

Los métodos robustos para analizar la distribución del ingreso, la desigualdad y la pobreza¹

MARÍA-PÍA VICTORIA-FESER²

Resumen

El análisis de la distribución del ingreso incluye una larga lista de temas de investigación de carácter económico. Es importante estudiar cómo los ingresos se distribuyen en una población; por ejemplo, para determinar las políticas impositivas de redistribución que permitan disminuir la desigualdad, o para llevar a cabo las políticas sociales que conduzcan a la reducción de la pobreza. La información disponible proviene, principalmente, de estudios (y no de censos, como frecuentemente se cree) y esto es causa habitual de largos debates sobre su confiabilidad, porque las fuentes de errores son numerosas. Es más, la forma en que los datos están disponibles no siempre es como uno esperaría, es decir completos y continuos (micro-datos). Además, uno puede tener sólo datos en forma agrupada (en las clases del ingreso) y/o datos truncados donde una porción de los datos originales se ha omitido de la muestra, o simplemente no se ha registrado.

Por las características de los datos, es importante complementar los procedimientos estadísticos clásicos con bases vigorosas. En el presente trabajo se presentan tales métodos, sobre todo la selección de modelos, su adaptación con varias clases de datos, análisis de la desigualdad y la pobreza, así como herramientas de ordenamiento. Una aproximación se basa en la función de influencia (IF) desarrollada por Hampel (1974), cuyo desarrollo ampliado lo encontramos en

1. Título original: "Robust Methods for the Analysis of Income Distribution, Inequality and Poverty", en *International Statistical Review*, vol. 68, No. 3, diciembre, 2000. Ed. International Statistical Institute, Voorburg, Países Bajos. Versión al castellano del Mtro. Jaime López Delgadillo con la colaboración de Patricia Fong González, alumna de Mercadotecnia del CUCEA. Correo electrónico: ldj3342@ucea.udg.mx
2. Universidad de Ginebra. CH-1211 Ginebra 4, Suiza.

Hampel, Ronchetti, Rousseeuw & Stahel (1986). También se muestra, a través del análisis de datos reales de Gran Bretaña y Túnez, que las técnicas vigorosas pueden dar otro panorama de la distribución del ingreso, de la desigualdad o de la pobreza, cuando se comparan con los análisis clásicos.

Introducción

En el campo de la economía del bienestar, la distribución del ingreso juega un papel central. Los diferentes análisis posibles son, por ejemplo, los datos que se embonan con los modelos apropiados que describen la distribución de ingresos, la opción del modelo apropiado, probando la bondad de los datos con el modelo, el recuento de los índices de desigualdad, las comparaciones entre las distribuciones, mediante la utilización de herramientas de ordenamiento como la curva de Lorenz y la evaluación de la pobreza. La inferencia estadística es un problema esencial, dado que uno trata con datos que, principalmente, vienen de estudios y que, además, influirán en las políticas económicas resultantes, como las políticas fiscales, de inversión, de trabajo y de redistribución.

Los datos del estudio están, a menudo, sujetos a fuertes debates, a propósito de su proximidad con la realidad, dadas las numerosas fuentes de errores a que están sujetos (véase, por ejemplo, Groves, 1989). En el contexto de distribución del ingreso, ocurre a menudo que los proveedores de los datos modifican los datos originales para eliminar ingresos negativos o en ceros (Jenkins, 1997) o censuran los altos ingresos por razones de confidencialidad (Fichtenbaum y Shahidi, 1988). Es más, lo que se considera un "ingreso" no es siempre percibido como tal por las personas que contestan los cuestionarios. El lapso en que se recibe el ingreso también es una fuente de confusión (semanalmente, mensualmente, anualmente). Lo que ocasionaría, por ejemplo, que algunas personas pudieran declarar los ingresos mensuales cuando son los ingresos semanales, de hecho, la información que se espera. La unidad económica (la persona, el núcleo familiar, la casa) no incluye el mismo tipo de las personas, ya que éstas dependen del país. El impacto en las estimaciones puede ser muy serio (véase, por ejemplo, Van Praag, Hagenaars y Van Eck, 1983).

Además de las fuentes de "errores" en los ingresos, podría haber también ingresos legítimos pero extremos, que no encajan en el cuadro presentado por la mayoría de los datos. Estos ingresos podrían ser muy pequeños (de las personas cuyas creencias las hacen escoger vivir con uno muy pequeño) o excepcionalmente grandes, por ejemplo los ganadores de la lotería (véase, por ejemplo, Prieto Alaiz y Victoria-Feser, 1996) o el famoso granjero alemán que informó un ingreso mensual de un millón de marcos en el Estudio de Ingresos de Luxemburgo 1981 (datos sobre Alemania). Aunque los ingresos extremos son claramente importantes, desde el punto de vista de la autoridad política, uno puede defender que ellos

deben tener una influencia limitada cuando la meta del estudio sea deducir la desigualdad o la medida de la pobreza o incluso otros indicadores económicos que deberían reflejar la situación económica y social de una región en conjunto. Esto es todavía más importante cuando dichos indicadores se alimentan (automáticamente) como parámetros de modelos econométricos más complejos, que se usan para comparar los diferentes países a través del tiempo.

También debe enfatizarse, además del problema de confiabilidad, la forma en que los datos son presentados. No siempre es como uno esperaría, es decir completa y continua (micro datos), además uno puede tener sólo datos en una forma agrupada (en las clases del ingreso) y/o datos truncados donde una porción de datos originales se ha omitido de la muestra o, simplemente, no se ha registrado. Esto produce serias dificultades cuando se comienza a estimar (véase, por ejemplo, Ben Horim, 1990; Nelson y Pope, 1990).

Dada la estructura de los datos y los posibles "poco usuales rasgos", los modelos que se supone los describen sólo pueden tomarse como aproximaciones a la realidad. Las técnicas robustas aparecen, por consiguiente, como los candidatos naturales para intentar aliviar los problemas parcialmente asociados con los datos de ingreso.

En este trabajo repasamos algunos aspectos de la aplicación de tales técnicas al problema del análisis de distribución del ingreso, desigualdad y pobreza. Más detalles pueden encontrarse en Cowell y Victoria-Feser (1996a, 1996b, 1996c), Victoria-Feser (1993, 1997) y Victoria-Feser y Ronchetti (1994, 1997).

Formulación del modelo

Dejemos a $F \in \mathfrak{F}$ ser la verdadera distribución del ingreso que pertenece a la familia \mathfrak{F} de funciones de la distribución. La distribución del ingreso podría depender de un juego de parámetros θ y, en ese caso, nosotros escribiríamos $F\theta$. Lo que nosotros observamos es una muestra de ingresos de n denotada por X_1, \dots, X_n generada independientemente por una aproximación de F . Para representar el impacto de la contaminación en una distribución del ingreso, necesitamos un modelo específico de contaminación. Considere la distribución elemental $G(z) \in \mathfrak{F}$ que tiene un punto de volumen a , un punto arbitrario z y cero en otros valores:

$$G^{(z)}(x) = t(x \geq z)$$

donde t es la función indicadora definida por

$$t(D) = \begin{cases} 1 & \text{si } D \text{ es verdadera} \\ 0 & \text{Si } D \text{ es falsa} \end{cases}$$

Podemos usar esto para modelar una forma elemental de contaminación. Supongamos que haya una cantidad pequeña de contaminación sin detectar en el punto z de la distribución del ingreso. Entonces, la muestra real observada no representará a la verdadera distribución F pero sí una mezcla de distribución $F_\varepsilon^{(z)}$ donde

$$F_g^{(z)}(x) := [1-g]F(x) + gG^{(z)}(x)$$

El parámetro ε se usa para capturar la importancia de la contaminación relativa a los verdaderos valores. Una observación deducida de $F_\varepsilon^{(z)}$ tiene la probabilidad $(1-\varepsilon)$ de ser generado por F y la probabilidad ε de ser igual a z . Debe notarse que $F_\varepsilon^{(z)}$ es un caso particular de una mezcla de distribución más general del tipo $[1-\varepsilon]F(x) + \varepsilon G(x)$ donde G es cualquier distribución. Sin embargo, para nuestro propósito, esta mezcla de distribución particularmente no es restrictiva.

