10	-1.80	-2.30	0.29	0.88	12.00	12.10	9.20	9.30	
		0.20	-0.80	-1.00	0.25	0.51	8.40	8.40	6.90	6.90	

Fuente: Elaboración propia.

Con el análisis de los resultados obtenidos en las Tablas 5 y 6 podremos confirmar que la asimetría hacia la derecha que presentan, en general, variables de perfil económico como son los ingresos, favorece que las estimaciones de cuantiles en la población de pobres y para órdenes pequeños sean más precisas en comparación con distribuciones más simétricas. En efecto, en la Tabla 5 (caso de muestreo aleatorio simple) observamos que tan sólo en una ocasión (cuando $\sigma^2 = 20$, $\gamma = 0.1$ y $f = 0.1$) se obtiene un valor de SR elevado (-17.6 %). Este sesgo puede deberse a que el coeficiente de asimetría de Fisher no es muy elevado ($g_1 = 0.84$), puesto que en el resto de poblaciones con distribución Gamma los sesgos están dentro de un rango razonable, con valores, en general,

menores al 5% en valor absoluto. Respecto al ECMR, si se puede observar que no hay ningún caso con valores extremadamente elevados, tal como ocurría en el caso de las poblaciones con distribución Normal. Estas conclusiones pueden derivarse de nuevo a partir de la Tabla 6, en la cual las muestras son seleccionadas con probabilidades desiguales.

Tabla 5

Valores de SR y ECMR de los estimadores de tipo Hájek para Y_γ^* y $\hat{F}_H^*(t)$, con $t = Y_\gamma^*$ en la población Gamma, $\alpha = 0.6$, $\beta = 0.5$

Muestras seleccionadas mediante muestreo aleatorio simple

σ^2	γ	f	SR (%)		ECMR (%)	
			\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
20	0.1	0.10	-17.60	0.73	34.80	79.30
		0.20	1.70	0.99	14.40	51.60
	0.3	0.10	0.80	1.47	10.30	43.70
		0.20	0.60	1.23	6.30	28.80
	0.5	0.10	0.50	1.93	9.60	31.60
		0.20	0.80	1.08	6.70	21.40
	0.7	0.10	-2.00	1.93	9.20	23.90
		0.20	-1.50	1.11	6.70	16.70
0.9	0.10	-1.90	-1.86	7.30	15.40	
	0.20	-1.00	-0.69	5.00	11.70	
40	0.1	0.10	4.20	2.77	34.00	61.50
		0.20	3.80	0.73	22.60	40.10
	0.3	0.10	0.80	2.13	19.20	33.20
		0.20	0.70	0.33	13.20	22.00
	0.5	0.10	-0.50	2.01	13.60	23.70
		0.20	0.90	0.52	9.00	15.50
	0.7	0.10	0.20	2.17	10.60	18.20
		0.20	1.10	0.63	7.10	11.90
0.9	0.10	-1.30	0.12	8.50	11.40	
	0.20	-0.20	0.31	5.60	8.40	
60	0.1	0.10	5.20	4.55	36.70	56.90
		0.20	1.60	2.57	24.40	37.00
	0.3	0.10	-0.40	4.60	17.70	31.70
		0.20	-0.20	2.10	10.70	20.40
	0.5	0.10	-5.20	4.53	15.70	23.40
		0.20	-3.40	2.25	11.00	15.20
	0.7	0.10	-1.60	4.26	11.50	18.30
		0.20	-0.40	2.18	7.40	12.10
0.9	0.10	-2.10	1.01	10.30	10.30	
	0.20	-0.90	1.05	6.70	8.10	

Fuente: Elaboración propia.

