

Callealta Barroso, Francisco Javier; Fedriani Martel, Eugenio M.; Martín Caraballo, Ana M.; Sánchez Sánchez, Ana María

Article

Análisis de la evolución temporal de las desigualdades con datos irregulares

Revista de Métodos Cuantitativos para la Economía y la Empresa

Provided in Cooperation with:

Universidad Pablo de Olavide, Sevilla

Suggested Citation: Callealta Barroso, Francisco Javier; Fedriani Martel, Eugenio M.; Martín Caraballo, Ana M.; Sánchez Sánchez, Ana María (2012) : Análisis de la evolución temporal de las desigualdades con datos irregulares, Revista de Métodos Cuantitativos para la Economía y la Empresa, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 13, pp. 73-96

This Version is available at:

<https://hdl.handle.net/10419/113841>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-sa/3.0/es/>



Análisis de la evolución temporal de las desigualdades con datos irregulares

CALLEALTA BARROSO, FRANCISCO JAVIER
Departamento de Estadística Económica, Estructura Económica y O.E.I.
Universidad de Alcalá de Henares
Correo electrónico: franciscoj.callealta@uah.es

FEDRIANI MARTEL, EUGENIO M.
Departamento de Economía, Métodos Cuantitativos e Historia Económica
Universidad Pablo de Olavide, de Sevilla (España)
Correo electrónico: efedmar@upo.es

MARTÍN CARABALLO, ANA M.
Departamento de Economía, Métodos Cuantitativos e Historia Económica
Universidad Pablo de Olavide, de Sevilla (España)
Correo electrónico: ammarcar@upo.es

SÁNCHEZ SÁNCHEZ, ANA MARÍA
Departamento de Economía, Métodos Cuantitativos e Historia Económica
Universidad Pablo de Olavide, de Sevilla (España)
Correo electrónico: amsansan@upo.es

RESUMEN

En este artículo se proporciona una herramienta que permite estudiar las distribuciones de renta en una población a lo largo del tiempo, sin importar que los datos se distribuyan de forma irregular. Incluso, se posibilita la aproximación de curvas de Lorenz en instantes en los que no se cuenta con los datos necesarios. Finalmente, se presenta un ejemplo de aplicación de la técnica en un país europeo en que las desigualdades son especialmente relevantes: España.

Palabras clave: curva de Lorenz; índice de Gini; series temporales irregulares; distribución de la renta.

Clasificación JEL: D31; D63; C81.

MSC2010: 91B82; 60E15; 52A38.

Analyzing the Income Inequalities with Irregular Time Series

ABSTRACT

This article provides a tool to study over time the distributions of income in a population, regardless the irregular distribution of data. We even approach Lorenz curves at the times when we lack the required data. Finally we apply the present technique to a European country where inequalities are particularly relevant: Spain.

Keywords: Lorenz curve; Gini index; irregular time series; income distribution.

JEL classification: D31; D63; C81.

MSC2010: 91B82; 60E15; 52A38.



1. Introducción

Las curvas de Lorenz, complementadas con el índice de Gini, constituyen probablemente la herramienta más utilizada para estudiar el grado de desigualdad económica que se manifiesta en las distribuciones de renta en una población. Su solidez está bien documentada por los numerosos autores que se refieren a ellas. A este respecto pueden consultarse, por ejemplo, [13] y [15]. No obstante, aún no se ha resuelto satisfactoriamente el problema de la medición del grado de desigualdad registrado globalmente a lo largo de un intervalo de tiempo. Si bien, cuando se dispone de datos regularmente espaciados y fiables en todas las unidades de análisis, se pueden utilizar a este fin numerosas técnicas de estadística descriptiva básica o de análisis funcional; sin embargo, cuando los datos están irregularmente espaciados la aplicación de tales técnicas no resulta tan inmediata, debiéndose adoptar otras alternativas como, por ejemplo, considerar variantes de las ideas propuestas en [14], que fueron posteriormente utilizadas en [4] para analizar las desigualdades en cuatro países europeos. En lo que sigue, además de introducir algunas mejoras a la técnica original y de analizar su aplicación a las series que presentan específicamente problemas de regularidad, se llega a proponer otra posible utilidad de la misma: la obtención de curvas de Lorenz en casos en los que no se dispone de suficiente información para construirla por los procedimientos habituales.

En esencia, el primer objetivo de este trabajo es introducir un instrumento que permita expresar el grado de desigualdad de una población medido globalmente a lo largo de un período de tiempo; y, en consecuencia, también se pretende comparar los grados de desigualdad de varias poblaciones (por ejemplo, regiones) medidos globalmente a lo largo de un mismo período. Para ello, se propondrá un indicador que generaliza el índice de Gini, incorporando la dimensión temporal. El segundo objetivo, como ya se ha insinuado, es proporcionar un procedimiento que permita obtener curvas de Lorenz en instantes no observados, a partir de la información observada en el resto del período de tiempo del estudio.

Este artículo consta de cuatro secciones, además de esta breve introducción. La siguiente se dedica a la presentación de los conceptos esenciales para desarrollar la técnica. Sigue un resumen de las propiedades que pueden resultar más interesantes y útiles al investigador que desee aplicarla. Después, la parte teórica se ilustra con datos de España en el período 2003-2009 y se propone su aplicación en diferentes situaciones. Finalmente, se resumen algunas conclusiones. Para facilitar la lectura del artículo, se ha decidido incorporar las cuestiones más técnicas en un Anexo, tras las referencias bibliográficas.

2. Definiciones básicas

El interés de la comunidad científica en la problemática de la desigualdad económica experimentó un notable incremento tras la publicación de los emblemáticos trabajos de Atkinson en 1970 [1] y de Sen en 1973 [19]. Desde entonces los artículos y libros sobre este campo han proliferado enormemente, derivando hacia otros cercanos de gran interés social, como los estudios sobre pobreza, movilidad, polarización y privación.

En concreto, Rawls definió la desigualdad en [16] a partir de las diferencias entre los beneficios y cargas vinculados directa o indirectamente a ellos, tales como prestigio o riqueza, o sujeción a imposición fiscal y a servicios obligatorios. Por tanto, para medir la desigualdad, beneficios y cargas han de ser cuantificados, lo que resulta problemático; habitualmente, y quizás de una forma excesivamente simplista, esta cuantificación se suele hacer a través de las distintas componentes y variantes de la renta. Pese a la existencia de voces críticas al respecto, es evidente que la distribución personal de la renta está relacionada, lógicamente, con la desigualdad y el reparto de la riqueza. Así, al estudiar la distribución de la renta, podemos ver qué parte de la población concentra más o menos recursos (acercándonos al estudio de la desigualdad); o qué otra parte de la población no tiene acceso a niveles de vida que se consideren aceptables en una región (lo que nos acerca al estudio de la pobreza). Este trabajo se centrará estrictamente en el campo de la desigualdad y se basa en la concepción de desigualdad subyacente en los trabajos de Gini y Lorenz, por lo que se parte de los instrumentos que específicamente aportaron ellos y que se exponen a continuación.

La curva de Lorenz de una distribución de rentas se define como una función que, para cualquier proporción p ($0 \leq p \leq 1$), indica la proporción de renta, respecto del total, poseída por la subpoblación cuyas rentas no superan el valor del cuantil de orden p (obviamente, dicha subpoblación está formada por la proporción p de la población con menores rentas). Dicha curva suele denotarse por $L_x(p)$ y, desde un punto de vista empírico-descriptivo, para un vector de rentas observadas, $x = (x_1, x_2, \dots, x_n)$, la curva de Lorenz se aproxima mediante una poligonal cuyos vértices son los puntos $(p, L_x(p))$, siendo:

$$L_x(p) = L_x\left(\frac{k}{n}\right) = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^k x_i}{n\mu_x},$$

donde μ_x es la media de la distribución x , $k = 0, 1, 2, \dots, n$ y n representa el tamaño de la población (ver [3], [17] y [18]).