El problema central que nos concierne puede declararse, entonces, como sigue. Supongamos que se quiere adaptar el modelo F a los datos o escoger entre los posibles modelos en \mathfrak{S} para encajar los datos, o computar el índice de desigualdad, o alinear dos distribuciones $F1$ y $F2 \in \mathfrak{S}$, o bien comparar los pobres en las poblaciones $F1$ y $F2 \in \mathfrak{S}$. ¿Será el análisis resultante en el que realmente estará basado $F1_\varepsilon^{(z)}$ (y $F2_\varepsilon^{(z)}$) el que dé respuestas engañosas? Si la cantidad de contaminación es relativamente grande respecto a los verdaderos valores, lo que podemos concluir razonablemente es que casi nada se puede esperar del análisis estadístico. Sin embargo, si la cantidad de contaminación es relativamente pequeña podríamos esperar que esa inferencia estadística sea razonablemente robusta, aun bajo la contaminación, y pudiera incumbir cuando éste no sea el caso.

Para cualquier estadístico T , esta idea puede ser más precisa al introducir la función de influencia (IF). Esto se obtiene tomando la derivada con respecto a ε en $F_\varepsilon^{(z)}$ cuando

$$IF(z; T, F) := \frac{\delta}{\delta_g} T(F_g^{(z)})|_{g \rightarrow 0}$$

La IF para el estadístico T mide el impacto en la estimación de una cantidad infinitesimal de la contaminación en el punto z . Representa una función de z , en el punto en que la contaminación ocurre. Si la IF es ilimitada para algún valor de z significa que el estadístico T puede afectar catastróficamente los datos contaminados en el valor de ingreso cerca de z .

La IF fue introducida, primero, por Hampel (1968, 1974) y puede concebirse como una aproximación de primer orden del sesgo de la estadística, debido a la introducción de la contaminación. En otras palabras, si uno plasma el máximo

(absoluto) en el sesgo de la estadística como una función de ε , entonces el máximo (absoluto) valor de IF es la inclinación de la tangente en $\varepsilon=0$. Esto significa que la IF puede ser infinita, y entonces también lo puede ser el sesgo de la estadística (vea Hampel, Ronchetti, Rousseeuw y Stahel, 1986). Aun más, la IF describe el peor de los sesgos sobre T cuando la contaminación de cualquier tipo ha sido introducida en el modelo (véase Hampel *et al.*, 1986).

La IF tiene dos funciones principales: puede usarse para estudiar propiedades de robustez estadística, y para construir estimadores robustos y procedimientos de análisis robustos. En el contexto de distribución del ingreso, ha sido usada para evaluar las propiedades de robustez de las herramientas de medición y cálculo de desigualdad y pobreza (véase Cowell y Victoria-Feser 1996a, 1996b, 1996c). En el contexto general de estimación paramétrica, la IF ha sido usada para definir estimadores robustos. De hecho, el primer problema es encontrar una clase general de estimadores y, entonces, dentro de esta clase escoger una robusta de manera óptima, tal y como será explicado posteriormente. La clase de M-estimadores (Huber, 1964) constituye una buena base. Se da para un modelo paramétrico general F_y para una muestra de n ingresos independientes X_1, \dots, X_n dada la solución en θ de

$$\sum_{i=1}^n \psi(x_i; \theta) = 0$$

donde ψ es una función muy general. Un caso particular se da por el MLE cuando $\psi(x; \theta) = s(x; \theta)$. Otro caso particular se da con el óptimo B estimatorio robusto (OBRE) de Hampel *et al.* (1986). El OBRE es óptimo porque tiene la mínima matriz de covarianzas asintótica entre los estimadores de la IF. Hay varios OBRE dependiendo de cómo la elasticidad esté determinada. El OBRE estándar está definido dentro de la solución de θ en

$$\sum_{i=1}^n [s(x_i; \theta) - a] w_c(x_i; a, A) = 0$$

donde

$$w_c(x; a, A) = \min \left\{ 1; \frac{c}{\|A[s(x; \theta) - a]\|} \right\}$$

son los pesos de la función de Huber (Huber 1964, 1981), $s(\cdot; \theta)$ es la función determinante y el vector $a = a(\theta)$ y matriz $A = A(\theta)$ están definidos implícitamente por

$$(6) \quad E [s(x; \theta) - a] [s(x; \theta) - a]^T w_c(x; a, A)^2 = A^{-1} A^{-T}$$

$$(7) \quad E [s(x;\theta)w_c(x;a,A)] = a$$

A y a aseguran la eficiencia y consistencia del estimador. Uno puede ver en (4) que el estimador robusto es relativamente simple. Puede verse como un MLE robusto cuyo peso lo otorga (5) con datos distantes al volumen central. Se puede deducir que el OBRE es robusto por naturaleza, no sólo por errores sino también por pequeños descuidos en los modelos. Por tanto tiene una IF que mide cualquier influencia de error en el modelo hacia los estimadores.

El grado de robustez se controla por c . Para incrementar la robustez deberemos decrementar c y es fácil verlo cuando tiende a infinito $c \rightarrow \infty, w_c(x;a,A) \rightarrow 1$, entonces $a \rightarrow 0, \psi(x;\theta) \rightarrow s(x;\theta)$ y uno obtiene así el MLE. De cualquier modo, esto tiene un precio y este precio es una pérdida de eficacia. En la práctica se puede escoger c para que así el OBRE tenga 95% de eficiencia comparado con el MLE y aún tenga buenas propiedades de robustez.

Al analizar los datos del ingreso, uno no siempre tiene los datos continuos y completos. En este caso fue necesario desarrollar nuevos estimatorios de robustez y una aproximación a lo dicho por Hampel *et al.* (1986). Lo mismo puede aplicarse sobre los procedimientos de análisis. En las siguientes secciones, repasamos algunos de los recientes desarrollos en estadística robusta para la distribución del ingreso por la clase de datos y/o clase de análisis.

Ajustar los datos continuos y completos

El problema aquí es cómo estimar θ dado un modelo paramétrico F_θ que supuestamente describe la distribución de ingresos. Se sabe ahora ampliamente que el estimador de probabilidad máxima (MLE) no es, en general, un estimador robusto. Esto es verdad para todos los modelos conocidos de distribución del ingreso (véase Victoria-Feser, 1993). Con los datos continuos y completos, el OBRE es un candidato conveniente para los parámetros de modelos de ingreso, como sucede con el de distribución Gamma, que depende de dos parámetros, y los Dagum tipo 1 (Dagum, 1980) que dependen de tres parámetros. El primer parámetro normalmente es para la escala, el segundo es para la forma de la distribución y un tercer parámetro, en ocasiones, se agrega para modelar la anchura de los valores extremos del lado derecho.

Como un ejemplo, intentamos ajustar una clase de las distribuciones de ingreso mencionada anteriormente usando el MLE y el OBRE estándar. Nuestra clase de datos será una muestra ($n = 746$) de un conjunto de datos estándar de ingreso en el Reino Unido, 1979, el receptor del ingreso es la casa donde se reciben beneficios sociales (véase *Sección de seguridad social*, 1992). Se usará

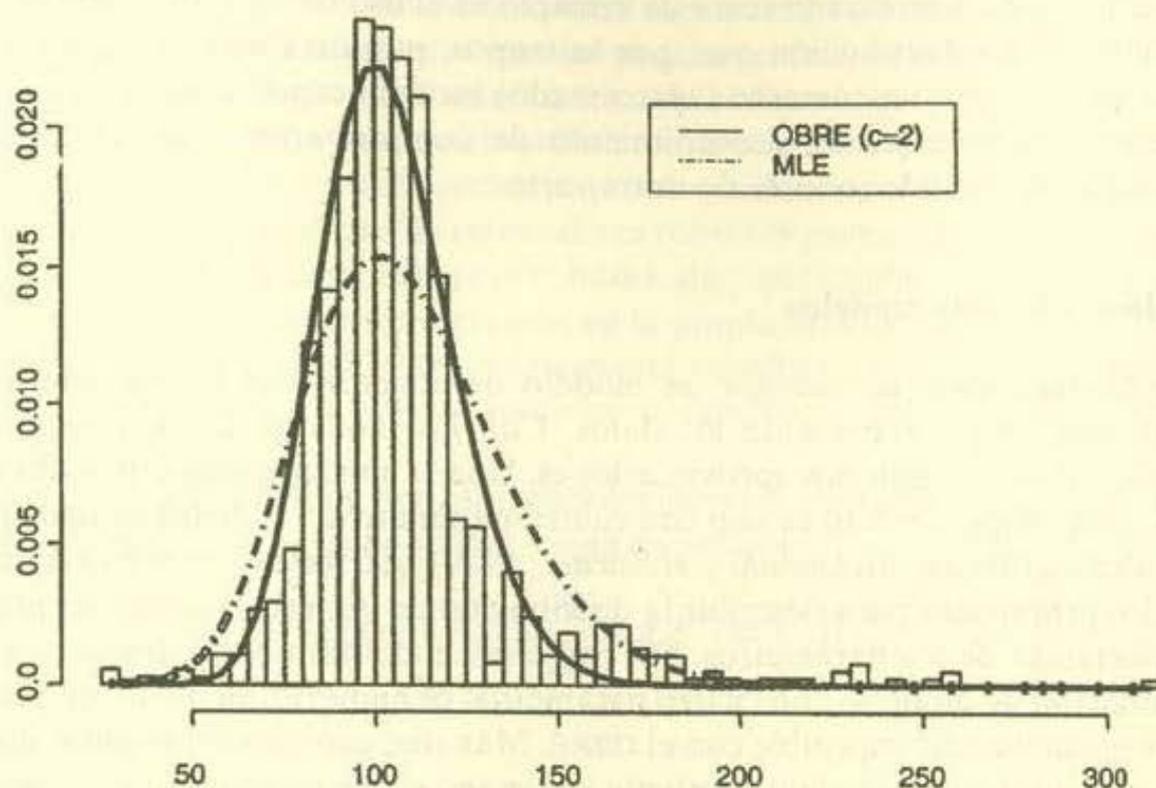


Figura 1. Ajuste gamma de datos para el Reino Unido.

primero una distribución Gamma. El histograma de los datos y las densidades estimadas (MLE y OBRE, $c = 2$) se presenta en la Figura 1.