En conclusión, cuando se asume un diseño muestral con probabilidades de selección desiguales, los dos estimadores para la estimación de los parámetros Y_γ^* y $F^*(t)$ sugeridos en este trabajo y para las poblaciones en estudio tienen un comportamiento muy similar, tanto en términos de SR como en términos de ECMR. En la práctica, los estimadores basados en \tilde{N}^* son más simples al no depender de las probabilidades de inclusión. Sin embargo, los estimadores basados en \hat{N}^* poseen propiedades teóricas deseables. Por ejemplo, según Hájek (1964) los estimadores basados en \hat{N}^* son asintóticamente insesgados. Por tanto, en caso de que resulte posible, se recomienda utilizar los estimadores basados en \hat{N}^* , aunque el comportamiento empírico con respecto a los estimadores basados en \tilde{N}^* es bastante similar. Por otra parte, se ha observado que los estimadores de cuantiles en la población de pobres pueden ser poco precisos con sesgos realmente importantes cuando γ sea pequeño, pero esta pérdida de precisión es más factible de aparecer cuando la distribución no sea asimétrica a la derecha. Afortunadamente, esta no es la situación más común para variables, como los ingresos por unidad de consumo, que son utilizadas para la determinación de la población de pobres.

Tabla 6

Valores de SR y ECMR de los diferentes estimadores de Y_γ^* y $\hat{F}^*(t)$, con $t = Y_\gamma^*$ en la población Gamma, $\alpha = 0.6$, $\beta = 0.5$

Muestras seleccionadas con probabilidades desiguales mediante el Método de Midzuno

σ^2	γ	f	SR (%)				ECMR (%)			
			\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$
20	0.1	0.10	-18.50	-18.50	0.67	1.20	34.90	34.90	78.00	78.40
		0.20	1.10	1.00	0.77	1.00	15.60	15.60	51.80	51.90
	0.3	0.10	0.60	0.50	0.95	1.47	10.30	10.30	43.60	43.80
		0.20	0.30	0.30	1.21	1.43	6.30	6.30	28.90	28.90
	0.5	0.10	0.80	0.60	1.34	1.86	9.70	9.70	31.60	31.80
		0.20	0.80	0.80	0.84	1.06	6.80	6.80	21.40	21.50
	0.7	0.10	-1.80	-1.80	1.59	2.11	9.10	9.20	23.80	24.00
		0.20	-1.40	-1.40	0.74	0.97	6.70	6.70	16.70	16.70
	0.9	0.10	-1.90	-2.00	-2.20	-1.70	7.40	7.50	15.90	15.90
		0.20	-1.00	-1.00	-0.92	-0.70	5.00	5.00	11.80	11.80

Tabla 6 (continuación)

Valores de SR y ECMR de los diferentes estimadores de Y_γ^* y $\hat{F}^*(t)$, con $t = Y_\gamma^*$ en la población Gamma, $\alpha = 0.6$, $\beta = 0.5$

Muestras seleccionadas con probabilidades desiguales mediante el Método de Midzuno

σ^2	γ	f	SR (%)				ECMR (%)				
			\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}^*(t)$	$\tilde{F}^*(t)$	
40	0.1	0.10	1.80	0.70	1.45	2.00	34.30	34.40	61.70	62.00	
		0.20	3.90	3.70	0.35	0.59	22.50	22.50	40.40	40.50	
	0.3	0.10	0.20	0.10	1.97	2.52	19.30	19.30	33.20	33.40	
		0.20	0.30	0.30	0.49	0.73	13.30	13.30	21.80	21.90	
	0.5	0.10	0.00	-0.30	1.72	2.27	13.60	13.60	23.90	24.10	
		0.20	0.90	0.90	0.65	0.88	8.90	8.90	15.50	15.50	
	0.7	0.10	0.40	0.30	1.63	2.18	10.50	10.50	18.10	18.30	
		0.20	1.00	1.00	0.67	0.91	7.10	7.10	11.80	11.90	
	0.9	0.10	-1.10	-1.30	-0.19	0.35	8.50	8.50	11.40	11.50	
		0.20	-0.20	-0.30	0.07	0.31	5.60	5.60	8.30	8.40	
	60	0.1	0.10	5.70	5.00	4.19	4.76	36.80	36.50	56.30	56.70
			0.20	2.40	2.20	1.73	1.98	24.70	24.60	36.80	36.90
0.3		0.10	-0.70	-0.80	4.00	4.57	17.70	17.70	31.60	31.80	
		0.20	-0.30	-0.40	1.68	1.93	10.80	10.80	20.50	20.60	
0.5		0.10	-4.50	-4.80	3.99	4.56	15.50	15.70	23.20	23.40	
		0.20	-3.00	-3.10	1.92	2.17	10.90	10.90	15.20	15.20	
0.7		0.10	-1.50	-1.80	3.54	4.11	11.50	11.50	18.00	18.20	
		0.20	-0.40	-0.50	1.82	2.07	7.40	7.40	11.90	12.00	
0.9		0.10	-1.50	-1.90	0.55	1.10	10.30	10.30	10.50	10.60	
		0.20	-0.60	-0.70	0.75	0.99	6.70	6.70	8.10	8.10	

Fuente: Elaboración propia.

