

ESTUDIOS SOBRE LA ECONOMÍA ESPAÑOLA - 2014/12

**Renta personal de los municipios españoles y su distribución:
Metodología de estimación a partir de microdatos tributarios**

Miriam Hortas Rico

Universidad Complutense de Madrid y GEN

Jorge Onrubia Fernández

Universidad Complutense de Madrid, FEDEA y GEN

fedea

Renta personal de los municipios españoles y su distribución:

Metodología de estimación a partir de microdatos tributarios

Miriam Hortas Rico ¹

Jorge Onrubia Fernández ²

Noviembre, 2014

Abstract

En este documento se presenta la base de datos *Renta personal de los municipios españoles y su distribución*, así como la metodología seguida para su elaboración. Esta base de datos proporciona una estadística anual para los municipios españoles de más de 5.000 habitantes pertenecientes a las Comunidades y Ciudades Autónomas de Régimen Fiscal Común. Para llevar a cabo las estimaciones tanto de la renta personal como de su distribución se utilizan las Muestras Anuales de Declarantes de IRPF difundidas por el Instituto de Estudios Fiscales (IEF) y la Agencia Estatal de Administración Tributaria (AEAT), así como las estadísticas agregadas de IRPF por municipios de la AEAT. La representatividad a nivel local de los microdatos tributarios es obtenida mediante un proceso computacional de corrección de los factores de elevación poblacional.

Palabras clave: renta personal local, distribución, *reweighting*, desigualdad, rentas altas

Códigos JEL: C42, C61, D31, D63, O15

¹Universidad Complutense de Madrid y GEN. e-mail: miriamhortas@ccee.ucm.es

²Universidad Complutense de Madrid, FEDEA y GEN. e-mail: jorge.onrubia@ccee.ucm.es

Agradecimientos

Con la habitual cláusula de exoneración de responsabilidad, agradecemos la ayuda prestada en el desarrollo del trabajo por Daniele Pacifico (Directorate for Employment, Labour and Social Affairs, OECD, y Center for the Analysis of Public Policies, University of Modena and Reggio Emilia, Modena). Asimismo queremos agradecer los útiles comentarios y sugerencias realizados por Ángel de la Fuente. Jorge Onrubia agradece el soporte financiero del Ministerio de Economía y Competitividad, a través del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011, VI Plan Nacional, Proyecto ECO2012-37572.

RESUMEN (NON TECHNICAL SUMMARY)

¿Qué municipios poseen una renta personal per cápita más alta? ¿Son los municipios más ricos los que acumulan una mayor proporción de residentes con altos ingresos? ¿Y qué diferencias existen respecto de los municipios que presentan una menor renta personal? ¿Cuáles son las causas de estas posibles diferencias? ¿Qué municipios presentan una distribución de la renta personal más desigual? ¿Existe un patrón distributivo espacial? ¿Cuál es el impacto de los ganadores de ingresos más altos en el crecimiento económico y la desigualdad local? ¿Cómo pueden diseñarse políticas redistributivas, tanto por la vía del gasto público como de los impuestos, para reducir la desigualdad de la renta local?

La información sobre el nivel de renta personal en los municipios y su distribución es esencial para responder a todas estas preguntas. De hecho, los datos sobre los ingresos obtenidos anualmente por los residentes en una localidad constituyen un indicador económico muy importante, ampliamente utilizado en los estudios relacionados con la economía urbana, el federalismo fiscal, la vivienda o el análisis territorial de los diferentes mercados de bienes y servicios. Además, los aspectos de la desigualdad de la renta y la pobreza a nivel local son temas que reciben en la actualidad una atención creciente de los investigadores y analistas económicos y sociales.

Sin embargo, a pesar de esta relevancia, los datos sobre los ingresos personales a nivel local siguen siendo una asignatura pendiente dentro de las estadísticas oficiales de muchos países desarrollados, incluida España. La explicación de esta carencia radica, fundamentalmente, en la complejidad del diseño de encuestas estadísticamente representativas en este ámbito territorial reducido y su elevado coste de elaboración, dado el gran número de entrevistas a realizar en cada municipio. No obstante, unos pocos países, como Estados Unidos, Reino Unido o Australia, ofrecen regularmente bases de datos con información sobre la renta personal local, ya sea individual o de los hogares, y su distribución. Estas bases de datos son elaboradas con una amplia gama de técnicas estadísticas desarrolladas en las últimas dos décadas con el objetivo de proporcionar estimaciones fiables de los ingresos locales y su distribución en áreas reducidas.

La disponibilidad de muestras de microdatos procedentes de las declaraciones de IRPF proporciona una atractiva alternativa para poder afrontar este reto, al ofrecer información bastante completa sobre los ingresos personales y con unos costes de elaboración muy reducidos. Este tipo de muestras anuales, como sucede en España con las facilitadas por el Instituto de Estudios Fiscales y la Agencia Estatal de Administración Tributaria, suelen incluir un número enorme de observaciones –entre 1

y 2 millones, de un total poblacional de entre 18 y 20 millones de declaraciones–, lo que asegura disponer de microdatos para la práctica totalidad de municipios a partir de un umbral de población. Puesto que estas bases de microdatos se obtienen con una representatividad territorial de ámbito provincial, resulta necesario aplicar una metodología de tratamiento estadístico que permita llevar a cabo estimaciones fiables de la renta personal en áreas geográficas menores, como son los municipios. En el caso de la base de datos que se presenta, esta representatividad a nivel local es obtenida mediante un proceso computacional de corrección de los factores de elevación poblacional (*reweighting*), cuya metodología consiste en una adaptación de las propuestas de Deville y Särndal (1992), Creedy (2003) y Creedy y Tuckwell (2004).

La base de datos *Renta personal de los municipios españoles y su distribución*, que a partir de ahora difundirá fedea, proporciona, en formato abierto, una estadística para los municipios españoles de más de 5.000 habitantes pertenecientes a las Comunidades y Ciudades Autónomas de Régimen Fiscal Común. Inicialmente, por razones metodológicas, esta base de datos arranca con la información correspondiente al año 2007 –1.109 municipios–, con el propósito de completar, en un plazo de tiempo razonablemente corto, los años comprendidos entre 2002 y 2009, para los cuales se dispone actualmente de la información estadística necesaria. Una vez culminada esta primera etapa, la base de datos irá incorporando los sucesivos años, a medida que la información primaria se difunda.

En cada año, la estadística incluye información para cada municipio de la renta personal per capita y por declarante de IRPF, medidas de desigualdad (índices de Gini y de Atkinson), así como diversos indicadores de concentración de la renta entre los declarantes con mayores ingresos (1%, 0,5% y 0,1% más rico). Además, se ofrece para cada localidad la estructura de su distribución de renta personal, mediante el cálculo de quintiles.

1. Introducción

La información sobre el nivel de renta personal existente en los municipios y su distribución resulta esencial para muchos campos de la investigación y el análisis económicos. De hecho, los datos sobre los ingresos obtenidos anualmente por los residentes en una localidad constituyen un indicador económico ampliamente utilizado en los estudios relacionados con la economía urbana, el federalismo fiscal, la vivienda o el análisis territorial de los diferentes mercados de bienes y servicios. Además, la desigualdad de la renta y la pobreza a nivel local son temas que reciben en la actualidad una atención creciente de los investigadores y analistas económicos y sociales.

A pesar de esta relevancia, en muy pocos países desarrollados las estadísticas oficiales incluyen información sobre los ingresos personales a nivel local y, mucho menos sobre su distribución. Detrás de esta carencia se encuentran, fundamentalmente, la complejidad de diseñar encuestas representativas en este ámbito territorial reducido y, sobre todo, su elevado coste de elaboración, dado el gran número de entrevistas que es necesario realizar en cada municipio. No obstante, unos pocos países, como Estados Unidos, Reino Unido o Australia, ofrecen regularmente bases de datos con información sobre la renta personal local, ya sea individual o de los hogares y su distribución. Estas bases de datos son elaboradas aplicando una variedad de técnicas estadísticas desarrolladas en las últimas dos décadas con el objetivo de proporcionar estimaciones fiables por localidades de los ingresos personales y su distribución en áreas reducidas¹. La mayoría suelen combinar información procedente de muestras de microdatos (generalmente encuestas, aunque también se utilizan registros administrativos) y datos poblacionales agregados correspondientes al ámbito municipal.

La disponibilidad de amplias muestras de declarantes del impuesto sobre la renta personal (IRPF) ofrece una posibilidad atractiva para obtener estimaciones de las distribuciones de ingresos personales por municipios, con un coste de elaboración muy reducido. Aunque estas muestras suelen ser representativas solamente a nivel regional o, como mucho, provincial, su utilización en el ámbito municipal es factible. Para ello, resulta necesario aplicar una metodología de tratamiento estadístico que permita llevar a cabo estimaciones fiables de la renta personal en áreas geográficas menores, como son los municipios.

En este documento se presenta la base de datos *Renta personal de los municipios españoles y su distribución*, que difundirá fedea, así como la metodología seguida para su elaboración. Esta base de datos proporciona una estadística anual para los municipios

¹ Haslett *et al.* (2010) distinguen tres métodos estadísticos principales: estimación en áreas pequeñas, técnicas de imputación y modelos de microsimulación espacial.

españoles de más de 5.000 habitantes pertenecientes a las Comunidades y Ciudades Autónomas de Régimen Fiscal Común. Para llevar a cabo las estimaciones tanto de la renta personal como de su distribución se utilizan las Muestras Anuales de Declarantes de IRPF difundidas por el Instituto de Estudios Fiscales (IEF) y la Agencia Estatal de Administración Tributaria (AEAT), así como las estadísticas agregadas de IRPF por municipios elaboradas por la AEAT. La representatividad a nivel local de los microdatos tributarios es obtenida mediante un proceso computacional de corrección de los factores de elevación poblacional (*reweighting*), basado en una adaptación de las propuestas de Deville y Särndal (1992), Creedy (2003) y Creedy y Tuckwell (2004).

El presente documento de trabajo se organiza de la siguiente manera. Tras esta introducción, en la sección dos se expone el problema de la medición de la renta local y de su desigualdad, con una revisión de las principales experiencias comparadas. En la sección tres se muestra la alternativa basada en el uso de microdatos tributarios. La cuarta sección se encarga de presentar la metodología seguida en la elaboración de la base de datos de los municipios españoles, así como la validación de los resultados de su aplicación. En la quinta sección se recoge una descripción breve de las fuentes estadísticas primarias que se emplean en las estimaciones realizadas. Por último, en la sexta sección presentamos la estructura de la nueva base de datos y el contenido de las variables incluidas.

