

# MICROSIMULACIÓN MEDIANTE FUSIÓN DE PHOGUE Y PANEL DE DECLARANTES PARA EVALUAR REFORMAS FISCALES\*

FIDEL PICOS SÁNCHEZ

Universidad de Vigo

Desde principios de la década de los noventa, la investigación en Economía Pública en nuestro país ha utilizado de manera reiterada técnicas de microsimulación, asociadas, sobre todo, a la medición de efectos de cambios fiscales. Con esta finalidad, una de las fuentes de datos más utilizadas es el Panel de Hogares Europeo (PHOGUE). Pese a acopiar información de gran riqueza, los datos de rentas de capital que proporciona esta fuente tienen un nivel reducido de desagregación y están infravalorados. Este problema debe ser resuelto si queremos dar certidumbre a los ejercicios de microsimulación. La propuesta de este trabajo consiste en imputar a PHOGUE, mediante técnicas de fusión estadística, valores de rentas de capital provenientes del Panel de Declarantes de IRPF. El resultado es una base de datos más completa y robusta y, por tanto, más fiable a la hora de evaluar los impactos recaudatorios y distributivos de las reformas fiscales. Estas conclusiones se ilustran mediante la realización de varias simulaciones fiscales.

*Palabras clave:* IRPF, microsimulación, Panel de Hogares Europeo, fusión estadística.

*Clasificación JEL:* H24 C81.

Desde principios de la década de los noventa, la investigación en Economía Pública en nuestro país ha hecho una utilización creciente de las técnicas de microsimulación. Esta tendencia viene marcada por la conjunción de dos factores. De un lado, las crecientes prestaciones de las herramientas informáticas, que han puesto a disposición de los investigadores individuales potencialidades hasta hace poco reservadas a grandes centros de investigación. De otro, la creciente disponibilidad de datos, marcada por las mejoras de calidad de los ya existentes y por la introducción de nuevas fuentes.

---

(\*) El autor agradece los inestimables consejos y recomendaciones de Alberto Gago y Juan Castañer, los comentarios y correcciones de dos evaluadores anónimos, las sugerencias de Belén Martín, María Antequera y María Gómez Rúa y la financiación del Ministerio de Ciencia y Tecnología (proyecto SEC2002-03095), Xunta de Galicia (proyectos PGIDT01SCX30002PR y PGIDIT03PXIC30008PN) e Instituto de Estudios Fiscales (proyecto "Estudio de viabilidad y análisis mediante microsimulación de la aplicación en España del Modelo Dual de reforma del IRPF").

No obstante, las técnicas de microsimulación siguen tropezando con dificultades estadísticas, y es habitual que la información proporcionada por una fuente no tenga la calidad o el detalle suficiente para un objetivo determinado, y deba ser completada con información proveniente de otras fuentes. Esto es precisamente lo que ocurre con el Panel de Hogares Europeo (PHOGUE), cuyos datos de ingresos presentan una deficiente calidad en lo concerniente a las rentas de capital, tanto en términos cuantitativos (están infravaloradas) como en términos cualitativos (presentan una desagregación escasa). Lo primero hace que los individuos parezcan más pobres de lo que realmente son, mientras que la combinación de ambos factores dificulta la simulación de políticas públicas que afecten de manera específica a las rentas de capital.

El objetivo de este trabajo es buscar una solución al problema planteado, para lo que se emplean técnicas de fusión estadística mediante las que se imputan a PHOGUE valores de rentas de capital provenientes del Panel de Declarantes de IRPF del Instituto de Estudios Fiscales. Esta corrección se realiza sólo en la tercera ola de PHOGUE, pero no sería difícil extenderla para facilitar la simulación de políticas públicas cuyos efectos quieran analizarse a lo largo de varios años.

Con este propósito, el resto del trabajo se organiza como sigue. En el apartado 1 se describe el problema de infravaloración de las rentas de capital de PHOGUE, y se analiza mediante una comparación con el Panel de Declarantes. El apartado 2 describe y aplica la metodología utilizada, ofreciéndose los principales resultados en el apartado 3, e ilustrándose en el apartado 4 su utilidad mediante varias simulaciones. Finalmente, en el apartado 5 se resumen las principales conclusiones.

