

DISEÑO, CONTENIDO Y APLICACIONES DEL NUEVO PANEL DE DECLARANTES DE IRPF 1999-2007*

JORGE ONRUBIA FERNÁNDEZ

Universidad Complutense de Madrid

FIDEL PICOS SÁNCHEZ

REDE-Universidad de Vigo

El nuevo Panel de Declarantes de IRPF 1999-2007 es un panel expandido representativo de la población declarante en cada año. Su diseño estadístico (con selección muestral amplia, estratificada por Comunidad Autónoma, nivel de renta y fuente mayoritaria, y afijación de mínima varianza con error inferior al 1%) asegura estimaciones muy precisas de las principales magnitudes del impuesto. Esto le convierte en una base de microdatos muy adecuada para el análisis de la distribución y redistribución de la renta, la evaluación de reformas fiscales y la microsimulación. Los objetivos de este artículo son: a) presentar las características técnicas del Panel construido, mostrando sus potencialidades para la investigación económica aplicada, destacando su complementariedad con otras bases utilizadas en este ámbito de estudios; b) ofrecer una comparación con otras bases de microdatos fiscales disponibles en España; y c) analizar su fiabilidad en estimaciones poblacionales de renta, base liquidable y cuota líquida.

Palabras clave: Panel de declarantes de IRPF, microdatos tributarios, registros administrativos.

Clasificación JEL: C81, C83, H24.

Inicialmente, los estudios empíricos sobre desigualdad y distribución de la renta se realizaban utilizando microdatos procedentes de las encuestas sobre el consumo de las familias. La implantación de encuestas específicas sobre los ingresos percibidos por los miembros de los hogares, diferenciando sus fuentes de procedencia, supuso un salto cualitativo para la investigación so-

(*) Los autores agradecen la financiación recibida del Instituto de Estudios Fiscales y el inestimable apoyo de César Pérez, responsable de su Área Estadística. Jorge Onrubia agradece además el soporte financiero recibido del MICINN a través del Proyecto ECO2009-10003 del Programa Nacional de Proyectos de Investigación Fundamental (VI Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011), así como la financiación recibida por el Grupo Consolidado de Investigación 940392 del Programa Santander-UCM 2010-2011 GR35/10A. Este trabajo debe mucho a la ayuda inestimable en el manejo de los datos de Mari Carmen González Queija, Carmen Gallego y Sara Huete. Los autores agradecen los comentarios y sugerencias realizados por dos evaluadores anónimos y por el editor, que han contribuido a mejorar la versión inicialmente enviada. Los errores son exclusivamente imputables a los autores.

cioeconómica y, en concreto, para la economía aplicada. De hecho, hasta la aparición a mediados de la década de los noventa del Panel de Hogares de la Unión Europea (PHOGUE), las Encuestas de Presupuestos Familiares (EPF), primero las decenales de 1980-81 y 1990-91 y posteriormente las continuas de base 1985, de referencia anual, constituían la principal fuente de información en España para estos estudios¹. La implantación del PHOGUE, con sus ocho olas de 1994 a 2001 y con ingresos detallados de los distintos miembros del hogar, permitió la realización de un considerable número de trabajos de investigación que analizaban la distribución de la renta personal y familiar, incluso de su redistribución a través de distintos programas de gastos e ingresos públicos². La sustitución a partir del año 2004 del PHOGUE por la Encuesta de Condiciones de Vida (ECV) ha permitido disponer de una nueva base de microdatos con información detallada de los ingresos personales de los hogares económicos, homogénea con las realizadas en la Unión Europea³. Este conjunto de bases de microdatos se ha visto ampliado con la Encuesta Financiera de las Familias, realizada por el Banco de España con carácter trienal (olas de 2002, 2005 y 2008), la cual incluye información sobre los activos en los que se materializa el ahorro de las familias españolas, el detalle de su endeudamiento, e información sobre su consumo y sus principales fuentes de ingresos.

Paralelamente, en los países más desarrollados de la OCDE, especialmente en los países escandinavos y anglosajones, existe una larga tradición en la confección de bases de microdatos con información procedente de registros de las administraciones públicas⁴. Entre éstas, ocupan un lugar destacado las elaboradas a partir de registros de carácter fiscal, algunas de ellas desde hace más de medio siglo como es el caso del *Individual Income Tax Return Panel* del *Internal Revenue Service* (IRS) estadounidense⁵. En estos países, la utilización de estas bases de datos es habitual tanto en el análisis empírico de la distribución y redistribución de la renta como en la construcción de modelos de microsimulación. Existe, además, en los países con mayor experiencia en este tipo de bases fiscales, un amplio consenso en considerar éstas como complementarias de las encuestas, utilizándolas a menudo en proyectos de fusión estadística con otras bases de microdatos (*statistical matching*).

(1) Las EPF son hoy en día fundamentales para el análisis del comportamiento de los consumidores, así como en la construcción de modelos de microsimulación dirigidos a evaluar reformas impositivas u otros programas de gasto que afecten a las cestas de consumo de los hogares (Rodríguez-Méndez, 2009). Desde 1997, la EPF mejoró sustancialmente su representatividad anual, dejándose de realizar las EPF decenales. En 2006 se definió una nueva EPF, con mayor tamaño muestral y número de variables.

(2) Fuenmayor y Granell (2009) hacen un inventario de los principales trabajos de microsimulación fiscal que han utilizado el PHOGUE como fuente de información. Por su parte, Andrés y Mercader (2001) estudian la fiabilidad de los datos de renta en el PHOGUE.

(3) La ECV es realizada por el Instituto Nacional de Estadística dentro del Proyecto *European Statistics on Income and Living Conditions* (EU-SILC). En Adiego *et al.* (2009) se revisan sus potencialidades para el análisis de la distribución y redistribución de la renta y su utilización en microsimulación.

(4) En Marcos (2007) se ofrece un amplio panorama nacional e internacional de la utilización de registros administrativos para la elaboración de bases de microdatos.

(5) Weber (2004).

Centrándonos en el ámbito de los microdatos de origen fiscal, una experiencia pionera en España fue la construcción, a finales de la década de los ochenta, del primer Panel de Declarantes de IRPF del Instituto de Estudios Fiscales (IEF), integrado por declaraciones de este impuesto aportadas por la Administración Tributaria. Inicialmente, este Panel abarcó los años de 1982 a 1994, siendo prolongado en 2005 hasta el ejercicio 1998. Su explotación fue bastante amplia, fundamentalmente en trabajos de índole hacendística, aunque algo condicionada por la intermitencia con la que se difundieron sus olas desde 1994 [Ayala *et al.* (2004)].

El creciente interés por la explotación de estos registros hizo que, a mediados de la década pasada, el IEF retomase la actividad de construcción de bases de microdatos fiscales, suscribiendo para ello un convenio de cesión de información con la Agencia Estatal de Administración Tributaria (AEAT), en el que el IEF actúa como receptor legal de la misma. El primer fruto de esta colaboración fue la creación de una Muestra Anual de Declarantes de IRPF, cuya primera entrega, el ejercicio 2002⁶, vio la luz en 2005, sucediéndose con continuidad hasta el ejercicio 2007, último disponible⁷.

Un segundo proyecto, más complejo, ha sido la construcción de un nuevo Panel de Declarantes de IRPF que diese continuidad al anterior. Este nuevo Panel que abarca actualmente los ejercicios de 1999 a 2007, construido por los autores en un proyecto de colaboración con la Unidad Estadística del IEF, se encuentra a disposición de los investigadores y analistas desde mediados de 2011⁸. El tiempo transcurrido desde la elaboración del primer Panel ha posibilitado la incorporación en el nuevo de importantes avances en su diseño estadístico, así como de mejoras de estructura y disposición de la información contenida. Asimismo, la generalización de la confección electrónica de las declaraciones de IRPF, junto con los avances en su registro informático, que prácticamente eliminan los errores de cálculo y las inconsistencias entre los datos declarados, han permitido mejorar su fiabilidad informativa.

Esta nueva base de microdatos se nutre, al igual que en el anterior Panel, de la información contenida en las declaraciones de IRPF extraídas por la AEAT a partir del diseño muestral facilitado desde la Unidad Estadística del IEF. Una vez recibida esta información “en bruto”, la construcción del Panel ha supuesto: a) la depuración de los errores identificados en los ficheros originales de datos; b) el diseño de una estructura interrelacional de ficheros, incluyendo la creación de dos nuevos, adicionales para un mejor tratamiento de los datos por los usuarios; c) la creación de una unidad observacional alternativa –el hogar fiscal–, con el cómputo de los correspondientes factores de elevación; y d) el establecimiento e implementación de un método de seguimiento longitudinal tanto de los declarantes como de los hogares fiscales en los que éstos se integran. En este proceso de construcción del Panel se han creado también variables adicionales, con el fin de facilitar su utilización.

(6) Véase Picos *et al.* (2005).

(7) Véase Picos *et al.* (2011).

(8) El nuevo Panel, lo mismo que las Muestras Anuales de Declarantes de IRPF, se facilitan gratuitamente mediante solicitud al IEF (<http://www.ief.es>).

Las características técnicas de esta nueva base de microdatos, unidas a la elevada capacidad de procesamiento de los ordenadores personales actuales, confieren al nuevo Panel de Declarantes de IRPF 1999-2007, a nuestro juicio, una gran potencialidad para la investigación económica, no sólo en el ámbito de la economía de la imposición, sino en general para cualquier estudio socioeconómico que requiera conocer la evolución de la renta personal a lo largo del tiempo, tanto desde una perspectiva nacional como territorial. En este sentido, el objetivo principal de este trabajo es la presentación por los autores de las características técnicas del Panel creado, con el propósito de mostrar a los investigadores y usuarios en general sus potencialidades para la investigación económica aplicada. El artículo incluye además una comparación con la otra base de microdatos fiscales del IEF disponible en la actualidad, las Muestras Anuales de IRPF 2002-2007.

El trabajo se organiza de la siguiente manera. En la segunda sección se exponen de forma sintética las especificaciones de su diseño estadístico, así como el proceso diseñado para la extracción anual de las muestras, de acuerdo con el carácter longitudinal del panel. La tercera sección detalla las características del panel, la organización de los ficheros anuales, así como la descripción de las variables incluidas. La sección cuarta recoge un análisis de fiabilidad poblacional de las principales variables del Panel, mientras que la sección quinta analiza la utilidad del panel en relación con otras bases de microdatos. Finalmente, la sección sexta ofrece una síntesis de conclusiones.

1. DISEÑO Y MUESTREO ESTADÍSTICO

El nuevo Panel de Declarantes de IRPF abarca en su actual configuración los ejercicios comprendidos entre 1999 y 2007. El arranque en el año 1999 da continuidad al anterior panel 1982-1998. Además el año 1999 fue el primer ejercicio de implantación de la Ley 40/1998, que reformó integralmente el IRPF, tras las vigencia del modelo previo, el de la Ley 18/1991, vigente entre 1992 y 1998. En cuanto al año que cierra actualmente el Panel, 2007, es el primer ejercicio de aplicación del modelo de IRPF actualmente vigente, establecido tras la reforma de la Ley 35/2006. En este momento, está previsto que el Panel siga ampliándose con sucesivos ejercicios a medida que la AEAT facilite las muestras de declaraciones, al menos hasta que su desgaste (*attrition*) deje de ser asumible.