2. El problema de la medición de la renta local y su desigualdad

La mayoría de los países de la OCDE no publican estadísticas oficiales de ámbito municipal sobre el nivel de la renta per cápita de los individuos o de los hogares residentes en las localidades ni sobre su distribución. Esta ausencia de información estadística supone una importante limitación para el análisis económico, en la medida que el nivel de renta personal de los municipios y su grado de desigualdad son variables con una relevancia muy importante en los trabajos de economía aplicada, lo mismo que en los estudios de mercado. No obstante, existen algunas excepciones, todas ellas situadas en el ámbito anglosajón, como es el caso de Estados Unidos, Reino Unido o Australia.

El caso de Estados Unidos es, posiblemente, el más destacable, tanto por la calidad de la información estadística como por su larga tradición. El *U.S. Bureau of Economic Analysis* (BEA) ofrece datos sobre la renta personal para 366 áreas metropolitanas (*Metropolitan Areas*) y sus correspondientes 3.113 condados (*Counties*) para un periodo comprendido entre 1969 y 2011². En este caso, la renta personal de una

² Véase: <http://www.bea.gov/regional/index.htm>

determinada entidad local es definida como la suma de rentas de cualquier fuente (salariales, del capital inmobiliario, dividendos, intereses y otras cantidades procedentes de transferencias corrientes recibidas) recibidas por todas las personas físicas residentes en la misma. Se trata, por tanto, de una medida en dólares corrientes de la renta personal antes de la aplicación del impuesto sobre la renta personal y de otros impuestos de carácter personal. La procedencia fundamental de la información estadística es de naturaleza fiscal, en la medida que esta procede de los datos de los contribuyentes registrados en el *Internal Revenue Service* (IRS).

En el Reino Unido, la *Office for National Statistics* (ONS) publica anualmente estimaciones de la Renta Familiar Bruta Disponible (RFBD) para 139 áreas locales definidas como NUTS3, las cuales incluyen tanto condados metropolitanos como no metropolitanos. Actualmente la serie disponible abarca el periodo 1997-2011. En este caso, la mayoría de los datos e indicadores utilizados para determinar la cuantía de la RFBD para estas entidades locales proviene de una amplia variedad de fuentes, de estudios y de bases de datos administrativas. La definición de la renta utilizada es de carácter macroeconómico, coincidiendo con el concepto de RFBD de la Contabilidad Nacional, a partir del cual se procede a su territorialización a nivel local. Esta definición se identifica con la cantidad de dinero que todos los individuos, en el sector de los hogares, tienen disponible para gastar o ahorrar después de los impuestos corrientes sobre la renta y la riqueza, las cotizaciones sociales pagadas y los beneficios sociales obtenidos.

Asimismo, la ONS ofrece también información de la renta personal media semanal bruta y neta, antes y después de considerar los gastos en vivienda, para los hogares pertenecientes a las denominadas *Middle Layer Super Output Areas* (MSOAs). Estas áreas locales, con población entre 5.000 y 15.000 habitantes, se definen espacialmente, a efectos estadísticos, a partir de las divisiones electorales para Inglaterra (6781 MSOAs) y Gales (413 MSOAs). La fuente de información procede de los microdatos de la *Family Resources Survey*, encuesta anual sobre las condiciones de vida y los ingresos de los hogares británicos, elaborada por el Departamento de Trabajo y Pensiones del Gobierno Británico. En este caso, la renta del hogar es definida como los ingresos percibidos por sueldos y salarios, beneficios generados por el trabajo por cuenta propia, pensiones y otras fuentes de renta. Los importes medios ofrecidos son obtenidos mediante un procedimiento estadístico de *Small Area Income Estimates*. La periodicidad de las estimaciones es trienal. La serie arranca con la estimación para 2001/02, siendo los últimos datos disponibles los correspondientes a la estimación para 2007/08, si bien está prevista la difusión de nuevos resultados para 2011/12 en el otoño

de 2014. La información está disponible en la web con acceso directo a través de Google Maps³.

En Australia, desde 2005, el *Australian Bureau of Statistics* proporciona estimaciones de la renta personal, con desagregación por fuentes, para cada estado y territorio, así como para las áreas de gobierno locales, de acuerdo con diferentes niveles establecidos en la *Australian Standard Geographical Classification*. La información estadística es obtenida mediante técnicas de estimación en áreas pequeñas (*Small Area Estimates*) y están disponibles para el periodo comprendido entre 1995-96 y 2010-11. Los datos utilizados proceden de las declaraciones individuales del impuesto sobre la renta personal australiano, proporcionados por la *Australian Taxation Office*. La renta es definida a partir de la suma de las rentas salariales, los beneficios de las empresas no societarias, rentas del capital, pensiones y rentas vitalicias, además de otras rentas sometidas a gravamen. También se añade información sobre subsidios monetarios facilitada por el *Commonwealth Department of Family and Community Services*.

En España, a diferencia de las experiencias internacionales que acabamos de mencionar, el Instituto Nacional de Estadística (en adelante INE) no proporciona datos sobre la renta familiar o personal para el ámbito local. No obstante, existen algunas estimaciones realizadas por varios Institutos y Servicios Regionales de Estadística, si bien, por regla general, se trata de bases de datos carentes de homogeneidad y regularidad temporal en su elaboración. En este sentido, hay que señalar que en la última década estos Institutos han hecho importantes esfuerzos por ofrecer este tipo de información estadística. Entre otros, por su mayor tradición y regularidad, hay que mencionar los trabajos realizados por el IDESCAT de Cataluña, el EUSTAT en el País Vasco, el SADEI en Asturias o el IAEST en Aragón, que han aportado, aunque no siempre de manera continuada, información estadística referida, fundamentalmente, a la Renta Familiar Bruta Disponible de los municipios incluidos en sus ámbitos regionales. Por regla general, se trata de estimaciones basadas en la territorialización local de los datos de la Contabilidad Regional de España, de carácter provincial, facilitados por el INE. En prácticamente todos los casos, esta imputación territorial se lleva a cabo usando métodos de estimación indirecta a partir de indicadores de actividad económica. Como alternativa, existen métodos de estimación directa, en los que la territorialización es realizada en función de la localización espacial de los diferentes componentes de la renta bruta disponible, desde el punto de vista de los factores de producción. Sin embargo, en la práctica, estos métodos directos apenas son utilizados, dada la gran complejidad que requiere llevar a cabo dicha imputación factorial.

³ Véase: <http://www.neighbourhood.statistics.gov.uk/HTMLDocs/incomeestimates.html>

Los métodos indirectos más comúnmente utilizados se basan en técnicas econométricas que utilizan la RFBD regional o provincial, junto con otros indicadores socioeconómicos disponibles a nivel local (tales como, la población total del municipio, el número de residentes desempleados, los miembros de la población con un título de licenciatura o superior, el número de vehículos matriculados, el número de establecimientos comerciales e industriales, el precio medio de la vivienda, etc.). Debido a la falta de estadísticas oficiales, durante años, el Instituto de Investigación R. Klein Lawrence de la Universidad Autónoma de Madrid se ha convertido en el principal proveedor de datos estadísticos sobre la renta personal en los municipios españoles. Con esta información de base, hasta el año 2004, el Servicio de Estudios de la Caixa difundía, en su Anuario Económico de España, las estimaciones de la renta familiar disponible por habitante para municipios de más de 1.000 habitantes. Asimismo, en el ámbito académico se han desarrollado bastantes estimaciones puntuales de la renta personal de ámbito municipal, en trabajos por lo general de alcance regional, que también utilizaban los mencionados métodos indirectos de territorialización de la RFBD⁴.

A nuestro juicio, este uso de las variables macroeconómicas territorializadas para derivar medidas de renta local –como el valor añadido bruto (VAB en adelante) o la RFBD– presenta dos limitaciones principales para el análisis de la renta personal o familiar en el ámbito territorial municipal. En primer lugar, estas magnitudes no representan adecuadamente la capacidad personal o del hogar para pagar impuestos, considerada como los ingresos que los individuos o las familias pueden destinar al consumo o al ahorro. Esto es debido, fundamentalmente, a que, por lo general, la información contenida en la territorialización de estas macromagnitudes incluye las rentas del capital de naturaleza empresarial bajo el criterio del lugar dónde radica la actividad de producción o la sede efectiva del negocio, en lugar de vincularlas al municipio en el que residen sus propietarios, personas físicas. Por ejemplo, podemos pensar en un municipio residencial con un alto nivel de vida en el que residen los dueños de empresas cuyas actividades se localizan en otros municipios, posiblemente con bastante menor renta personal, incluso en otras regiones o países. Por supuesto, también habrá municipios cuyos residentes no tienen un alto nivel de vida, pero donde se encuentran empresas muy rentables, debido a, por ejemplo, la existencia de salarios

⁴ Entre otros, podemos mencionar los trabajos de Arcarons *et al.* (1994) y Oliver *et al.* (1995) en Cataluña, Esteban y Pedreño (1992) en la Comunidad Valenciana, Fernández y Sierra (1992) en La Rioja, De las Heras (1992) y De las Heras y Murillo (1998) en Cantabria, Herrero (1998) en Castilla y León, Remírez-Prados (1991), en Navarra, y Chasco y López (2004), en Murcia. Algunos de estos trabajos introducen métodos de estimación complejos, como el análisis de factores multivariante, el análisis de cluster o modelos econométricos multiecuacionales. Del mismo modo, Alañón (2002), utilizando técnicas de econometría espacial ofrece estimaciones del valor añadido bruto de los municipios españoles, y Chasco (2003) y Buendía *et al.* (2012) obtienen estimaciones de RFBD per cápita para la Comunidad Autónoma de Madrid y la Región de Murcia, respectivamente.

más bajos. Una segunda limitación importante del uso de los agregados macroeconómicos para estimar la renta local personal está relacionada con la imposibilidad de obtener distribuciones de la renta para los municipios y, en consecuencia, la de contar con los correspondientes índices de desigualdad y concentración de la renta. Esto es debido a que, cualquiera que sea el método estadístico o econométrico para estimar la renta per cápita de cada municipio, el resultado obtenido es un valor único, lo que hace imposible disponer de medidas sobre la dispersión de la renta estimada o simplemente de su distribución por cuantiles. Como ya dijimos, esta información distributiva tiene un indudable interés para el análisis económico aplicado.

Respecto a esta segunda limitación, la disponibilidad de muestras de microdatos representativas en el nivel local resulta indispensable para poder calcular medidas de desigualdad correspondientes a la distribución de la renta personal de un municipio. Hasta donde tenemos conocimiento, el *U.S. Census Bureau* es la única institución que desarrolla una actividad de producción estadística específica de este tipo de encuestas, pues, como hemos comentado, aunque la *Office for National Statistics* británica y el *Australian Bureau of Statistics* ofrecen información estadística a nivel local con origen en bases de microdatos, esta procede del tratamiento de encuestas de un ámbito territorial superior mediante técnicas de estimación en áreas pequeñas.