## 1. DIAGNÓSTICO DEL PROBLEMA Y JUSTIFICACIÓN DEL TRABAJO

La versión española del Panel de Hogares Europeo (PHOGUE) es una muestra representativa del conjunto de hogares, elaborada anualmente por el Instituto Nacional de Estadística desde 1994. Los datos se obtienen mediante encuestas y proporcionan información cuantitativa y cualitativa sobre numerosas características socioeconómicas para hogares e individuos, sumando unas 15.000 observaciones individuales con diferentes grados de ponderación. En lo referente a las variables de ingresos, PHOGUE ofrece una gran desagregación para rentas de trabajo y transferencias, pero proporciona sólo dos variables para las rentas de capital, que presentan además importantes problemas de definición. En cualquier caso, el problema principal es que ambas variables están muy infravaloradas respecto a la realidad (Andrés y Mercader, 2001)<sup>1</sup>.

Para confirmar este hecho compararemos los datos de la tercera ola de PHOGUE (correspondiente a 1996 y con datos de ingresos de 1995) con los del Panel de Declarantes de IRPF del Instituto de Estudios Fiscales de 1995 (en adelante

---

(1) Estas autoras también encuentran en PHOGUE una cierta infrarrepresentación de las rentas profesionales y empresariales. Dado que dichas actividades pueden guardar relación con una gran acumulación de capital, es probable que los problemas de ocultación de ambos tipos de rentas estén relacionados, al menos en la cola alta de la distribución.

Panel)<sup>2</sup>. Panel es una muestra representativa del conjunto de declarantes del IRPF, elaborada anualmente entre 1990 y 1995 por el Instituto de Estudios Fiscales, que recoge el detalle de los datos computados en las declaraciones de la renta de unos 300.000 contribuyentes. La razón de escoger como referencia esta fuente de datos es que presenta una información mucho más detallada que PHOGUE y, además, al ser una fuente fiscal sus datos no están sobrevalorados respecto a la realidad (en todo caso estarán infravalorados por el fraude fiscal). Esto garantiza que, si las cifras de Panel son mayores que las de PHOGUE, podamos tener la seguridad de que las primeras están más cerca de la realidad que las segundas<sup>3</sup>.

Para llevar a cabo esta comparación analizamos en primer lugar los conceptos de rentas de capital definidos en cada muestra. Como se dijo, PHOGUE ofrece solamente dos variables: “ingresos netos por rentas de la propiedad” (variable pi122a) e “ingresos netos por rentas del capital” (pi121). Aunque en ambos casos las definiciones son bastante ambiguas, entenderemos que la primera recoge las rentas provenientes de arrendamientos, mientras que en la segunda se incluyen el resto de rentas de capital (mobiliario e inmobiliario, explícitas e implícitas, regulares e irregulares, etc.). Por su parte, Panel proporciona una desagregación mucho mayor, ya que ofrece seis grupos diferentes de rentas de capital, según la legislación de IRPF vigente en 1995<sup>4</sup>. Las definiciones son en general más precisas que las de PHOGUE, aunque no están exentas de ambigüedades, sobre todo en el caso de los incrementos de patrimonio<sup>5</sup>.

Pese a las discrepancias en las definiciones, podemos crear dos grupos de rentas comparables entre ambas muestras. Así, los “ingresos netos por rentas de la propiedad” de PHOGUE los equiparamos a los “rendimientos del capital inmobiliario de inmuebles urbanos y rústicos arrendados y subarrendados de Panel”, y denominamos a esta variable común “renta de arrendamientos”. Para el resto de las rentas no es posible mantener ninguna desagregación, por lo que equiparamos los “ingresos netos por rentas del capital” de PHOGUE con la suma de todos los rendimientos restantes de Panel. A la nueva variable común la denominamos “renta mixta de capital”<sup>6</sup>.

---

(2) Actualmente están disponibles las posteriores de PHOGUE, pero Panel no se ha actualizado desde 1995. Esta es la razón por la que se utiliza la tercera ola de PHOGUE.