Con el fin de mantener su representatividad para cada año, el Panel se ha diseñado como un panel expandido. Así, en cada ejercicio, además de las observaciones cuyos declarantes ya lo fueron en años anteriores, se dispone de una submuestra adicional extraída aleatoriamente, con un doble objetivo: uno, compensar las bajas producidas respecto del ejercicio anterior; y dos, añadir una representación de las altas habidas en la población declarante. De este modo conseguimos que el conjunto de observaciones de cada corte transversal sea representativo de la población declarante de cada año. Paralelamente, el subconjunto integrado exclusivamente por aquellas observaciones correspondientes a declarantes que lo han sido en todos los años del Panel (panel puro) o en determinados años consecutivos posibilita el trabajo longitudinal.

La población representada en el Panel está integrada por los declarantes del IRPF en los ejercicios comprendidos entre 1999 y 2007. Hay que precisar que la unidad observacional no son los declarantes, sino las declaraciones presentadas por éstos. Esta diferencia radica en el número de declarantes que puede contener una declaración, de acuerdo con la normativa del IRPF: un único declarante en el régimen de tributación individual y dos en el caso de matrimonios que optan por el régimen de tributación conjunta⁹.

Para la selección de las declaraciones se ha utilizado el método de muestreo aleatorio con tres niveles de estratificación: a) Comunidad Autónoma (las quince de régimen fiscal común más las Ciudades Autónomas de Ceuta y Melilla); b) fuente mayoritaria de la renta de muestreo, diferenciando entre rentas del trabajo y el resto de rentas; y c) nivel de la renta de muestreo en una escala de diez tramos. Aunque existen 340 estratos teóricos ($17 \times 2 \times 10$), sólo 323 tienen observaciones, al haberse considerado arbitrariamente que las observaciones del primer nivel (rentas negativas y ceros) nunca tienen como fuente mayoritaria las rentas del trabajo¹⁰. La renta de muestreo utilizada está constituida por la suma de los ingresos íntegros del trabajo, de los rendimientos procedentes del resto de fuentes de renta y de las imputaciones de rentas. Se ha optado por no incluir las ganancias y pérdidas patrimoniales dado su carácter esporádico.

Para determinar el número de declaraciones que deben integrar la muestra de cada año, en primer lugar se ha procedido a clasificar todas las declaraciones registradas en la AEAT en cada uno de los h estratos, determinándose así su tamaño poblacional (N_h). A continuación se obtuvo en cada estrato la cuasivarianza poblacional de la renta de muestreo (S_h^2). Con estos valores, facilitados por la AEAT, mediante el criterio de afijación de mínima varianza de Neyman se estableció el tamaño de la muestra (n), para un error relativo de muestreo admitido ($e < 0,001$) y un nivel de confianza adicional ($\alpha = 0,001$). A partir de los valores de n , N_h y S_h^2 se determinó el número de observaciones a extraer aleatoriamente en cada estrato (n_h), tal que $\sum_h n_h = n$. El cuadro 1 recoge los tamaños muestrales total y por CC.AA.

Determinados estos tamaños óptimos, la AEAT procedió a la extracción de las muestras, comenzando por 2003, el año base¹¹. La extracción para este año fue realizada por muestreo aleatorio de las n_h observaciones muestrales en cada uno de los h estratos, asignándole a cada observación muestral su factor de elevación w_h , tal que, $w_h = N_h / n_h$. Para los años posteriores, las n_h observaciones de cada estrato se extrajeron en dos fases. En la primera, se buscaron en cada estrato

(9) El régimen de tributación conjunta también es de aplicación a las unidades familiares monoparentales, formadas por la madre o el padre solteros, viudos o separados junto con sus hijos menores de edad o incapacitados judicialmente. En este caso, el Panel no recoge identificación personal de los descendientes, mientras que en el caso de matrimonios en tributación conjunta sí figuran identificados ambos cónyuges.

(10) La asignación es arbitraria porque al ser la renta total nula o negativa, el cociente entre la renta de trabajo y la renta total deja de ser interpretable.

(11) Esta elección del año base atiende a que se trata del primer ejercicio en el que la AEAT grabó datos de las casillas de tres dígitos de los impresos de declaración, coincidiendo además con el primer año de aplicación de la reforma del impuesto aprobada por la Ley 46/2002.

Cuadro 1: TAMAÑOS MUESTRALES POR CC.AA.

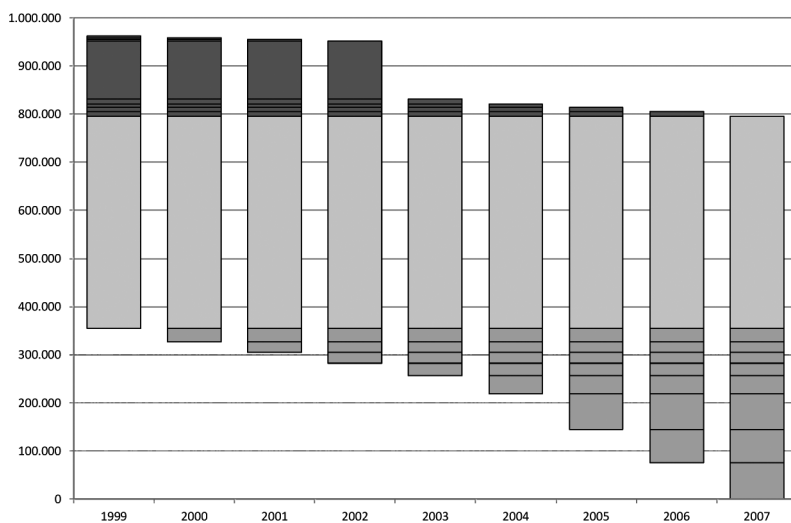
Comunidad Autónoma	1999	2000	2001	2002	2003	2004	2005	2006	2007
1. Andalucía	60.020	63.093	66.194	69.751	62.873	69.061	76.319	81.834	87.082
2. Aragón	14.642	15.003	15.285	15.622	13.420	14.257	14.749	15.336	15.944
3. Principado de Asturias	10.182	10.595	11.001	11.432	10.156	10.643	11.063	11.424	11.865
4. Islas Baleares	9.248	9.704	10.114	10.523	8.662	9.161	10.312	11.611	12.606
5. Canarias	13.672	14.542	15.433	16.461	15.006	16.298	18.518	20.586	22.540
6. Cantabria	5.336	5.621	5.907	6.187	5.504	5.923	6.967	7.997	8.922
7. Castilla-La Mancha	15.399	16.025	16.621	17.371	15.660	17.127	20.444	23.095	25.731
8. Castilla y León	26.006	26.836	27.522	28.421	24.573	26.211	29.933	34.042	37.379
9. Cataluña	78.817	82.233	85.051	87.723	73.352	77.393	87.591	97.502	106.155
10. Extremadura	8.406	8.881	9.148	9.530	8.806	9.612	11.240	13.413	15.192
11. Galicia	21.613	22.789	24.037	25.522	22.699	24.456	25.510	26.769	28.043
12. Comunidad de Madrid	70.950	74.304	77.549	80.504	68.615	72.055	82.940	91.025	99.454
13. Región de Murcia	9.953	10.539	11.189	11.991	10.388	11.325	13.134	15.533	17.590
16. La Rioja	3.507	3.585	3.690	3.780	3.085	3.283	3.920	4.414	4.896
17. Comunidad Valenciana	41.675	43.411	45.001	46.915	39.239	42.307	48.827	55.261	61.307
18. Ceuta	628	682	706	723	640	661	700	782	834
19. Melilla	559	596	592	634	559	578	626	678	771
Total	390.613	408.439	425.040	443.090	383.237	410.351	462.793	511.302	556.311

Fuente: Elaboración propia.

aquellas declaraciones cuyo declarante (o declarantes) figurase entre las declaraciones extraídas en el año anterior. En la segunda fase, y hasta completar los tamaños n_h prefijados, se extrajeron aleatoriamente declaraciones entre las altas de ese año. Para los años anteriores a 2003, el proceso fue simétrico: se buscaron las declaraciones de los declarantes seleccionados en el año posterior, mientras que la expansión se formó por declaraciones de declarantes que no lo hubiesen sido en el año posterior.

El resultado de este proceso de muestreo se ilustra en el gráfico 1, donde la franja central representa a los declarantes que figuran en todos los años del Panel (panel puro), entre un máximo del 70% y un mínimo del 53% de los incluidos en la muestra del año respectivo. Por su parte, las diferencias entre las franjas superiores de dos años consecutivos recogen la expansión anual correspondiente, mientras que las existentes entre las franjas inferiores de dos años consecutivos representan las bajas producidas¹².

Gráfico 1: PANEL DURO Y EXPANSIONES DEL PANEL (NÚMERO DE DECLARANTES)



Fuente: Elaboración propia.

(12) En cuanto a la bajas, vemos que un número importante de declarantes de 2002 dejaron de serlo en 2003. Esto es debido a que la reforma del IRPF de este año modificó al alza el umbral de renta que obliga a presentar declaración. Asimismo, las expansiones a partir de 2003 son todas muy importantes, explicándose por el gran crecimiento del número de declarantes habido hasta 2007.

Cuadro 2: PRINCIPALES CARACTERÍSTICAS DEL PANEL 1999-2007 EN RELACIÓN CON EL PANEL 1982-1998 Y LAS MUESTRAS ANUALES

	Panel 1982-1998	Panel 1999-2007	Muestras Anuales
Periodo	1982-1998	1999-2007	2002-2007
Tipo de microdatos	Panel expandido	Panel expandido	Muestras anuales sin vinculación, entre años, de declarantes
Población representada	Declaraciones de IRPF	Declaraciones de IRPF	Declaraciones de IRPF y no obligados-no declarantes de IRPF que soportaron retenciones por rendimientos del trabajo
Ámbito geográfico	CC.AA. de territorio fiscal común más las dos Ciudades Autónomas	CC.AA. de territorio fiscal común más las dos Ciudades Autónomas	CC.AA. de territorio fiscal común más las dos Ciudades Autónomas, más declaraciones de no residentes (art. 10 Ley 35/2006)
Unidad observacional	Unidades familiares, pero con posibilidad de construir unidades familiares fiscales	Declaración, pero con posibilidad de construir unidades familiares fiscales	Declaración
Tipo de muestreo	Proporcional (2%) para cada Delegación de Hacienda, por unidades familiares	Estratificado con afijación de mínima varianza, por declaraciones	Estratificado con afijación de mínima varianza, por declaraciones
Variables de estratificación	<ul style="list-style-type: none"> • Delegación de Hacienda 	<ul style="list-style-type: none"> • Nivel de renta (10 tramos) • Comunidad autónoma (15 de Territorio Fiscal Común más las Ciudades Autónomas de Ceuta y Melilla) • Fuente de renta (las rentas de trabajo representan más o menos de la mitad del total de rentas de muestreo) 	<ul style="list-style-type: none"> • Nivel de renta (12 tramos) • Provincia (46 de Territorio Fiscal Común más las Ciudades Autónomas de Ceuta y Melilla, más los residentes en el extranjero que son contribuyentes por art. 10 Ley 35/2006) • Modalidad de declaración (individual o conjunta)
Renta de muestreo	-	Suma de ingresos íntegros del trabajo personal, rendimientos netos e imputaciones de rentas	Suma de rendimientos netos, imputaciones de rentas, y parte especial de la renta del periodo (base imponible del ahorro en 2007)

Fuente: Elaboración propia.