En el caso norteamericano, esto es posible ya que su base de datos, la *American Community Survey (ACS) - Public Use Microdata Sample (PUMS)* proporciona datos anuales sobre la renta personal y familiar, representativa en diferentes niveles de entidades locales. Para la ACS-PUMS, la renta personal total es definida en términos de ingresos antes de impuestos, incluyendo la suma de las cantidades correspondientes a sueldos y salarios percibidos, el beneficio neto de actividades económicas realizadas por cuenta propia, intereses, dividendos, alquileres, royalties y otros derechos de explotación de la propiedad intelectual, ingresos de otras propiedades y fideicomisos, pensiones y rentas vitalicias de todo tipo, así como el resto de ingresos gravados por el impuesto sobre la renta personal en Estados Unidos.

En la actualidad, la ACS-PUMS publica datos anuales para todas las áreas con poblaciones de 65,000 o más habitantes. Entre las cerca de 7.000 áreas locales que cumplen con este umbral de población, se encuentran todos los estados, la totalidad de los *Congressional Districts* (circunscripciones electorales), más de 700 condados y más de 500 ciudades (*places*). Para las áreas con poblaciones de menos de 65.000 habitantes se requiere el uso de estimaciones plurianuales que permitan alcanzar un tamaño de muestra adecuado para la publicación de los datos. En 2008, el *U.S. Census Bureau* comenzó a publicar las estimaciones trienales para las áreas con población superior a

20.000 habitantes, y desde 2010 se encuentran en marcha las primeras estimaciones quinquenales para la totalidad de las secciones censales y los *block-groups* (barrios).

Las encuestas de hogares que incluyen información sobre los ingresos de sus miembros son la fuente estadística natural para proporcionar los microdatos que permiten calcular medidas de renta personal, tanto a nivel individual como familiar. Para los Estados miembros de la Unión Europea, estas encuestas son el Panel de Hogares de la Comunidad Europea (PHOGUE), con ocho olas disponibles, de 1994 a 2001, y La Encuesta sobre la Renta y las Condiciones de Vida (EU-SILC), disponible desde 2003. A pesar de su alto nivel de calidad, garantizada por la coordinación y supervisión de Eurostat, el diseño de la muestra no es válido para realizar estimaciones poblacionales de las zonas territoriales más pequeñas, como provincias y mucho menos, municipios. Hay que tener en cuenta que para cumplir con un aceptable grado de representatividad estadística a nivel municipal, la realización de la encuesta requeriría un gran número de entrevistas, muy superior al que se necesita para conseguir una buena representatividad a nivel nacional, incluso regional. Por tanto, podemos decir que la razón de esta falta de microdatos disponibles para realizar estimaciones representativas en áreas locales es, principalmente, un problema de coste de producción de las encuestas. Además, tanto la declaración errónea como la infradeclaración, problemas habituales en las encuestas de gastos e ingresos, son problemas difíciles de mitigar, que se acentúan a medida que es menor el ámbito de representatividad⁵. En este sentido, Deaton (2003) advierte que los datos de las encuestas sobre la renta personal muestran una subestimación importante, si se comparan con magnitudes equivalentes incluidas en las Cuentas Nacionales, haciéndolos inadecuados para la estimación en áreas pequeñas⁶.

3. El uso de microdatos tributarios como alternativa para estimar la renta local y su distribución

Las declaraciones de los impuestos sobre la renta personal recaudados por las administraciones tributarias son una interesante alternativa para superar las limitaciones de representatividad territorial mostradas por las encuestas de hogares para el análisis de la distribución la renta personal. Como señalan Atkinson y Piketty

⁵ Meyer y Sullivan (2011) evalúan las implicaciones de estos inconvenientes para el análisis de la desigualdad de ingresos. Por otra parte, Lohmann (2011) aborda la cuestión de la recogida de datos en la UE-SILC, la búsqueda de una ventaja mayor fiabilidad en los países que complementar la información de las entrevistas de la encuesta utilizando administrativa o registro de datos para una amplia gama de variables, como ocurre en los países nórdicos.

⁶ Utilizando datos para países desarrollados y de economías en transición, Ravallion (2003) analiza las desviaciones existentes entre las Cuentas Nacionales y el ingreso familiar medio y los gastos medios de las encuestas de hogares, corroborando estas importantes diferencias. Un estudio estadístico detallado de estas discrepancias se ofrece en el Informe del Canberra Expert Group (2001).

(2007), el uso de los datos fiscales para el estudio de la distribución de la renta de las personas se remonta a décadas atrás en el tiempo⁷. No obstante, en la mayoría de países de la OCDE, las bases de microdatos correspondientes a declaraciones de impuestos están disponibles sólo para ejercicios posteriores a los años setenta y, sobre todo, a partir de los ochenta, a excepción de los Estados Unidos, donde el *Internal Revenue Service* (IRS) comenzó a publicar anualmente este tipo de bases de microdatos en 1960 (Atkinson y Piketty, 2007). Mayoritariamente estas bases de microdatos corresponden al impuesto sobre la renta personal.

La representatividad de los microdatos del impuesto sobre la renta personal suele ser adecuada para realizar estimaciones para ámbitos territoriales reducidos, como en el caso de los municipios. En general, las declaraciones anuales de este impuesto recogen información detallada sobre las diferentes categorías de la renta gravada: rendimientos procedentes de la prestación del factor trabajo por cuenta ajena, rendimientos derivados del capital mobiliario, tanto financiero como no financiero, y del capital inmobiliario, rendimientos de la realización de actividades empresariales y profesionales de carácter no societario, los ingresos procedentes de todo tipo de pensiones tanto públicas como privadas, algunas prestaciones monetarias públicas como los subsidios de desempleo o algunas indemnizaciones y subvenciones compensatorias, y la mayor parte de las ganancias patrimoniales. En algunos países, también se incluyen en la renta gravable los alquileres nocionales imputados a los propietarios de viviendas no arrendadas. La suma de todos estos componentes de la renta gravable proporciona, por regla general, una medición adecuada de los ingresos personales antes de impuestos. En principio, la aplicación universal de los impuestos sobre la renta personal garantiza un amplio marco poblacional sobre el que obtener las muestras representativas de microdatos, sin tener que incurrir en los costes de producción estadística que antes hemos comentado en el caso de las encuestas. Estas cualidades explican el creciente auge que estas bases de microdatos fiscales están teniendo dentro del análisis económico aplicado, sin olvidar, como hemos dicho, su dilatada trayectoria en algunos países anglosajones y nórdicos.

Además, durante la última década, un buen número de trabajos han centrado su atención en la concentración de la renta y la riqueza en los denominados perceptores de altos, generalmente caracterizados por las particiones de la última centila, como el 0,1% o el 0,01% más rico de la población⁸. En esta literatura, la utilización de estas bases de microdatos fiscales para determinar la distribución de los ingresos personales resulta fundamental. En este sentido, es importante señalar que

⁷ Las primeras estimaciones se remontan a Bowley (1914) y a Stamp (1914, 1916) para el Reino Unido. No obstante, las estimaciones realizadas por Kuznets (1953, 1955) para los Estados Unidos pueden ser considerados como las distribuciones pioneras de ingresos a partir de datos fiscales.

⁸ Una completa revisión de esta línea de investigación puede verse en Atkinson *et al.* (2011).

esta definición fiscal de los ingresos se considera consistente con la noción de capacidad de pago usada comúnmente en los modelos microeconómicos (Piketty y Saez, 2003), además de constituir una medida razonable de bienestar individual (Leigh, 2007). En relación con la fiabilidad de los datos fiscales para medir los ingresos personales, como señalan Feldman y Slemrod (2007) y Slemrod y Weber (2012), los datos de encuestas a menudo no son demasiado creíbles, debido al problema de las respuestas falsas a preguntas delicadas. Por lo tanto, los datos de ingresos contenidos en las declaraciones impositivas son generalmente más fiables, sobre todo cuando se trata de sueldos y salarios, pensiones, subsidios, intereses y dividendos, categorías de renta gravable por regla general sometidos a retención en la fuente por el pagador.

Sin embargo, la estimación de distribuciones de la renta personal a partir de los microdatos del impuesto sobre la renta personal no está exenta de condicionantes, algunos de ellos que deben ser tenidos en cuenta. Posiblemente, la unidad de análisis – el individuo frente a la familia– sea la cuestión más relevante. Si bien es cierto que la delimitación de la condición de contribuyente –como máximo al ámbito del matrimonio y sus dependientes, generalmente los hijos menores y en algunas ocasiones, los ascendientes con rentas por debajo de un umbral– impide disponer del hogar familiar utilizado en las encuestas, como se señala en Atkinson (2007), el enfoque individual también resulta útil en el análisis de la distribución del ingreso personal y, como tal, ha sido de uso común en la literatura relacionada sobre la desigualdad de ingresos y la redistribución. Por supuesto, existen diferencias para el análisis cuando se elige al individuo como unidad de análisis en lugar del hogar económico, pero estas deben resolverse mediante la interpretación de los resultados, sin tener por ello que renunciar a las potencialidades estadísticas del enfoque individual.

Una segunda crítica tiene que ver con la población cubierta por las declaraciones del impuesto sobre la renta personal. Con frecuencia, las normas legales que regulan estos tributos fijan umbrales de renta mínimos para establecer la obligación de presentar declaración. Aunque en principio esto supone dejar fuera de la población a un buen número de declarantes, y consecuentemente de las muestras de microdatos, esto no supone que no existan contribuyentes que aun no estando obligados a declarar sí presenten declaración por el impuesto, bien para solicitar la devolución de retenciones o pagos a cuenta, bien para acreditar bases imponibles negativas o pérdidas patrimoniales a compensar en ejercicios futuros. Asimismo, la posibilidad de disponer de ficheros complementarios de contribuyentes que han soportado retenciones por el impuesto y no han presentado declaración permite extender este marco poblacional, si bien es necesario que estos registros cuenten con las variables mínimas imprescindibles para poder fusionarlos con los de las declaraciones.

Una tercera cuestión es que estos datos podrían estar sesgados debido a la evasión y elusión fiscal. Sin embargo, Atkinson *et al.* (2011) señalan que cuando los datos fiscales se comparan con otras fuentes de información, como las encuestas, la influencia de la evasión y elusión fiscal en los resultados distributivos no es lo suficientemente grande como para significar que deben ser rechazadas de plano. De hecho, Hurst *et al.* (2010) y Paulus (2013) también encontraron en las encuestas una infradeclaración de ingresos nada despreciable en el caso de los ingresos de actividades económicas por cuenta propia, en comparación con los datos fiscales. Para España, Picos (2006) y, recientemente, Domínguez-Barrero *et al.* (2014) también encuentran que la renta personal media contenida en los microdatos del IRPF es superior a la que se deriva de la información individual reflejada en las encuestas de ingresos (el PHOGUE en el primer trabajo y la ECV en el segundo).