(3) Las características mencionadas pueden hacer pensar que Panel es una muestra más apropiada para realizar simulaciones con datos de ingresos. Sin embargo, esta fuente presenta problemas de otro tipo: no es representativa de toda la población (sólo lo es de los declarantes existentes bajo una legislación fiscal determinada); los datos no están individualizados (para los casos en que los contribuyentes casados optan por declaración conjunta aparecen agregados sin posibilidad de diferenciación); no permite agrupar en hogares; y no proporciona características personales como la edad o el sexo de los individuos.

(4) Panel proporciona además el dato correspondiente a las aportaciones a planes de pensiones realizadas por el contribuyente que, si bien no tiene equivalente en PHOGUE y no tiene interés, por tanto, para la comparación, sí será útil para completar la información de PHOGUE.

(5) Para una explicación más detallada de los problemas de definición de ambas muestras, véase Picos (2004b).

(6) De esta definición sólo excluiríamos las rentas imputadas por la propiedad del hogar en Panel, recogidas como “rendimientos del capital inmobiliario de inmuebles urbanos, excepto arrendados

Una vez agrupadas las rentas en conceptos equiparables, debemos asegurarnos de que las magnitudes sean comparables. En lo referente a la renta de arrendamientos, tomamos las magnitudes netas, que son las que aparecen en ambas muestras. Adicionalmente asignamos valores nulos a las rentas de Panel negativas, ya que PHOGUE no ofrece rentas negativas en ningún caso (este criterio también será aplicado a las rentas mixtas de capital). En cualquier caso, el principal problema de comparabilidad lo encontramos en los perceptores de las rentas: mientras que en Panel las rentas de arrendamientos están asignadas a sus perceptores legales, en PHOGUE dichas rentas se reparten a partes iguales entre todos los miembros adultos del hogar. Dado que en Panel no tenemos datos sobre la agrupación de los declarantes en hogares económicos, sólo los datos agregados serán comparables.

El cuadro 1 muestra las cifras de una y otra muestra, así como las diferencias encontradas.

Cuadro 1: RENTA DE ARRENDAMIENTOS: COMPARACIÓN DE MAGNITUDES<sup>(1)</sup>

Concepto	PHOGUE	Panel	Diferencia
Masa	1.633.708.799	3.784.511.196	-56,83%
Perceptores	1.022.892	787.524	29,89%
Renta media	1.597	4.806	-66,76%

Fuente: Elaboración propia.

(1) Datos elevados a población.

Tal y como se podía prever, el número de perceptores es claramente mayor en PHOGUE, debido al reparto de rentas entre los miembros del hogar. Pero la masa de rentas es, por el contrario, mucho menor en PHOGUE (más de un 56%). Como resultado de estas dos diferencias, la renta media de PHOGUE es sólo un tercio de la de Panel. En definitiva, las rentas de arrendamientos de PHOGUE están infravaloradas respecto a las de Panel.

En cuanto a la renta mixta del capital, realizaremos la comparación con la renta disponible, que es la única magnitud que podemos obtener de ambas muestras, a partir de las variables de que disponemos en cada caso y que aparecen recogidas en el cuadro 2.

En lo referente a los perceptores, el problema no es aquí de reparto, sino de representatividad, ya que en 1995 no estaban obligados a declarar los contribuyentes cuyas rentas de capital mobiliario sumadas a los incrementos patrimoniales regulares no superasen 250.000 pesetas (1.502,53 euros) y, además, el resto de sus

o subarrendados". Dado que no se trata de una renta monetaria, entendemos que PHOGUE no ofrece este dato.