Podemos apreciar que el MLE usado en el modelo Gamma proporciona un ajuste muy pobre, mientras que el OBRE captura el volumen de los datos en el centro de la distribución. Estima de mejor manera las distribuciones, pero a costa de subvalorar los datos localizados en los extremos (es decir, los ingresos altos). Estos ingresos extremos representan una pequeña proporción de todos los ingresos. Mientras el MLE intenta (no siempre con éxito) acomodarlos, el OBRE se concentra en los ingresos más probables, es decir alrededor de la moda de la distribución. Uno podría defender que dos parámetros no son suficientes para describir esta clase de datos, y que se requiere un tercero. La pregunta es: ¿Valdrá la pena añadir un parámetro más para acomodar unos cuantos valores extremos y, entonces, estamos realmente a salvo con un parámetro más? Para dar una respuesta a este problema, acomodaremos una distribución Dagum tipo 1 a la misma clase de datos. El histograma de los datos y las densidades estimadas (MLE y OBRE, $c = 2$) se presentan en la Figura 2. Podemos encontrar que con un parámetro extra para medir la intensidad de los valores extremos de la derecha (para modelar los valores extremos), el MLE se basó en tres parámetros y da un mejor acomodo de los datos que con dos parámetros, pero todavía subvalora la moda de los datos, lo cual no pasa con el OBRE. Sin embargo, con un modelo de tres parámetros la diferencia entre los dos estimadores es reducida.

Lo que hemos mostrado con este ejemplo es el uso de un estimador robusto para ajustar una distribución que, por lo menos, proporcione otra figura. Esto sólo es posible con un correcto (aproximado) modelo específico. En la sección siguiente, discutiremos un procedimiento de comprobación que nos permita seleccionar un modelo robusto de entre varios.

Elección entre dos modelos

El problema radica en escoger un modelo de acuerdo con cierto criterio, el modelo que mejor represente los datos. Con los modelos de distribución de ingreso podemos seguir dos aproximaciones. Una es ajustar un supermodelo a los datos, dicho supermodelo es uno con cuatro parámetros, las distribuciones Beta generalizadas (véase *McDonald y el rescate*, 1979) que incluye la mayoría de los modelos propuestos para describir la distribución de ingresos, y entonces probar la importancia de los parámetros. Sin embargo, encontramos en la práctica que la estimación de modelos con cuatro parámetros es numéricamente difícil para el MLE, y se vuelve casi imposible con el OBRE. Más aún, este procedimiento supone que contamos con un modelo excelente a la mano y éste no siempre es el caso en otros tipos de análisis. Nosotros seguimos una aproximación seguida por Cox (1961, 1962) para hacer la prueba entre dos hipótesis.

En general, se asume que bajo H_0 (la hipótesis nula) el modelo es F_{α}^0 y que bajo H_1 (la hipótesis alterna) el modelo es F_{β}^1 con α y β , siendo dos juegos de parámetros. Atkinson (1970) y otros han mostrado que las pruebas estadísticas de Cox pueden interpretarse como un multiplicador de Lagrange o como valor de prueba sobre $H_0: \lambda = 1$ (contra $H_1: \lambda \neq 1$) para el modelo comprensivo:

$$|c(x; \alpha, \beta) = |^0(x; \alpha)^{\lambda} \cdot |^1(x; \beta)^{1-\lambda} \left[\int |^0(y; \alpha)^{\lambda} \cdot |^1(y; \beta)^{1-\lambda} dy \right]^{-1}$$

donde el $|^0$ y $|^1$ son las densidades de F_{α}^0 y F_{β}^1 . En la práctica uno necesita estimar α y β , lo que se hace usando el MLE (o el pseudo MLE, es decir el MLE bajo una de las hipótesis). La distribución nula asintótica de la prueba estadística es la distribución normal.

Las pruebas estadísticas de Cox se han criticado, a menudo, por varias razones. Una es la falta de exactitud en la aproximación de la distribución de la muestra estadística por su distribución asintótica. Otra razón importante es la falta de robustez del procedimiento en la comprobación. El vigor de la prueba ha sido introducido por Ronchetti (1982), Rousseeuw y Ronchetti (1979, 1981) quienes fueron los primeros en adaptar el problema de optimización de Hampel para probar los procedimientos. El problema de optimización de Hampel para probar los procedimientos puede declararse como: Bajo un término de influencia de contaminación pequeña en el nivel de prueba y poder (el requisito de robu-

tez), el poder de la prueba del modelo ideal se maximiza (el requisito de eficacia). Victoria-Feser (1997) demuestra que las pruebas estadísticas de Cox no son tan robustas, ya que una cantidad infinitesimal de contaminación puede conducir a una prueba significativa cuando el verdadero modelo es el que está bajo H_0 , o, en otras palabras, que la IF en el nivel de la prueba es ilimitada. El punto interesante es que no es suficiente el uso de estimadores robustos para α y β a fin de conseguir un procedimiento de comprobación robusto, sino que también se requiere eliminar la influencia de los datos extremos en la propia prueba estadística. Victoria-Feser (1997), por consiguiente, propuso una estadística de prueba de tipo Cox robusta basada en resultados de modelos paramétricos generales de Heritier y Ronchetti (1994).

La prueba estadística del multiplicador normal asintótico robusto de Cox o el generalizado de Lagrange (GLM), están dados por:

$$U_{GLM} = \frac{1}{n} \sum_{i=1}^n \left[A_{(21)} S^0(xi; \alpha) + A_{(22)} \left[\frac{\delta}{\delta \lambda} \log f^c(xi; \alpha, \beta) \Big|_{\lambda=1} - a_{(2)} \right] \right] w_c(xi; A, a)$$

donde $\hat{\alpha}^2$ es el MLE de α ,

$$w_c(x; A, a) = \min \left\{ 1; \frac{c}{\left| A_{(21)} S^0(x; \alpha) + A_{(22)} \left[\frac{\delta}{\delta \lambda} \log f^c(xi; \alpha, \beta) \Big|_{\lambda=1} - a_{(2)} \right] \right|} \right\}$$

y el vector $A_{(21)}$ ($1 \times \dim(\alpha)$) y los escalares $A_{(22)}$ y $a_{(2)}$ están implícitamente determinados (véase Victoria-Feser, 1997). Como en los estimadores, la prueba robusta estadística puede verse como una prueba ponderada estadística clásica tipo Cox. De hecho, cuando $c \rightarrow \infty$, conseguimos la prueba estadística tipo Cox clásica. Nótese también que, en la práctica, β requiere ser estimada de forma consistente por medio del MLE.

Un estudio de simulación muestra la actuación del U_{GLM} cuando se compara con el estadístico tipo Cox clásico. Una muestra de Pareto de 200 observaciones contaminadas y no contaminadas fueron simuladas (con parámetro de forma $\alpha = 3$ y el parámetro escalar fijo γ) y fue probada la distribución de Pareto con la distribución exponencial trunca, utilizando la prueba robusta estadística tipo Cox clásica (también, véase Victoria-Feser, 1997). Debiera notarse que la distribución de Pareto se usa para planear las distribuciones de ingreso con moda cero para los ingresos altos. Las muestras se contaminaron por medio de

$$(1 - \varepsilon 7 \sqrt{200}) F_{\alpha, \gamma} + \varepsilon / \sqrt{200 F_{\alpha, 10 \varepsilon \gamma \varepsilon}}$$

donde por cada cantidad de contaminación se generaron 1,000 observaciones. La cantidad de contaminación aquí es de tipo ε/\sqrt{n} la cual es apropiada para los procedimientos de pruebas (véase Heritier y Ronchetti, 1994). La Tabla 1 nos da los niveles reales de la estadística clásica y robusta de Atkinson, es decir, las que son obtenidas por simulación. Estos niveles son, en realidad, la proporción de veces que la prueba es significativa cuando se compara la prueba de Pareto con la exponencial truncada, y la prueba no es significativa cuando se prueba la exponencial truncada contra la de Pareto.