El proceso de muestreo seguido no garantiza la incorporación de los dos cónyuges de un matrimonio cuando éstos optan por presentar declaraciones separadas. Para solventar este problema y contar con información completa de todos los matrimonios, en cada año se ha extraído un fichero complementario integrado por las declaraciones individuales de aquellos cónyuges pertenecientes a matrimonios en los que la declaración del otro cónyuge resultó seleccionada en el proceso de muestreo¹³. Asimismo, hay que señalar que no es posible vincular las declaraciones de miembros de parejas no casadas ni conocer cuándo varias unidades familiares conviven en el mismo hogar, al no existir en ambos casos vinculación fiscal.

El diseño del nuevo Panel presenta diferencias con el anterior Panel, como respecto de las Muestras Anuales de Declarantes, tal y como se puede ver en el cuadro 2¹⁴.

De las características señaladas en el cuadro, quizá la diferencia más significativa entre el panel antiguo y el nuevo es la unidad de muestreo: la unidad familiar en el antiguo frente a la declaración en el nuevo. Hasta 1988, al ser obligatoria la declaración conjunta para matrimonios, ambos procesos son coincidentes, pero a partir de entonces, tras la sentencia del Tribunal Constitucional que dispuso la inconstitucionalidad de la obligación de tributación conjunta, el muestreo por unidades familiares obligaba a agrupar previamente las declaraciones de los cónyuges que optaban por tributar separadamente, si bien para configurar el fichero de observaciones ambas declaraciones se incluían separadamente. En cambio, en el Panel nuevo el muestreo se realiza directamente por declaraciones, y sólo una vez finalizado el proceso se crea un fichero adicional de cónyuges para los matrimonios que eligieron tributar en el régimen individual. Por su parte, las principales diferencias de las Muestras Anuales respecto al Panel son: su carácter estrictamente transversal, la ausencia de identificación de cualquier tipo en las observaciones y la falta de un fichero adicional de cónyuges, características que, como se verá más adelante, limitan en algunos aspectos su potencialidad respecto al panel, aunque la superan en otros.

Por último, el cuadro 3 recoge los tamaños muestrales de los paneles y de las Muestras Anuales. Como se ve, los dos paneles presentan cifras con un orden de magnitud similar, mientras que las Muestras, debido a la inclusión de la provincia como nivel de estratificación, presentan magnitudes mucho mayores, en torno al doble de los cortes transversales del panel nuevo.

(13) Existe, no obstante, un número importante de casos (entre un 3% y un 25%, mayor cuanto más lejos del año base) en el que el muestreo aleatorio ha conducido a la extracción de las declaraciones de los dos cónyuges, integrándose ambas en el fichero de principales.

(14) Descripciones detalladas de cada una de las Muestras pueden encontrarse en la colección de Documentos de Trabajo del IEF. Para la Muestra más reciente, la del ejercicio 2007, véase Picos *et al.* (2011). Sobre el Panel 1982-1998 véase Ayala *et al.* (2004).

Cuadro 3: TAMAÑOS MUESTRALES DE LOS PANELES Y LAS MUESTRAS ANUALES DE IRPF

Año	Panel 1982-1998	Panel 1999-2007	Muestras Anuales 2002-2007
1982	123.599		
1983	130.500		
1984	134.957		
1985	145.664		
1986	165.303		
1987	173.979		
1988	193.444		
1989	208.808		
1990	235.646		
1991	251.197		
1992	277.733		
1993	287.291		
1994	313.116		
1995	325.039		
1996	310.859		
1997	308.736		
1998	308.558		
1999		390.613	
2000		408.439	
2001		425.040	
2002		443.090	907.399
2003		383.237	936.678
2004		410.351	941.029
2005		462.793	952.578
2006		511.302	964.489
2007		556.311	1.351.802

Fuente: Elaboración propia.

2. CARACTERÍSTICAS Y CONTENIDO BÁSICO

El Panel dispone para cada año de dos ficheros bidimensionales: uno, denominado genéricamente fichero de “principales”, que contiene las declaraciones extraídas a través del proceso de muestreo descrito, y un fichero complementario de “conyuges”, con las declaraciones de cónyuges que no resultaron seleccionadas en el proceso de muestreo. La estructura de ambos ficheros es idéntica, ofreciéndose

para cada observación una serie de variables cuya fuente de información es, directa o indirectamente, el impreso de declaración de IRPF del año correspondiente.

Por su naturaleza, dividimos las variables en dos grupos: no monetarias, comunes a todos los cortes transversales del Panel y que recogen la información personal y cualitativa de cada declaración; y monetarias, que recogen los datos que figuran en las casillas del impreso de declaración anual de IRPF. Mientras la mayoría de las variables cuantitativas se ofrecen tal y como fueron proporcionadas por la AEAT, algunas de las variables no monetarias han sido creadas por los autores para enriquecer la información del Panel, así como, en otros casos, para facilitar la utilización del Panel por los usuarios. El cuadro 4 recoge estas variables, señalándose con un asterisco las creadas *a posteriori* por los autores.

En lo que se refiere a las variables no monetarias, el nuevo Panel proporciona información cualitativa más completa que el antiguo, pero menos completa que las Muestras. Éstas proporcionan información “tratada” de las declaraciones (número de descendientes, ascendientes, posibles minusvalías) que en el Panel sólo puede ser inferida a partir de las cuantías de los mínimos personales y familiares y de algunas reducciones¹⁵, y además proporcionan información cualitativa sobre actividades económicas (tipo de actividad y epígrafes del IAE, desde 2004) y sobre inmuebles (tenencia de vivienda, imputaciones de rentas, etc., desde 2005). El cuadro A.1 del anexo compara las variables disponibles para el Panel nuevo y las Muestras.

En cuanto a las variables monetarias (última fila del cuadro 4), su número está condicionado por el detalle de la información grabada por la AEAT. Así, hasta 2002 el panel nuevo ofrece un grado de información similar al antiguo, mientras que a partir de 2003, debido a la grabación de casillas de tres dígitos por parte de la AEAT, las cifras se duplican. En comparación con las Muestras, éstas proporcionan en general más variables que el Panel, aunque las diferencias no son significativas salvo en el año 2002. El cuadro A.2 del anexo compara el número de variables de los dos paneles y las Muestras.

En cuanto a la estructura del nuevo Panel, debemos comenzar diciendo que la principal limitación de los ficheros de “principales” y “conyuges” originalmente extraídos por la AEAT es que no proporcionan información directa sobre la pertenencia de cada declaración a una unidad familiar concreta. Tampoco, en su configuración original, estos ficheros permiten realizar de forma sencilla un seguimiento de observaciones a lo largo del tiempo. En buena medida, estos problemas se deben a que la unidad observacional elegida, la declaración de IRPF, no es estable a lo largo del tiempo, ya que los declarantes en condiciones de formar unidades familiares pueden decidir, en cada año, modificar su opción de tributar individual o conjuntamente.

(15) Un método sencillo para obtener el número de hijos o ascendientes en el Panel se basa en cruzar los valores de los mínimos y reducciones correspondientes, que figuran en el Panel, con los de la Muestra anual para un año determinado, teniendo en cuenta el valor asignado en esta última para esa variable “tratada”. Aunque las Muestras sólo están disponibles desde 2002, la estabilidad normativa del IRPF entre 1999 a 2002 permite utilizar este ejercicio de 2002 para los años comprendidos entre 1999 y 2002.

Cuadro 4: VARIABLES INCLUIDAS EN LOS FICHEROS DE “PRINCIPALES” Y DE “CONYUGES”

Tipo	Nombre	Descripción campo	Formato y/o valores
Númérico	<i>ndecla*</i>	Número identificativo único de la declaración	–
Númérico	<i>nhogar*</i>	Número del hogar al que pertenece la declaración	–
Númérico	<i>nppal</i>	Numero interno del primer declarante	–
Númérico	<i>ncyge</i>	Numero interno del cónyuge o del segundo declarante	–
Alfanumérico	<i>expediente</i>	Número interno de expediente	–
Númérico	<i>municipio</i>	Código de municipio de residencia	–
Númérico	<i>dh</i>	Delegación de Hacienda	–
Númérico	<i>CC.AA.</i>	Comunidad Autónoma	–
Númérico	<i>tramo</i>	Tramo de renta	1 = Negativas y cero 2 = De 0,01 a 3.000 3 = De 3.000,01 a 6.000 4 = De 6.000,01 a 12.000 5 = De 12.000,01 a 18.000 6 = De 18.000,01 a 30.000 7 = De 30.000,01 a 60.000 8 = De 60.000,01 a 120.000 9 = De 120.000,01 a 240.000 10 = De 240.000,01 en adelante
Númérico	<i>marcfuente</i>	Fuente mayoritaria de renta	0 = rentas del trabajo (variable <i>ingresotrabajo</i>) representan el 50% o menos del total de rentas (variable <i>rentamuestreo</i>) 1 = rentas del trabajo representan más del 50% del total de rentas
Númérico	<i>factor</i>	Factor de elevación	–
Númérico	<i>fechappal</i>	Fecha de nacimiento del primer declarante	Formato AAAAMMDD
Númérico	<i>sexoppal</i>	Sexo del primer declarante	1 = hombre 2 = mujer

Cuadro 4: VARIABLES INCLUIDAS EN LOS FICHEROS DE “PRINCIPALES” Y DE “CONYUGES” (continuación)

Tipo	Nombre	Descripción campo	Formato y/o valores
Numérico	<i>falleceppal</i>	Fecha de fallecimiento del primer declarante	Formato AAAAMMDD
Numérico	<i>sexocgye</i>	Sexo del cónyuge o del segundo declarante	1 = hombre 2 = mujer
Numérico	<i>fallececyge</i>	Fecha de fallecimiento del cónyuge o del segundo declarante	Formato AAAAMMDD
Alfanumérico	<i>tipotributacion</i>	Tipo de tributación	1 = declaración individual C = declaración conjunta
Numérico	<i>categoriadecla*</i>	Categoría de la declaración	1 = Individuos no casados 2 = Familias monoparentales 3 = Matrimonios que presentan declaración conjunta 4 = Matrimonios que presentan declaraciones separadas
Numérico	<i>minimosdecla*</i>	Mínimos aplicados en la declaración	1 = Mínimo personal individual, mínimo descendientes = 0 2 = Mínimo personal individual, mínimo descendientes > 0 3 = Mínimo personal monoparental, mínimo descendientes > 0 4 = Mínimo personal conjunta 9 = Casos no identificados
Numérico	<i>incluirdecla*</i>	Variable que indica si la declaración debe incluirse a la hora de trabajar sin las duplicidades derivadas de errores de muestreo	0 = No debe incluirse 1 = Debe incluirse
Numérico	<i>rentamuestreo*</i>	Renta de muestreo de la declaración	Céntimos de euro
Numérico	<i>ingresostrabajo*</i>	Ingresos íntegros de trabajo de la declaración	Céntimos de euro
Numérico	<i>cXXX</i>	Valores de las casillas del impreso	Céntimos de euro

Nota: (*) Variables creadas *a posteriori*, no proporcionadas por la AEAT.

Fuente: Elaboración propia.

Con el fin de superar esta limitación, en el proceso de construcción del Panel hemos definido dos unidades de observación alternativas a la declaración: el hogar fiscal y el declarante. El concepto de hogar fiscal es identificable con el de unidad familiar definida en la normativa del IRPF. Por consiguiente, el concepto de hogar fiscal es equivalente al de declaración para los declarantes no casados y para los casados que realizan declaración conjunta, mientras que difiere en el caso de los casados que realizan dos declaraciones individuales, donde el hogar fiscal está formado por la unión de las declaraciones de ambos. De esta manera estamos en condiciones de superar la diferencia existente en el Panel entre matrimonios en función de su régimen de tributación, irrelevante en términos económicos para muchos análisis.