**Cuadro 2: RENTA MIXTA DE CAPITAL: COMPARACIÓN DE DEFINICIONES**

Concepto	PHOGUE	Panel			
		Capital mobiliario sujeto a retención	Capital mobiliario no sujeto a retención	Incrementos de patrimonio (regulares e irregulares)	
Rendimientos íntegros	A	–	V12	V13	–
Gastos deducibles	B	–		V14	–
Reducción legal	C	–		V15	–
Rendimiento neto	A-B-C	–		V16	V32+V46
Retenciones	D	–	V90	–	–
Renta disponible	A-B-C-D	pi121		–	–

Fuente: Elaboración propia.

rentas cumplieren otros límites. Por tanto, es de esperar que en Panel no aparezcan todos los perceptores que aparecen en PHOGUE, aunque sí aparecerán individuos no obligados a declarar pero que, por una razón u otra, declaran. De esta situación cabe esperar que las magnitudes agregadas (masa y número de perceptores) sean menores en Panel, mientras que la renta media debería ser mayor, dado que quedarían fuera la mayoría de los no obligados a declarar por sus rentas bajas.

Con estas premisas realizamos la comparación que se recoge en el cuadro 3.

**Cuadro 3: RENTA MIXTA DE CAPITAL: COMPARACIÓN DE MAGNITUDES<sup>(1)</sup>**

Concepto	PHOGUE	Panel	Diferencia
Masa	4.688.032.087	11.769.390.558	-60,17%
Perceptores	4.960.595	9.226.683	-46,24%
Renta media	945	1.276	-25,91%

Fuente: Elaboración propia.

(1) Datos elevados a población.

Los datos no se corresponden con lo previsto. Las cifras agregadas de Panel (sólo declarantes) son mucho mayores que las de PHOGUE (toda la población), por lo que podemos pensar que existe una infravaloración muy importante de estas rentas en PHOGUE. Por el contrario, la diferencia encontrada en rentas medias a favor de Panel sí responde a lo previsto, lo que nos hace pensar que el principal problema de PHOGUE es la ocultación de rentas por parte de un número importante de contribuyentes, y no la infradeclaración por parte de los que declaran.

## 2. LA FUSIÓN ESTADÍSTICA DE PHOGUE Y PANEL

La constatación de que las rentas de capital de PHOGUE están infravaloradas y presentan poco detalle justifica la búsqueda de algún procedimiento de mejora de la calidad de las mismas. Este objetivo puede lograrse mediante fusión estadística (*statistical matching*), técnica que consiste en la creación, a partir de dos muestras representativas de la misma población, de una nueva base de datos que recoja variables de ambas. La fusión se realiza observación a observación, es decir, se emparejan aquellas observaciones individuales que tengan unas características similares, en función de variables de fusión comunes a ambas muestras. No obstante, pese a que la técnica ha venido utilizándose desde principios de los setenta (Cohen, 1991), sigue sin existir una técnica estándar, ya que en cada caso particular el investigador se enfrenta a un gran número de problemas ad hoc (Sutherland *et al.*, 2001)<sup>7</sup>.

A la hora de afrontar una fusión estadística deben tenerse en cuenta varias cuestiones (Sutherland *et al.*, 2001). En primer lugar, es necesario que las dos muestras presenten un cierto número de variables comunes, proporcionando además cada una de ellas variables propias no comunes, cuestiones ambas que se cumplen en nuestro caso. En segundo lugar, deberemos distinguir entre muestra principal y muestra complementaria: la primera de ellas será la que nos proporcione la estructura, los datos y el tamaño muestral, mientras que la segunda se limitará a suministrar las variables adicionales necesarias. En tercer lugar, deberán homogeneizarse las muestras para corregir la posible heterogeneidad en distintos aspectos (unidad de análisis, territorio, población representada). En cuarto lugar, deberán escogerse las variables de fusión apropiadas que permitan emparejar observaciones de una y otra muestra. Una vez dados estos pasos, se lleva a cabo la fusión propiamente dicha, que suele realizarse en dos fases: en la primera de ellas se crean “celdas” de observaciones similares, que se agrupan según criterios de distancia establecidos para varias variables de fusión; y en la segunda se lleva a cabo el emparejamiento sólo entre observaciones de la misma celda, a través de otra variable de fusión, generalmente la renta.

Veamos a continuación como se han resuelto cada una de estas cuestiones en la fusión llevada a cabo entre PHOGUE y Panel.