Algunas de las propiedades más relevantes que verifica la curva de Lorenz son que $L_x(0) = 0$, que $L_x(1) = 1$ y que, además, es siempre no-decreciente y convexa en el intervalo $[0, 1]$. Por otro lado, prácticamente cualquier curva que verifique estas condiciones

puede considerarse una curva de Lorenz de una población hipotética.

Las curvas de Lorenz se pueden emplear para comparar poblaciones, desde el punto de vista de la desigualdad que presentan, en el conocido como “sentido de Lorenz”. Así, al comparar dos distribuciones, x e y , se dice que la distribución x domina en el sentido de Lorenz a la distribución y cuando la curva de la primera distribución queda siempre por encima de la segunda, en todo el intervalo abierto $(0, 1)$. Y si la distribución x domina en el sentido de Lorenz a la y , eso significa que la primera población presenta menos desigualdad que la segunda. Pero llegar a una conclusión de este tipo exige que las respectivas curvas de Lorenz no se corten, lo que suele ser bastante infrecuente en la gran mayoría de situaciones de interés.

Tratando de dar una respuesta parcial al problema de comparar la desigualdad entre distribuciones que no se dominan estrictamente en el sentido de Lorenz y empleando la información que se recoge en sus respectivas curvas de Lorenz, se puede medir numéricamente el grado de desigualdad dentro de cada distribución mediante el índice de Gini, que mide el grado de concentración de la renta. Usualmente se calcula multiplicando por 2 la llamada “área de desigualdad” (área de la región comprendida entre la curva de Lorenz y la línea de distribución igualitaria $L_x(x) = x$). Si G es el índice de Gini asociado a la curva $L_x(p)$:

$$G = 2 \int_0^1 (p - L_x(p)) dp = 1 - 2 \int_0^1 L_x(p) dp.$$

En el caso extremo de desigualdad absoluta, el índice de Gini sería igual a 1; en el caso contrario, en el que se da una distribución totalmente igualitaria de la renta, el índice de Gini sería igual a 0. Cuanto más cercano a 0 se encuentre el índice de Gini, más igualitario será el reparto de los ingresos, mientras que cuanto más cercano esté a 1, menos igualitario resultará ese reparto y la renta estará más concentrada.

Debe tenerse en cuenta que, además de la anterior, hay muchas formas alternativas de expresar el índice de Gini, como por ejemplo las utilizadas en [5], [2], [6] o [12]. A los efectos de este artículo, no hay problema en considerar cualquiera de dichas definiciones.

Hay que reconocer que el Índice de Gini no es el único indicador de desigualdad posible y que hay investigadores que prefieren utilizar en su lugar otros índices que verifican ciertas propiedades deseables, que el índice de Gini no cumple, como ocurre por ejemplo con: el índice de Theil [20], que verifica la propiedad de descomposición aditiva (es decir, que el índice de desigualdad de una población se puede obtener a partir de los índices de desigualdad de cada una de las subpoblaciones en que se subdivide); o el de Atkinson [1], que recoge de forma flexible la llamada “aversión al riesgo”, que puede variar de una sociedad a otra; entre otros. Sin embargo, por su simple interpretación y aceptación más

generalizada, este artículo se ocupa especialmente del índice de Gini, como una primera experiencia, con la esperanza de que los resultados aquí obtenidos puedan servir para alumbrar otras extensiones a otros índices en un futuro. Así pues, a continuación se presenta el índice de desigualdad en el tiempo (*IDT*), que se propone como instrumento para medir el grado de desigualdad de las distribuciones de renta, que se aprecia globalmente a lo largo de un intervalo de tiempo fijado.

2.1. Indicador de las desigualdades a lo largo del tiempo

Inicialmente supondremos que se dispone de sendos conjuntos de datos conocidos (de renta) para cada una de las unidades de análisis del estudio (por ejemplo, regiones) y para diferentes instantes de tiempo del período que se desea analizar. Para facilitar la siguiente exposición, se fijará una de dichas unidades de análisis y se considerarán sus distribuciones de renta a lo largo del tiempo (posteriormente se generalizaría el mismo proceso para las otras unidades de análisis).

Se puede representar conjuntamente las correspondientes curvas de Lorenz de dichas distribuciones de renta en un espacio de 3 dimensiones, determinadas por tres variables (x, y, z) . Así, las variables x e y (sobre sus correspondientes ejes OX y OY) representarán, respectivamente, a modo de variables independientes, las proporciones p y el instante de tiempo t que determinarán la ordenada de la curva de Lorenz para la proporción p (de población con menor renta) sobre la distribución de renta observada en el instante t ; reservando la variable z (sobre su correspondiente eje OZ) para representar, a modo de variable dependiente, la ordenada de dicha curva de Lorenz.

Así pues, fijada una de las unidades de análisis, se dispone de datos en diferentes instantes de tiempo que, desde un punto de vista empírico-descriptivo, determinan nodos sobre el plano (X, Y) , cuyas coordenadas (x, y) se componen con el instante de observación (y) y el orden (x) del cuantil de renta con el que se corresponde cada dato observado para la distribución de renta en dicho instante. Dichos nodos, junto con los valores de la variable dependiente en ellos (ordenadas de las curvas de Lorenz), se utilizarán para construir una superficie poliédrica a partir de técnicas sencillas de interpolación (en concreto, se trata de aproximaciones funcionales lineales). Una descripción más detallada de este proceso puede encontrarse en [14]. Si procedemos paso a paso, se tiene la siguiente descripción algorítmica:

Supóngase que, para cada unidad de análisis, se observa la distribución de la renta en $k + 1$ instantes de tiempo t (t_0, t_1, \dots, t_k), obteniéndose $k + 1$ conjuntos de datos con tamaños n (n_0, n_1, \dots, n_k) no necesariamente iguales. A cada una de las k curvas de

Lorenz, en cada instante t , las representaríamos por $L_t(x)$ para $t \in \{t_0, t_1, \dots, t_k\}$, con $x \in [0, 1]$ (por ser $L_t(x)$ una curva de Lorenz). Sin embargo, desde una perspectiva empírico-descriptiva, basta con conocer sus valores en los nodos de cada corte temporal (sobre cada recta $y = t$). Estos nodos vienen dados por $\Pi_X \equiv \{0 = x_0 < x_1 < \dots < x_n = 1\}$, una partición del intervalo $[0, 1]$. Obviamente, Π_X viene recomendado por los datos de los que se disponga o por los que fueron utilizados para construir las curvas de Lorenz. Realmente no existe ninguna limitación para elegir los datos, salvo que es necesario contar con algún x_i distinto del cero y el uno. Más adelante, en la Figura 1, se presentará un ejemplo de conjunto de nodos irregular. Fijado el valor (disponible) en $x_j \in \Pi_X$ para cada instante de tiempo t_i , se tienen los puntos $(x_j, t_i, L_{t_i}(x_j))$ para $i = 1, \dots, k$ y $j = 1, \dots, n$, donde $\Pi_T \equiv \{t_0 = 0 < t_1 < \dots < t_{k-1} < t_k = T\}$, es una partición del intervalo $[0, T]$, que normalmente vendrá determinada por la disponibilidad de los datos.

A partir de aquí, se puede calcular una función de interpolación para los puntos de \mathbb{R}^3 que surgen de los nodos y sus respectivas ordenadas. Para ello, sería necesario utilizar una interpolación en dos variables y en la malla rectangular $[0, 1] \times [0, T]$. Supongamos que decidimos resolver este problema de interpolación mediante técnicas sofisticadas, como el método de interpolación de Lagrange (luego se recomendará un procedimiento mucho más sencillo). A la función de interpolación obtenida utilizando la interpolación de Lagrange para los puntos $(x_j, t_i, L_{t_i}(x_j))$ se representa ahora por $L(x, t)$ (que es un polinomio en las variables x y t de grado n para la variable x y de grado k para la variable t , luego es un polinomio de grado $n + k$), donde $x \in [0, 1]$ y $t \in [0, T]$. Por supuesto, este procedimiento debería repetirse para cuantas unidades de análisis compongan el estudio.

A partir de las funciones interpoladas $L(x, t)$ (que luego propondremos reemplazar por otras más simples, que producirán una o varias superficies poliédricas), se va a definir el indicador propuesto IDT , que permitirá comparar la desigualdad en distintas zonas durante un período temporal de amplitud T .