Por su parte, entendemos por declarante a la persona física que declara IRPF en calidad de contribuyente. Este concepto coincide con el de declaración para los individuos no casados y para los casados que hacen declaración individual. Sin embargo, a cada declaración conjunta de individuos casados le corresponden dos declarantes. El cuadro 5 resume las equivalencias entre las unidades observacionales utilizadas.

Cuadro 5: EQUIVALENCIAS ENTRE UNIDADES OBSERVACIONALES EN EL PANEL

Situación	Número de declaraciones	Número de hogares	Número de declarantes
Individuo no casado	1	1	1
Matrimonio que realiza declaración conjunta	1	1	2
Matrimonio que realiza declaración individual	2	1	2

Fuente: Elaboración propia.

Una vez que hemos definido estas dos nuevas unidades observacionales, ya resulta posible realizar un seguimiento de las mismas a lo largo del tiempo. El seguimiento de un declarante resulta inmediato: basta que su NIF figure como declarante de IRPF, ya sea en una declaración individual o conjunta. A partir de esta información, hemos creado una variable de seguimiento con nueve posiciones, una por cada año del Panel, tomando cada posición los valores uno o cero según exista o no declaración de ese declarante ese año. Por ejemplo, si esta variable toma el valor “001110111”, está informando de que el declarante figura en el Panel en los años 2001, 2002, 2003, 2005, 2006 y 2007. Las observaciones con valor “111111111” identifican el subconjunto de panel puro.

De forma paralela realizamos un seguimiento de hogares fiscales, a través de la identificación por el NIF de un declarante soltero, o por los dos NIF de sus componentes concatenados, en el caso de declarantes casados. Consideramos que

uno de estos hogares se mantiene en el tiempo cuando mantiene su estado civil, siendo irrelevante el tipo de declaración –individual o conjunta– que presente. Esto implica que un declarante soltero forma el mismo hogar tanto si realiza declaración individual como si pasase a realizar declaración conjunta monoparental, mientras que un matrimonio es el mismo hogar fiscal tanto si se opta por realizar dos declaraciones individuales como si presenta sólo una conjunta. A partir de esta información creamos una variable de seguimiento, con idéntica estructura a la creada para el seguimiento de declarantes.

Las unidades observacionales alternativas que hemos creado se recogen en dos ficheros adicionales: un fichero maestro de “hogares” y un fichero maestro de “declarantes”. El primero recoge los hogares disponibles en el año correspondiente (uno por fila), indicándose qué declaraciones forman parte del mismo e incluyéndose la variable de seguimiento y las variables adicionales que identifican el tipo de hogar, además de su factor de elevación, calculado por los autores con información poblacional proporcionada por la Agencia Tributaria. Paralelamente, el fichero maestro de “declarantes” recoge los individuos que declaran en el año correspondiente (uno por fila), indicando para cada declarante en qué declaraciones figura su NIF, además de la variable de seguimiento.

El cuadro 6 y el cuadro 7 ofrecen el diseño de registro de los ficheros maestros de “hogares” y “declarantes”, respectivamente, mientras que el cuadro 8 recoge para cada año el número de observaciones incluidas en cada uno de los cuatro ficheros¹⁶.

El gráfico 2 muestra esquemáticamente la estructura final del Panel y las variables contenidas en cada fichero. Como puede verse, esta estructura que hemos creado permite establecer enlaces entre todos los ficheros. Esto es factible mediante la numeración de todas las observaciones en las diversas variables de enlace. Así, a cada observación de los ficheros básicos de “principales” y “cónyuges” le hemos asignado un número único de declaración (*ndecla*), que no se repite de un año para otro; uno o dos NIF (*nppal* y *ncyge*), según sean declaraciones de no casados o de casados, respectivamente; y el número de hogar al que pertenecen (*nhogar*). Por su parte, cada observación del fichero maestro de “hogares” ha sido identificada por la variable *nhogar*, y se indica, para cada año, qué declaraciones (*ndecla*) forman parte de ese hogar (variables *ndecla1* a *ndecla4*¹⁷). Paralelamente, las observaciones del fichero maestro de “declarantes” se identifican mediante la variable *ndeclarante*, que contiene los NIF de las personas que declaran en cada año, obtenidas a través de las variables *nppal* y *ncyge* de los ficheros de principales y cónyuges. Igual que en el fichero maestro de hogares, las variables *ndecla1* a *ndecla4* indican las declaraciones de ese año en las que aparecen los correspondientes declarantes.

(16) Para una descripción detallada del proceso de creación de ficheros y variables véase Onrubia *et al.* (2011).

(17) Teóricamente en cada hogar no puede haber más de dos declaraciones, pero se crean cuatro variables porque en los tres primeros años se han encontrado casos en los que un mismo NIF figuraba en varias declaraciones, aunque nunca en más de cuatro. Estos errores representaban el 0,17% en 1999, el 0,35% en 2000 y el 0,02% en 2001.

Cuadro 6: VARIABLES INCLUIDAS EN EL FICHERO MAESTRO DE “HOGARES”

Tipo	Nombre	Descripción campo	Formato y/o valores
Numérico	<i>nhogar*</i>	Número identificativo único del hogar	–
Numérico	<i>factorhogar*</i>	Factor de elevación del hogar	–
Numérico	<i>estructurahogar*</i>	Estructura relacional del hogar	Cuatro dígitos con formatos 1XXX, 3XXX, 40XX, 45XX y 9999
Numérico	<i>ccaa</i>	Comunidad Autónoma	Véase Cuadro 4
Numérico	<i>marcafuelle</i>	Tramo de renta	Véase Cuadro 4
Numérico	<i>tramo</i>	Fuente mayoritaria de renta	Véase Cuadro 4
Numérico	<i>tipohogar*</i>	Tipo de hogar fiscal	1 = Individuos no casados 2 = Familias monoparentales 3 = Matrimonios que presentan declaración conjunta 4 = Matrimonios que presentan declaraciones separadas 9 = Tipos no identificados
Numérico	<i>declaporphog*</i>	Número de declaraciones que debe haber por cada hogar	1 = una declaración 2 = dos declaraciones 9 = desconocido
Alfanumérico	<i>seguimientohog*</i>	VARIABLE que indica en qué años figura el hogar	Nueve dígitos que toman valor 0 o 1
Numérico	<i>ndecla1*</i>	Número de la primera declaración que forma parte del hogar	–
Numérico	<i>ndecla2*</i>	Número de la segunda declaración que forma parte del hogar	–
Numérico	<i>ndecla3*</i>	Número de la tercera declaración que forma parte del hogar	–
Numérico	<i>ndecla4*</i>	Número de la cuarta declaración que forma parte del hogar	–

Nota: (*) Variables creadas *a posteriori*, no proporcionadas por la AEAT.

Fuente: Elaboración propia.

Cuadro 7: VARIABLES INCLUIDAS EN EL FICHERO MAESTRO DE “DECLARANTES”

Tipo	Nombre	Descripción campo	Formato y/o valores
Númérico	<i>ndeclarante*</i>	Número identificativo único del declarante	Toma valores de las casillas <i>nppal</i> y <i>ncyge</i>
Alfanumérico	<i>seguimientodeclarante*</i>	Variable que indica en qué años figura el declarante	Nueve dígitos que toman valor 0 o 1
Númérico	<i>ndecla1*</i>	Número de la primera declaración en la que figura el declarante	–
Númérico	<i>ndecla2*</i>	Número de la segunda declaración en la que figura el declarante	–
Númérico	<i>ndecla3*</i>	Número de la tercera declaración en la que figura el declarante	–
Númérico	<i>ndecla4*</i>	Número de la cuarta declaración en la que figura el declarante	–

Nota: (*) Variables creadas *a posteriori*, no proporcionadas por la AEAT.

Fuente: Elaboración propia.

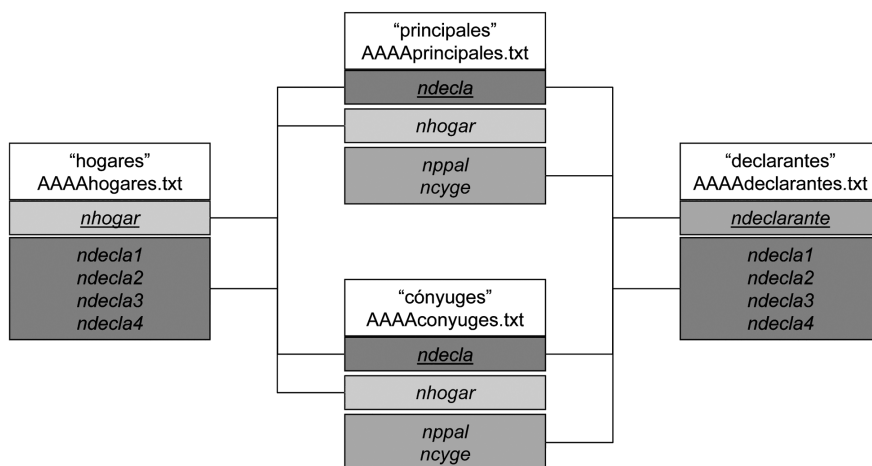
Cuadro 8: NÚMERO DE OBSERVACIONES EN CADA FICHERO DEL PANEL

	Principales	Cónyuges	Hogares	Declarantes
1999	390.613	89.796	366.467	624.564
2000	408.439	101.563	385.394	660.791
2001	425.040	114.266	405.517	690.424
2002	443.090	129.645	429.169	724.432
2003	383.237	119.490	378.830	625.594
2004	410.351	113.297	396.618	656.124
2005	462.793	116.109	440.143	721.473
2006	511.302	118.325	478.692	777.020
2007	556.311	120.858	514.952	827.958

Fuente: Elaboración propia.

Si comparamos esta estructura con la del panel antiguo, aquella presentaba un esquema mucho más simple. Como hemos visto, no existía más que un único fichero de declaraciones. Por lo que respecta a las Muestras Anuales, en la medida que se trata de cortes transversales totalmente independientes entre años, sin vinculación entre declaraciones de cónyuges, su estructura consta también de un único fichero.

Gráfico 2: ESTRUCTURA DE FICHEROS DEL PANEL, UNIDADES DE OBSERVACIÓN Y VARIABLES DE ENLACE

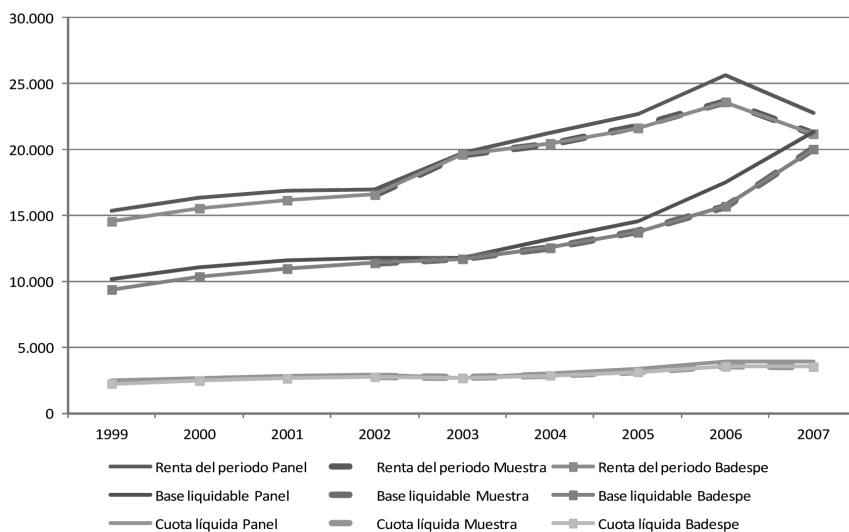


Fuente: Elaboración propia.