6. ESTIMACIÓN CON INFORMACIÓN AUXILIAR

Cuando se lleva a cabo una encuesta por muestreo es común recabar información de más de una variable, además de la variable y objeto de estudio. Estas variables auxiliares pueden estar correlacionadas con la variable de interés y en este caso tales variables auxiliares pueden utilizarse en la fase de estimación de un determinado parámetro con el fin de mejorar la precisión de los correspondientes estimadores. Supongamos que existe una variable auxiliar x asociada con la variable principal y . En nuestro caso, esta variable auxiliar podría ser los gastos familiares o los propios ingresos familiares por unidad de consumo observados en un periodo anterior. Con el fin de utilizar la información auxiliar en la fase de estimación, es necesario conocer en primer lugar la relación existente entre ambas variables, y en este caso dicha relación se podrá estudiar a partir de un modelo econométrico. Chambers y Dunstan (1986) y Rao et al. (1990), entre

otros autores, analizaron estimadores basados en información auxiliar considerando el modelo

$$y_i = \beta_1 x_i + v(x_i) u_i \quad (i = 1, \dots, N), \quad (1)$$

donde β_1 es el parámetro desconocido, $v(x) = x^{1/2}$ y u_i son variables aleatorias independientes e idénticamente distribuidas con media cero. Notamos que por simplicidad consideraremos el modelo (1) en este trabajo, aunque fácilmente y a partir de la bibliografía existente se podría extender este estudio al caso de otros modelos más complejos.

Asumiendo el modelo (1) y para el problema de la estimación de la media poblacional, $\bar{Y} = N^{-1} \sum_{i=1}^N y_i$, el estimador comúnmente utilizado es el estimador de tipo razón, el cual viene dado por

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}} \bar{X},$$

donde \bar{X} denota la media poblacional de la variable auxiliar y \bar{y} y \bar{x} denotan, respectivamente, los estimadores de tipo Horvitz-Thompson para \bar{Y} y \bar{X} . Podemos observar que el estimador de tipo razón \bar{y}_r requiere conocer la media poblacional de la variable auxiliar, o bien que dicha cantidad se pueda estimar sin error. Notamos que esta suposición es comúnmente utilizada en el contexto de estimación con información auxiliar.

Para el problema de la estimación de la función de distribución, $F(t)$, y considerando de nuevo el modelo (1), Rao et al. (1990) propusieron el siguiente estimador de tipo razón

$$\hat{F}_r(t) = \frac{1}{N} \frac{\sum_{i=1}^n d_i \Delta(t - y_i)}{\sum_{i=1}^n d_i \Delta(t - \hat{R}x_i)} \sum_{i=1}^N \Delta(t - \hat{R}x_i),$$

donde $\hat{R} = \bar{y}/\bar{x}$ es el estimador basado en el diseño de la razón poblacional $R = \bar{Y}/\bar{X}$. Respecto al problema de estimación de un cuantil de orden β , Y_β , Rao et al. (1990) propusieron el siguiente estimador de tipo razón

$$\hat{Y}_{\beta,r} = \frac{\hat{Y}_\beta}{\hat{X}_\beta} X_\beta.$$

En resumen, los métodos de estimación existentes basados en información auxiliar se han centrado en la estimación de parámetros como medias, totales,

funciones de distribución, cuantiles, proporciones, varianzas, etc. Actualmente no se ha planteado la estimación de la función de distribución y cuantiles para la población de pobres incorporando información auxiliar en la etapa de estimación. A continuación se plantea esta cuestión, y se describen los problemas relacionados con la estimación con información auxiliar de estos parámetros.