Un cuarto condicionante tiene que ver con las definiciones fiscales de los distintos componentes de la renta gravable, magnitud empleada en los análisis distributivos realizados a partir de estos microdatos tributarios. Como hemos señalado anteriormente, las bases imponibles del impuesto sobre la renta suelen incluir, en principio, todos los ingresos obtenidos por los residentes en un territorio, independientemente de su fuente. Pero, hay que tener en cuenta que aunque lo ideal sería medir el ingreso bruto antes de cualquier deducción o exención establecida en la normativa fiscal, esto no siempre es posible. El problema surge porque la información estadística disponible proviene de las declaraciones presentadas por los contribuyentes de acuerdo con las normas fiscales, y en muchos casos estas sólo incorporan el resultado neto de aplicar dichas minoraciones, sin que sea posible conocer el importe bruto identificable con la capacidad económica real y no con la considerada a efectos del cálculo del gravamen. Algo similar sucede cuando los ingresos son estimados, a efectos fiscales, mediante métodos indiciarios o *à forfait*, como sucede en ocasiones con los rendimientos de las actividades económicas individuales, las rentas inmobiliarias imputadas a sus propietarios o las ganancias de capital cuando se aplican coeficientes reductores para su gravamen.

En resumen, hay que decir que, a pesar de estas limitaciones, cuando nos fijamos en la medición de la renta bruta de los hogares a nivel nacional proporcionada por las encuestas se observa que esta fuente estadística ofrece niveles de ingresos medios para los adultos que perciben rentas más bajas que los derivados de los datos fiscales. A nuestro juicio, para finalizar, puede defenderse que la agregación de los diferentes componentes de los ingresos recogidos en las declaraciones del impuesto sobre la renta personal, antes de aplicar las deducciones personales y el resto de las deducciones aplicadas antes de obtener la base liquidable, permite, razonablemente,

aproximarse a una magnitud conceptualmente equivalente al concepto de renta bruta personal que se puede obtener a partir de las encuestas de ingresos.

4. Metodología para obtener estimaciones de renta municipal y su distribución a partir de los microdatos del IRPF español

4.1. El modelo

Sea $F(Y)$ la distribución de la renta personal (medida por la variable base imponible) para un año t dado y correspondiente a la población de referencia N . A su vez, $F(y)$ es la función de distribución de la misma variable pero para una muestra n de esa población obtenida del censo administrativo de declaraciones de IRPF.

Para cada una de las n observaciones (declaraciones) muestrales, los microdatos correspondientes contienen información sobre esta variable de renta personal, así como otras variables de identificación territorial, como los códigos regionales (autonómicos), provinciales y municipales. En la medida en que esta muestra anual se ha obtenido utilizando una técnica de muestreo estratificado, a cada observación i extraída en el proceso le corresponde un determinado peso w_i^h que es su factor de elevación poblacional. Sea y_i la base imponible correspondiente al declarante muestral i . Para el total de la población declarante N , la base imponible agregada (\hat{Y}) se puede estimar utilizando los pesos originales proporcionados en la muestra de declarantes, tal que:

$$\hat{Y} = \sum_{i=1}^n y_i w_i \quad [1]$$

En la medida en que la variable de estratificación espacial se fijó en el nivel provincial, las estimaciones poblacionales tanto para el nivel provincial como para el conjunto nacional mantienen el nivel de confianza establecido en el diseño de la muestra (99,9% con un error del 3 por 1000). Sin embargo, para obtener estimaciones a nivel municipal, es necesario calcular unas nuevas ponderaciones de la población, en la medida en que estas estimaciones se enfrentan a ámbitos espaciales más pequeños que los utilizados como estrato territorial en la extracción de la muestra.

Definimos este "nuevo peso", como $z_{i,j}$, de manera que la base imponible total estimada para la población del municipio j puede obtenerse de la siguiente forma:

$$\hat{Y}_{j|z} = \sum_{i=1}^{n_j} y_{i,j} z_{i,j} \quad [2]$$

De acuerdo con Creedy y Tuckwell (2004), podemos utilizar el criterio de distancia matemática para evaluar la proximidad entre $z_{i,j}$ y $w_{i,j}$ en cada una de las j áreas territoriales contempladas (en nuestro caso, los municipios). En términos

generales, definimos esta distancia a través de la función $\phi(w_{i,j}, z_{i,j})$, la cual debe verificar, en términos agregados, que:

$$D = \phi(w_{i,j}, z_{i,j}), D \in \mathbb{R}_+ \quad [3]$$

Por consiguiente, a partir de este concepto de distancia podemos obtener unos nuevos pesos que nos permitan obtener estimaciones de las bases imponibles agregadas para cada municipio utilizando la misma muestra de declaraciones del IRPF. Para ello basta con resolver el programa de optimización consistente en minimizar la función de distancia [3] sujeta a la restricción por municipio recogida en [2]. Para poder calcular estas nuevas ponderaciones tan solo necesitamos información sobre los valores poblacionales verdaderos para cada uno de los j municipios de la variable renta que queremos estimar en ese ámbito territorial (en nuestro caso, la base imponible). De este modo, para cada municipio, el valor $\hat{Y}_{j|z}$ se sustituye por el valor realmente observado para toda la población declarante del municipio j , Y_j . Esta información adicional procede del censo administrativo de declarantes de IRPF, del que se extrajo la muestra⁹.

3.2. Método computacional: el enfoque de calibración

En este apartado se presenta un resumen del método que utilizamos para ajustar los pesos originales de la muestra de microdatos proporcionada por la AEAT, con el fin de conseguir la representatividad de la base imponible media de cada uno de los municipios del territorio común de aplicación del IRPF español, así como de su número total de contribuyentes. La metodología computacional que se presenta es una adaptación de los métodos propuestos en Creedy (2003), Creedy y Tuckwell (2004) y Deville y Särndal (1992). Para su aplicación empírica, los algoritmos de optimización han sido programados utilizando el software Stata[®] versión 13.

Siguiendo a Creedy (2003), consideremos una muestra de n contribuyentes y K variables con información a nivel individual (para cada observación), tanto monetarias (como p.e. la base imponible o la cuota impositiva) y no monetarias (como p.e. la edad del declarante, sexo, región, provincia o municipio de residencia). El conjunto de estas variables para el contribuyente genérico i se recoge en el siguiente vector: $x_i = [x_{i1}, x_{i2}, \dots, x_{ik}]'$. Si definimos los pesos originales de la muestra para las n observaciones a través del vector $w = [w_1, w_2, \dots, w_i, \dots, w_n]$, el valor agregado poblacional de cada una de las K variables puede estimarse tal que:

$$\hat{X}_{k|w} = \sum_{i=1}^n w_i x_{ik} \quad [4]$$

⁹ Para el desarrollo de las estimaciones realizadas, esta información poblacional ha sido facilitada por la Agencia Estatal de Administración Tributaria (AEAT).

A partir de la información suministrada por la AEAT correspondiente al valor agregado poblacional de la base imponible en cada uno de los J municipios (X_j) y del número total de contribuyentes residentes en cada municipio (N_j), es posible calcular un nuevo vector de pesos de la muestra para cada municipio, $z_j = [z_{1j}, \dots, z_{n_jj}]$, siendo las observaciones muestrales disponibles de cada municipio n_j , de manera que debe cumplirse que $\sum_{j=1}^J n_j = n$. Ahora se trata de calcular este nuevo vector de factores de elevación poblacional tal que cumpla la condición de ser el más cercano al vector que contiene los pesos originales de la muestra. El nuevo vector de pesos así calculado debe satisfacer el conjunto de las K ecuaciones de calibración (en nuestro caso una sólo, al ser $K=1$, pues estimamos una única variable, la base imponible:

$$X_k^j = \sum_{i=1}^{n_j} z_{i,j} x_{ik} \quad [5]$$

donde X_k^j representa el verdadero valor poblacional del agregado de la variable k para el municipio j . Por tanto, si denominamos con la función $\phi(w_{i,j}, z_{i,j})$ la distancia entre los pesos poblacionales originales y las nuevos factores de elevación aplicables a la muestra, estos se pueden obtener minimizando con respecto a z la siguiente función Lagrangiana:

$$L = \sum_{i=1}^n \phi(w_{i,j}, z_{i,j}) + \sum_{k=1}^K \lambda_k [X_k^j - \sum_{i=1}^{n_j} z_{i,j} x_{ik}] \quad [6]$$

donde $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_K]'$ son los respectivos multiplicadores de Lagrange. En nuestro caso, al existir una única variable x , la expresión quedaría simplificada tal que,

$$L = \sum_{i=1}^n \phi(w_{i,j}, z_{i,j}) + \lambda [X^j - \sum_{i=1}^{n_j} z_{i,j} x_i] \quad [6']$$

Es evidente que la solución del anterior problema de minimización depende fuertemente de las propiedades de la función distancia $\phi(w_{i,j}, z_{i,j})$. Por ello, vamos a exigir que la especificación que adopte esta función $\phi(\cdot)$ respete dos propiedades fundamentales:

- Que la primera derivada de $\phi(\cdot)$ con respecto a $z_{i,j}$ pueda ser expresada como una función del ratio entre los nuevos pesos y los originales de la muestra:

$$\frac{\partial \phi(w_{i,j}, z_{i,j})}{\partial z_{i,j}} = \phi' \left(\frac{z_{i,j}}{w_{i,j}} \right) \quad [7]$$

- Que la inversa de la primera derivada de $\phi(\cdot)$ sea explícitamente invertible.

Si ambas propiedades se cumplen, entonces las primeras n condiciones de primer orden para el problema de optimización mostrado en [6] son:

$$\phi' \left(\frac{z_{i,j}}{w_{i,j}} \right) - x'_i \lambda = 0 \quad i = 1, 2, \dots, n \quad [8]$$

Entonces, ahora ya podemos obtener directamente los nuevos pesos, tal que:

$$z_{i,j} = w_{i,j} \phi'^{-1}(x'_i \lambda) \quad i = 1, 2, \dots, n \quad [9]$$

Asimismo, podemos disponer de una solución para los multiplicadores de Lagrange, que se pueden obtener a través de un procedimiento iterativo (el método de Newton) después de algunas manipulaciones algebraicas. En concreto, si sustituimos la ecuación [9] en la ecuación [5] y luego restamos de ambos lados la ecuación [1], después de algunas reordenaciones, obtenemos la siguiente ecuación,

$$\left(X - \hat{X} \Big|_w \right) - \sum_{i=1}^n w_i [\phi'^{-1}(x'_i \lambda) - 1] x_i = 0 \quad [10]$$

cuya raíz se puede calcular iterativamente mediante el siguiente algoritmo recursivo,

$$\lambda^{(l+1)} = \lambda^{(l)} - \left[\frac{\partial f(\lambda)}{\partial \lambda} \right]^{-1} f(\lambda) \quad [11]$$

donde $f(\lambda)$ está dado por el lado izquierdo de la ecuación [10] y $\lambda^{(l+1)}$ es evaluado, para cada $l + 1$ iteración, utilizando el valor del multiplicador de Lagrange en la iteración l previa, $\lambda^{(l)}$. Por lo tanto, dado un conjunto de valores iniciales para λ , la ecuación [11] se puede evaluar varias veces hasta que se alcanza, siempre que sea posible, la convergencia.