3. FIABILIDAD

En esta sección realizamos un breve análisis de la representatividad poblacional del nuevo Panel de Declarantes de IRPF, examinando en qué medida logra replicar las principales magnitudes del IRPF, y comparando los resultados con los proporcionados por las Muestras Anuales. Para ello escogemos tres variables básicas: la renta gravable del periodo, la base liquidable y la cuota líquida del impuesto. Se trata de tres conceptos con una definición fiscal clara y homogénea en las normativas aplicables del IRPF. En el gráfico 3 comparamos con los datos de Badespe¹⁸ tanto los valores medios estimados con el Panel como con las Muestras Anuales.

Gráfico 3: VARIABLES FISCALES MEDIAS ESTIMADAS CON EL PANEL Y LAS MUESTRAS ANUALES



Fuente: Elaboración propia.

El gráfico muestra con claridad cómo las variables analizadas se comportan de forma similar a lo largo del tiempo en las tres fuentes comparadas. Hay que tener en cuenta que la evolución no se explica sólo en función del crecimiento económico del periodo, sino también de las reformas del IRPF producidas en dos de los años analizados, 2003 y 2007. Así, la reforma de 2003 provocó un fuerte incremento de la renta media (ya que, por una parte, se modificaron al alza de los

(18) Base de Datos del Sector Público Español, disponible en <http://www.estadief.meh.es/>.

umbrales de tributación y, por otra, se trasladó a la base liquidable la reducción por rendimientos de trabajo), a la vez que generaba una ligera reducción de la cuota líquida (ya que implicó un menor gravamen para todos los contribuyentes, lo que compensó incluso el fuerte incremento de la renta media). Por su parte, la reforma de 2007 volvió a situar la reducción de trabajo en el cómputo particular de estas rentas (lo que provoca la reducción observada de la renta media), cambiando además la definición de la base imponible, que ya no aparece minorada por el mínimo personal y familiar (lo que explica el fuerte incremento producido respecto a 2006, pese a la reducción de la renta media en el mismo periodo)¹⁹.

En cuanto a la capacidad de representación poblacional, el gráfico revela también cómo las Muestras Anuales ajustan perfectamente las magnitudes a los datos reales. En cambio, en el caso del Panel sí observamos algunas diferencias visibles con los datos de Badespe, salvo en el año base, 2003. Tal y como cabía esperar, estas diferencias van aumentando al separarnos del año base, aunque las curvas mantienen las mismas tendencias. La explicación de estas divergencias es la propia de la construcción de paneles, incluso de tipo expandido, donde existe una pérdida de representatividad transversal impuesta por la necesidad de mantener una proporción suficiente de observaciones de panel puro para el análisis longitudinal.

Para analizar estas diferencias con más detalle, en el cuadro 9 desagregamos rentas y bases entre sus componentes general y especial, comparándolas con los datos de Badespe (base 100)²⁰.

En lo que se refiere a las rentas, la parte general presenta errores de estimación mucho menores que la especial, tanto en el Panel como en las Muestras Anuales, lo que se debe sin duda a que la segunda tiene una relevancia mucho menor en el impuesto, tanto en cuantía total como en número de perceptores, así como una mayor dispersión entre el total de declaraciones. No obstante, existen diferencias entre el Panel y las Muestras Anuales. En la renta general, la Muestra presenta desviaciones mínimas, que oscilan entre el 0,01% de 2006 y el 0,29% de 2005, mientras que en el Panel se observa el efecto del sesgo visto en el gráfico 3 a medida que nos alejamos del año base (en el que sólo hay una desviación del 0,47%), más se desvían las cifras de las reales (llegando al 4,43% en 1999). La parte general de la base liquidable se comporta de manera muy similar.

(19) El concepto de Renta del Periodo sólo existió así definido entre 2003 y 2006. No obstante, Badespe lo utiliza también en años anteriores y posteriores, mediante la oportuna homogeneización, recogiendo la suma de rendimientos netos reducidos, imputaciones de rentas y ganancias y pérdidas patrimoniales antes de la aplicación del mínimo personal y familiar. Como se ha visto, los datos de 1999-2002 y 2007 no son estrictamente comparables con los de 2003-2006, pero hemos respetado las definiciones de Badespe para poder hacer comparaciones. Tampoco es homogéneo entre periodos el concepto de Base Imponible, pero en este caso el concepto sí que existe todos los años. Para analizar la evolución de la renta gravable mediante la definición de una variable homogénea, véase Onrubia y Picos (2011).

(20) Dado que tanto las Muestras anuales como el Panel elevan a población con exactitud (es decir, la suma de los factores de elevación es idéntica a la población real), las comparaciones son válidas tanto para agregados como para medias.

Cuadro 9: VARIABLES FISCALES ESTIMADAS CON EL PANEL Y LAS MUESTRAS ANUALES (BADESPE = 100)

Año	Renta del periodo						Base liquidable						Couta líquida	
	Parte general		Parte especial*		Parte general		Parte especial*		Parte general		Parte especial*		Panel	Muestra
	Panel	Muestra	Panel	Muestra	Panel	Muestra	Panel	Muestra	Panel	Muestra	Panel	Muestra	Panel	Muestra
1999	104,43		132,46		106,64		133,42		109,25				109,25	
2000	103,75		131,34		105,02		131,92		108,16				108,16	
2001	103,19		131,74		104,17		132,01		107,34				107,34	
2002	101,95	99,89	122,76	103,28	102,47	99,93	122,72	103,46	105,07				105,07	100,19
2003	100,47	100,13	101,88	95,52	100,70	100,27	101,73	95,08	100,57				100,57	100,00
2004	101,92	100,03	146,69	95,79	102,10	100,07	147,53	95,76	105,66				105,66	100,09
2005	102,17	100,29	143,20	102,36	102,24	100,49	143,60	102,40	106,66				106,66	101,05
2006	102,66	100,01	163,61	99,77	102,68	100,03	164,16	99,78	110,53				110,53	100,02
2007	102,83	100,02	140,18	101,76	101,63	100,00	140,23	101,77	109,38				109,38	100,25

Nota: (*) En 2007, Base del Ahorro.

Fuente: Elaboración propia.

En cuanto a la renta especial, ésta oscila mucho más, siendo en todo caso las diferencias mucho menores en las Muestras Anuales (con un máximo del 4,48% en 2003) que en el Panel (más de un 63% en 2006), lo que se explica por el mayor tamaño muestral de las primeras, que captura mejor la gran variabilidad de este tipo de rentas. Se detecta además que, mientras que las desviaciones de las Muestras difieren de signo según los años (es decir, unos años son mayores que 100 y otros menores), las del Panel son siempre positivas (mayores que 100), y parecen mostrar un sesgo similar al de la renta general (aunque no tan claro entre 2004 y 2007 como entre 1999 y 2002). También en este caso la parte especial de la base liquidable se comporta de manera simétrica.

En lo referente a las cuotas líquidas, las diferencias entre Muestras Anuales y Panel se acentúan. Así, mientras que la Muestra mantiene una precisión muy similar a la mostrada en la estimación de rentas y cuotas (entre el 0% de 2003 y el 1,05% de 2005), el Panel presenta los sesgos ya explicados, pero agravados por el carácter progresivo del impuesto, que hace que los errores en las bases se transformen en errores mayores en las cuotas (hasta el 10,53% en 2006).

4. UTILIZACIÓN DEL PANEL: POSIBILIDADES Y LIMITACIONES

A nuestro juicio, el nuevo Panel de IRPF tiene un potencial grande de aplicación en diversos campos de la investigación económica, con especial orientación hacia el ámbito de la Economía Pública. La utilización de esta nueva base de microdatos vendría recomendada por una comparación de sus potencialidades y limitaciones respecto de otras bases, bien las tradicionales basadas en encuestas, bien respecto de otras bases de microdatos de carácter fiscal alternativas, como son las Muestras Anuales de IRPF.

Entre las diversas potencialidades de utilizar la información contenida en los registros administrativos fiscales para trabajos orientados hacia la distribución y redistribución de la renta o en el desarrollo de microsimuladores, destacan especialmente las asociadas al carácter censal de la población de referencia y el grado de detalle y precisión con el que suelen reflejarse las variables que se incluyen en estas bases. Su dependencia de la normativa tributaria les confiere, no obstante, algunas limitaciones, destacando las asociadas a la unidad observacional, distinta de la habitual definición de hogar económico usada en las encuestas, o las derivadas de la definición fiscal de las magnitudes²¹. No obstante, una regla recomendable es considerar, de forma previa a la elección, los objetivos de la investigación, teniendo en cuenta la adecuación del diseño informativo de la base de microdatos a los conceptos o definiciones de las magnitudes que deben utilizarse en el desarrollo del trabajo de investigación, así como la calidad, precisión y representatividad de esa información.

Centrándonos en el estudio de la distribución y redistribución de la renta como área amplia de investigación, esta comparación puede realizarse respecto de las en-

(21) En Wallgren y Wallgren (2007) puede consultarse una amplia revisión de estas potencialidades y sus implicaciones estadísticas.

cuestas disponibles, como la Encuesta de Condiciones de Vida, la Encuesta de Presupuestos Familiares -ambas realizadas por el INE- o la Encuesta Financiera de las Familias, realizada por el Banco de España. Así, vemos, en primer lugar, que el origen informativo del Panel proporciona gran precisión a la hora de reflejar los distintos componentes de la renta personal, si bien es cierto, que bajo definiciones derivadas de la normativa fiscal. En el caso de las encuestas, el grado de detalle depende de cada una de ellas. Así, mientras las EPF ofrecen una única cifra de renta del hogar, neta de impuestos y transferencias, la ECV detalla la composición de esa renta neta según se trata de rentas salariales, empresariales o del capital. La consideración en términos netos de los ingresos, es decir, después del pago de IRPF y de las cotizaciones sociales del empleado, introduce una limitación alternativa en las encuestas, pues exigen la elevación al importe íntegro de los mismos, con la dificultad que entraña dada la progresividad y ajustes de la carga fiscal del IRPF. Además, hay que tener en cuenta, frente al carácter censal de los registros administrativos, la infrarrepresentación y la falta de respuesta que son problemas habituales presentes en las encuestas. En cuanto a la riqueza informativa de sus datos, las bases de microdatos fiscales permiten realizar muchos análisis que están vedados a otras muestras de origen no fiscal, especialmente en el ámbito impositivo, dada la precisión con la que se definen los distintos componentes monetarios de la estructura del impuesto. Este hecho resulta crucial para la construcción de modelos de microsimulación impositiva, donde la liquidación de impuestos de estructura compleja como el IRPF, a partir de la información contenida en bases de microdatos como el PHOGUE o la ECV, o en la EFF, introducen sesgos importantes en las proyecciones generadas²².