### 2.1. *Papel de cada muestra*

La primera decisión que debe tomarse en relación con la fusión es el papel que va a desempeñar cada una de las bases de datos utilizadas. Lo razonable en nuestro caso será utilizar PHOGUE como muestra principal, siendo Panel la muestra complementaria o donante. De esta forma, la estructura y el tamaño

---

(7) Tradicionalmente el método de fusión se ha utilizado para disponer simultáneamente en una misma base de datos de datos de ingreso y gasto de individuos o de familias. Ejemplos de este tipo de trabajos son los de Dayal *et al.* (2000) o Sutherland *et al.* (2001) para el Reino Unido. En España la técnica de fusión estadística ha sido utilizada sólo en los trabajos de Alegre *et al.* (2001) y Calonge y Manresa (2001). También con el objetivo de disponer en una misma muestra de datos de gasto y de ingreso de las familias, estos autores realizaron una fusión entre Panel y la Encuesta de Presupuestos Familiares de 1990.

muestral de la muestra fusionada serán los de PHOGUE, y el papel de Panel consistirá exclusivamente en proporcionar a cada observación PHOGUE unas rentas de capital más fiables y desagregadas.

## 2.2. Homogeneización de muestras

En lo referente a la homogeneización, será necesario realizar ajustes en relación con la unidad de análisis, el territorio y la población representada, tal y como se explica en el cuadro 4.

Cuadro 4: HOMOGENEIZACIÓN DE MUESTRAS

	PHOGUE	Panel	Solución
Unidad de análisis	Individuos y hogares	Declaraciones individuales y conjuntas (unidades familiares fiscales), sin posibilidad de separación individual	En ambas muestras se construyen unidades familiares fiscales, lo que implica la agrupación de: (a) los matrimonios, con sus hijos menores si los tienen (b) los padres y madres solteros con sus hijos menores  El resto de las observaciones individuales no se agrupan, formando unidades familiares individuales, independientemente de que convivan con otras en un mismo hogar.
Territorio	Toda la población, agrupada en 7 regiones	Población de territorio común, identificada por Delegación de Hacienda	Se eliminan en PHOGUE las observaciones de la región 2 (País Vasco, Navarra, Aragón y La Rioja) y en Panel las de Aragón y La Rioja
Población representada	Toda la población	Declarantes de IRPF	Se mantienen todas las observaciones con renta, pero para evitar sesgos se tendrá en cuenta la obligación de declarar como variable ( <i>dummy</i> ) de fusión

Fuente: Elaboración propia.

## 2.3. Selección de variables de fusión

Las variables de fusión son aquellas variables comunes a ambas muestras que se utilizarán para emparejar observaciones. Para que la fusión se realice correctamente, debe exigirse que las variables no comunes estén relacionadas entre sí sólo a través de las variables de fusión, lo que se conoce como propiedad de independencia condicional<sup>8</sup>. Dadas las dificultades prácticas que plantea la búsqueda de variables de fusión que sean buenos predictores de las variables no comunes de ambas muestras, se ha optado por restringir la búsqueda a aquellas variables comunes que sean buenos predictores de las rentas de capital de Panel, ya que son

(8) Este principio suele asumirse como supuesto, dada la dificultad para comprobar su existencia.

éstas las que queremos imputar a las observaciones PHOGUE. Por ello, se han seleccionado las variables que recogen conceptos que presumiblemente inciden en las pautas de ahorro de los individuos y que son, siguiendo la literatura teórica y empírica sobre la cuestión<sup>9</sup>, el tipo de unidad familiar, la renta de la misma, el lugar de residencia y el régimen de propiedad de la vivienda<sup>10</sup>.

Tanto en PHOGUE como en Panel existen variables que recogen, directa o indirectamente, estos conceptos. Sin embargo, sus definiciones no son directamente comparables, por lo que será necesario realizar modificaciones, restringir los valores e incluso calcular nuevas variables a partir de las existentes<sup>11</sup>. El cuadro 5 recoge el resultado final de este proceso de selección y homogeneización de las variables de fusión discretas, que son las que se utilizarán para la creación de celdas homogéneas en la primera fase de la fusión<sup>12</sup>.