Construida dicha superficie, $z = L(x, t)$, se calcula el volumen que queda por debajo, aunque nuestra propuesta es aproximarlos por la suma de los volúmenes de cada uno de los prismas triangulares truncados en que se puede descomponer el volumen que queda por debajo de la superficie poliédrica.

Definición 2.1 *En las condiciones anteriores y con la notación establecida,*

$$IDT = T - 2 \int_0^1 \int_0^T L(x, t) dt dx.$$

En cualquier caso,¹ el indicador IDT se puede interpretar como una generalización del índice de Gini, cuyo valor se corresponde con el doble del área comprendida entre

¹Como se verá enseguida, también tiene sentido definir el IDT sin multiplicar por dos la integral.

la diagonal del cuadrado $[0, 1] \times [0, 1]$ y la curva de Lorenz. Por su parte, el indicador IDT , tal y como queda determinado por la Definición 2.1, representa el doble del volumen comprendido entre el plano $z = x$ (la “diagonal” del cubo $[0, 1] \times [0, T] \times [0, 1]$) y la superficie interpolada por las curvas de Lorenz en cada instante. Lógicamente, cuanto menor sea tal volumen, menor desigualdad existirá en el período de tiempo estudiado.

Recordemos que el índice de Gini es un ejemplo de indicador de desigualdad construido a partir de la curva de Lorenz, el cual calcula la desigualdad de cierta distribución en un determinado instante de tiempo. En este caso, al definir el indicador IDT , se persigue un objetivo distinto, ya que una vez calculadas (o aproximadas) las curvas de Lorenz para una misma distribución en instantes de tiempo distintos, proponemos caracterizar el grado de desigualdad globalmente observado, o “típico”, en el período de observación, a partir de cómo han evolucionado dichas curvas; es decir, de cómo ha evolucionado la desigualdad a lo largo de un período de tiempo. Este es el primer objetivo del IDT .

3. Algunas propiedades

Cuando el número de unidades de análisis (por ejemplo, zonas geográficas) objetos del estudio que se realiza es reducido y el número de nodos es suficiente, sucede que el indicador IDT se comporta de forma similar a la media de los índices de Gini en cada instante de tiempo para cada una de las zonas. Esto puede hacer pensar que el IDT no es suficientemente útil como para justificar su estudio, pero la ventaja más importante del uso de este indicador frente a la media aritmética de los índices de Gini es que permite la realización de un análisis temporal comparativo incluso cuando no se tengan datos con estructuras análogas en todas las unidades de análisis; es decir, cada unidad de análisis puede tener distinto número de datos, estos pueden estar desigualmente distribuidos (diferente espaciamiento o un espaciamiento no uniforme), pueden incluirse nuevos datos conforme se vayan obteniendo, etc. y, además, el proceso propuesto es independiente de la irregularidad temporal de los datos con los que se cuenta, algo que no se verifica para otras técnicas estadísticas.

Pero, ciertamente, una desventaja del indicador IDT (en su primera versión, tal y como se presentó en la Definición 2.1) es su coste computacional, ya que para su cálculo sería necesario determinar las curvas de Lorenz para varios instantes de tiempo, además de calcular una función de interpolación polinómica en dos variables.

Por otro lado, es posible normalizar este indicador dividiendo por T . Pudiera, incluso, considerarse la posibilidad de dividir por T antes de realizar la triangulación, para que el rectángulo de la malla base sea un cuadrado.

Hemos detectado, además, que el cálculo del *IDT* (en su primera versión) puede presentar algunos problemas adicionales, como pueden ser: la sensibilidad a que algunos datos sean incorrectos (por ejemplo, si los datos son agregados, pueden haberse producido errores significativos por falta de información al agregarlos); errores significativos de pérdida de información; se incrementa el riesgo de errar el valor del indicador por la complejidad de su cálculo; y, finalmente, puede ocurrir que la “superficie de Lorenz”, $z = L(x, t)$, obtenida mediante interpolación, y en la que se basa el cálculo del *IDT*, no se ajuste convenientemente conforme a lo deseado (esto es bastante probable, pues tal superficie, en general, no es convexa). Para resolver estos problemas, se propondrán pautas alternativas de cálculo del *IDT* que evitan las complejas interpolaciones y que permiten calcular el *IDT* mediante operaciones muy sencillas y fácilmente implementables con un programa de computación simbólica o, incluso, en una hoja de cálculo, lo que mejora y dota de mayor potencia a este indicador de desigualdad.

Una primera pauta sería, en vez de trabajar con los datos agregados, utilizar directamente todos los datos (o microdatos, procedentes de las encuestas), de forma que se estarían reduciendo los posibles errores por pérdida de información en los datos. No obstante, es obvio que un gran número de datos, además de proporcionar una mejor aproximación final, también genera un mayor coste computacional. A continuación se verá cómo solucionar este problema gracias a las superficies poliédricas de las que ya se había comentado algo con anterioridad.

Consideremos que se tiene una malla de puntos en el rectángulo $[0, 1] \times [0, T]$ (los nodos en los que se cuenta con datos para la correspondiente curva de Lorenz); si los denotamos por (x_j, t_i) , los datos son $L_{t_i}(x_j)$ y la función $L(x, t)$ debe pasar por todos ellos. Recordemos que (considerando las curvas de Lorenz en tantos por uno y no en tantos por ciento) siempre se verifica que: $L(0, t) = 0$ para todo $t \in [0, T]$; y $L(1, t) = 1$ con $t \in [0, T]$. Por tanto, siempre podrían considerarse nodos informados, tanto los $(1, t_i)$ como los $(0, t_i)$, cualquiera que sea el $t_i \in [0, T]$.

A continuación se desea establecer una triangulación en el rectángulo $[0, 1] \times [0, T]$ que utilice todos los nodos anteriores y que pueda ser útil para dividir el volumen bajo la superficie $z = L(x, t)$ en prismas triangulares truncados. De todas las triangulaciones posibles, las más convenientes para nuestros intereses son aquellas que minimizan las áreas de los triángulos (más precisamente, con el mayor número posible de triángulos y, a la vez, minimizando el área del mayor triángulo) y, de entre todas ellas, las que minimizan sus perímetros, porque todo ello supondrá un mejor ajuste a la “superficie de Lorenz” por parte de la superficie poliédrica que se va a definir. Además, merece la pena tener en cuenta que dicha superficie poliédrica no es necesariamente convexa (a pesar de serlo todas las curvas de Lorenz en cada instante), por lo que tampoco es una buena idea definir

triángulos con vértices en dos instantes temporales no consecutivos. Según todo lo anterior, para encontrar estas triangulaciones óptimas, se puede recurrir al siguiente algoritmo, que presenta una posible variante en algunos de los pasos, por lo que puede llegar a producir varias triangulaciones distintas en la misma malla:

Algoritmo para triangularizar la malla de puntos:

1. Dos nodos de la forma (x_j, t_i) y (x_{j+1}, t_i) están siempre conectados para todos los instantes.
2. Por las restricciones explicadas anteriormente, se reduce el problema a obtener una triangulación entre los nodos correspondientes a dos instantes consecutivos, esto es, t_h y t_i , donde $h = i - 1$ o $h = i + 1$.
3. Se ordenan los valores de la primera variable de los nodos correspondientes en t_h y t_i ; es decir, todos los nodos en esos dos instantes de tiempo determinarían un conjunto de valores distintos en la primera variable: $\Pi_X = \{x_0 = 0, x_1, \dots, x_n = 1\}$. Téngase en cuenta que algunos x_j pueden corresponder a nodos en ambos instantes y otros pueden corresponder a uno solo.
4. Se conectan (x_0, t_i) y (x_0, t_h) .
5. Supongamos conectados los puntos (x_j, t_i) y (x_k, t_h) , donde j y k pueden ser iguales o distintos, pero consecutivos; esto significa que en el instante t_i se tienen conectados todos los nodos hasta el x_j (pongamos que el siguiente nodo en el instante t_i , aún no conectado, se llama x_r) y que en el instante t_h tenemos conectados todos los nodos hasta el x_k (pongamos que el siguiente, aún no conectado, se llama x_s).
6. Se calcula la distancia entre (x_j, t_i) y (x_s, t_h) y la distancia entre (x_k, t_h) y (x_r, t_i) . Se elige la menor de esas distancias y se conectan los nodos correspondientes. En el caso en que ambas distancias sean iguales, hay que elegir una de las parejas de nodos y la elección de la otra pareja produciría otra triangulación distinta.
7. Se vuelve al paso 5 mientras queden nodos por conectar (es decir, hasta que j y k sean los dos 1).