Sin duda, la principal limitación de las bases de microdatos fiscales, que está presente en el nuevo Panel y en las Muestras Anuales de Declarantes de IRPF, es la unidad observacional considerada, distinta del concepto tradicional de hogar económico empleado por las encuestas. Aunque como hemos visto en las secciones anteriores de este trabajo, estas bases fiscales cuentan con información sobre descendientes y ascendientes, éstos siempre están vinculados a una unidad contribuyente, bien sea un matrimonio o un individuo no casado. En cambio, no es factible vincular distintos contribuyentes adultos cuando éstos no están casados, aunque exista convivencia. Sin embargo, esta limitación no tiene por qué ser invalidante, como puede verse en muchos de los trabajos que utilizan este tipo de bases de microdatos fiscales, tanto referidos a la realidad española como a la de otros países con bases disponibles similares. En unos casos, la definición de un concepto de hogar fiscal en el que agregar las rentas y cuotas impositivas de los cónyuges resulta suficiente para evaluar las distribuciones correspondientes, incluso pudiéndose aplicar al mismo escalas de equivalencia referidas a los adultos e hijos menores presentes en cada unidad familiar contribuyente. En otros casos, en los que resulta indispensable considerar el hogar económico como unidad de análisis, la utilización complementaria del Panel junto a encuestas como la ECV o la EPF, consideradas como bases receptoras en procesos de fusión estadística (*statistical matching*), cada vez más extendidos, permiten salvar esta limitación, sin renunciar a la riqueza informa-

(22) Véase Sanz *et al.* (2009).

tiva de los microdatos fiscales²³. En cualquier caso, defendemos aquí la idea cada vez más extendida de la complementariedad entre ambos tipos de bases de datos.

Por lo que respecta a la opción entre el nuevo Panel y las Muestras anuales, la elección depende fundamentalmente de dos limitaciones. La primera tiene que ver con la población representada. En el caso de las Muestras Anuales, a partir de 2002, éstas recogen un fichero complementario de personas que no estando obligadas a presentar declaración de IRPF (al estar sus rentas por debajo del umbral legal fijado) no lo hacen, aunque soportan retenciones a cuenta del impuesto. Este fichero de “no obligados-no declarantes”, que registra individualmente los distintos pagos retenidos ofrece información interesante sobre los contribuyentes de rentas bajas²⁴. La complejidad de la elaboración longitudinal del Panel imposibilita contar con un fichero análogo en esta base de datos, si bien es cierto que para el análisis transversal cuando se utiliza el Panel, resulta consistente, en cada año desde 2002, complementar el fichero de “principales” del Panel con el de “no obligados-no declarantes” de las Muestras. La segunda limitación se refiere a la unidad de análisis. Como hemos visto, las Muestras Anuales no proporcionan ninguna información que permita vincular a los cónyuges que declaran individualmente. Este hecho imposibilita la consideración de la renta de los matrimonios de forma homogénea, pues en el caso de tributación conjunta sí aparecen agregadas las rentas de ambos cónyuges. Por tanto, en aquellos trabajos en los que sea recomendable el análisis bajo una definición de hogar fiscal, la utilización del Panel se convierte en la mejor opción. En el ámbito del análisis impositivo, la simulación de formas alternativas de tratamiento de la unidad contribuyente, como el *income splitting* o el *quotient familial* no podrían simularse con las Muestras Anuales.

Una vez expuestas las potencialidades y limitaciones genéricas del Panel y de las Muestras Anuales de Declarantes de IRPF para su utilización en el análisis económico aplicado, a continuación ofrecemos una revisión sintética de las áreas más concretas donde estas bases de microdatos fiscales pueden ser de aplicación. Para ello seguimos la clasificación realizada en Ayala *et al.* (2004) en relación con la explotación del Panel anterior respecto a cuatro grandes tipos de trabajos de investigación: efectos de la aplicación del sistema impositivo, evaluación de propuestas de reforma fiscal, investigaciones de ámbito territorial, y estudios socioeconómicos no relacionados directamente con el sistema fiscal.

(23) Como ejemplo ilustrativo de este tipo de fusiones, puede verse la realizada por Picos (2006) entre el Panel de IRPF (año 1995) y el Panel de Hogares de la Unión Europea (ola de 1996, con rentas de los hogares de 1995). Un trabajo pionero en España, incluyendo una fusión estadística entre la EPF y el antiguo Panel de Declarantes de IRPF, fue el de Calonge y Manresa (2001).

(24) Estas personas, que hasta 1998 no representaban una cifra significativa, representan hoy en día una proporción importante de los contribuyentes. Esto es debido a que el sistema de retenciones está muy ajustado para eximir de la obligación de declarar a los perceptores las rentas de trabajo medias y bajas obtenidas de un solo pagador, aunque la existencia de deducciones, principalmente la de adquisición de vivienda habitual o la de alquiler, llevan a presentar declaración a un número importante de “no obligados”.

• *Análisis de los efectos de aplicación del sistema fiscal*

En este ámbito el diseño del Panel es ideal para realizar análisis longitudinales que impliquen el seguimiento de declarantes y de su comportamiento a lo largo del tiempo. En este sentido las Muestras no son útiles, al extraerse de manera independiente cada año, sin ninguna información identificativa de los declarantes²⁵. Si el análisis no exige el seguimiento de declarantes, entonces el uso del Panel puede presentar tanto ventajas como inconvenientes respecto del uso de las Muestras Anuales. En cuanto a las primeras, el Panel abarca un periodo mayor de tiempo (tres años más, incluyendo periodos completos de aplicación de modelos de reforma, como el incorporado en 1999), y además ofrece la importante ventaja de permitir construir unidades familiares fiscales para los casos en los que los cónyuges del matrimonio declaran por separado. Pese a que no es posible reconstruir hogares económicos, esta posibilidad representa un avance importante respecto a las Muestras, especialmente en los análisis de desigualdad y redistribución. Además esto es también atractivo para aquellos casos en los que se contemple la realización de fusiones estadísticas en las que la base de datos receptora tiene como unidad observacional el hogar económico, dado que la agrupación de rentas del matrimonio será fundamental.

El inconveniente principal del Panel, habitual en este tipo de bases de microdatos longitudinales, es la pérdida de representatividad a medida que nos alejamos del año base, 2003. Su diseño como panel expandido palia esta limitación, pero no la elimina. El primer problema es que excepto en el año base, en el que todas las observaciones se extraen aleatoriamente, en el resto de ejercicios todas las observaciones son seleccionadas con criterios no estrictamente aleatorios. Esto es debido a que primeramente se seleccionan aquellas observaciones que han sido muestreadas en el año anterior, y el resto son extraídas exclusivamente del subconjunto de la población que no presentó declaración en el año inmediato anterior. Para los ejercicios anteriores al año base, sucede lo mismo, pues primero se incorporan las observaciones de aquellos declarantes que fueron seleccionados en el año inmediatamente posterior, y la extensión se realiza a partir del subconjunto poblacional de declarantes que no lo fueron en el año siguiente. En ambos casos el proceso introduce sesgos, que afectan principalmente a las variables no utilizadas en la estratificación, ya que de alguna manera se “arrastra” hacia años anteriores y posteriores la distribución que estas variables tenían en el año base²⁶. No obstante, el sesgo también afecta a la variable renta utilizada en el muestreo, ya que los tamaños óptimos de cada estrato fueron calculados a partir de la cuasivarianza de esa variable para la población total de cada uno de ellos, pero al selec-

(25) No obstante, la reciente aplicación de técnicas de *pooling data* como el *propensity score matching* permiten utilizar en análisis longitudinales este tipo de Muestras transversales independientes, aunque con limitaciones respecto de las bases de datos que incluyen un subconjunto de panel puro (véase Caliendo y Kopeinig, 2008).

(26) Por ejemplo, si la proporción de declarantes varones fuese en 2003 mucho mayor que en 2007, el corte transversal de 2007 infrarrepresentaría a las mujeres, en la medida en que buena parte de las observaciones no habrían sido obtenidas aleatoriamente, sino como consecuencia de su inclusión previa en 2003.

cionarse sólo las observaciones que cumplan una serie de condiciones, el tamaño óptimo debería venir dado por las cuasivarianzas de esos subgrupos. Este problema no surge, obviamente, con la utilización de las Muestras Anuales, donde la ausencia de sesgos supone una mayor precisión en las estimaciones de variables.

- *Evaluación de propuestas de reforma fiscal*

En el campo de las reformas fiscales el Panel presenta, al igual que para el análisis del sistema fiscal actual, una importante ventaja: permite analizar los efectos de una reforma a lo largo del tiempo sobre los mismos individuos, así como estimar el comportamiento de los mismos ante el impuesto. Sin embargo, la utilización del Panel para este tipo de trabajos presenta también dos inconvenientes. El primero es que las predicciones de cambios recaudatorios y distributivos serán menos precisas que las obtenidas a partir de las Muestras, debido a los sesgos explicados en el apartado anterior. Y el segundo es que, tal y como se puede ver en el cuadro A.1 del Anexo, las Muestras presentan una gran cantidad de variables personales y familiares (descendientes, ascendientes y minusvalías) que permiten simular con facilidad cambios en el tratamiento de las circunstancias personales y familiares. En el Panel, por el contrario, no siempre pueden deducirse las características personales y familiares a partir de mínimos y reducciones (especialmente entre 1999 y 2002), y en los casos en los que se puede, esto requiere un trabajo de cómputo adicional.

- *Análisis de naturaleza territorial*

La consideración en los trabajos de investigación de la cuestión territorial también pone de manifiesto diferencias entre ambas bases de microdatos. El origen de las mismas radica en que la variable de estratificación escogida es diferente: la Comunidad Autónoma en el Panel y la provincia en las Muestras Anuales. Para estudiar cualquier fenómeno de ámbito autonómico o superior, ambas fuentes presentan la misma precisión, pero para ámbitos *territoriales* más reducidos, en principio serán preferibles las Muestras. Otra ventaja de las Muestras Anuales, aunque sólo desde 2004 en adelante, es que éstas presentan para cada observación el código postal de residencia del declarante, mientras que el Panel sólo ofrece un código de municipio específico de la AEAT, siendo por tanto menor la desagregación. Antes del ejercicio 2004, las Muestras sólo ofrecían la provincia, con lo cual para esos años la información geográfica está más desagregada en el Panel. En todo caso, debe señalarse que cualquier estimación para un nivel geográfico menor que el usado en la estratificación, presentará un cierto grado de error que puede ser calculado y en su caso corregido con métodos de estimación en áreas pequeñas²⁷.

- *Estudios socioeconómicos de carácter no fiscal*

En lo que se refiere a los estudios no relacionados directamente con el sistema fiscal, la principal variable de interés que aportan las bases con microdatos del IRPF es la renta, en sus distintas definiciones. Aunque no es posible calcular un

(27) Sobre estas técnicas de estimación véase Pfefferman (2002).

valor de renta que no venga condicionado en su definición por el sistema fiscal (debido fundamentalmente, a que, por una parte, no hay información sobre rentas exentas, y por otra, las rentas gravadas sólo lo son parcialmente y los datos no permiten en algunos casos reconstruir la renta íntegra original), sí es posible obtener una estimación razonable de la renta bruta de los declarantes²⁸. En este sentido, tanto el Panel como las Muestras Anuales ofrecen una cantidad de información similar, excepto en el año 2002, en el que la Muestra ofrece muchas más variables monetarias que el Panel (véase cuadro A.2: Número de variables monetarias en el Panel y las Muestras Anuales).