Tabla 1. Alternativas de funciones distancia

	$D(w,z)$
1. Chi-cuadrado (χ^2)	$\frac{(z - w)^2}{2w}$
2. Mínima Entropía	$-w \ln \left(\frac{z}{w} \right) + z - w$
3. Mínima Entropía Modificada	$z \ln \left(\frac{z}{w} \right) - z + w$
4. Deville y Särndal (1992)	$\left(u - \frac{z}{w} \right) \ln \left(\frac{u - \frac{z}{w}}{u - 1} \right) + \left(\frac{z}{w} - l \right) \ln \left(\frac{\frac{z}{w} - l}{1 - l} \right) + \frac{u - l}{\alpha} w$

Nota: u y l son constantes conocidas en el intervalo $l < 1 < u$, tal que, $\alpha = \frac{u-l}{(1-l)(u-1)}$.

En la Tabla 1, se presentan las cuatro funciones de distancia que hemos utilizado para obtener las estimaciones de la renta personal de los municipios españoles. La primera, la función de distancia Chi-cuadrado (χ^2), es probablemente una de las opciones más habituales en la literatura aplicada, debido a que su uso en la

especificación del problema de minimización restringida de la ecuación [6] tiene una solución explícita, lo que permite obtener de forma inmediata los nuevos pesos buscados. Como salvedad, hay que señalar que esta función no impone limitaciones en el tamaño del ajuste realizado para cada peso, lo que puede provocar que algunos de los nuevos factores tomen valores negativos. Para evitar este problema, las otras tres funciones de distancia consideradas (ver Tabla 1), se incorpora una restricción de no negatividad para los nuevos valores de los pesos. No obstante, para estas otras funciones alternativas, la ausencia de una solución explícita al problema de minimización restringida obliga a tener que utilizar el procedimiento iterativo explicado anteriormente. Por supuesto, esto conlleva la posibilidad de que surjan problemas de no convergencia, consecuencia de la combinación de la forma específica de la función distancia con los pesos originales o con los valores concretos que entren en la primera iteración del algoritmo recursivo.

En este sentido, en las funciones distancia de mínima entropía y mínima entropía modificada se obliga a que el nuevo factor de elevación obtenido sea positivo, aunque no se impone un límite superior para el ajuste. Esto puede dar lugar a que, en algunos casos, tras el proceso de calibración aparezcan unos pesos con elevados valores, realmente inverosímiles. Precisamente, para paliar este problema, consideramos una cuarta función distancia, la propuesta por Deville y Särndal (1992), que limita los valores de los nuevos factores poblacionales dentro de un rango definido por el usuario. En particular, la restricción contempla límites inferiores y superiores para el ratio entre el valor de cada nuevo peso y su respectivo valor original, tal que,

$$l < \frac{z}{w} < u \quad [12]$$

siendo tanto l como u parámetros conocidos incorporados en la función distancia antes del proceso recursivo de calibración¹⁰.

4.3. Selección óptima del vector de pesos y validación

Para realizar la estimación de la renta personal a nivel municipal en el año base de 2007, contamos con los microdatos de la Muestra Anual de Declarantes de IRPF elaborada por el IEF y la AEAT¹¹. Esta muestra de microdatos es representativa de la población declarante del IRPF en el territorio de aplicación común estatal de este impuesto. Por tanto, cubre 5.346 de los 7.588 municipios españoles, es decir, aquellos que tienen régimen fiscal común. En el análisis empírico realizado, se han descartado

¹⁰ Los valores iniciales contemplados para estos dos parámetros son, respectivamente, 0,2 y 3,0. respectivamente. Si la convergencia no se alcanza después de 100 iteraciones con diferentes valores iniciales, consideramos unos nuevos límites para estos dos parámetros extraídos de dos distribuciones uniformes, con soportes, 0,1–1 y 1–6.

¹¹ Sobre su contenido, véase Picos *et al.* (2011).

18 municipios para los que solamente se disponía de una observación en la muestra (lo que impide aplicar cualquiera de los métodos de reponderación presentados en la sección anterior). Para aplicar la metodología descrita, la AEAT ha proporcionado dos magnitudes de población total de declarantes de IRPF del ejercicio 2007: el número de declaraciones presentadas y registradas en cada municipio y la renta imponible bruta total de cada municipio. Por tanto, con esta información, ya se está en condiciones de definir y computar el conjunto de ecuaciones de calibración recogidas en la sección metodológica.

En la Tabla 2 se muestra el porcentaje de los 5.328 municipios para los cuales se logra la convergencia al utilizar el algoritmo iterativo que proporciona los nuevos factores de elevación. La tabla también refleja el porcentaje de municipios para los cuales se observaron pesos no negativos después de la calibración con cada una de las cuatro funciones distancia.

Tabla 2. Porcentaje de municipios con un nuevo vector de pesos no negativos

Función Distancia	Porcentaje
Chi-cuadrado (χ^2)	82,2%
Mínima Entropía	91,6%
Mínima Entropía Modificada	94,8%
Deville y Särndal (1992)	73,3%

Fuente: Elaboración propia.

Para 250 municipios (para los cuales se dispone de 1.953 declaraciones muestrales de IRPF) ninguna de las funciones distancia utilizadas proporciona un nuevo vector de pesos. Esto es debido bien a cuestiones de no convergencia del algoritmo empleado, bien a que la función de la distancia Chi-cuadrado produce pesos negativos. Hay que tener en cuenta que la metodología empleada elimina todas las observaciones muestrales de un municipio si algoritmo aplicado no proporciona para todas las observaciones valores positivos para los nuevos pesos. Se trata, no obstante, de municipios bastante pequeños, todos ellos de menos de 1.000 habitantes y con un número poblacional de declaraciones de IRPF por debajo de 500 y por debajo de 30 declaraciones muestrales. En Hortas-Rico *et al.* (2014) se grafican las distribuciones de densidad Kernel para estos municipios en los que no ha sido posible obtener un nuevo vector de factores poblacionales.

Para seleccionar el vector de pesos óptimo de cada municipio, de entre los cuatro disponibles obtenidos para las respectivas funciones distancias, utilizamos el criterio propuesto en Särndal (2007), basado en los siguientes requisitos a cumplir por el vector elegido para el municipio j :

(i) No negatividad de los nuevos pesos:

$$z_{i,j} \geq 0, \forall i \quad [13]$$

(ii) Limitación de las distancias entre los pesos del nuevo vector (**Z**) y del vector original (**W**), lo cual es contrastado mediante el criterio de bondad del ajuste a través de la minimización del cuadrado de las diferencias:

$$\min \sum_{i=1}^n (w_{i,j} - z_{i,j})^2 \quad [13]$$

(iii) Consecución en el ejercicio de calibración de una convergencia lo más suave posible.

Tabla 3. Función distancia elegida en cada municipio

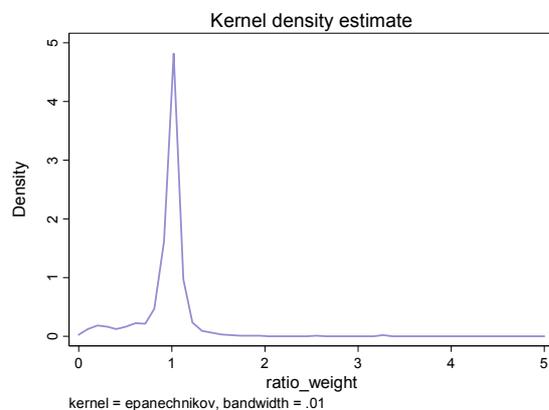
Función Distancia	Número de municipios	Porcentaje
Chi-cuadrado (χ^2)	1.607	82,2%
Mínima Entropía	2.496	91,6%
Mínima Entropía Modificada	473	94,8%
Deville y Särndal (1992)	502	73,3%
Total	5.078	100,0%

Fuente: Elaboración propia.

De acuerdo con los criterios que acabamos de exponer, la distancia de Mínima Entropía es la función más utilizada, seguida por la distancia Chi-cuadrado, tal y como se refleja en la Tabla 3. No obstante, hay que aclarar que tal y como prueban Deville y Särndal (1992), las cuatro funciones generan calibraciones de los estimadores asintóticamente equivalentes, es decir, que, por regla general, un cambio en la función distancia elegida apenas tiene efectos sobre la varianza del estimador, incluso si el tamaño muestral del municipio es bastante pequeño.

En la Figura 1 mostramos la distribución del ratio entre los nuevos factores de elevación poblacional obtenidos de la calibración por municipios y los pesos originales proporcionados en la muestra de microdatos de IRPF del IEF-AEAT para el ejercicio 2007. Como podemos observar, la enorme mayoría de los valores del ratio $z_{i,j}/w_{i,j}$ se encuentran alrededor de uno, lo que supone que los nuevos pesos que hacen representativa la muestra por municipios son muy cercanos a los originales que aseguraban la representatividad a nivel provincial. El rango de los valores del ratio va de 0,06 a 1,80, situándose su media y su mediana muy próximos a 1, con una desviación estándar de 0,98. La distribución de este ratio por percentiles se ofrece en la Tabla 4.

Figura 1. Ratio de los nuevos factores de elevación poblacional y los pesos originales



Fuente: Elaboración propia

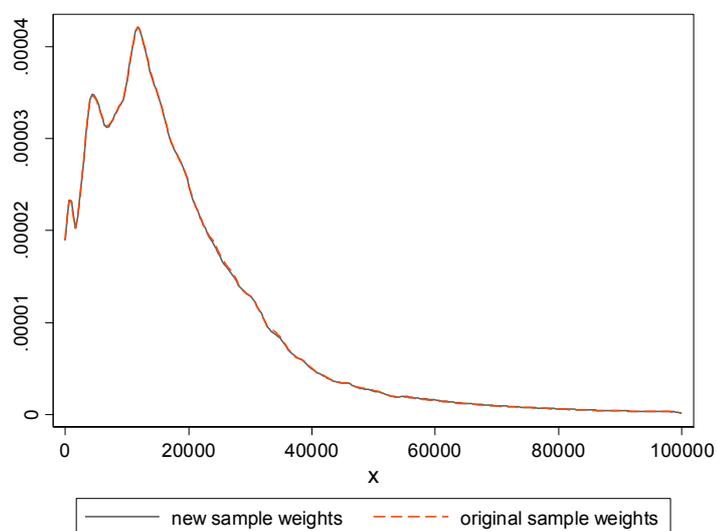
Tabla 4. Distribución del ratio de los nuevos factores de elevación poblacional y los pesos originales por percentiles.