En la Figura 1 se representa gráficamente la triangulación en un ejemplo en el que surgen únicamente dos posibilidades si se aplica el algoritmo anterior. El que el algoritmo de triangulación anterior pueda proporcionar varias triangulaciones diferentes no parece producir graves problemas desde el punto de vista aplicado, pese a que después de la Proposición 3.1 será evidente que las triangulaciones obtenidas no son equivalentes para nuestros propósitos. Como en general no será posible determinar cuál de las triangulaciones proporciona un mejor ajuste de la superficie poliédrica a la superficie teórica que se desea aproximar, se podría elegir una de ellas, pero lo más razonable parece ser realizar los cálculos siguientes para cada una de las triangulaciones (lo que permitirá alcanzar varias

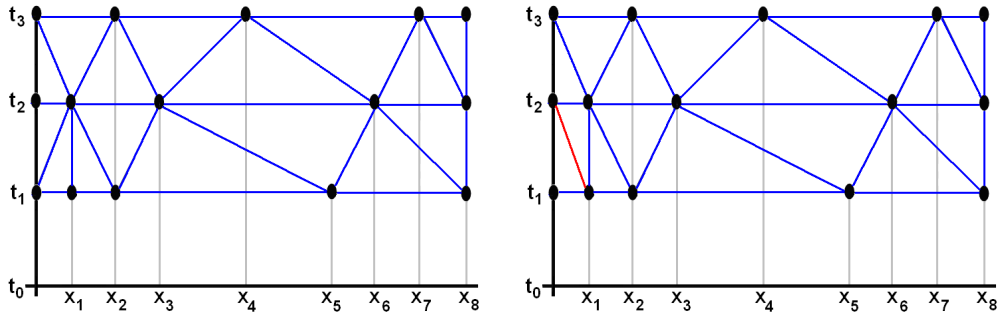


Figura 1: Dos posibles triangulaciones para el mismo conjunto de nodos.

aproximaciones al volumen real) y al final calcular una media aritmética entre todos los volúmenes obtenidos.

La función bidimensional quedaría definida en cada triángulo como aquella función lineal cuya representación gráfica es el plano que pasa por los datos de los tres vértices. Una vez conocida dicha función (para cada una de las triangulaciones, si así se ha decidido), se debería calcular el volumen bajo la función lineal bidimensional en cada uno de los triángulos, para después calcular las sumas de todos los volúmenes de los prismas truncados y, finalmente, la media de las aproximaciones al volumen (utilizando las distintas triangulaciones).

Por supuesto, el cálculo de los volúmenes puede realizarse mediante integrales dobles en regiones triangulares, pero este procedimiento es tedioso y puede provocar errores de computación (aparte de que, a pesar de que las funciones son lineales en cada triángulo, puede tener un coste elevado si el número de nodos es considerable). Por eso, se trata de evitar errores en los cálculos de los volúmenes aplicando la Proposición 3.1, que permitirá calcular el volumen de cada prisma truncado multiplicando el área del triángulo (base) por una determinada altura, también fácil de calcular.

Proposición 3.1 *En la situación anterior, consideremos el prisma triangular truncado cuya base viene determinada por los nodos $N_1(x_1, t_1)$, $N_2(x_2, t_2)$ y $N_3(x_3, t_3)$ y cuya altura en cada nodo (x_j, t_i) es el valor de $L_{t_i}(x_j)$. Entonces, se tiene que el volumen de dicho prisma se puede calcular como el producto del área de la base por la media de las tres alturas en los nodos; o, lo que es lo mismo, el producto del área de la base por la altura que tiene el prisma en el baricentro del triángulo.*

La demostración de la Proposición 3.1 puede consultarse en el Anexo.

Corolario 3.2 *Dos triangulaciones distintas (de las generadas por el algoritmo anterior) pueden producir prismas cuyas sumas de volúmenes son distintas.*

La demostración del Corolario 3.2 también puede consultarse en el Anexo.

Por otra parte, el cálculo de las áreas de los triángulos es bastante sencillo en nuestro caso, por las características de la malla en la que trabajamos. No obstante, por si en algún caso fuera conveniente el cálculo del área del triángulo de la base de cada prisma, en el Anexo se recuerda una fórmula clásica: el Lema 5.1.

Pese a lo anterior, la Proposición 3.1 proporciona una forma alternativa y directa de calcular los volúmenes de los prismas triangulares, que son fácilmente identificables y sus volúmenes fácilmente acumulables, a partir de la representación gráfica de la malla de nodos. En lo que resta, nos centraremos en las propiedades más útiles del *IDT* en cuanto a los objetivos de este artículo: su significado (en el siguiente párrafo) y su relación con la estimación de curvas de Lorenz (en el apartado 3.1).

Evidentemente, el *IDT* que se ha propuesto presentará un mayor valor, cuanto mayor sea la desigualdad en la distribución de la renta en los distintos instantes del período de tiempo observado. El valor 0 se corresponde con el caso de total equidad y el valor T (o 1, si se normaliza convenientemente el *IDT*) corresponde con el caso teórico de total desigualdad. Finalmente, conviene tener en cuenta que el *IDT* determina mejor la desigualdad cuanto mayor sea el número de nodos (datos) con los que se cuente, aunque también es relevante que los nodos estén suficientemente distribuidos en el rectángulo $[0, 1] \times [0, T]$.

3.1. Estimación de una curva de Lorenz

Otra posible aplicación de esta técnica puede ser la de estimar la curva de Lorenz en un instante de tiempo t dentro del período en estudio, aunque no se disponga de observaciones en dicho instante; para ello, una vez calculada la superficie poliédrica, se realizará el corte de dicha superficie poliédrica en el instante de tiempo t con un plano vertical de la forma $y = t$. Esta idea proporciona una forma relativamente asequible de obtener estimaciones de curvas de Lorenz sin necesidad de aplicar exigentes técnicas de interpolación funcional, lo que hace aún más interesante la utilización de la metodología empleada para el cálculo del *IDT* y, en concreto, de la superficie poliédrica subyacente. A continuación se verán las propiedades más relevantes que verifica la curva que se obtiene mediante el corte.

Consideremos la superficie poliédrica $z = L(x, y)$ obtenida por el procedimiento descrito en esta sección. Para que el corte de dicha superficie con $y = \tilde{t}$ (curva de Lorenz estimada) sea realmente una curva de Lorenz, deben verificarse varias condiciones ya comentadas:

La primera propiedad se cumple trivialmente y es que la curva de Lorenz estimada

pasa por $(0, 0)$ y por $(1, 1)$, ya que la superficie $z = L(x, y)$ contiene, por construcción al eje OY y la recta que pasa por $(1, 0, 1)$ y $(1, 1, 1)$.

La segunda cuestión que hay que comprobar es que la curva estimada es no decreciente. La verificación de este aspecto puede consultarse en el Anexo.

Finalmente, en lo que respecta a la tercera propiedad, sería muy deseable que la curva estimada fuese convexa. Sin embargo, es necesario exigir condiciones adicionales a los datos (aunque habitualmente se verifican cuando no hay grandes modificaciones al variar el instante temporal considerado) para poder garantizar algebraicamente la convexidad. A modo de boceto del razonamiento necesario para establecer las condiciones adicionales, se puede consultar la última parte del Anexo.