Una vez definidas las variables discretas de fusión, en el cuadro 6 las utilizamos para caracterizar las dos muestras, con el fin de ver en qué medida la población representada es la misma en ambos casos. Además de las cifras totales, se ofrece desagregación entre obligados y no obligados a declarar. En principio esto debe permitir que, al menos en el primer caso, los valores de las variables sean similares, ya que los obligados a declarar están representados en la misma medida en una y otra muestra. Por el contrario, la escasa cantidad de observaciones no obligadas a declarar que hay en Panel hace que la distribución de las variables pueda ser muy diferente a la de los no obligados de PHOGUE.

Como se puede observar, entre los obligados a declarar las cifras son muy similares en las dos muestras, excepto en lo que concierne al régimen de propiedad de la vivienda<sup>13</sup>. En cuanto a los no obligados a declarar se aprecia que, pese a

---

(9) Para una síntesis de la literatura teórica y alguna aplicación empírica para España véase Argandoña (1986) y Analistas Financieros Internacionales y DFC (2000).

(10) También es significativa la edad, pero en Panel sólo está disponible para asalariados cotizantes a la Seguridad Social incluidos desde 1992. No obstante, la edad está correlacionada con otras de las variables escogidas, por lo que es de esperar que su exclusión no provoque efectos importantes.

(11) Algunos de estos ajustes también están motivados por la necesidad de tener suficientes observaciones en cada celda en el posterior proceso de emparejamiento.

(12) Además de las variables potencialmente relacionadas con las pautas de ahorro, se han creado dos variables *dummy* adicionales. La primera de ellas distingue entre observaciones que tienen renta de trabajo o de autónomos y las que no tienen, ya que estas últimas no podrán fusionarse según la renta de fusión definida más adelante. La segunda variable distingue entre individuos obligados y no obligados a declarar, según la normativa del IRPF de 1995. De no realizarse esta distinción, muchas observaciones PHOGUE no obligadas a declarar se emparejarían con observaciones Panel obligadas a declarar, lo que en los niveles bajos de renta implicaría un sesgo a escoger individuos Panel obligados a declarar por el tipo de rentas que obtienen, fundamentalmente autónomos y arrendamientos.

(13) La mayor diferencia estriba entre los dos primeros grupos (alquiler y vivienda propia pagada), lo que puede ser debido a la forma de obtener los datos en cada caso: mientras que en PHOGUE la clasificación corresponde directamente a las respuestas dadas por los encuestados, en Panel los individuos con vivienda propia se han identificado a través de la declaración de renta imputada. Así, todos aquellos contribuyentes que no hayan declarado esta renta imputada habrán sido identificados como habitantes de vivienda en alquiler. En cualquier caso, a pesar de los posibles errores se ha optado por mantener esta variable de fusión, dada su especial influencia sobre las pautas de ahorro de los individuos.



haber diferencias mucho mayores, existe cierta proximidad en algunas variables como la región o el número de perceptores de renta de trabajo. No obstante, es necesario insistir en que la necesidad de separar obligados y no obligados a declarar tiene como objetivo evitar ciertos sesgos, y no lograr poblaciones homogéneas, ya que a través de la fusión es posible emparejar observaciones similares, independientemente de la proporción entre observaciones del mismo tipo que haya en una y otra muestra.

**Cuadro 5: CREACIÓN Y SELECCIÓN DE VARIABLES DE FUSIÓN**

Concepto	Variable	Tipo de variable	Método de obtención		Valores
			PHOGUE	Panel	
Lugar de residencia	Región	Discreta Cualitativa	No modificada (hg015)	Se asigna a cada delegación de hacienda su región PHOGUE correspondiente	1: Noroeste 3: Madrid 4: Centro 5: Este 6: Sur 7: Canarias
Características de la unidad familiar	Estado civil	Dummy	No modificada (pd005)	Son casados los que tienen asignado un código de cónyuge o presentan declaración conjunta	Nueva variable: tipo de u. familiar 0: soltero y sin hijos 1: soltero o casado con hijos
	Número de descendientes que dan derecho a deducción	Discreta Cuantitativa	Cálculo a partir de los datos de relaciones entre individuos PHOGUE	Cálculo a partir de las deducciones por descendientes solteros (V58)	
	Número de perceptores de rentas de trabajo	Discreta Cuantitativa	Cálculo a partir de los datos de perceptores individuales	Cálculo a partir de las deducciones por trabajo dependiente (V79)	0 1 2 o más
	Principales ingresos provenientes de pensiones	Dummy	Cálculo a partir de los datos de ingresos por tipo de renta	Se deduce a partir de los datos de ingresos de trabajo (V1), cotizaciones sociales (contenidas en V3) y edad (variable externa adicional): la variable valdrá 1 si hay ingresos de trabajo, no se cotiza a la Seguridad Social y los que obtienen las rentas tienen más de 65 años	0: no 1: sí