CUADRO 1
Niveles reales de la prueba estadística clásica y robusta
del tipo Cox ($c=2$). Pareto vs. Exponencial.

Cant. de contaminación ε	Estadística clásica Niveles nominales				Estadística robusta Niveles normales			
	1%	3%	5%	10%	1%	3%	5%	10%
0%	2.1	3.1	3.5	5.2	1.3	3.5	5.5	10.2
3%	6.3	8.7	10.3	14.7	1.2	3.3	5.1	10.3
6%	13.1	18.5	22.5	27.6	1.4	3.6	5.4	10.7
10%	24.4	31.3	35.2	43.9	1.3	3.0	5.6	11.4
15%	35.6	44.6	49.9	58.1	1.4	4.1	7.9	14.5
20%	46.3	54.2	58.6	67.1	0.9	4.1	7.6	14.5

Podemos observar que la estadística clásica tiene un comportamiento muy extraño, ya que cuando no hay contaminación la hipótesis nula está por debajo del rechazo y con sólo cantidades pequeñas de contaminación la hipótesis nula está por encima del rechazo. La primera situación se debe, probablemente, al hecho de que la aproximación del actual estadístico a la distribución de tipo Cox, por medio de su distribución asintótica, no es exacta; y en la segunda situación es la falta de robustez. Por otro lado, encontramos que con la prueba estadística robusta no sólo la distribución asintótica es una buena aproximación de la distribución muestral, sino también que las salidas pequeñas del modelo bajo la hipótesis nula no influyen en el nivel de prueba, por lo menos cuando las cantidades de contaminación están cerca de $= 10\%$. Con más contaminación, la hipótesis nula tiende a estar levemente sobre el rechazo, a niveles de 5% y 10%, pero esto no es demasiado drástico comparado con el caso clásico. En otras palabras, la prueba robusta es muy estable.

Ajustar los datos truncados

Por datos truncados se entiende al subconjunto de los ingresos (típicamente la línea real), de los cuales no se cuenta con información sobre la naturaleza a la que pertenecen en ese subconjunto. Las razones por las cuales no se tiene información pueden ser diferentes; como, por ejemplo, las razones de confidencialidad, cuando los ingresos altos (es decir, aquellos que excedan un nivel fijo) no son proporcionados por los informantes; o por razones prácticas, mencionadas ya en la introducción. Con los datos de ingreso de los estudios podría haber también el problema de no obtener contestación, que es un problema más complejo que el supuesto aquí. De hecho, es discutible que las proporciones de los que no contestaron puedan variar, dependiendo del nivel de ingreso, así que es necesario un modelo apropiado que considere esta relación: véase, por ejemplo, Little y Rubin (1987).

Para abordar el problema de datos truncados, Victoria-Feser (1993) propone usar una distribución marginal o una generalización del algoritmo de EM (Dempster, Laird y Rubin, 1977) para M estimadores (el algoritmo de EMM). Dejemos que $\mathcal{K} \subseteq \mathcal{R}$ sea el subconjunto de valores de ingreso para los casos observados y el subconjunto para aquellos que no lo son. El algoritmo de EMM define una clase de estimador OBRE con los datos truncados y se da la solución en θ de:

$$(8) \int_{\mathcal{K}} [s(x;\theta) - a] w_c(x; A, a,) dF_{\theta}(x) + \int_{\mathcal{K}^c} dF_{\theta}(x) \frac{1}{n} \sum_{i=1}^n [s(x_i;\theta) - a] w_c(x_i; a, A) = 0$$

donde w_c , A , y (a) se definen en (5), (6) y (7). La primera parte de (8) estima la parte perdida (el paso de la expectativa) mientras la segunda parte define el paso de maximización.

Otra aproximación sería considerar la distribución marginal, es decir:

$$\frac{F_{\theta}}{\int s dF_{\theta}(X)}$$

y considerar esta función y usar el OBRE para datos continuos. Esta aproximación y el EMM son, en general, diferentes (son equivalentes al MLE), pero su diferencia es pequeña y radica en los supuestos del modelo subyacente. De hecho, contrariamente a la aproximación de la distribución marginal, el algoritmo de EMM asume que el modelo subyacente es el completo, es decir F_{θ} y no (9). Esto nos lleva a una diferencia en la manera en que se estiman las ponderaciones asociadas a cada observación (véase Victoria-Feser, 1993). Para mostrar las propiedades del EMM comparadas con los datos truncados, presentamos un estudio de simulación. La Tabla 2 da el sesgo y el MSE para el MLE y el OBRE (EMM) en el caso de

datos contaminados y truncados (por debajo de un valor mínimo), que fueron generados por una distribución Gamma con escala paramétrica $\lambda = 1$ y un parámetro de la forma $\alpha = 3$. La contaminación consiste en haber escogido al azar 1% de datos diez veces su valor.

CUADRO 2
Bias y MSE del OBRE ($c=2$ para datos gamma truncados)

Pérdida de información (%)	Parámetro	OBRE		MLE	
		Bias	MSE	Bias	MSE
7.5	α	0.07	0.24	-2.39	5.71
	λ	0.02	0.03	-0.75	0.57
2.5	α	0.06	0.17	-2.39	5.71
	λ	0.02	0.0202	-0.78	0.39

El comportamiento del MLE no es satisfactorio, considerando que el OBRE tiene un sesgo pequeño. De hecho, el MLE es aún más parcial que en el caso de datos completos (véase Victoria-Feser y Ronchetti, 1994). Esto no es sorprendente, truncar significa información incompleta y los estimadores robustos son construidos con base en modelos aproximados.

Ajustar los datos agrupados

Otro rasgo de los datos del ingreso es que sólo pueden estar disponibles en una forma agrupada, es decir en forma de frecuencias por clase. En estos casos el OBRE de Hampel *et al.* (1986) no puede usarse como tal. El problema puede considerarse como problema de datos discretos con un modelo subyacente continuo. Además de los efectos usuales de contaminación, tenemos en esta situación los efectos de agrupamiento, donde algunas observaciones pueden cambiarse de una clase a otra debido a errores de redondeo o por la definición de clase. Una clase general de estimadores (Estimadores divergentes de mínima fuerza o MPE) para los parámetros del modelo subyacente basados en los datos agrupados fueron definidos por Cressie y Lea (1984). El modelo incluye el MLE, el estimador de distancia mínima Hellinger (MHDE), los estimadores basados en Pearson χ^2 , etc. No obstante que su IF es limitado, las desviaciones del modelo subyacente pueden causar un sesgo amplio, sobre todo en las clases con probabilidades bajas. Victoria-Feser y Ronchetti (1997) proponen una clase más general de estimadores que contengan el MPE, llamado MGP, el cual puede verse como un estimador M para los datos agrupados. Un estimador óptimo (OBRE) que minimiza la matriz de covarianzas asintótica bajo una limitación de su IF se deriva, y lo que se

encuentra es una pérdida de eficiencia pequeña; este estimador es más robusto que el clásico MPE. Dejemos que p_j sea las frecuencias relativas observadas en un J de las clases desoladas I_1, \dots, I_J y $k_j(\theta) = \int_{I_j} dF_\theta(x)$ las verdaderas frecuencias correspondientes, los estimadores óptimos están dados por la solución en a partir de:

$$(10) \sum_{j=1}^J \left(\frac{P_j}{k_j(\theta)} \right)^\gamma A \left[\frac{\delta}{\delta\theta} k_j(\theta) - a k_j(\theta) \right] w_{c,j}(A, a) = 0$$

donde es una constante arbitraria (por ser la definida),

$$w_{c,j}(A, a) = \min \left\{ 1; \frac{c}{\left\| A \left[\frac{\delta}{\delta\theta} \log k_j(\theta) - a \right] \right\|} \right\}$$

y donde la matriz $A = A\theta$ y el vector $a = a\theta$ están implícitamente determinados por las ecuaciones

$$\sum_{j=1}^J A \left[\frac{\delta}{\delta\theta} k_j(\theta) - a k_j(\theta) \right] w_{c,j}(A, a) = 0$$

$$\sum_{j=1}^J A \left[\frac{\delta}{\delta\theta} k_j(\theta) - a k_j(\theta) \right] \frac{\delta}{\delta\theta^T} \log k_j(\theta) w_{c,j}(A, a) = I$$

Nótese que (10) define un conjunto de estimadores robustos. Cada estimador clásico tiene su homólogo robusto. Por ejemplo, la versión robusta del MLE se obtiene poniendo $\gamma=1$, y el robusto MHDE poniendo $\gamma=0.5$. El c limitado permite el control sobre el grado de robustez. A medida que c decrece, se incrementa la robustez pero, al mismo tiempo, la eficacia disminuye. En la práctica encontramos, sin embargo, que un buen grado de robustez puede lograrse sin perder demasiada eficacia.