4. Aplicación al análisis de la desigualdad en España

A continuación se presenta, a modo de ilustración de la técnica anteriormente descrita, una valoración de la desigualdad en España en el período 2003-2009. Para ello, hemos empleado la información sobre la distribución personal de la renta que se encuentra, concretamente, en la Encuesta de Presupuestos Familiares y en la Encuesta Continua de Presupuestos Familiares (la Encuesta de Presupuestos Familiares, iniciada en enero de 2006, sustituye a la Encuesta Continua de Presupuestos Familiares base 1997, que se realizó con periodicidad trimestral desde 1997 hasta 2005), utilizando como unidad básica de análisis el hogar. Sin embargo, tienen características muy diferentes respecto al número de individuos que los forman, sus edades, el número de componentes que realizan un trabajo remunerado, los niveles educativos, etc. No obstante, sin perjuicio de las múltiples formas en que se pueden aplicar las escalas de equivalencia, por motivos didácticos y de simplicidad, en este ejemplo se utilizará la renta per cápita, al dividir la renta del hogar por el número de individuos del mismo.

Las Encuestas de Presupuestos Familiares (en adelante EPF) a lo largo del tiempo han ido sufriendo diversos cambios, tanto en lo que se refiere a su periodicidad como a cambios de tipo metodológicos, pero todas ellas informan sobre diversas características (o variables) relativas a las condiciones de vida de los hogares así como a los ingresos y gastos de ellos. Tales cambios se han realizado por recomendaciones metodológicas de distintos organismos internacionales en general y de EUROSTAT en particular.

Desde hace unos años el INE pone a disposición de los usuarios ficheros de microdatos para que se puedan analizar los datos contenidos en esta encuesta, siendo estos ficheros los que se han utilizado en este trabajo, explotando los datos que en ellos se proporcionaban.

En particular, interesa la información recogida referente a los ingresos de los hogares.

La variable ingreso del hogar se ha construido de la siguiente forma: en la encuesta se pregunta el ingreso mensual neto regular de todos los miembros del hogar y, en caso de que no sepan dar el valor puntual, el intervalo al cual pertenece; así, en algunos casos se dispone del punto y en otros del intervalo. Por ello, se ha imputado a todos los hogares de los cuales se dispone de intervalo pero no de un valor puntual, unos ingresos netos regulares, condicionado a que dicho valor esté situado dentro del intervalo que se ha dado. Esto se ha llevado a cabo según el programa de imputación IVE (Imputation and Variance Estimator), del Institute of Social Research de la Universidad de Michigan, corregidos después de falta de respuesta e inflación.

Una vez obtenidos tales datos se ha tenido en cuenta el tamaño del hogar y se ha calculado el ingreso mensual neto regular per cápita del hogar, que es la variable que se utiliza en el presente estudio, a modo de ejemplo. De forma análoga se podría replicar el caso usando otras escalas de equivalencia si el investigador lo creyera más conveniente para su propósito de análisis.

En los años estudiados (2003 y de 2006 a 2009) la encuesta cambia sustancialmente en el año 2006. Las principales diferencias entre la encuesta del año 2003 [7] y las de los años de 2006 a 2009 [8, 9, 10, 11] son las siguientes:

- Cambia la periodicidad, que pasa de ser trimestral a anual.
- Aumenta el tamaño muestral; en la encuesta base 1997 (la utilizada en el año 2003) se entrevistaban unos 8 000 hogares cada trimestre, lo que suponía (al tener en cuenta el diseño de panel rotante de la encuesta) una muestra anual de aproximadamente 11 000 hogares. En la EPF base 2006 (la utilizada a partir del año 2006) la muestra de hogares está en torno a los 24 000.
- El tiempo de colaboración de los hogares seleccionados para formar parte de la muestra cambia sutilmente; pasan de colaborar en esta durante ocho trimestres a colaborar durante dos años consecutivos.

En las encuestas utilizadas, la representatividad de la muestra es buena tanto a nivel nacional como de comunidades autónomas, por lo que, aunque en esta ilustración se utilizan datos a nivel nacional, análogamente se podría replicar para cualquiera de las comunidades autónomas del país.

Sin embargo, los datos también presentan irregularidades. La primera que salta a la vista es la temporal, puesto que la serie comienza en el año 2003 y después faltan dos años. Aunque en 2004 sí se realizó la encuesta, no incluía la variable ingresos, mientras que en 2005 no se realizó dicha encuesta. Además, los datos son irregulares en cuanto al número,

Año	Hogares	Índice de Gini
2003	9 132	0,2998
2006	19 391	0,3177
2007	21 514	0,3149
2008	22 027	0,3137
2009	22 272	0,3211

Tabla 1: Número de hogares de la muestra e índices de Gini calculados para España en el período 2003-2009.

ya que cada año se realizó la encuesta a un número distinto de individuos (los tamaños muestrales con los que se trabaja son exactamente los que se recogen en la Tabla 1).

Antes de calcular el *IDT* de España para el período de análisis, decidimos realizar el ejercicio de determinar las curvas de Lorenz y los índices de Gini anuales. El procedimiento seguido para determinar las curvas de Lorenz para cada año, así como su correspondiente índice de Gini, fue el siguiente:

1. Se calculan los ingresos regulares netos mensuales per cápita para cada hogar, en cada año de la encuesta.
2. Se ordenan, para cada año, los hogares según la variable anterior y se le asigna un “orden” (entre 1 y el tamaño muestral).
3. Se añade un hogar con orden 0 e ingreso 0.
4. Se calcula una columna de ingresos acumulados.
5. En el eje X se colocan los valores orden/tamaño muestral; la altura viene dada por el ingreso acumulado hasta el hogar/suma total del ingreso acumulado. Con estos datos, se dibuja la curva (poligonal) de Lorenz.
6. También se puede aproximar el índice de Gini: primero se calcula la distancia entre dos puntos consecutivos del eje X (por construcción, son equiespaciados para cada año) y se multiplica por la diferencia entre el valor de X y el de la altura; después se suman todos esos productos y se multiplica por dos.

En la Figura 2 se muestran las curvas de Lorenz obtenidas para cada período (nótese que son muy similares al nivel de resolución de la representación, pero los índices de Gini demuestran que son distintas). El valor correspondiente de los índices de Gini calculados se puede consultar en la Tabla 1.

Después de este ejercicio, calculamos el *IDT* para el período. Puesto que las operaciones eran relativamente sencillas, decidimos realizarlas con la hoja de cálculo Excel 2003,