**Cuadro 5: CREACIÓN Y SELECCIÓN DE VARIABLES DE FUSIÓN (continuación)**

Concepto	Variable	Tipo de variable	Método de obtención		Valores
			PHOGUE	Panel	
Régimen de propiedad de la vivienda	Régimen de propiedad de la vivienda	Discreta Cualitativa	Cálculo a partir de régimen de propiedad de la vivienda (ha023) y de existencia de préstamo o hipoteca para adquirirla (ha024a)	Cálculo a partir de los rendimientos del capital inmobiliario de inmuebles urbanos (excepto arrendados o subarrendados) (V6) y de las deducciones por adquisición de vivienda habitual (V67)	0: Alquiler o disfrute gratuito 1: En propiedad, pago realizado. 2: En propiedad, pago pendiente.
Variables adicionales	Observación con rentas de trabajo o de autónomos	Dummy	Mediante observación directa de las variables	Mediante observación directa de las variables	0: no 1: sí
	Observación obligada a declarar	Dummy	Aplicación de la normativa del IRPF de 1995	Aplicación de la normativa del IRPF de 1995	0: no 1: sí

Fuente: Elaboración propia.

Una vez creadas y analizadas las variables discretas necesarias en la primera fase, la segunda consiste en crear una variable de fusión continua, para la que no sólo es necesario que exista una definición única para ambas muestras, sino que las magnitudes en una y otra sean comparables. Al igual que ocurría con las rentas de capital, los niveles y formas de agregación de una y otra muestra son diferentes: mientras que Panel no presenta ninguna desagregación en rentas de trabajo, pero sí distingue entre varios tipos de rentas de autónomos, en PHOGUE ocurre lo contrario. Por tanto, sólo distinguiremos dos grandes tipos de renta, a las que denominaremos “renta del trabajo” y “renta de autónomos”, cuya composición y comparabilidad se analizan en el apéndice A<sup>14</sup>.

Dado que la fusión se realiza observación a observación, lo que nos interesa es la comparación de los valores medios individuales de estas dos variables. En este sentido, las diferencias entre una y otra muestra son, en general, pequeñas. La

(14) En este análisis no entran las rentas de capital, puesto que ya hemos constatado que las magnitudes de una y otra muestra no son comparables, siendo además esta la principal razón que nos lleva a realizar la fusión.

Cuadro 6: VALORES DE LAS VARIABLES DE FUSIÓN CUALITATIVAS

Variable	Todas las observaciones		Obligados a declarar		No obligados a declarar		
	PHOGUE	Panel	PHOGUE	Panel	PHOGUE	Panel	
Número de observaciones	6.814	262.170	4.144	233.641	2.820	28.488	
Población representada	14.685.140	11.562.792	8.955.096	10.304.545	5.745.739	1.256.440	
Región	1	13%	12%	12%	15%	12%	
	3	14%	16%	18%	17%	10%	
	4	15%	15%	14%	14%	18%	17%
	5	32%	35%	35%	35%	28%	32%
	6	21%	19%	18%	19%	27%	24%
	7	4%	4%	3%	4%	5%	4%
	Tipo de unidad familiar	Soltero sin hijos	41%	31%	29%	28%	60%
Casado sin hijos		18%	23%	18%	22%	19%	32%
Con hijos		40%	46%	52%	50%	22%	17%
Pensionistas	27%	14%	14%	13%	47%	21%	
Perceptores de renta de trabajo	0	8%	11%	12%	12%	0%	3%
	1	73%	72%	66%	70%	85%	88%
	2	19%	17%	22%	18%	15%	9%
Régimen de propiedad vivienda	Alquiler	17%	32%	16%	30%	19%	55%
	Propia	66%	43%	63%	43%	71%	39%
	Hipoteca	16%	25%	21%	27%	10%	6%
Con rentas de trab. o aut.	98%	98%	98%	98%	99%	97%	
Obligados a declarar	61%	89%	100%	100%	0%	0%	

Fuente: Elaboración propia.