Ilustramos la actuación de los estimadores MGP con una simulación pequeña de una distribución de Pareto. Los datos son contaminados al tomar una proporción ε al azar y multiplicándolos por 10, y luego se agrupan en 22 clases. El MPE y OBRE (para el MLE y MHDE) es computado y su sesgo y MSE reportados en la tabla 3 (Los errores normales para los valores del sesgo son menores a 0.02).

Podemos ver que aunque el IF del MLE se limita cuando el modelo subyacente se contamina, el MLE tiene un sesgo grande. Por otro lado, con el estimador robusto correspondiente podemos ver que este sesgo y los MSE globales son considerablemente más pequeños. El MHDE tiene mejores propiedades de robustez que el MLE, pero puede mejorarse usando la versión robusta correspondiente que tiene el mejor sesgo y el MSE global en el ejemplo.

El procedimiento robusto también se aplica a los datos del Reino Unido presentados anteriormente. Los datos se agrupan en 58 clases, igualmente clasificadas según el tamaño.

CUADRO 3
Bias y MSE de estimadores clásicos y robustos
para datos agrupados en datos simulados de Pareto.

ϵ	MPE				OBRE			
	MLE		MHDE		Rob MLE		Rob MHDE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0%	0.01	0.0101	0.05	0.013	0.02	0.0202	0.01	0.011
1%	0.19	0.05	0.06	0.015	0.03	0.02	0.01	0.015
3%	0.55	0.27	0.18	0.05	0.13	0.04	0.04	0.015
5%	0.77	0.61	0.29	0.11	0.25	0.06	0.06	0.02

La primera clase se extiende al cero y la última clase a ∞ . La distribución Gamma se escoge como candidata para modelar los datos y los MLE, MHDE, y MHDE robustos ($c = 175$ que corresponde a 95% de eficacia) se computan. El histograma de los datos y las densidades estimadas se presentan en la Figura 3.

Podemos apreciar que mientras el MLE intenta acomodar las colas de la distribución, pierde la descripción del volumen de los datos en el centro. El MHDE mejora el ajuste de la mayoría de los datos, pero es su versión robusta da una mejor descripción. El último tiene una eficiencia de 95% respecto al MPE clásico. Finalmente, Victoria-Feser y Ronchetti (1997) muestran que el OBRE, para datos agrupados, tiene un cambio perceptible de límite local y, por consiguiente, controla los efectos de los errores de agrupamiento también. Éste no es el caso con el MPE clásico.

Hasta ahora hemos revisado los métodos estadísticos para describir la distribución de ingresos. Los parámetros de estimación pueden ser usados para dirigir los análisis extensos, como la desigualdad o el análisis de pobreza. Esto se desarrollará en las secciones siguientes.

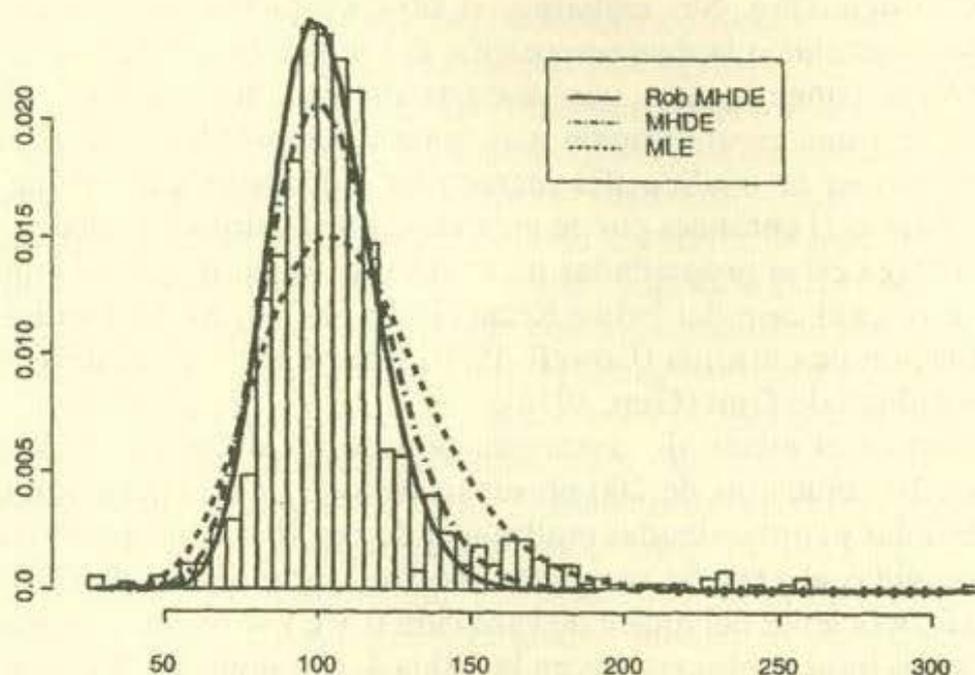


FIGURA 3

El modelado Gamma de los datos agrupados para el Reino Unido.

Calcular la desigualdad

La medición de la desigualdad y otras herramientas relacionadas con el análisis de la distribución de ingreso se utilizan, a menudo, para resumir la información sobre las distribuciones del ingreso. Ambos juegan una parte importante en los debates políticos sobre las tendencias económicas y sociales y en la economía de bienestar. Teóricamente, las medidas de desigualdad han sido derivadas por el requerimiento de un número esencial de propiedades. La más importante (y de mayor aceptación) propiedad es el principio de traslados (Dalton, 1920) en que los estados que trasladan una cantidad positiva arbitraria de ingreso de un receptor de ingreso más pobre a uno más rico (tal que la media de la distribución se conserva) deben aumentar el valor de la medida de desigualdad. Por otro lado, se estiman las medidas de desigualdad de los datos del ingreso. Por consiguiente, es importante que se entienda la relación entre las propiedades económicas de la medida de desigualdad y las propiedades estadísticas del estimador correspondiente. En esta sección resumimos algunos de los resultados obtenidos en Victoria-Feser (1993) y Cowell y Victoria-Feser (1996b) sobre la relación existente entre las propiedades económicas y las medidas de desigualdad vigorosas. El objetivo es encontrar una manera simple y conveniente de verificar si una medida de desigualdad es resistente contra la contaminación de los datos, y deducir una conclusión sobre las clases generales de los índices de desigualdad. Puede verse que el principio de traslados no necesariamente implica la no-resistencia de una

medida de desigualdad. Sin embargo, si otra propiedad se agrega, como la independencia escalar o la descomposición del índice (véase Cowell y Victoria-Feser, 1996b) restringe la clase, pero todavía abarca la mayoría de los índices de desigualdad, o si una especificación más realista del problema de estimación es considerada (como la medida del ingreso ha sido estimada, en lugar de ser especificada *a priori*) entonces puede mostrarse que cualquier índice de desigualdad que satisfaga estas propiedades no es robusto en su IF que es ilimitado. En particular, éste es el caso del índice Kolm (Kolm, 1976a, b). La familia generalizada en situación de entropía (Cowell, 1980) incluye los índices de Theil (Theil, 1967), y los índices de Gini (Gini, 1910).

Para mostrar el efecto de contaminación de los datos en el estimador de desigualdad, 100 muestras de 200 observaciones de una distribución de Pareto fueron generadas y contaminadas multiplicando por 10 una proporción de observaciones escogidas al azar. Se computan tres índices: el índice de Gini, el índice de Theil y el coeficiente del índice de variación (CV), y se encontraron los valores medios de estos índices observados en la Tabla 4. (La segunda fila representa los verdaderos valores de los índices y SD representa la desviación estándar).