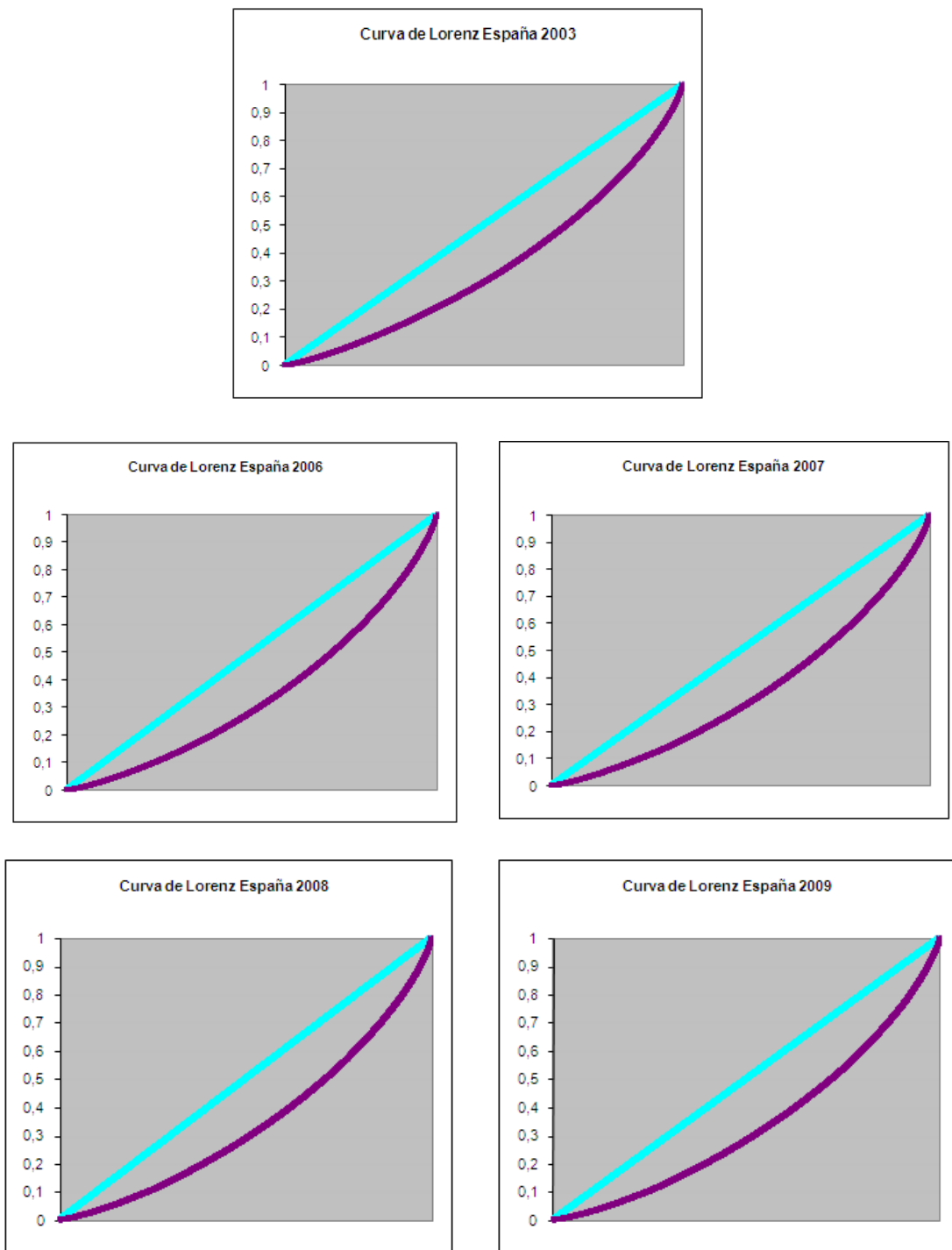


Figura 2: Curvas de Lorenz calculadas para España durante los años 2003, 2006, 2007, 2008 y 2009.

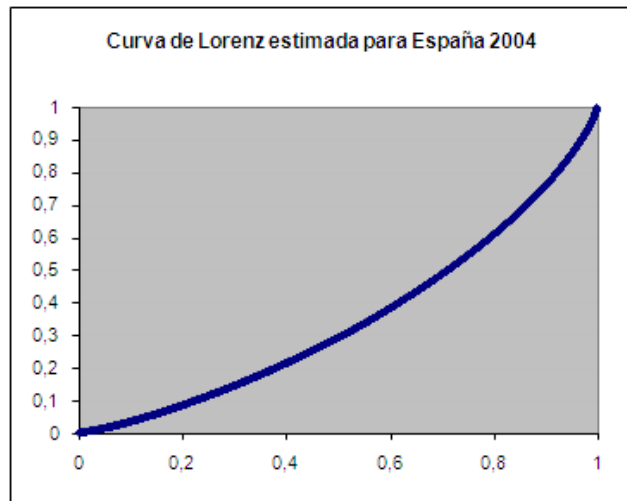


Figura 3: Estimación de la curva de Lorenz para España en el año 2004.

calculando el volumen para cada par de instantes temporales consecutivos. El resultado final fue de 1,0368 para el período 2003-2006, de 0,3418 para 2006-2007, de 0,3428 para 2007-2008 y de 0,3413 para 2008-2009, con lo que el *IDT* global resultó 2,0628. Este resultado se mueve en el rango de $[0, 6]$, siendo 0 el valor de máxima igualdad y 6 el de mayor desigualdad.

A continuación, utilizamos la misma hoja de cálculo para estimar la curva de Lorenz en un año del que no se tenían datos. En concreto, se eligió 2004. Para realizar dicha estimación, se calculó el corte del plano vertical que representaba a 2004 (a un tercio de distancia de 2003 respecto de la distancia total entre 2003 y 2006, instantes de los que sí disponíamos de datos) con la “superficie de Lorenz” calculada anteriormente y eso produjo la curva de la Figura 3. A partir de la curva de Lorenz estimada, repitiendo el ejercicio anteriormente realizado para los años de los que se tenían datos, se calculó el índice de Gini para 2004, que resultó ser de 0,3058, muy similar a los estimados oficialmente para dicha fecha.

5. Conclusiones

En lo que respecta a la cuestión metodológica, en este artículo se ha presentado un conjunto de conceptos y técnicas que consideramos que pueden ser útiles para las futuras investigaciones sobre desigualdad, en muy diversos ámbitos y zonas geográficas. Como se ha visto, además de proporcionar una forma de medir las desigualdades en un período de tiempo, se posibilita una estimación de la curva de Lorenz en un instante del que no se tengan datos suficientes. A pesar de que el comportamiento del indicador y del método de estimación parecen adecuados, hay que advertir de que su funcionamiento correcto depende de la coherencia de los datos a los que se aplica, pues en ocasiones el *IDT* puede compensar o enmascarar situaciones anómalas puntuales y, por otro lado, la estimación de una curva de Lorenz, aunque aproxime a la función efectiva, puede no ser totalmente convexa en casos extremos. Queda para el futuro la determinación explícita de las condiciones que deben verificarse para garantizar el funcionamiento correcto de los procedimientos presentados.

Asimismo, queda pendiente la utilización de las ideas anteriores para generalizar (en cuanto a su aspecto temporal) otros índices de desigualdad distintos del de Gini.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el proyecto 2012DEC002 de AACID.

Referencias

- [1] Atkinson, A.B. “On the Measurement of Inequality”, *Journal of Economic Theory*, **2** (1970), 244–263.
- [2] Cowell, F.A. “Measuring inequality”, *LSE Handbooks in Economics. Prentice Hall/Harvester Wheatsheaf*, 2^a ed., Baltimore, 1995.
- [3] Cowell, F.A. y Victoria-Feser, M.P. “Modeling Lorenz Curves: Robust and Semi-Parametric Issues”, *Distributional Analysis Research Programme*, Londres, 2007.
- [4] Domínguez, J., Fedriani, E.M., Martín, A.M. y Núñez, J.J. “Una herramienta para el estudio de la evolución temporal de la desigualdad con datos incompletos. Una

- aplicación a los países del sur de Europa”, *Anales de Economía Aplicada*, **XXV** (2011), 737–749.
- [5] Dorfman, R. “A Formula for the Gini Coefficient”, *Review of Economics and Statistics*, **61** (1979), 146–149.
- [6] Duclos, J.Y. y Araar, A. “Poverty and Equity: Measurement, Policy and Estimation with DAD”, *Springer*, Nueva York, 2006.
- [7] INE. “Encuesta de Presupuestos Familiares. Base 1997”, *Ficheros longitudinales del usuario*, 2003.
- [8] INE. “Encuesta de Presupuestos Familiares. Base 2006”, *Ficheros del usuario*, 2006.
- [9] INE. “Encuesta de Presupuestos Familiares. Base 2006”, *Ficheros del usuario*, 2007.
- [10] INE. “Encuesta de Presupuestos Familiares. Base 2006”, *Ficheros del usuario*, 2008.
- [11] INE. “Encuesta de Presupuestos Familiares. Base 2006”, *Ficheros del usuario*, 2009.
- [12] Liao, T.F. “Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-Based Clustering”, *Sociological Methodology*, **36** (2006), 201–224.
- [13] Lorenz, N.C. “Methods of measuring the concentration of wealth”, *Publications of the American Statistical Association*, **9** (1905), 209–219.
- [14] Martín, A. “Valoración de la pobreza mediante técnicas de agregación de datos de diferente naturaleza”, *Tesis doctoral*, Universidad Pablo de Olavide, Sevilla, 2005.
- [15] Núñez, J.J. “La desigualdad económica medida a través de las curvas de Lorenz”, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, **2** (2006), 67–108.
- [16] Rawls, J. “Justicia como equidad”, *Tecnos*, Madrid, 1986; traducido de Rawls, J. “Justice and Fairness”, *Philosophical Review*, Vol. **LXVII** (1958).
- [17] Schluter, C. y Trede, M. “Statistical Inference for Inequality and Poverty Measurement with Dependent Data”, *International Review of Economics*, **43** (2002), 493–508.
- [18] Schluter, C. y Trede, M. “Tails of Lorenz Curves”, *Journal of Econometrics*, **109** (2002), 151–166.
- [19] Sen, A. “On Economic Inequality”, *Clarendon Press, Paperbacks*, Oxford, 1973.
- [20] Theil, H. “Economics and Information Theory”, *North-Holland*, Ámsterdam, 1967.