Percentiles	Ratio z/w
1%	0.06013
5%	0.31796
10%	0.62805
25%	0.91277
50%	0.99691
75%	1.04968
90%	1.14317
95%	1.24089
99%	1.80791
Media	0.97445
Desviación estándar	0.98183

Fuente: Elaboración propia.

A partir de los nuevos factores de elevación poblacional de los microdatos de la muestra de IRPF 2007 que han sido obtenidos mediante el proceso computacional descrito, ya estamos en condiciones de estimar las funciones de distribución de la renta personal correspondientes a todos los municipios españoles del territorio común de aplicación de este impuesto. Un primer ejercicio de validación que se ha realizado ha sido comparar la estimación de la función de distribución de la renta gravable para el total de observaciones de la muestra anual de IRPF antes y después de este proceso de refactorización. El resultado se muestra en la Figura 2.

Figura 2. Distribución de la renta gravable



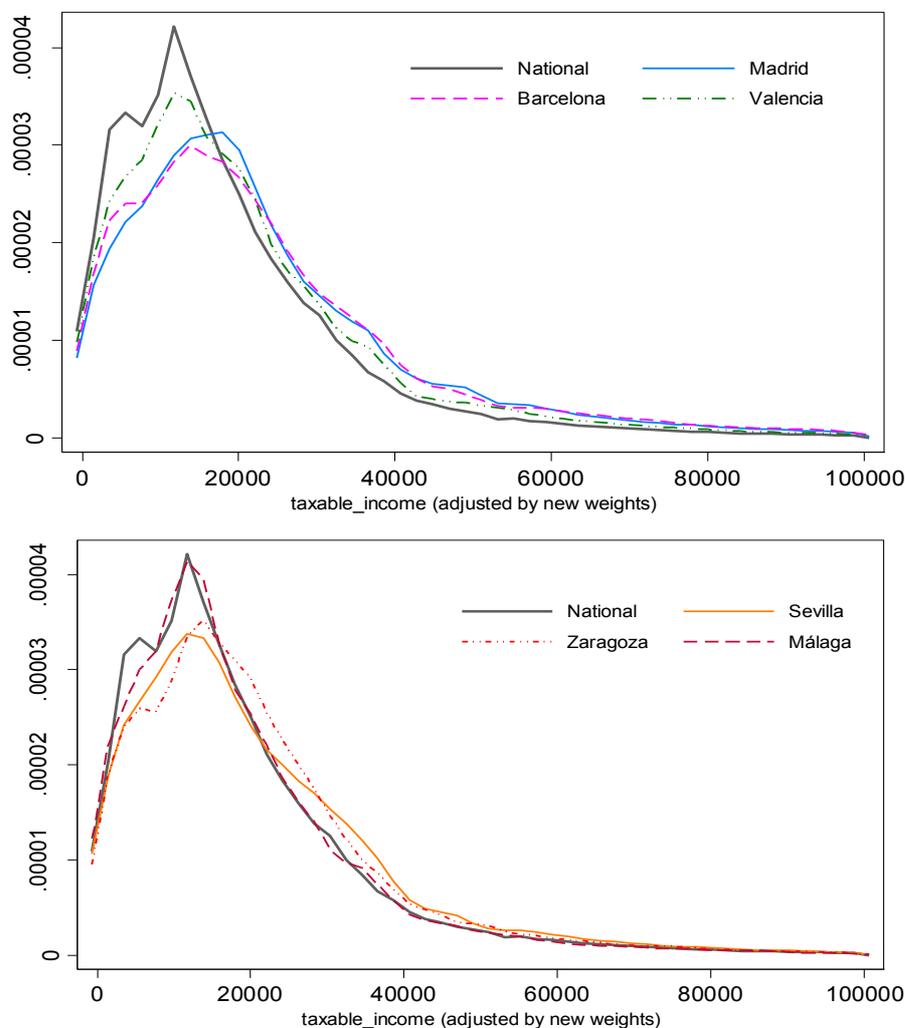
Fuente: Elaboración propia.

Tal y como era de esperar, la distribución, expresada en términos de densidad de frecuencia, obtenida con los nuevos pesos replica perfectamente a la función proporcionada por los pesos originales. En cambio, al realizar esta comparación por municipios, cabe esperar la existencia de diferencias entre ambas estimaciones de las funciones de densidad de frecuencia, puesto que los factores de elevación originales únicamente son representativos en el nivel de agregación territorial provincial, mientras que los nuevos pesos calculados lo son en el nivel municipal. Conviene aclarar aquí que, en todo caso, la muestra utilizada, con los nuevos factores, es representativa de la población completa, de acuerdo con las condiciones exigidas en el marco metodológico expuesto, tal y como se comprueba en la Figura 2.

En la Figura 3 mostramos las distribuciones de la renta personal (funciones de densidad de frecuencia) de los seis municipios con mayor población de España (si exceptuamos Bilbao, al no estar incluido en el territorio fiscal común) estimadas con los nuevos pesos, comparadas con la correspondiente al total nacional (en la línea continua más gruesa). Estos gráficos revelan, como era de esperar, la existencia de patrones distributivos heterogéneos, especialmente entre las tres ciudades más grandes (Madrid, Barcelona y Valencia), si bien, en todos los casos, la renta local aparece más uniformemente distribuida que cuando nos fijamos en la distribución para la población de todos los municipios incluidos en la muestra. Hay que destacar también que estas ciudades presentan, en comparación con el total de población nacional, distribuciones

de la renta personal con mayores concentraciones de población relativa en la cola alta de la distribución y menores en la parte baja.

Figura 3. Distribuciones de renta personal para los 6 municipios con mayor población de España



Fuente: Elaboración propia.

Nota: Distribución truncada en los 100.000 euros.

5. Los datos utilizados

5.1. Las Muestras Anuales de Declarantes de IRPF (IEF-AEAT)

Para llevar a cabo la estimación de la renta personal de los municipios españoles en este proyecto se utilizan los microdatos de las Muestras Anuales de Declarantes de IRPF proporcionadas y difundidas públicamente, actualmente para el periodo 2002-2010, por el Instituto de Estudios Fiscales, organismo dependiente del

Ministerio de Hacienda y Administraciones Públicas, y la Agencia Estatal de Administración Tributaria, entidad encargada de la extracción de las muestras a partir de sus registros administrativos censales. En particular, como año base de la base de datos *Renta personal de los municipios españoles* que elaboramos para FEDEA se utiliza la muestra anual de IRPF del ejercicio 2007. Esta muestra incluye 1.351.802 registros extraídos de una población de 18.702.875 declaraciones del IRPF español (Picos *et al.*, 2011), si bien, finalmente, la selección de municipios realizada en el proceso de aplicación de la metodología de refactorización reduce a 1.337.957 las observaciones muestrales empleadas.

Para la construcción de estas muestras anuales de Declarantes de IRPF se utiliza el método de muestro estratificado aleatorio de mínima varianza (método de Neyman). Los tres estratos considerados en su selección son: a) las provincias españolas del Territorio Fiscal Común (48, además de Ceuta y Melilla que se han considerado de forma conjunta); b) el nivel de renta de los declarantes, considerado mediante 12 tramos de renta; y c) el tipo de declaración, según se trate de tributación individual o conjunta. Por tanto, el número de estratos de último nivel es de 1.152 ($48 \times 2 \times 12$), no existiendo estratos vacíos. La variable utilizada para definir los tramos de renta se construye como la suma del “Saldo neto de los rendimientos a integrar en la base imponible general y de las imputaciones de rentas” (casilla 452 del impreso de declaración de IRPF de 2007) más la “Base imponible del ahorro” (casilla 465). Esta variable es utilizada en el procedimiento de selección como renta de muestreo, a efectos de calcular la varianza mínima en los estratos. Esta noción de renta incluye la suma de todos los rendimientos netos, de los ingresos imputados, así como las ganancias y pérdidas patrimoniales declaradas por los contribuyentes del impuesto.

Para seleccionar la muestra, inicialmente, el censo de declaraciones de IRPF que posee la AEAT para este ejercicio fiscal (N) fue clasificado en los 1.152 estratos definidos (h). Por lo que respecta a la muestra, en primer lugar se calculó el número total de observaciones a obtener (n), para un error específico relativo de muestreo ($e < 0,011$) y un nivel de confianza del 3 por 1.000. A continuación, con la población de cada estrato (N_h) se calculó la cuasi-varianza poblacional para la renta de muestreo acumulada en el mismo (S_h^2). Por último, a partir de N_h y S_h^2 se determinó el número de observaciones muestrales a extraer aleatoriamente para cada estrato (n_h), de forma que $\sum_h n_h = n$. Por consiguiente, el factor analítico de elevación poblacional de cada estrato queda definido como $w_h = N_h/n_h$. En la Tabla 3 se muestra el número final de observaciones muestrales disponibles por provincias.

Los registros originales de las muestras proporcionados por la AEAT se incorporan por el IEF en un archivo bidimensional que contiene las declaraciones de

IRPF obtenidas mediante el proceso de muestreo descrito (una por fila). Para cada observación, este archivo ofrece una serie de variables para las que la fuente de la información es, directa o indirectamente, el formulario de declaración de IRPF para el año correspondiente. De acuerdo con la naturaleza de las variables incluidas en el archivo, estas pueden dividirse en dos grupos: variables no monetarias, las cuales contienen las principales características cualitativas y personales de los declarantes; y variables monetarias, que contienen la información recogida en las casillas con importes del formulario de IRPF.

Entre las variables no monetarias, entre otras se incluyen la fecha de nacimiento el sexo, el estado civil, el número de descendientes (menores de 25 años con derecho a los beneficios contemplados en la normativa del IRPF), el número de ascendientes, o el grado de discapacidad de las personas afectadas por la declaración. Asimismo, se incluyen en el fichero variables territoriales relacionadas con la residencia fiscal de los declarantes, como la Comunidad Autónoma, la Provincia y el Código Postal del municipio de residencia. En total, en la muestra de 2007 se han incorporado 76 variables no monetarias, mientras que las variables monetarias son 295. A partir de estas variables monetarias nosotros establecemos la definición de renta personal gravable que se utilizará en la base de datos, la cual incorpora los siguientes conceptos de renta: rentas salariales y otros rendimientos del trabajo personal por cuenta ajena, pensiones de jubilación, pensiones de discapacidad y otras no exentas en el IRPF, subsidios de desempleo, rendimientos netos de actividades empresariales, profesionales y agropecuarias de carácter individual, intereses, dividendos, cánones y otras rentas de cesión de propiedad no inmobiliaria, pagos de planes y seguros de pensiones privados, alquileres netos y rentas inmobiliarias imputadas a los propietarios, excepto las de la vivienda habitual, y ganancias de capital realizadas. Por tanto, se trata de una renta personal de carácter fiscal definida en términos brutos (renta antes de aplicar el IRPF), calculada sin tener en cuenta las posibles minoraciones por circunstancias personales y familiares, las reducciones aplicadas para los rendimientos del trabajo personal, así como las exenciones por aportación a planes de pensiones privados y otras minoraciones aplicables para calcular la base liquidable del impuesto.