A partir de la variable x definimos los n^* valores muestrales

$$\hat{x}_i^* = \{x_i : y_i \leq \hat{L}\} \quad i = 1, \dots, n,$$

los cuales se pueden utilizar para estimar la función de distribución y cuantiles en la población de pobres y para la variable auxiliar. Para el caso de parámetros poblacionales de la variable auxiliar utilizaremos el conjunto de valores

$$x_i^* = \{x_i : y_i \leq L\} \quad i = 1, \dots, N. \quad (2)$$

Una vez definida la variable auxiliar para la población de pobres, se podrían aplicar los métodos existentes basados en información auxiliar para la estimación de la función de distribución y cuantiles en la población de pobres. Por ejemplo, siguiendo la definición del estimador $\hat{F}_r(t)$ anteriormente definido, se podría estimar la función de distribución para la población de pobres, $F^*(t)$, del siguiente modo

$$\hat{F}_r^*(t) = \frac{1}{N^*} \frac{\sum_{i=1}^{n^*} d_i \Delta(t - \hat{y}_i^*)}{\sum_{i=1}^{n^*} d_i \Delta(t - \hat{R}^* \hat{x}_i^*)} \sum_{i=1}^{N^*} \Delta(t - \hat{R}^* x_i^*),$$

donde $\hat{R}^* = \bar{y}^* / \bar{x}^*$. Podemos observar que el estimador $\hat{F}_r^*(t)$ no podría calcularse en la práctica, puesto que depende de los valores x_i^* definidos en (2), los cuales dependen a su vez de la variable y a nivel poblacional, y esta información poblacional es desconocida. Del mismo modo, $\hat{F}_r^*(t)$ depende del número de pobres en la población, N^* , y esta cantidad también es desconocida en la práctica. Una posible solución para no tener que utilizar N^* podría ser utilizar los valores

$$\hat{x}_i^* = \{x_i : y_i \leq \hat{L}\} \quad i = 1, \dots, N,$$

en lugar de x_i^* . Suponiendo que N_x^* es la longitud de la variable \hat{x}^* , podríamos utilizar esta cantidad en lugar de N^* . Sin embargo, la variable \hat{x}^* de nuevo depende de la variable y a nivel poblacional, y el estimador de la función de

distribución para la población de pobres así definido no se podría tampoco utilizar en la práctica.

Respecto a la estimación de cuantiles en la población de pobres se tiene el mismo problema. En este caso, el estimador de tipo razón para Y_γ^* vendría dado por

$$\hat{Y}_{\gamma,r}^* = \frac{\hat{Y}_\gamma^*}{\hat{X}_\gamma^*} X_\gamma^*,$$

donde $X_\gamma^* = \inf \{t : F_x^*(t) \geq \gamma\}$ y

$$F_x^*(t) = \frac{1}{N^*} \sum_{i=1}^{N^*} \Delta(t - x_i^*).$$

Se puede comprobar fácilmente que el estimador de tipo razón $\hat{Y}_{\gamma,r}^*$ para Y_γ^* no podría calcularse en la práctica, puesto que al igual que en el caso de la función de distribución, el mencionado estimador depende la variable principal a nivel poblacional a través de los valores x_i^* .

7. CONCLUSIONES

En este trabajo se ha planteado la estimación de la función de distribución y cuantiles en la población de pobres, definida según distintos organismos como el conjunto de individuos cuyos ingresos familiares por unidad de consumo se encuentran por debajo del umbral de pobreza. La estimación de estos parámetros se ha llevado a cabo considerando un diseño muestral arbitrario, es decir, los dos estimadores planteados en este trabajo se pueden aplicar a cualquier muestra en la cual se conozcan las probabilidades de inclusión de primer orden. No obstante, se han definido los estimadores para muestreo aleatorio simple, el diseño muestral más conocido y utilizado en la práctica.

Se ha evaluado el comportamiento empírico de los estimadores de la función de distribución y cuantiles en la población de pobres para distintas poblaciones con características diferentes. Este estudio se ha llevado a cabo mediante muestras con probabilidades iguales (muestreo aleatorio simple) y con probabilidades desiguales (utilizando el Método de Midzuno). Además se ha analizado el comportamiento de los estimadores en función de distintos factores: según valores de g_1 , α , γ y f ; aspectos que pueden afectar al comportamiento de los estimadores. Se ha considerado siempre $\beta = 0.5$ puesto que es el caso que siempre se tiene de referencia en los estudios de pobreza. Para distribuciones con una importante asimetría hacia la derecha se ha comprobado que, en gene-

ral, los estimadores tienen un buen comportamiento en términos de SR y de ECMR. En este trabajo puede consultarse además la evolución de la precisión de los estimadores en función de α , γ y f , lo que nos permite hacernos una idea de lo que puede ocurrir en otras poblaciones con características similares. En este trabajo también se ha observado que para distribuciones con un determinado grado de simetría, los estimadores de cuantiles con órdenes pequeños ($\gamma \leq 0.3$) para la población pobre pueden presentar sesgos relativamente elevados, con valores de ECMR también bastante elevados.