Anexo: algunos resultados geométricos y demostraciones

Demostración de la Proposición 3.1. Sin pérdida de generalidad, y a fin de simplificar la demostración analítica, podemos descomponer el prisma triangular truncado en dos volúmenes: un prisma triangular regular que se obtiene desde su base al truncarlo regularmente a la altura del menor de sus aristas verticales (que suponemos la del nodo N_1) y la pirámide irregular que quedaría en el extremo desgajado al truncar, que en general puede tener base trapezoidal. Al realizar esta operación pueden presentarse dos casos, A y B, como ilustra la Figura 4.

En primer lugar calcularemos el volumen de la pirámide trapezoidal, para lo que, sin pérdida de generalidad, podemos suponer que el vértice sobre el nodo N_1 está situado en el origen del sistema de referencia, por lo que sus coordenadas serían $(0, 0, 0)$. Análogamente, aplicando un movimiento conveniente, podemos suponer que las coordenadas de los otros dos vértices situados sobre los nodos N_2 y N_3 son (e, c, a) y $(0, c + d, b)$, respectivamente.

En cualquiera de los dos casos, A y B, la base trapezoidal de la pirámide irregular desgajada tendría por área $\frac{a+b}{2} \cdot \sqrt{e^2 + d^2}$, mientras que su altura h se podría obtener aplicando dos veces el Teorema de Pitágoras (en el caso A aparece “+ x ” en algunas de las expresiones siguientes, mientras que en el caso B aparece “- x ”; por lo que, para presentar los dos casos a la vez, escribimos el símbolo “ \pm ” delante de la x):

$$\begin{cases} h^2 + x^2 = c^2 + e^2 \\ h^2 + (\pm x + \sqrt{e^2 + d^2})^2 = (c + d)^2 \end{cases} \quad \begin{cases} h^2 + x^2 = c^2 + e^2 \\ h^2 + (\pm x + \sqrt{e^2 + d^2})^2 = c^2 + d^2 + 2cd \end{cases}$$

Luego $\pm 2x\sqrt{e^2 + d^2} + e^2 = 2cd - e^2$ y, por tanto, $x = \left| \frac{cd - e^2}{\sqrt{e^2 + d^2}} \right|$, lo que nos lleva a que $h^2 = c^2 + e^2 - \left(\frac{cd - e^2}{\sqrt{e^2 + d^2}} \right)^2 = \frac{e^2(c^2 + d^2 + 2cd)}{e^2 + d^2} = \frac{e^2(c + d)^2}{e^2 + d^2}$.

Con esto, la altura h de la pirámide trapezoidal resulta ser $h = \frac{e(c+d)}{\sqrt{e^2 + d^2}}$. De ahí que el volumen de la pirámide se puede calcular, multiplicando el área de la base por la altura y dividiendo por tres, como sigue:

$$V = \frac{\frac{a+b}{2} \sqrt{e^2 + d^2} \cdot \frac{e(c+d)}{\sqrt{e^2 + d^2}}}{3} = \frac{e(a+b)(c+d)}{6}.$$

O, lo que es lo mismo:

$$V = \frac{e(c+d)}{2} \cdot \left(\frac{a+b}{3} \right).$$

Para finalizar la demostración, el volumen total del prisma truncado es la suma del volumen de la pirámide anterior más el volumen de un prisma (no truncado) de base

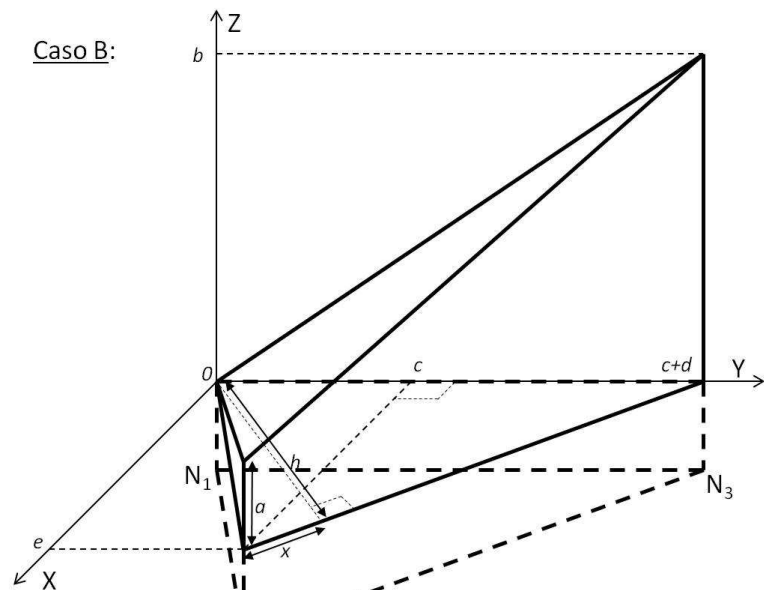
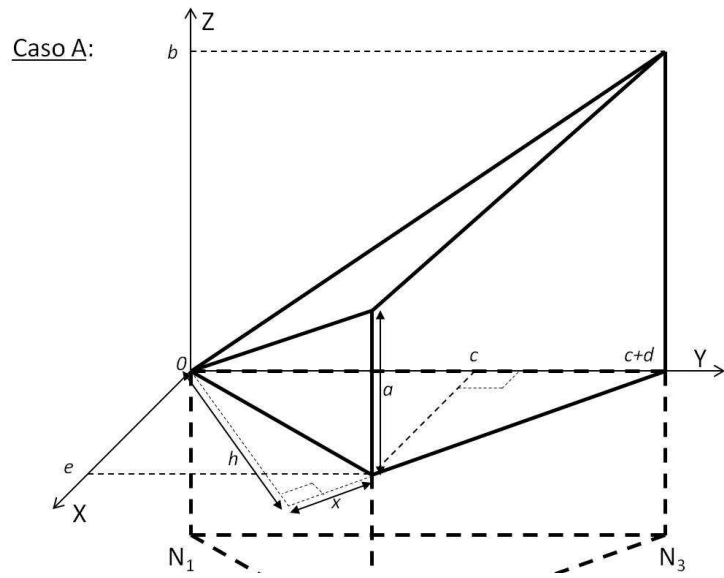


Figura 4: Dos casos posibles de pirámide irregular en la Proposición 3.1.

triangular y altura la de la menor de las aristas sobre los tres nodos:

$$\begin{aligned}
\text{Vol. prisma truncado total} &= \frac{e(c+d)}{2} \cdot \left(\frac{a+b}{3}\right) + \text{Vol. prisma no truncado} \\
&= \frac{(x_2 - x_1) \cdot (t_3 - t_1)}{2} \cdot \frac{(L_{t_2}(x_2) - L_{t_1}(x_1)) + (L_{t_3}(x_3) - L_{t_1}(x_1))}{3} \\
&\quad + \frac{(x_2 - x_1) \cdot (t_3 - t_1)}{2} \cdot L_{t_1}(x_1) \\
&= \frac{(x_2 - x_1) \cdot (t_3 - t_1)}{2} \cdot \frac{L_{t_1}(x_1) + L_{t_2}(x_2) + L_{t_3}(x_3)}{3}. \quad \square
\end{aligned}$$

La demostración anterior se podría haber afrontado desde un punto de vista geométrico (y no tan analítico), con solo comprobar que el prisma triangular tiene el mismo volumen que otro de igual base y que tenga por altura un determinado “punto intermedio” del triángulo inclinado de la base superior. Pero dicho punto ha de dejar el mismo volumen por encima y por debajo de él, lo que solo puede conseguirse en la intersección de las tres medianas del triángulo inclinado, es decir, en su baricentro, cuya proyección en el plano horizontal es también el baricentro del triángulo de la base. Para ver esto, conviene recordar que las medianas unen cada vértice con el punto medio del lado opuesto, por lo que cada mediana divide al triángulo en dos triángulos de igual área.