Por lo que respecta a la unidad observacional del IRPF, esta se identifica con la declaración del impuesto. Desde el ejercicio fiscal de 1988, por Sentencia del Tribunal Constitucional, el IRPF español es concebido como un impuesto de carácter individual, si bien, optativamente se permite a los cónyuges presentar declaración conjunta. No obstante, los sucesivos diseños legales del impuesto han supuesto que en los matrimonios que optan por el régimen de tributación conjunta sólo exista un ganador

principal de rentas, puesto que si estas se obtuviesen de forma significativa por ambos cónyuges, el régimen de tributación individual permitiría a estos una rebaja sustancial de la cuota tributaria total del matrimonio. Este escenario permite identificar, sin demasiado error, al declarante individual como unidad observacional en la muestra.

En cuanto a la representación territorial, la muestra anual de microdatos de 2007 incluye declaraciones de IRPF para 5.346 municipios sobre los 7.588 existentes en las 15 Comunidades Autónomas con régimen fiscal común, incluidas las dos Ciudades Autónomas de Ceuta y Melilla. No obstante, por razones de anonimato de los contribuyentes, la AEAT ha limitado la disponibilidad de aquellos microdatos para los que no se cumple que exista un número mínimo de observaciones que garanticen, en el último nivel de desagregación, la salvaguarda del secreto y la confidencialidad de los contribuyentes.

5.2. Los datos poblacionales de las Declaraciones de IRPF por municipios (AEAT).

El Departamento de Tecnología de la Información de la AEAT ha proporcionado, previa petición oficial, una estadística anual que contiene información sobre el número de declaraciones de IRPF presentadas en 7.024 municipios españoles de Régimen Fiscal Común, junto con su renta imponible media –concepto este asimilable a la definición de renta gravable personal construida con los microdatos de la muestra anual de IRPF (suma de las casillas 455 y 465 del impreso de IRPF de 2007)– y la cuota tributaria promedio. Para fines de identificación, esta base de datos incluye un código municipal tributario específico establecido por la AEAT, así como el nombre del municipio correspondiente.

Para poder establecer la vinculación necesaria entre la muestra de microdatos de IRPF y esta estadística poblacional de IRPF por municipios, ha sido necesario vincular estos códigos tributarios municipales a los correspondientes códigos postales que figuran en las declaraciones de la muestra de microdatos. Una vez identificado el municipio en ambas bases de datos, se ha procedido a incluir en cada observación el código municipal definitivo, coincidente con el de cinco dígitos proporcionado por el Instituto Nacional de Estadística (en adelante, “código poblacional INE”).

6. La base de datos y su contenido

Los nuevos vectores de pesos se calculan para la totalidad de los municipios con representación en la Muestra de Declarantes de IRPF, con la excepción, ya comentada, de aquellos para los cuales el procedimiento no ha permitido obtener los

nuevos factores de elevación para la totalidad de observaciones del municipio¹² (250 municipios en 2007). No obstante, la creación de una base de datos de difusión abierta, que vaya incorporando la información sobre la renta personal y su distribución para cada año a medida que se tenga disponibilidad de los datos necesarios para la estimación, nos ha llevado a incluir en la misma un subconjunto de estos municipios para los cuáles el número de observaciones muestrales previsiblemente se mantendrá suficientemente estable a lo largo del tiempo. El criterio elegido ha sido el de la población con residencia en el municipio, según el dato proporcionado por el INE para cada ejercicio. En concreto, la base de datos *Renta personal de los municipios españoles y su distribución* incorpora la información para las localidades con una población residente, en el año correspondiente, de más de 5.000 habitantes y para las cuales la Muestra Anual de IRPF IEF-AEAT ofrece microdatos. Bajo este criterio, en el año 2007, son 1.114 municipios los incluidos en la base de datos, repartidos por todas las CC.AA., si bien para cinco de ellos el proceso de refactorización no proporciona los nuevos pesos buscados, quedando por tanto excluidos¹³. En la Tabla 5 se muestra el número de municipios finalmente incorporados a la estadística de 2007.

6.1. Estructura y contenido de la base de datos

La base de datos elaborada presenta, para cada año, una estructura bidimensional, en la que los municipios se disponen en las filas, mientras que las columnas recogen la información estadística vinculada a cada uno de ellos. En concreto, esta información se agrupa en cinco grandes bloques. En primer lugar, un conjunto de variables que, bajo la rúbrica “Información básica”, recogen los datos correspondientes a la identificación del municipio, su ubicación regional y provincial, así como a su población censal y fiscal. El segundo bloque, “Renta personal local”, incorpora los datos correspondientes a la estimación de la renta personal municipal que hemos realizado a partir de los nuevos pesos, recogiendo el montante agregado para cada municipio, así como los importes medios por habitante y por declarante de IRPF. El tercer bloque, “Índices de desigualdad de la renta local”, incluye las medidas de desigualdad que calculadas con los nuevos vectores de pesos a partir de los microdatos de IRPF. El cuarto bloque, “Concentración de la renta local”, muestra el porcentaje de concentración de la renta personal en los niveles más altos de la distribución: para el 1%, el 0,5% y el 0,1% más rico de la población declarante. Por último, la información presentada en el quinto bloque, “Distribución de la renta local por quintiles” tiene como propósito mostrar la estructura de la función de distribución

¹² Por supuesto, para municipios de pequeño tamaño pueden producirse algunos cambios entre ejercicios.

¹³ El número de municipios de Régimen Fiscal Común con más de 5.000 habitantes, en 2007, es de 1.178.

de la renta personal de cada municipio, ofreciendo para ello el porcentaje de renta personal que se concentra en cada quintil.

Tabla 5. Número de municipios por Provincias de Régimen Común. Año 2007

<i>Provincia</i>	<i>Número de municipios</i>		<i>Provincia</i>	<i>Número de municipios</i>	
	<i>Total INE</i>	<i>Estadística FEDEA (*)</i>		<i>Total INE</i>	<i>Estadística FEDEA (*)</i>
Albacete	87	7	León	211	11
Alicante	141	52	Lleida	231	13
Almería	102	18	Lugo	67	12
Asturias	78	31	Madrid	179	69
Ávila	248	5	Málaga	100	24
Badajoz	164	28	Melilla	1	1
Baleares	67	37	Murcia	45	32
Barcelona	311	115	Ourense	92	10
Burgos	371	5	Palencia	191	5
Cáceres	219	11	Palmas de G.C.	34	29
Cádiz	44	31	Pontevedra	62	38
Cantabria	102	17	Rioja, La	174	9
Castellón	135	18	Salamanca	362	7
Ceuta	1	1	S.C. Tenerife	53	36
Ciudad Real	102	23	Segovia	209	4
Córdoba	75	24	Sevilla	105	59
Coruña, A	94	57	Soria	183	3
Cuenca	238	7	Tarragona	183	26
Girona	221	26	Teruel	236	3
Granada	168	34	Toledo	204	22
Guadalajara	288	5	Valencia	266	69
Huelva	79	17	Valladolid	225	9
Huesca	202	7	Zamora	248	3
Jaén	97	29	Zaragoza	293	10
<i>Total Régimen Común con población mayor de 5.000 habitantes</i>				<i>1.178</i>	<i>1.109</i>
<i>Total Régimen Común</i>				<i>7.588</i>	

Fuente: Elaboración propia

NOTA: (*) Municipios con población INE mayor de 5.000 habitantes y con observaciones en la Muestra Anual de Declarantes de IRPF y para los que ha convergido el algoritmo proporcionando su vector de nuevos pesos.

6.2. Descripción de variables

- "Código INE Comunidad Autónoma"

Variable numérica que recoge el código de dos dígitos del INE para la identificación de las Comunidades Autónomas. Sus valores presentes en la base de datos son:

1	Andalucía
2	Aragón
3	Asturias, Principado de
4	Balears, Illes
5	Canarias
6	Cantabria
7	Castilla y León
8	Castilla-La Mancha
9	Cataluña
10	Extremadura
11	Galicia
12	Madrid, Comunidad de
13	Murcia, Región de
16	Rioja, La
17	Comunitat Valenciana

- "Nombre Comunidad Autónoma"

Recoge el nombre oficial de la Comunidad Autónoma, según figura en el listado de la variable anterior.

- "Código INE Provincia"

Variable numérica que recoge el código de dos dígitos del INE para la identificación de las Provincias. Sus valores presentes en la base de datos son:

02	Albacete
03	Alicante/Alacant
04	Almería
05	Ávila
06	Badajoz
07	Balears, Illes
08	Barcelona
09	Burgos
10	Cáceres
11	Cádiz
12	Castellón/Castelló
13	Ciudad Real
14	Córdoba
15	Coruña, A
16	Cuenca
17	Girona
18	Granada
19	Guadalajara
21	Huelva
22	Huesca
23	Jaén
24	León
25	Lleida
26	Rioja, La
27	Lugo
28	Madrid
29	Málaga

30	Murcia
32	Ourense
33	Asturias
34	Palencia
35	Palmas, Las
36	Pontevedra
37	Salamanca
38	Santa Cruz de Tenerife
39	Cantabria
40	Segovia
41	Sevilla
42	Soria
43	Tarragona
44	Teruel
45	Toledo
46	Valencia/València
47	Valladolid
49	Zamora
50	Zaragoza
51	Ceuta
52	Melilla

- “Nombre Provincia”

Recoge el nombre oficial de la Provincia, según figura en el listado de la variable anterior.

- “Código INE municipio”

Variable numérica que recoge el valor del código asignado por el INE a cada municipio. Al ser tratado numéricamente, aquellos municipios con códigos inferiores a las decenas de millar sólo presentan cuatro dígitos.

- “Código INE municipio (5 dígitos)”

Variable alfanumérica de cinco posiciones que recoge los cinco dígitos del código de municipio asignado por el INE. A diferencia de la anterior, todos los municipios presentan los cinco dígitos.

- “Nombre municipio”

Recoge el nombre oficial del municipio de acuerdo con el listado disponible en el INE.

- “Número de observaciones muestrales”

Variable numérica correspondiente al número de observaciones de la Muestra Anual de Declarantes de IRPF IEF-AEAT disponibles y utilizadas en la estimación de la renta municipal.

- “Población (INE)”

Variable numérica que recoge el número de habitantes del municipio a 31 de diciembre del año de la estadística, de acuerdo con la información sobre población ofrecida por el INE.