CUADRO 4
Los valores del promedio de los índices de desigualdad
estándar de datos de Pareto contaminados
y simuladamente incontaminados

Contaminación	Gini (0.2000)	SD	Theil (-0.0945)	SD	CV (-0.1667)	SD
0%	0.1974	0.002	0.0908	0.003	0.1545	0.02
1%	0.2577	0.002	0.2068	0.005	0.4974	0.02
2%	0.2845	0.002	0.2584	0.006	0.6466	0.03
3%	0.3319	0.002	0.3485	0.007	0.8993	0.04
4%	0.3542	0.002	0.36920	0.007	1.0180	0.04
5%	0.3891	0.002	0.4500	0.007	1.1150	0.04

Podemos ver que los tres índices son muy sensibles a una pequeña contaminación en los datos. Cowell y Victoria-Feser (1996b) desarrollaron un estudio más completo que involucra a otros modelos de distribución de ingreso. Para evitar el efecto de contaminación en los índices estimados, una manera es calcular la desigualdad de las estimaciones robustas de los parámetros de los modelos de distribución de ingreso. Cowell y Victoria-Feser (1996b) muestran, por ejemplo, que para los índices generalizados entrópicos, la IF de los miembros de esta clase es proporcional a la IF del estimador de los parámetros del modelo de ingreso subyacente. Por consiguiente, si se usan estimadores robustos para los parámetros del modelo del ingreso, entonces la medida de desigualdad resultante

también será robusta. Estos estimadores robustos de desigualdad llevan mucha información en ellos, ya que proporcionan una revisión de las estimaciones clásicas; donde las diferencias entre los resultados surgen y son atribuibles a las pequeñas desviaciones del modelo supuesto, cuya información debiera tenerse en cuenta para deducir las conclusiones sobre el "verdadero" cuadro de desigualdades. Por ejemplo, usando el conjunto de datos del Reino Unido, los tres índices de desigualdad anteriormente presentados son estimados a través de MLE y OBRE de una distribución Gamma y comparado con los estimados directos (empíricamente) en la muestra completa y en una muestra truncada en que 2% de los ingresos superiores han sido removidos. Esto es para mostrar hasta qué punto las medidas de desigualdad consideradas aquí son sensibles a los valores extremos. Los resultados se presentan en la Tabla 5. Estos resultados revelan dos puntos interesantes. Primero, podemos apreciar que el OBRE produce una distribución que exhibe la más baja desigualdad uniforme que la producida por el MLE (¡Incluso por un factor 2!). Después, el mismo fenómeno puede observarse al comparar las dos estimaciones directas; el truncamiento superior de la muestra tiene un impacto en la desigualdad que es similar al cambiar del MLE al OBRE. Esto no sorprende cuando uno vuelve a revisar la Figura 1.

CUADRO 5
Índices de la desigualdad en los datos de Reino Unido,
usando la distribución Gamma

	Gini	Theil	CV
MLE	0.1375	0.0299	0.0302
OBRE (c=2)	0.0983	0.0152	0.0153
Estimaciones directas sobre	0.1287	0.0320	0.0358
Estimaciones directas sobre	0.1120	0.0222	0.0228

Debe enfatizarse que aquí el punto no es afirmar que las estimaciones robustas sean mejores que las clásicas al estimar la desigualdad, más bien podemos decir que nos proporcionan otra "fotografía". El supuesto básico en el cómputo de los índices es que los datos siguen una distribución Gamma. Este no es exactamente el caso, desde que hay diferencias entre el MLE y el OBRE, pero uno puede sostener eso para la mayoría de los datos de la distribución Gamma apropiada. Desgraciadamente, los datos que son considerados como "extremos" por la distribución Gamma también son aquellos que tienen una influencia importante en la medición de la desigualdad, y también corresponden a los datos truncados que, en conjunto, explican los resultados presentados en la Tabla 5. El uso de un estimador robusto, en este caso particular, revela que unos cuantos ingresos representan una gran diferencia en la medición de la desigualdad, pero cuál medida sea la más apropiada, no es una pregunta fácil de contestar.

Evaluar la pobreza

A partir del artículo escrito por Sen (1976), la evaluación de la pobreza se ha vuelto un área prolífica de investigación para la economía. Los temas teóricos importantes son: la determinación de la línea de pobreza bajo la cual una unidad económica, por ejemplo una casa, es considerada pobre; las funciones (índices) que debieran ser utilizadas para agregar los ingresos de los pobres; y cómo puede manejarse la inferencia estadística. Todos estos temas están claramente unidos. Nos enfocamos aquí sobre las propiedades robustas para la medición de pobreza, como fue desarrollado en Cowell y Victoria-Feser (1996a).

Una clase muy importante de índices de pobreza (además, índices de pobreza separable) se define como la distribución de ingreso F por la integral

$$P(F; z) = \int p(z(F), x) dF(x)$$

donde el $z(F)$ es la línea de pobreza, la cual se determina por la propia distribución del ingreso (endógeno). Varias medidas de pobreza importantes surgen de esta clase. Una simple aunque muy conocida medida es la proporción por persona (la proporción de pobres) cuando $p(z(F), x) = i(x < z(F))$, otra es la diferencia en pobreza (la proporción modificada de pobres, la cual tiene en cuenta el tamaño relativo de la pobreza) cuando $p(z(F), x) = i(x < z(F)) (z(F) - x) / z(F)$, y uno más general es el índice de pobreza Foster-Greer-Thorbecke (Foster, Greer y Thorbecke, 1984) cuando $p(z(F), x) = i(x < z(F)) [(z(F) - x) / z(F)]^\gamma$ (véase Cowell y Victoria-Feser, 1994).

En la práctica, la pobreza se estima usando la distribución empírica $F(n)$. Uno podría preguntarse si el efecto de datos contaminados puede tener una influencia importante en el nivel estimado de pobreza. Cowell y Victoria-Feser (1996a) han abordado el problema y han encontrado que una cantidad infinitesimal de contaminación puede tener un efecto catastrófico en la medición de la pobreza solamente a través de la evaluación de la línea de pobreza. Por consiguiente, si lo anterior se estima robustamente, entonces el analista estará muy seguro.

Ayadi, Matoussi & Victoria-Feser (1998) han llegado a este resultado en un análisis de pobreza en Túnez. Los problemas centrales de la investigación eran evaluar el diferencial de pobreza entre regiones urbanas y rurales, de manera robusta, en las diferentes partes económicas de Túnez. Esto se ubica en una especificación correcta de la línea de pobreza. Los datos disponibles fueron los de consumo familiar (estudio de 1990), así que se utilizó la aproximación de Ravallion y Bidani (1994) que permite evaluar la pobreza con esta clase de datos. Túnez se dividió en 3 regiones económicas, a saber: el Gran Túnez (alrededor de la capital); el litoral del mar mediterráneo, que ha tenido prosperidad económica

desde la independencia; y el interior (la parte occidental), que tiene varios problemas sociales y económicos. El litoral y el interior también pueden ser divididos en una parte rural y una urbana. Para determinar la pobreza, la línea de pobreza se dividió en dos partes, a saber: la cantidad de ingreso para satisfacer las necesidades de comida y la cantidad destinada a no-comida (como vivienda, vestido, etc.). Para tener en cuenta las diferencias regionales, el comportamiento del consumo de los pobres se estima determinando los productos alimenticios o la canasta básica, en comparación con la gran cantidad de productos alimenticios diferentes. Esto proporciona el centro multivariado de datos. Es importante estimar este centro en una moda robusta, porque la canasta básica del consumidor se evalúa con los precios de los mercados locales y constituye una parte importante de la línea de pobreza. La parte de no-alimentos es estimada usando un modelo llamado AIDS (Ravallion, 1993) que es un modelo de regresión de ingreso de comida proporcional al ingreso total y las variables socio-económicas relacionadas con el sostenimiento de la casa. Aquí, de nuevo, para evitar las desviaciones debido a un número pequeño de datos extremos, un enfoque robusto es adoptado (véase Ayadi, Matoussi y Victoria-Feser, 1998).

Para resumir el análisis, la Tabla 6 presenta tres índices de pobreza calculados para cada región de Túnez, mismos que siguen un enfoque robusto para la evaluación de la línea de pobreza comparado con un clásico (llamados líneas de pobreza más bajas. Véase Ayadi, Matoussi y Victoria-Feser, 1998). Los índices son la proporción del *headcount* (HCR), el hueco de pobreza (PG) y el Foster-Greer-Thorbecke (FGT) índice de pobreza con el parámetro $\gamma=2$. Podemos concluir lo siguiente: En las regiones más desarrolladas (Gran Túnez y el Litoral Urbano) no sólo la pobreza es menos intensa sino que no hay una diferencia real entre el enfoque clásico y el robusto. Esto es porque los estimadores clásicos y robustos de las líneas de pobreza dan valores muy similares. En las otras regiones las diferencias entre el clásico y el robusto son grandes (arriba de 45%), porque el estimador clásico y el robusto de las líneas de pobreza difieren sustancialmente (véase Ayadi, Matoussi y Victoria-Feser, 1998). Cuando regiones urbanas y litorales son comparadas, uno puede ver que la pobreza es principalmente rural, lo que contradice algunas creencias anteriores sobre la situación de los pobres. Esta diferencia es menos importante cuando el análisis robusto es considerado.