Demostración del Corolario 3.2. Basta con darse cuenta de que en los cálculos de los volúmenes interviene el valor en un nodo concreto tantas veces como el número de triángulos a los que pertenece dicho nodo y ese número de triángulos no se mantiene invariante en las diferentes triangulaciones (el área de los triángulos sí puede ser la misma, pero la altura variará en general). □

Lema 5.1 *Considérese el triángulo cuyos vértices (A , B y C) tienen por coordenadas respectivas (x_A, y_A) , (x_B, y_B) y (x_C, y_C) . Entonces, el área de dicho triángulo viene dada por:*

$$\begin{aligned}
\frac{1}{2} \cdot \left| \text{Det} \begin{pmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{pmatrix} \right| &= \frac{1}{2} |x_A y_B + x_B y_C + x_C y_A - x_C y_B - x_B y_A - x_A y_C| \\
&= \frac{1}{2} |(x_A - x_C) \cdot (y_B - y_A) - (x_A - x_B) \cdot (y_C - y_A)|,
\end{aligned}$$

siendo Det el determinante e indicando el valor absoluto por líneas verticales. □

Comprobación de que la curva estimada en el apartado 3.1 es no decreciente

Para confirmar esta cuestión, conviene comentar algunos aspectos geométricos e introducir una notación sencilla y coherente con lo anterior. Como antes, supondremos los valores x_j (en los que se tienen datos para algún instante) ordenados de menor a mayor. Llamemos t_i y t_{i+1} a dos instantes en los que se tienen datos o las correspondientes curvas de Lorenz; esto es, en dichos instantes se conocen los valores $L_t(x)$ o, lo que es lo mismo, se conocen las coordenadas $(x_j, t_i, L_{t_i}(x_j))$. Supongamos que se desea calcular las coordenadas de un punto de corte de la superficie poliédrica con el plano $y = \tilde{t}$, siendo $\tilde{t} \in (t_i, t_{i+1})$. En esta situación, consideremos tres nodos conectados por una triangulación (por lo que dos de ellos corresponden a los dos instantes consecutivos). Sin pérdida de generalidad, se puede suponer que dos de dichos nodos son de la forma (x_j, t_i) , (x_{j+1}, t_{i+1}) (el que $x_j = x_{j+1}$ no es más que el caso límite y tampoco presenta mayor dificultad que el analizado) y el tercer nodo, para comprender la situación más fácilmente, podría ser (x_{j+2}, t_i) o (x_{j+2}, t_{i+1}) ; en el primer caso, el triángulo lo formarían los nodos (x_j, t_i) , (x_{j+1}, t_{i+1}) y (x_{j+2}, t_i) , mientras que en el segundo caso los tres vértices serían (x_j, t_i) , (x_{j+1}, t_{i+1}) y (x_{j+2}, t_{i+1}) (como antes, hay otros casos posibles, pero su comportamiento es similar). En ambas situaciones, el que las curvas de Lorenz sean no decrecientes en t_i y en t_{i+1} nos asegura que $L_{t_i}(x_j) \leq L_{t_i}(x_{j+2})$ y $L_{t_{i+1}}(x_{j+1}) \leq L_{t_{i+1}}(x_{j+2})$, respectivamente. Por consiguiente, la altura $L_{\tilde{t}}(x)$ vendría dada por la pendiente de un plano inclinado (el que pasa por $(x_j, t_i, L_{t_i}(x_j))$, $(x_{j+1}, t_{i+1}, L_{t_{i+1}}(x_{j+1}))$ y $(x_{j+2}, t_i, L_{t_i}(x_{j+2}))$ o, en el segundo caso, por $(x_j, t_i, L_{t_i}(x_j))$, $(x_{j+1}, t_{i+1}, L_{t_{i+1}}(x_{j+1}))$ y $(x_{j+2}, t_{i+1}, L_{t_{i+1}}(x_{j+2}))$) en el que $L_{t_i}(x_j) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L_{t_{i+1}}(x_{j+1}) - L_{t_i}(x_j)) \leq L_{t_i}(x_{j+2}) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L_{t_{i+1}}(x_{j+1}) - L_{t_i}(x_{j+2}))$ (equivalentemente, en el segundo caso sería $L_{t_i}(x_j) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L_{t_{i+1}}(x_{j+1}) - L_{t_i}(x_j)) \leq L_{t_i}(x_j) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L_{t_{i+1}}(x_{j+2}) - L_{t_i}(x_j))$) y, por tanto, la curva estimada siempre será no decreciente.

Condiciones para que la curva estimada en el apartado 3.1 sea convexa

Para comprobar que las condiciones que hay que exigir a los datos son naturales, comentaremos lo que sigue. Utilizando la notación anterior, podemos suponer, sin pérdida de generalidad (renumerando los índices y obviando casos similares), que todas las configuraciones posibles de nodos se pueden encuadrar en una de las dos representadas en la Figura 5: (x_1, t_i) , (x_2, t_{i+1}) , (x_3, t_i) y (x_4, t_{i+1}) (en la representación de la izquierda) o, en el segundo caso (representado a la derecha), (x_1, t_i) , (x_2, t_{i+1}) , (x_3, t_i) y (x_4, t_i) . En ambos casos, debido a la linealidad de la superficie poliédrica, probar la convexidad de la curva

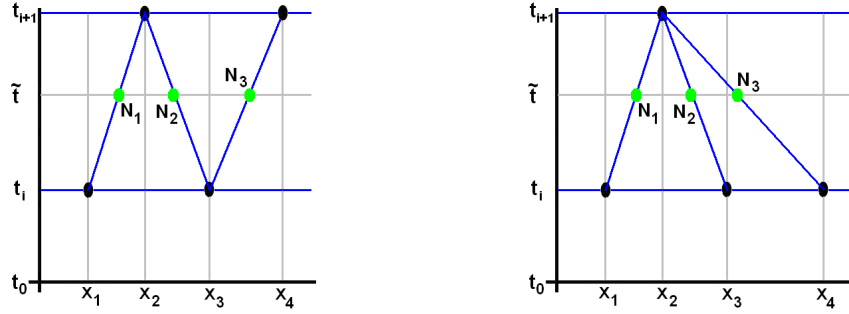


Figura 5: Dos triangulaciones de interés para comprobar la convexidad de las curvas de Lorenz estimadas.

de Lorenz estimada es equivalente a demostrar que se verifica lo siguiente:

$$\frac{L(N_2) - L(N_1)}{x(N_2) - x(N_1)} \leq \frac{L(N_3) - L(N_2)}{x(N_3) - x(N_2)},$$

donde, para $i = 1, 2, 3, 4$, $x(N_i)$ denota la primera coordenada en el nodo N_i , mientras que $L(N_i)$ es la altura propuesta para dicho nodo. En particular, $x(N_i)$ se puede calcular fácilmente a partir de la Figura 5; además, $L(N_1) = L(x_1, t_i) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L(x_2, t_{i+1}) - L(x_1, t_i))$ y $L(N_2) = L(x_3, t_i) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L(x_2, t_{i+1}) - L(x_3, t_i))$; en cuanto a $L(N_3)$, en el primer caso es $L(N_3) = L(x_3, t_i) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L(x_4, t_{i+1}) - L(x_3, t_i))$, mientras que en el segundo vale $L(N_3) = L(x_4, t_i) + \frac{\tilde{t}-t_i}{t_{i+1}-t_i} \cdot (L(x_2, t_{i+1}) - L(x_4, t_i))$. En la configuración de la izquierda de la Figura 5 sí es posible concluir la convexidad de la superficie poliédrica; sin embargo, dicha desigualdad no tiene que verificarse siempre en la configuración de la derecha, aunque es una condición muy razonable si las curvas de Lorenz tienen un comportamiento similar en i e $i + 1$.