En relación con otras variables, las Muestras ofrecen información mucho más rica en tres aspectos, especialmente a partir del ejercicio 2005. El primero de ellos es la información personal y familiar, ya que las Muestras Anuales proporcionan una serie de variables que no se incluyen en el Panel, como estado civil, minusvalías y fechas de nacimiento de declarante principal y cónyuge, y sexo del declarante principal, además de información sobre número, edad y minusvalías de ascendientes y descendientes. De estas variables, como vimos, el Panel sólo ofrece la fecha de nacimiento del principal, si bien ofrece el sexo de ambos cónyuges (información no trivial a partir de 2006 por la aprobación del matrimonio homosexual) y, en su caso, las fechas de fallecimiento de principal y cónyuge. El segundo aspecto se refiere a los inmuebles, donde la Muestra proporciona, a partir de 2005, información detallada sobre la vivienda habitual y sobre el resto de inmuebles a disposición de los titulares. El Panel no ofrece información de este tipo, salvo la que indirectamente pueda derivarse de las imputaciones de rentas para propietarios que no arriendan sus viviendas y la deducción por inversión en vivienda habitual. La tercera ventaja de la Muestra Anual, también desde 2005, es la inclusión de información cualitativa sobre actividades económicas, que tampoco ofrece el Panel. En concreto, se proporciona información sobre el tipo de actividades realizadas (empresariales o profesionales) y la modalidad de declaración (normal o simplificada) en estimación directa, más el epígrafe de actividad correspondiente al Impuesto de Actividades Económicas (IAE), en este caso tanto para estimación directa como objetiva del rendimiento de las actividades empresariales.

5. CONCLUSIONES

La reciente puesta a disposición de los investigadores y analistas de un nuevo Panel de Declarantes de IRPF 1999-2007 viene a completar la oferta de microdatos fiscales disponible hasta ahora en España. Con el objetivo de familiarizar a los investigadores con el nuevo Panel, en el presente trabajo hemos realizado una presentación del mismo, describiendo sus características técnicas y estructura, mostrando su potencialidad para la investigación económica aplicada y analizando sus puntos fuertes y débiles, comparándolo con otras bases de microdatos fiscales con el anterior Panel de Declarantes de IRPF 1982-1998 y con las Muestras Anuales

(28) En Ayala y Onrubia (2001) se discute la relación entre la renta económica y la renta fiscal proporcionada por los de microdatos del IRPF.

de Declarantes de IRPF 2002-2007. También hemos tratado de poner en relación las posibilidades de utilización del nuevo Panel con el uso de bases de microdatos procedentes de encuestas, destacando, a nuestro juicio, su complementariedad.

Del análisis realizado cabe destacar dos conclusiones. La primera es que el nuevo Panel representa un importante avance respecto al anterior, debido básicamente a dos motivos. Uno, que el diseño muestral utilizado hace que su grado de representatividad sea mucho mayor. Y dos, que la creación *a posteriori* de ficheros complementarios de hogares y declarantes, así como de variables adicionales de caracterización y seguimiento de las observaciones, facilitan de forma importante el trabajo de los investigadores, especialmente en estudios de carácter longitudinal.

Una segunda conclusión es que el Panel, más que como una base de datos sustitutiva, se presenta como un buen complemento de las Muestras anuales de IRPF. Dependiendo de los objetivos buscados, será mejor utilizar una u otra fuente de datos. En términos generales, el Panel es una mejor opción para realizar análisis longitudinales que impliquen el seguimiento de individuos declarantes, bien sea para analizar su comportamiento ante cambios en el sistema impositivo o, más genéricamente, para analizar cambios producidos en la renta o en cualquier otro aspecto socioeconómico que esté reflejado en las variables proporcionadas por el Panel. Además, el Panel también permite realizar análisis por hogares fiscales (ya que permite reconstruir unidades familiares en los casos de matrimonios que optaron por declarar separadamente) y cubre un periodo de tiempo mayor que las Muestras (éstas no proporcionan datos para el periodo 1999-2001). Se trata sin duda de una ventaja muy relevante en el análisis económico aplicado, pues permite aproximarse mejor a la renta de las familias, si bien con las limitaciones propias de la naturaleza fiscal de la unidad familiar frente a la concepción del hogar económico utilizada habitualmente en las encuestas de tipo económico. No obstante, las posibilidades de desarrollar ejercicios de fusión estadística entre las bases fiscales y estas encuestas, hacen más recomendable contar con la posibilidad de agregar las rentas en todos los matrimonios.

Por su parte, las Muestras anuales presentan en general una mayor precisión, debida al muestreo independiente realizado para cada año. De este modo, son útiles para el desarrollo de ejercicios de microsimulación, además de incorporar directamente información personal y familiar de los declarantes. Asimismo, permiten la realización de análisis sectoriales que el Panel no permite, ya que incorporan información cualitativa sobre tenencia de vivienda y actividades económicas.

Para concluir, es necesario destacar la complementariedad que, a nuestro juicio, presenta el nuevo Panel en relación con otras fuentes de datos. Como ya se apuntó, la posibilidad de su combinación con otras bases de microdatos, bien procedentes de registros administrativos, bien procedentes de encuestas, permiten al investigador disponer de una información muy precisa en relación con la estructura y elementos del IRPF, además de una información alternativa sobre la renta personal. La generalización en el uso de fuentes de información de naturaleza administrativa, que como el Panel que aquí se ha presentado presentan elevados tamaños muestrales –gracias a sus relativos bajos costes de extracción– dependerá en buena medida, como sucede, por ejemplo, en los países nórdicos, de su creciente disponibilidad, facilitada por las distintas Administraciones Públicas. Un siguiente reto es hacer realidad la posibilidad de vinculación individual entre distintas fuentes administrativas, lo que sin duda ofrecerá un enorme potencial informativo.

ANEXO

Cuadro A.1: VARIABLES NO MONETARIAS EN EL PANEL Y EN LAS MUESTRAS ANUALES

Variable	Panel	Muestras								
		2002	2003	2004	2005	2006	2007			
Identificación										
Número de declaración	<i>ndecla</i>									
Número de hogar	<i>nhogar</i>									
Número del declarante principal	<i>nppal</i>									
Número del cónyuge	<i>ncyge</i>									
Número de expediente	<i>expediente</i>									
Factor de elevación	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>	<i>factor</i>
Características de la declaración										
Tipo de tributación	<i>tipotributacion</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>	<i>dec</i>
Tipo de impreso	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>	<i>limodelo</i>
Categoría de declaración	<i>categoriadecla</i>									
Tipo de mínimos personal y familiar aplicado	<i>minimosdecla</i>									
Declaración duplicada	<i>incluirdecla</i>									
Renta										
Tramo de renta	<i>tramo</i>								<i>tramorenta</i>	
Tramo de renta agrupado (*)									<i>tramosrenta</i>	
Fuente principal de renta	<i>marcafuelle</i>									
Localización geográfica										
Comunidad Autónoma	<i>CC.AA.</i>								<i>ca</i>	
Provincia									<i>prov</i>	<i>prov</i>
Delegación de hacienda	<i>dh</i>								<i>prov</i>	<i>prov</i>
Municipio	<i>municipio</i>									
Código postal									<i>cdpost</i>	<i>cdpost</i>

(*) Clasificación en tres tramos: valores negativos, positivos entre 0 y 60.000 euros, y mayores de 60.000 euros, solo disponible para 2005.

Cuadro A.1: VARIABLES NO MONETARIAS EN EL PANEL Y EN LAS MUESTRAS ANUALES (continuación)

Variable	Panel	Muestras						
		2002	2003	2004	2005	2006	2007	
Características personales de declarante y cónyuge								
Estado civil		<i>estcv</i>	<i>estcv</i>	<i>estcv</i>	<i>estcv</i>	<i>estcv</i>	<i>estcv</i>	<i>estcv</i>
Sexo del declarante principal	<i>sexppal</i>	<i>sexo</i>	<i>sexo</i>	<i>sexo</i>	<i>sexo</i>	<i>sexo</i>	<i>sexo</i>	<i>sexo</i>
Sexo del cónyuge	<i>sexocyge</i>							
Fecha de nacimiento del declarante principal	<i>fechappal</i>	<i>ejnacd</i>	<i>ejnacd</i>	<i>ejnacd</i>	<i>ejnacd</i>	<i>ejnacd</i>	<i>ejnacd</i>	<i>ejnacd</i>
Fecha de nacimiento del cónyuge		<i>ejnacc</i>	<i>ejnacc</i>	<i>ejnacc</i>	<i>ejnacc</i>	<i>ejnacc</i>	<i>ejnacc</i>	<i>ejnacc</i>
Grado de minusvalía del declarante		<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>
Grado de minusvalía del cónyuge		<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>	<i>minusc</i>
Fecha de fallecimiento del declarante principal	<i>falleceppal</i>							
Fecha de fallecimiento del cónyuge	<i>fallececyge</i>							
Descendientes								
Número total de descendientes		<i>nmdesc</i>	<i>nmdesc</i>	<i>nmdesc</i>	<i>nmdesc</i>	<i>nmdesc</i>	<i>nmdesc</i>	<i>nmdesc</i>
Número de descendientes <3 años		<i>nmdesc0</i>	<i>nmdesc0</i>	<i>nmdesc0</i>	<i>nmdesc0</i>	<i>nmdesc0</i>	<i>nmdesc0</i>	<i>nmdesc0</i>
Número de descendientes >= 3 y < 16 años		<i>nmdesc3</i>	<i>nmdesc3</i>	<i>nmdesc3</i>	<i>nmdesc3</i>	<i>nmdesc3</i>	<i>nmdesc3</i>	<i>nmdesc3</i>
Número de descendientes >= 16 y < 18 años		<i>nmdesc1618</i>	<i>nmdesc1618</i>	<i>nmdesc1618</i>	<i>nmdesc1618</i>	<i>nmdesc1618</i>	<i>nmdesc1618</i>	<i>nmdesc1618</i>
Número de descendientes >= 18 y < 25 años		<i>nmdesc1825</i>	<i>nmdesc1825</i>	<i>nmdesc1825</i>	<i>nmdesc1825</i>	<i>nmdesc1825</i>	<i>nmdesc1825</i>	<i>nmdesc1825</i>
Número de descendientes >= 25 años		<i>nmdescr</i>	<i>nmdescr</i>	<i>nmdescr</i>	<i>nmdescr</i>	<i>nmdescr</i>	<i>nmdescr</i>	<i>nmdescr</i>

Cuadro A.1: VARIABLES NO MONETARIAS EN EL PANEL Y EN LAS MUESTRAS ANUALES (continuación)