- “Población declarante (IRPF)”

Variable numérica que informa del número total de declaraciones del IRPF de los residentes con domicilio fiscal en el municipio registradas en el ejercicio de la estadística por la AEAT.

- “Renta imponible agregada (IRPF)”

Variable numérica, expresada en euros, que recoge la suma de la renta imponible que juntan todos los declarantes del IRPF con domicilio fiscal en el municipio. Esta magnitud es definida en términos de renta gravable total, conforme se indicó en la sección 5 de este documento.

- “Renta imponible media (por declarante)”

Variable numérica, expresada en euros y con dos decimales, que incorpora el valor de la renta imponible media por declarante del IRPF con domicilio fiscal en el municipio.

- “Renta imponible media (por habitante)”

Variable numérica, expresada en euros y con dos decimales, que recoge el valor medio de la renta imponible resultante de dividir la renta imponible agregada del municipio entre la población residente en el mismo.

- “Índice de Gini”

Valor del índice de Gini (1912) correspondiente a la distribución de la renta imponible del municipio, calculado a partir de los microdatos de la Muestra Anual de Declarantes de IRPF con los nuevos factores de elevación poblacional calculados. El valor del índice se expresa en el rango 0 – 1.

El cómputo de este índice se realiza mediante el programa “ineqdec0” para Stata[®] (Jenkins, 2008).

- “Índice de Atkinson 0,5”

Valor del índice de desigualdad de Atkinson (1970) para un parámetro de aversión a la desigualdad $\alpha = 0,5$, correspondiente a la distribución de la renta imponible del municipio, calculado a partir de los microdatos de la Muestra Anual de Declarantes de IRPF con los nuevos factores de elevación poblacional calculados. El valor del índice se expresa en el rango 0 – 1.

El cómputo de este índice se realiza mediante el programa “ineqdec0” para Stata[®] (Jenkins, 2008).

- “Top 1%”

Porcentaje de la renta imponible agregada del municipio obtenida por el 1 por 100 de sus habitantes declarantes de IRPF con mayor renta imponible en el año de la estadística.

- “Top 0,5%”

Porcentaje de la renta imponible agregada del municipio obtenida por el 0,5 por 100 de sus habitantes declarantes de IRPF con mayor renta imponible en el año de la estadística.

- “Top 0,1%”

Porcentaje de la renta imponible agregada del municipio obtenida por el 0,1 por 100 (1 por 1000) de sus habitantes declarantes de IRPF con mayor renta imponible en el año de la estadística.

- “Quintil 1, 2, 3, 4, 5”

Porcentaje de la renta imponible agregada del municipio obtenida por los declarantes de IRPF que residen en el mismo pertenecientes, respectivamente, a los quintiles 1, 2, 3, 4 y 5 de la distribución de la renta imponible de la localidad.

Todos los anteriores porcentajes de distribución de la renta imponible de cada municipio (Top 1%, Top 0,5%, Top 0,1%, Quintil 1, 2, 3, 4 y 5) se han calculado con el programa Stata[®] versión 13, expandiendo la muestra con los nuevos factores poblacionales calculados con la metodología expuesta. Para el cómputo de los índices de desigualdad de Gini y Atkinson se han utilizado directamente los datos muestrales con los nuevos pesos analíticos calculados.

Referencias

- Arcarons, J., A. Castells, G. García y M. Parellada (1992). *Estimació de la renda familiar disponible a les comarques i municipis de Catalunya.1989*, Generalitat de Catalunya, Departament d’Economia i Finances, Barcelona.
- Alañón, A. (2002). “Estimación del valor añadido per cápita de los municipios españoles en 1991 mediante técnicas de econometría espacial”, *Ekonomiaz*, 51: 172-194.
- Atkinson, A. B. (1970). “On the measurement of inequality”, *Journal of Economic Theory*, 2: 244-263.
- Atkinson, A. B. (2007). “Measuring Top Incomes: Methodological Issues”, en A. B. Atkinson and T. Piketty (eds.), *Top Incomes over the Twentieth Century*. Oxford, UK: Oxford University Press. Ch. 2, pp. 18-42.
- Atkinson, A. B. y T. Piketty (2007). *Top Incomes over the Twentieth Century*. Oxford, UK: Oxford University Press.
- Atkinson, A. B., T. Piketty y E. Saez (2011). “Top Incomes in the Long Run of History”, *Journal of Economic Literature*, 49 (1): 3-71.
- Bowley, A. L. (1914). “The British Super-Tax and the Distribution of Income”, *Quarterly Journal of Economics*, 28: 255–68.

- Buendía, J. D., M. Esteban y J. C. Sánchez de la Vega (2012). "Estimación de la renta bruta disponible municipal mediante técnicas de econometría espacial: Un ejercicio de aplicación", *Revista de Estudios Regionales*, (93): 119-142.
- Canberra Expert Group on Household Income Statistics (2001). *Final Report and Recommendations*. Ottawa: Canberra Expert Group on Household Income Statistics.
- Chasco, C. (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Consejería de Economía e Innovación tecnológica de la Comunidad de Madrid.
- Chasco, C. y F. López (2004). "Modelos de regresión espacio-temporales en la estimación de la renta municipal: el caso de la región de Murcia", *Estudios de Economía Aplicada*, 22-3:1-24
- Creedy, J. (2003). "Survey reweighting for tax microsimulation modelling". *New Zealand Treasury Working Paper Series*, 17/03.
- Creedy, J. e I. Tuckwell (2004). "Reweighting household surveys for tax microsimulation modelling: An application to the New Zealand household economic survey", *Australian Journal of Labour Economics* 7 (1): 71–88.
- De las Heras, A. (1992). "Un Modelo General de Estimación Indirecta de la Renta Familiar Disponible Municipal. Su Aplicación a la Comunidad Autónoma de Cantabria". *Tesis Doctoral*. Universidad de Cantabria.
- De las Heras, A. y C. Murillo (1998). "Información fiscal y estimación indirecta de la renta familiar disponible municipal en España". I Congreso de Economía Aplicada. Barcelona.
- Deaton, A. (2003). "Household Surveys, Consumption, and the Measurement of Poverty", *Economic Systems Research*, 15 (2): 135-159.
- Deville, J. y C. Särndal (1992). "Calibration estimators in survey sampling", *Journal of the American Statistical Association*, 87: 376–382.
- Domínguez-Barrero, F., J. López-Laborda y F. Rodrigo (2014). "El hueco que deja el diablo': una estimación del fraude en el IRPF con microdatos tributarios", *Estudios sobre la Economía Española*, EEE2014-01, FEDEA.
- Esteban, J. y A. Pedreño (1992). "La Articulación Territorial de la Economía Valenciana", en J. A. Martínez-Serrano, A. Pedreño y E. Reig (dirs.), *Estructura Económica de la Comunidad Valenciana*. Madrid: Espasa Calpe. pp. 73-112.

- Feldman, N. E., y J. Slemrod (2007). "Estimating tax noncompliance with evidence from unaudited tax returns", *The Economic Journal*, 117 (518): 327-352.
- Fernández, C. y Y. Sierra (1992). "Estimación de la Renta Familiar Disponible a Nivel Municipal. Una Aplicación a La Rioja. Año 1985". Actas de la VI Reunión Asepelt España. Granada.
- Gini, C. (1912) "Variabilità e mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche", *Studi Economico-Giuridici dell' Università di Cagliari*, 3, (2): 1-158.
- Haslett, S., G. Jones, A. Noble y D. Ballas (2010). "More of Less? Comparing small area estimation, spatial microsimulation, and mass imputation", *JSM 2010 Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association. pp. 1584-1598.
- Herrero, L. C. (1998). "Perspectivas de Desarrollo Territorial: Renta Municipal y Desarrollo Económico en las Comarcas de Castilla y León". Junta de Castilla y León, Consejería de Economía y Hacienda.
- Hortas-Rico, M, Onrubia, J. y Pacifico, D. (2014) "Estimating the Personal Income Distribution in Spanish Municipalities using Tax Micro-Data", *Working Paper*, 14-19, International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.
- Hurst, E., G. Li y B. Pugsley (2010). "Are Household Surveys like Tax Forms: Evidence from Income Underreporting of the Self Employed," *NBER Working Paper*, no.16527.
- Jenkins, S. (2008). <http://fmwww.bc.edu/repec/bocode/i/ineqdeco.html>
- Kuznets, S. (1953). *Shares of Upper Income Groups in Income and Savings*. National Bureau of Economic Research, New York.
- Kuznets, S. (1955). "Economic growth and income inequality", *American Economic Review*, 65: 1-28.
- Leigh, A. (2007). "How Closely Do Top Income Shares Track Other Measures of Inequality?", *The Economic Journal*, 117 (524): F619-F633.
- Lohmann, H. (2011). "Comparability of EU-SILC survey and register data: The relationship among employment, earnings and poverty", *Journal of European Social Policy*, 21 (1): 1-18.

- Meyer, B. y J. X. Sullivan (2011). "Viewpoint: Further Results on Measuring the Well-Being of the Poor using Income and Consumption", *Canadian Journal of Economics*, 44 (1): 52-87.
- Oliver, J., I. Busó y J. Trullén (1995). *Estimació de la renda familiar disponible per càpita de Barcelona, els seus districtes i els 27 municipis de la Corporació Metropolitana de Barcelona*, Ajuntament de Barcelona, Barcelona.
- Paulus, A. (2013). "Tax Evasion and Measurement Error: A Comparison of Survey Income with Tax Records", 69th Congress of the International Institute of Public Finance, Taormina, Italy, 22-25, August 2013.
- Picos, F. (2006). "Microsimulación mediante fusión de Phogues y Panel de Declarantes para evaluar reformas fiscales". *Revista de Economía Aplicada*, 41: 33-60.
- Picos, F., C. Pérez, y M. C. González (2011). "La muestra de declarantes de IRPF de 2007: descripción general y principales magnitudes", *Documentos de Trabajo del Instituto de Estudios Fiscales*, 01/11.
- Piketty, T. y E. Saez (2003). "Income inequality in the United States, 1913–1998", *The Quarterly Journal of Economics*, 118 (1): 1-41.
- Ravallion, M. (2003). "Measuring aggregate welfare in developing countries: how well do national accounts and surveys agree?", *Review of Economics and Statistics*, 85 (3): 645-652.
- Remírez-Prados, J. A. (1991). *Una estimación de la renta familiar disponible a nivel municipal. El caso de Navarra*. Madrid: Confederación Española de Cajas de Ahorro.
- Särndal, C. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33 (2): 99–119.
- Slemrod, J. y C. Weber (2012). "Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy", *International Tax and Public Finance*, 19 (1): 25-53.
- Stamp, J. C., Lord (1914). "A New Illustration of Pareto's Law", *Journal of the Royal Statistical Society*, 77: 200–204.
- Stamp, J. C., Lord (1916). *British Incomes and Property*. London: P. S. King.