Ordenar las distribuciones de ingreso

El uso de medidas de desigualdad, consideran muchos, resulta bastante restrictivo porque resume información sobre la distribución de ingresos en un solo número. Para comparar dos o más distribuciones de ingreso se usan, a menudo, herramientas de ordenamiento, de las cuales la curva de Lorenz es la más conocida. Estas herramientas están basadas, principalmente, en cuantiles y en

distribuciones acumulativas. Por ejemplo, los criterios dominantes de primer orden están simplemente definidos para una distribución F por el cuantil funcional $Q = Q(F, q)$, es decir por q^{th} cuantiles de la distribución F . Por ejemplo $Q(F, 0.5)$ es la mediana. Dos (o más) distribuciones F_1 y F_2 se comparan entonces en cada nivel de q y si para F_1 todos los cuantiles son más grandes o igualan a aquellos de F_2 , entonces se dice que F_1 de primer orden domina a F_2 . La implicación económica es que dada una función de bienestar (con algunas condiciones muy bondadosas), el nivel de bienestar en la población con distribución del ingreso F_1 es más alto. En la práctica, uno puede llegar a tres conclusiones: la dominación de una u otra distribución, o también ninguna dominación porque el orden se invierte en algunos cuantiles. Un concepto más fuerte de dominación (por lo que se refiere a las condiciones de la función de bienestar) es la herramienta de dominación llamada de segundo orden, basada en la distribución acumulativa funcional de C .

$$C(F, q) = \int^{Q(F, q)} x dF(x)$$

En otras palabras, $C(F, q)$ es el ingreso total ganado por la proporción q del más pobre. La función define la curva de Lorenz, la cual para cada proporción de q traza la proporción del ingreso total de la población ganada por la proporción q de los más pobres.

CUADRO 6

Índices de pobreza robustos y clásicos para las regiones de Túnez

	Análisis Robusto			Análisis Clásico		
	HCR	PG	FGT	HCR	PG	FGT
Gran Túnez	4.4	0.81	0.24	4.3	0.79	0.24
Litoral urbano	3.6	0.57	0.15	3.6	0.61	0.16
Interior urbano	9.1	2.12	0.75	13.1	3.36	1.26
Litoral rural	8.4	1.63	0.54	10.9	2.42	0.82
Interior rural	11.9	3.15	1.18	16.0	4.28	1.66

El tema que nos interesa aquí es investigar si las herramientas pueden ser influidas por la contaminación de los datos, que lleva a revertir las órdenes cuando se comparan las distribuciones empíricas. Cowell y Victoria-Feser (1996c) realizaron esta investigación usando el IF y concluyeron que cantidades infinitesimales de contaminaciones (en las colas de la distribución) pueden distorsionar el resultado obtenido para el criterio de dominación de segundo orden. No queda claro cómo se pueda resolver el problema. El recurso de los modelos paramétricos no es una buena solución, ya que uno puede demostrar que en

algunos modelos las curvas de ordenamiento nunca se cruzan. Cowell y Victoria-Feser (2000a, b) proponen dos enfoques: uno pragmático, basado en muestras arregladas y otro, semi-paramétrico. El primero necesita un poco más de investigación, considerando que el semi-paramétrico parece muy prometedor. La idea es usar un modelo paramétrico para acomodar la cola de la distribución y, entonces, construir una curva de ordenamiento empírica para el volumen de la distribución mezclada con una herramienta de ordenamiento para la cola de la distribución. El modelo apropiado es el de Pareto y el ajuste debiera ser robusto. Una pregunta que queda abierta es: ¿Qué proporción de los datos en la cola debe ser considerada?

Para ilustrar este punto, Cowell & Victoria-Feser (2000a) hicieron el ejercicio de simulación siguiente: Dos muestras de 10,000 observaciones donde simularon una distribución Dagunm I, cuya propiedad es que, para los ingresos grandes, la distribución converge en la distribución de Pareto. Los parámetros fueron escogidos en orden para conseguir dos distribuciones en las que una domine exactamente a la otra. Dejemos que sea la distribución empírica que domina y el dominado. fue contaminado al multiplicar 0.25% de las observaciones más grandes por 10, que da la mezcla empírica de distribución. Las tres distribuciones empíricas se muestran en la Figura 4.

Podemos ver que con sólo 0.25% de datos extremos, la clasificación de las distribuciones es completamente invertida. Cuando la cola superior se modela usando la distribución de Pareto (5% de la cola superior), la situación cambia. La Figura 5 describe la curva de Lorenz de junto con las curvas semi-paramétricas de Lorenz de , usando el MLE clásico y un estimador robusto para el parámetro del modelo de Pareto (véase Cowell y Victoria-Feser, 2000a). Podemos apreciar que en una curva de Lorenz semi-paramétrica robusta el ordenamiento se conserva. No hay ninguna diferencia visual real entre la curva de Lorenz semi-paramétrica de ordenamiento robusta con datos contaminados y la curva de Lorenz empírica con datos no contaminados para la misma distribución original (véase Cowell y Victoria-Feser, 2000a). La curva semi-paramétrica de Lorenz, usando el MLE, también se distorsiona por los datos extremos, aunque menos que la curva empírica de Lorenz con datos contaminados. Esto no es sorprendente dado que es bien conocido que el MLE no es robusto.

Conclusiones

Hemos mostrado que las técnicas robustas pueden jugar un papel útil en el análisis de la distribución del ingreso. Proveen estudios más confiables, medidas más estables de desigualdad y herramientas de ordenamiento. No defendemos el uso único de técnicas robustas, pero creemos que proporcionan información muy útil cuando se combinan con el enfoque clásico. Éstas también tienen la ventaja

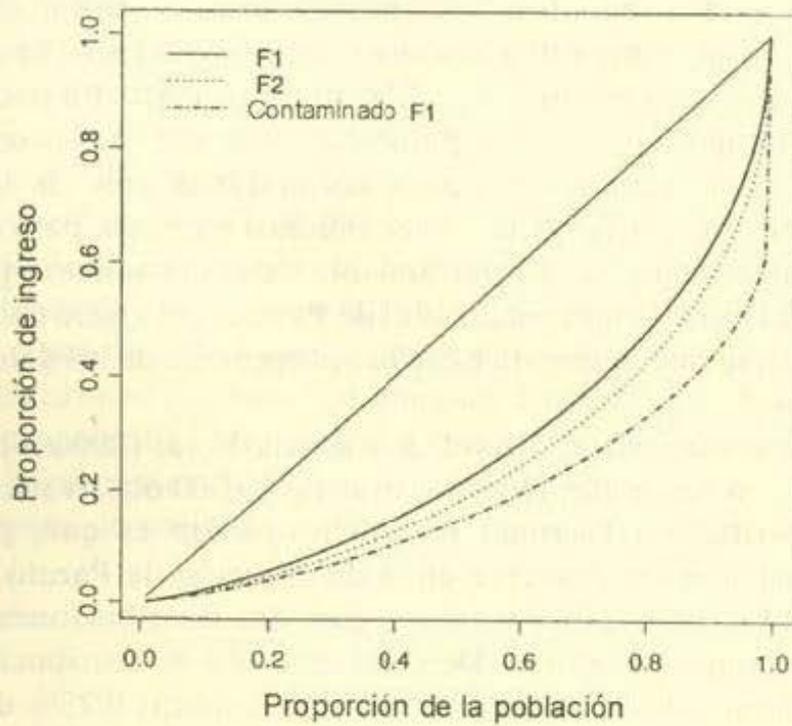


Figura 4. Comparaciones entre distribuciones empíricas contaminadas y no contaminadas, en la curva de Lorenz.