El uso de información auxiliar en la etapa de estimación es una técnica muy utilizada que puede mejorar la precisión de un estimador. Sin embargo, en este trabajo se comentan los problemas que surgen al utilizar esta técnica en la población de pobres y para los parámetros objeto de estudio. Como futura línea de investigación se podría solucionar el problema del uso de la información auxiliar en la estimación de parámetros relacionados con la población de pobres.

REFERENCIAS BIBLIOGRÁFICAS

- ATKINSON, A.B. (1987): "On the measurement of poverty". *Económica*, 55, (4), pp.749-764.
- BÁRCENA, E.; COWELL, F. (2006): "Static and dynamic poverty in Spain, 1993-2000". *Hacienda Pública Española*, 179, pp. 51-98.
- CHAMBERS, R.L.; DUNSTAN, R. (1986): "Estimating distribution functions from survey data". *Biométrica*, 73, pp. 597-604.
- DOMÍNGUEZ, J.; MARTÍN, A.M. (2006): "Medición de la pobreza: una revisión de los principales indicadores". *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 2, pp. 27-66.
- DOMÍNGUEZ, J.; NÚÑEZ, J.J. (2007): "A proposal of a synthetic indicador to measure poverty intensity, with an application to EU-15 countries". ECINEQ Working Paper 2007-81.
- EUROSTAT (2000): Low-wage employees in EU countries. In Statistics in Focus: Population and Social Conditions. Luxembourg: Office for Official Publications of the EC.
- FOSTER, J.E. (2007): "A class of chronic poverty measures". Working Paper, 07-W01. Vanderbilt University.
- GRADÍN, C.; CANTÓ, O.; DEL RIO, C. (2008) "Inequality, poverty and mobility: choosing income or consumption as welfare indicators". *Investigaciones Económicas*, XXXII, (2), pp. 169-200.

- HÁJEK J. (1964): "Asymptotic theory of rejective sampling with varying probabilities from a finite population". *Annals of Mathematical Statistics*, 35, pp. 1491–1523.
- HORVITZ, D.G.; THOMPSON, D.J. (1952): "A generalization of sampling without replacement from a finite universe". *Journal of the American Statistical Association*. 47, pp. 663–685.
- INE (2005): "Estudio descriptivo de la pobreza en España. Resultados basados en la Encuesta de Condiciones de Vida 2004". Documento de trabajo del Instituto Nacional de Estadística.
- KAKWANI, N.C. (1980): *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. Oxford University Press.
- NÚÑEZ, J.J. (2009): "Estado actual y nuevas aproximaciones a la medición de la pobreza". *Estudios de Economía Aplicada*, 27, pp. 325–344.
- PÉREZ-MAYO, J.; FAJARDO, M.A. (2003): "Un método alternativo para identificar los hogares en situación de pobreza". *Estadística Española*, 45, (152), pp. 115-134.
- PÉREZ-MORENO, S. (2009): "El estudio de la pobreza en España desde una óptica económica: medición y políticas". *Estudios de Economía Aplicada*, 27, pp. 349 - 372.
- RAO, J.N.K.; KOVAR, J.G.; MANTEL, H.J. (1990): "On estimating distribution function and quantiles from survey data using auxiliary information". *Biometrika* 77, pp. 365–375.
- SÄRNDAL, C.E.; SWENSSON, B.; WRETMAN, J. (1992): *Model Assisted Survey sampling*. Springer Verlag.
- SILBER, J. (2007): "Measuring poverty: Taking a multidimensional perspective". *Hacienda Pública Española*, 182, 3m pp. 29-73.
- SINGH, S. (2003): *Advanced sampling theory with applications: How Michael Selected Amy*. Kluwer Academic Publishers, The Netherlands.