Variable	Panel	Muestras					
		2002	2003	2004	2005	2006	2007
Número de descendientes con edad desconocida		<i>nmdescd</i>	<i>nmdescd</i>	<i>nmdescd</i>	<i>nmdescd</i>	<i>nmdescd</i>	<i>nmdescd</i>
Número de descendientes sin minusvalía		<i>nmdesm0</i>	<i>nmdesm0</i>	<i>nmdesm0</i>	<i>nmdesm0</i>	<i>nmdesm0</i>	<i>nmdesm0</i>
Número de descendientes con minusvalía > 0 y < 33 %		<i>nmdesm33</i>					
Número de descendientes con minusvalía >= 33 y < 65 % sin movilidad reducida		<i>nmdesm65</i>	<i>nmdesm65</i>	<i>nmdesm65</i>	<i>nmdesm65</i>	<i>nmdesm65</i>	<i>nmdesm65</i>
Número de descendientes con minusvalía >= 33 y < 65 % con movilidad reducida		<i>nmdesm65mr</i>	<i>nmdesm65mr</i>	<i>nmdesm65mr</i>	<i>nmdesm65mr</i>	<i>nmdesm65mr</i>	<i>nmdesm65mr</i>
Número de descendientes con minusvalía >= 65 %		<i>nmdesm</i>	<i>nmdesm</i>	<i>nmdesm</i>	<i>nmdesm</i>	<i>nmdesm</i>	<i>nmdesm</i>
Número de descendientes con minusvalía		<i>nmdiscd</i>	<i>nmdiscd</i>	<i>nmdiscd</i>	<i>nmdiscd</i>	<i>nmdiscd</i>	<i>nmdiscd</i>
Número de ascendientes	Ascendientes	<i>nmasc</i>	<i>nmasc</i>	<i>nmasc</i>	<i>nmasc</i>	<i>nmasc</i>	<i>nmasc</i>
Número de ascendientes con minusvalía		<i>nmdisca</i>	<i>nmdisca</i>	<i>nmdisca</i>	<i>nmdisca</i>	<i>nmdisca</i>	<i>nmdisca</i>
Número de ascendientes > 65 años		<i>nmnm65a</i>	<i>nmnm65a</i>	<i>nmnm65a</i>	<i>nmnm65a</i>	<i>nmnm65a</i>	<i>nmnm65a</i>
Número de ascendientes > 75 años		<i>nmnm75a</i>	<i>nmnm75a</i>	<i>nmnm75a</i>	<i>nmnm75a</i>	<i>nmnm75a</i>	<i>nmnm75a</i>
Número de ascendientes sin minusvalía		<i>nmascm0</i>	<i>nmascm0</i>	<i>nmascm0</i>	<i>nmascm0</i>	<i>nmascm0</i>	<i>nmascm0</i>
Número de ascendientes con minusvalía >= 33 y < 65 % sin movilidad reducida		<i>nmascmh65</i>	<i>nmascmh65</i>	<i>nmascmh65</i>	<i>nmascmh65</i>	<i>nmascmh65</i>	<i>nmascmh65</i>
Número de ascendientes con minusvalía >= 33 y < 65 % con movilidad reducida		<i>nmascmh65mr</i>	<i>nmascmh65mr</i>	<i>nmascmh65mr</i>	<i>nmascmh65mr</i>	<i>nmascmh65mr</i>	<i>nmascmh65mr</i>
Número de descendientes con minusvalía >= 65%		<i>nmascmr</i>	<i>nmascmr</i>	<i>nmascmr</i>	<i>nmascmr</i>	<i>nmascmr</i>	<i>nmascmr</i>

Cuadro A.1: VARIABLES NO MONETARIAS EN EL PANEL Y EN LAS MUESTRAS ANUALES (continuación)

Variable	Panel	Muestras					
		2002	2003	2004	2005	2006	2007
Vivienda habitual	Titularidad				c/2	c/2	c/2
	% participación primer declarante				c/3	c/3	c/3
	% participación cónyuge				c/4	c/4	c/4
Inmuebles a disposición del contribuyente (hasta seis)	Contribuyente titular				p/40	p/40	p/40
					p/46	p/46	p/46
					p/52	p/52	p/52
				p/58	p/58	p/58	p/58
				p/64	p/64	p/64	p/64
				p/70	p/70	p/70	p/70
				p/41	p/41	p/41	p/41
				p/47	p/47	p/47	p/47
				p/53	p/53	p/53	p/53
				p/59	p/59	p/59	p/59
				p/65	p/65	p/65	p/65
				p/71	p/71	p/71	p/71
Uso				p/44	p/44	p/44	
				p/50	p/50	p/50	
				p/56	p/56	p/56	
				p/62	p/62	p/62	
				p/68	p/68	p/68	
				p/74	p/74	p/74	

Fuente: Elaboración propia.

Cuadro A.1: VARIABLES NO MONETARIAS EN EL PANEL Y EN LAS MUESTRAS ANUALES (continuación)

Variable	Panel	Muestras					
		2002	2003	2004	2005	2006	2007
Actividades económicas (hasta seis)	Tipo de actividad/es realizada/s				<i>par101_1</i>	<i>par101_1</i>	<i>par101_1</i>
	en estimación directa				<i>par101_2</i>	<i>par101_2</i>	<i>par101_2</i>
					<i>par101_3</i>	<i>par101_3</i>	<i>par101_3</i>
					<i>par101_4</i>	<i>par101_4</i>	<i>par101_4</i>
					<i>par101_5</i>	<i>par101_5</i>	<i>par101_5</i>
					<i>par101_6</i>	<i>par101_6</i>	<i>par101_6</i>
Epígrafe(s) IAE en estimación directa					<i>par102_1</i>	<i>par102_1</i>	<i>par102_1</i>
					<i>par102_2</i>	<i>par102_2</i>	<i>par102_2</i>
					<i>par102_3</i>	<i>par102_3</i>	<i>par102_3</i>
					<i>par102_4</i>	<i>par102_4</i>	<i>par102_4</i>
					<i>par102_5</i>	<i>par102_5</i>	<i>par102_5</i>
					<i>par102_6</i>	<i>par102_6</i>	<i>par102_6</i>
Modalidad aplicable para la determinación del rendimiento neto en estimación directa					<i>par103_4_1</i>	<i>par103_4_1</i>	<i>par103_4_1</i>
					<i>par103_4_2</i>	<i>par103_4_2</i>	<i>par103_4_2</i>
					<i>par103_4_3</i>	<i>par103_4_3</i>	<i>par103_4_3</i>
					<i>par103_4_4</i>	<i>par103_4_4</i>	<i>par103_4_4</i>
					<i>par103_4_5</i>	<i>par103_4_5</i>	<i>par103_4_5</i>
					<i>par103_4_6</i>	<i>par103_4_6</i>	<i>par103_4_6</i>
Epígrafe IAE en estimación objetiva (excepto agrícolas y ganaderas)					<i>par151_1</i>	<i>par151_1</i>	<i>par151_1</i>
					<i>par151_2</i>	<i>par151_2</i>	<i>par151_2</i>
					<i>par151_3</i>	<i>par151_3</i>	<i>par151_3</i>
					<i>par151_4</i>	<i>par151_4</i>	<i>par151_4</i>
					<i>par151_5</i>	<i>par151_5</i>	<i>par151_5</i>
					<i>par151_6</i>	<i>par151_6</i>	<i>par151_6</i>

Fuente: Elaboración propia.

Cuadro A.2: NÚMERO DE VARIABLES MONETARIAS
EN EL PANEL Y LAS MUESTRAS ANUALES

Año	Panel	Muestras Anuales
1999	113	
2000	108	
2001	108	
2002	121	200
2003	237	199
2004	262	256
2005	282	267
2006	291	275
2007	291	295

Fuente: Elaboración propia.



REFERENCIAS BIBLIOGRÁFICAS

- Adiego, M., Méndez, J.M. y M. Paniagua (2009): “La Encuesta de Condiciones de Vida (ECV)”, en F. Picos y S. Díaz de Sarralde (coord.), *Las reformas fiscales bajo el microscopio. Microsimulación fiscal en España: datos, metodología y aplicaciones*. Madrid, Instituto de Estudios Fiscales, págs. 25-38.
- Andrés, L. y M. Mercader (2001): “Sobre la fiabilidad de los datos de renta en el Panel de Hogares de la Unión Europea (PHOGUE, 1994)”, *Estadística Española*, nº 43, págs. 241-280.
- Ayala, L. y J. Onrubia (2001): “La distribución de la renta en España según datos fiscales”, *Papeles de Economía Española*, nº 88, págs. 89-112.
- Ayala, L., Onrubia, J. y J. Ruiz-Huerta (2004): “Modelos de microsimulación: aplicaciones a partir del Panel de Declarantes por IRPF del Instituto de Estudios Fiscales”, *Cuadernos Económicos de ICE*, nº 68, págs. 111-137.
- Caliendo, M. y S. Kopeinig (2008): “Some practical guidance for the implementation of Propensity Score Matching”, *Journal of Economic Surveys*, vol. 22, nº 1, págs. 31-72.
- Calonge, S. y A. Manresa (2001): *Incidencia fiscal y del gasto público social sobre la distribución de la renta en España y sus CC. AA.* Bilbao: Fundación BBVA.
- Fuenmayor, A. y R. Granell (2009): “Registros fiscales y encuestas: ventajas e inconvenientes”, en F. Picos y S. Díaz de Sarralde (coord.), *Las reformas fiscales bajo el microscopio. Microsimulación fiscal en España: datos, metodología y aplicaciones*. Madrid, Instituto de Estudios Fiscales, págs. 39-52.
- Marcos, C. (dir.) (2007): *El papel de los registros administrativos en el análisis social y económico y el desarrollo del sistema estadístico*. Madrid: Instituto de Estudios Fiscales.
- Onrubia, J. y F. Picos (2011): “Comportamiento recaudatorio y redistributivo del IRPF en la última década”, *Economistas*, nº 126, págs. 191-197.
- Onrubia, J., Picos, F. y C. Pérez (2011): *Panel de Declarantes de IRPF 1999-2007: diseño, metodología y guía de utilización*. Madrid: Instituto de Estudios Fiscales.

- Pfefferman, D. (2002): "Small area estimation: New developments and directions", *International Statistical Review*, vol. 70, nº 1, págs. 125-143.
- Picos, F. (2006): "Microsimulación mediante fusión de PHOGUE y Panel de Declarantes para evaluar reformas fiscales", *Revista de Economía Aplicada*, vol. XIV, nº 41, págs. 33-60.
- Picos, F., Antigueira, M., Pérez, C., Moreno, A., Marcos, C. y S. Díaz de Sarralde (2005): "La Muestra de declarantes de IRPF de 2002: descripción general y principales magnitudes", *Documento de Trabajo del Instituto de Estudios Fiscales*, nº 15/2005.
- Picos, F., Pérez, C. 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*, nº 01/2011.
- Rodríguez-Méndez, M. (2009): "Las Encuestas de Presupuestos Familiares y los modelos de microsimulación", en F. Picos y S. Díaz de Sarralde (coord.), *Las reformas fiscales bajo el microscopio. Microsimulación fiscal en España: datos, metodología y aplicaciones*. Madrid, Instituto de Estudios Fiscales, págs. 53-65.
- Sanz, J.F., Castañer, J.M. y D. Romero (2009): *Análisis de reformas del impuesto sobre la renta personal a partir de microdatos tributarios. El simulador de la Fundación de las Cajas de Ahorros (FUNCASim)*. Madrid: Fundación de las Cajas de Ahorros.
- Wallgren, A. y Wallgren, B. (2007): *Register-based Statistics: Administrative Data for Statistical Purposes*. Chippenham, UK: John Wiley & Sons.
- Weber, M.E. (2004): «The Statistics of Income 1979-2002 Continuous Work History Sample Individual Income Tax Return Panel», *2004 Proceedings of the American Statistical Association*, American Statistical Association, Alexandria, VA.

Fecha de recepción del original: abril, 2011
Versión final: enero, 2012

ABSTRACT

The new 1999-2007 Personal Income Tax Return Panel is an expanded panel that represents the Spanish population of taxpayers for that period. Its statistical design (large sample stratified by region, income level and main source of income, with minimum variance sampling that leads to income estimations with errors of less than 1%) guarantees very precise estimations of the main variables of the Spanish Personal Income Tax. These characteristics make the Panel suitable for distribution and redistribution analysis, evaluation of tax reforms and microsimulation. The objectives of this paper are: a) to provide information on the technical characteristics of the Panel, showing its usefulness for applied economic research and its complementarity with other microdata bases; b) to compare it to other tax microdata files available in Spain; and c) to analyse its reliability in population estimates of income, tax base and tax liability.

Key words: Spanish personal income tax return panel, tax microdata, administrative records.

JEL Classification: C81, C83, H24.