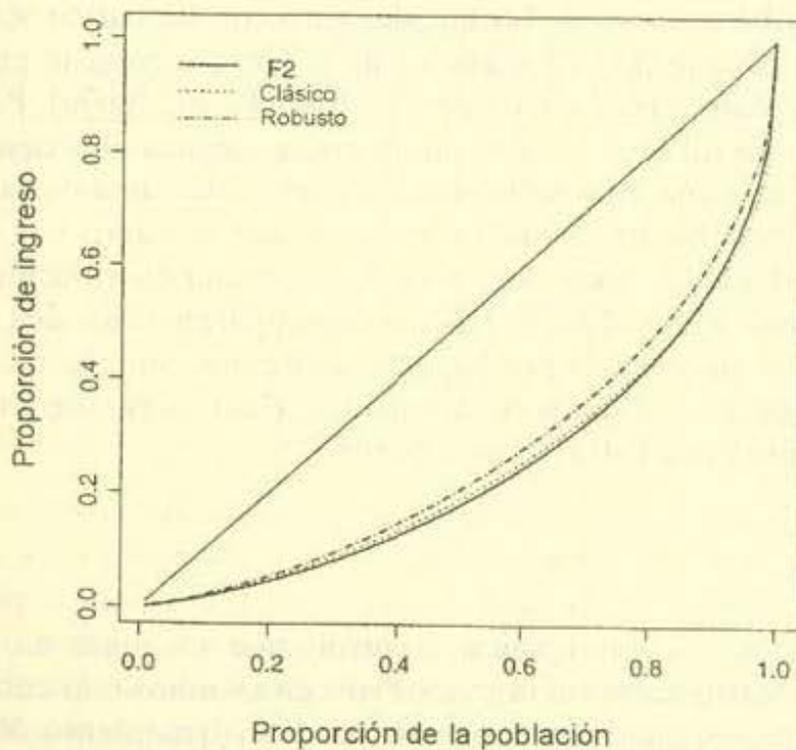


Figura 5. Curvas semi-paramétricas de Lorenz: clásicas y robustas.

de ser más bien métodos objetivos porque no confían en el juicio del analista quien decide, a menudo, qué datos podrían ser considerados como extremos. Queda mucho todavía por desarrollar en este campo de investigación, sobre todo los aspectos computacionales. De hecho, aunque las técnicas robustas son muy importantes son también bastante difíciles de programar y computacionalmente intensivas. Una solución podría encontrarse al desarrollar estimadores robustos menos eficaces, que serían más fáciles de computar. Pero ésa es otra historia.

Referencias

- Atkinson, A. C. (1970). "A method for discriminating between models, en *Journal of the Royal Statistical Society*, serie B, No. 32, pp. 323-353.
- Ayadi, M., Matoussi, M. S. y Victoria-Feser, M. P. (1998). *Urban rural poverty comparisons in Tunisia: A robust statistical approach*, Cahiers du département d'économétrie, University of Geneva.
- Ben Horim, M. (1990). "Stochastic dominance and truncated sample data, en *Journal of Financial Research*, No. 13, pp. 105-116.
- Cowell, F. A. (1980). "Generalized entropy and the measurement of distributional change, en *European Economic Review*, No. 13, pp. 147-159.
- Cowell, F. A. y Victoria-Feser, M. P. (1994). *Robustness properties of poverty indices*, DARP discussion papers, No. 8, STICERD, London School of Economics, London WC2A 2AE.
- Cowell, F. A. y Victoria-Feser, M. P. (1996a). "Poverty measurement with contaminated data: A robust approach, en *European Economic Review*, No. 40, pp. 1761-1771.
- Cowell, F. A. y Victoria-Feser, M. P. (1996b). "Robustness properties of inequality measures, en *Econometrica*, No. 64, pp. 77-101.
- Cowell, F. A. y Victoria-Feser, M. P. (1996c). "Welfare judgments in the presence of contaminated data, DARP discussion paper, No. 13, STICERD, London School of Economics, London WC2A 2AE.
- Cowell, F. A. y Victoria-Feser, M. P. (2000a). "Robust Semi-Parametric Lorenz Curves, en DARP discussion paper, No. 50, STICERD, London School of Economics, London WC2A 2AE.
- Cowell, F. A. y Victoria-Feser, M. P. (2000b). "Distributional Dominance with Dirty Data, DARP discussion paper, No. 51, STICERD, London School of Economics, London WC2A 2AE.
- Cox, D. R. (1961). "Tests of separate families of hypotheses, en *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability I*, Berkeley, pp. 105-123, University of California Press.

- Cox, D. R. (1962). "Further results on tests of separate families of hypotheses, en *Journal of the Royal Statistical Society*, serie B, No. 24, pp. 406-424.
- Cressie, N. A. C. y Read, T. R. C. (1984). "Multinomial goodness-of-fit tests, en *Journal of the Royal Statistical Society*, serie B, No. 46, pp. 440-464.
- Dagum, C. (1980). "Generating systems and properties of income distribution models, en *Metron*, 38(3-4), 3-26.
- Dalton, H. (1920). "The measurement of the inequality of incomes, en *Economic Journal*, No. 30, pp. 348-361.
- Dempster, A. P., Laird, M. N. y Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm, en *Journal of the Royal Statistical Society*, serie B, No. 39, pp. 1-22.
- Department of Social Security (1992). *Households below Average Income*, 1979-1988/89, London, HMSO.
- Gini, C. (1955). *Memorie di Metodologia Statistica*, Roma: Pizetti e Salvemini Libreria Eredi Virgilio Veschi.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*, John Wiley (ed.), New York.
- Hampel, F. R. (1968). *Contribution to the theory of Robust Estimation*. Ph. D. thesis, University of California. Berkeley.
- Hampel, F. R. (1974). "The influence curve and its role in robust estimation, en *Journal of the American Statistical Association*, No. 69, pp. 383-393.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P.J. y Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence functions*, John Wiley (ed.), New York.
- Heritier, S. y Ronchetti, E. (1994). "Robust bounded. Influence tests in general parametric models, en *Journal of the American Statistical Association*, No. 89, p. 427.
- Huber, P. J. (1964). "Robust estimation of a location parameter, en *Annals of Mathematical Statistics*, No. 35, pp. 73-101.
- Huber, P. J. (1981). *Robust Statistics*, John Wiley (ed.), New York.
- Jenkins, S. P. (1997). "Trends in real income in Britain: A microeconomic analysis, en *Empirical Economics*, No. 22, pp. 483-500.
- Kolm, S. C. (1976a). "Unequal inequalities, en *Journal of Economic Theory*, No. 12, pp. 416-442, (parte I).
- Kolm, S.C. (1976b). "Unequal inequalities, en *Journal of Economic Theory*, No. 13, pp. 82-111, (parte II).
- Little, R. J. A. y Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

- McDonald, J. B. y Ransom, M. R. (1979). "Functional forms, estimation techniques and the distribution of income, en *Econometrica*, No. 47, pp. 1513-1525.
- Nelson, R. D. y Pope, R. D. (1990). "Imprecise tail estimation and the empirical fail of stochastic dominance, ASA Proceedings, Business and Economic Statistics Section, pp. 374-379.
- Prieto Alaiz, M. y Victoria-Feser, M. P. (1996). "Modelling income distribution in Spain: A robust parametric approach, *DARP* discussion paper, No. 20, STICERD, London School of Economics, London WC2A 2AE.
- Ravallion, M. (1993). *Poverty comparison. Volume 56 of Fundamentals of Pure and Applied Economics*, Chur, Switzerland Harwood Academic Press.
- Ravallion, M. y Bidani, B. (1994). "How robust is a poverty profile?, en *The World Bank Economic Review*, No. 8, pp. 75-102.
- Ronchetti, E. (1982). *Robust Testing in Linear Models: The Infinitesimal Approach*. Ph. D. thesis. ETH, Zürich, Switzerland
- Rousseeuw, P. J. y Ronchetti, E. (1979). "The influence curve for tests, en *Research Report*, No. 21, ETH, Zürich, Switzerland.
- Rousseeuw, P. J. y Ronchetti, E. (1981). "Influence curves for general statistics, en *Journal of Computational and Applied Mathematics*, No. 7, pp. 161-166.
- Sen, A. K. (1976). "Poverty: An ordinal approach to measurement, en *Econometrica*, 44, 219-231.
- Theil, H. (1967), *Economics and Information Theory*, North-Holland, Amsterdam.
- Van Praag, B., Hagenaars, A. y Van Eck, W. (1983). "The influence of classification and observation errors on the measurement of income inequality, en *Econometrica*, No. 51, pp. 1093-1108.
- Victoria-Feser, M. P. (1993). *Robust Methods for Personal Income Distribution Models*. Ph. D. thesis, University of Geneva, Switzerland. Thesis no. 384.
- Victoria-Feser, M. P. (1997). "Robust model choice test for non-nested hypothesis, en *Journal of the Royal Statistical Society*, serie B, No. 59, pp. 715-727.
- Victoria-Feser, M. P. y Ronchetti, E. (1994). "Robust methods for personal income distribution models, en *The Canadian Journal of Statistics*, No. 22, pp. 247-258.
- Victoria-Feser, M. P. y Ronchetti, E. (1997). "Robust estimation for grouped data, en *Journal of the American Statistical Association*, No. 92, pp. 333-340.