única variable que presenta unas diferencias claramente mayores es la de las rentas de los desempleados, pero, dada su importancia relativamente pequeña en el conjunto de rentas de trabajo, no consideramos que provoque problemas graves.

En resumen, el análisis anterior nos permite afirmar que tanto la renta de trabajo como la renta de autónomos son adecuadas como variables de fusión, por lo que definiremos la renta de fusión como la suma de ambas. Esto no puede aplicarse a aquellas observaciones que sólo tienen rentas de alquileres y/o rentas mixtas de capital, para las cuales, como excepción, la renta de fusión será la suma de estas dos últimas<sup>15</sup>.

(15) Aunque esto puede dar lugar a muchas inexactitudes, afectará a muy pocos miembros de la muestra (en torno a un 3%), por lo que es de esperar que los efectos globales sean pequeños. Esta decisión implica que, para este tipo de observaciones, la fusión sirva no para otorgar rentas de capital diferentes (ya que se fusionarán observaciones con rentas iguales), sino para otorgar una desagregación determinada a dichas rentas.

#### 2.4. *Emparejamiento de observaciones y asignación de variables*

Siguiendo la metodología explicada anteriormente, el emparejamiento de observaciones se produce en dos fases. En la primera de ellas creamos las celdas de observaciones homogéneas. El número total es de 490, de los cuales 120 no tienen ningún elemento de PHOGUE. Esto implica que, en la práctica, tenemos 370 grupos con características diferentes<sup>16</sup>. Como resultado de todo el proceso, globalmente tenemos 37 observaciones Panel por cada observación PHOGUE, aunque esta cifra varía para cada celda.

Culminado este proceso, la segunda fase consiste en emparejar las observaciones una a una (PHOGUE con Panel), dentro de cada celda, según su renta de fusión. El criterio utilizado será el de mínima distancia, según el cual se emparejarán aquellas observaciones de la misma celda que tengan su renta de fusión lo más próxima posible, sin repetir observaciones de la muestra donante. En el caso de que en una celda existan pocas observaciones de la muestra donante, es posible que no se puedan encontrar observaciones no repetidas que se mantengan en unos límites razonables de distancia, por lo que en estos casos se permitirán emparejamientos con repetición<sup>17</sup>.

Como resultado del proceso anterior, a cada observación PHOGUE se le ha asignado una observación Panel con iguales variables discretas y una renta casi idéntica, como lo demuestra el hecho de que la renta de fusión media de las observaciones PHOGUE sea del 98,9% de la observación de Panel con la que se han emparejado. Por su parte, la mediana es del 99,99%, y la desviación típica de 8,76%, lo que indica que el valor medio es representativo.

El siguiente paso es traspasar las variables de capital de Panel a PHOGUE. El primer problema viene dado por el hecho de que el emparejamiento se produce entre observaciones cuyas rentas de fusión no son exactamente iguales. En este caso se ha optado por prorratear las rentas de capital Panel en función del ratio de renta de fusión PHOGUE/Panel de cada observación, impidiéndose así una sobrevaloración o infavaloración de las rentas de capital traspasadas.

Otro problema que surge es que, mientras que los emparejamientos se han hecho entre unidades familiares fiscales, el PHOGUE original está formado por

---

(16) Para realizar la fusión se eliminan todas las observaciones con renta de fusión igual a cero. Asimismo, se ha decidido eliminar también aquellas observaciones de ambas bases de datos con menos de 100.000 pesetas (601,01 EUR) de renta de fusión, ya que introducían muchas dificultades en la búsqueda de parejas, mientras que su exclusión no afecta en lo fundamental a los resultados de la fusión. Para poder seguir utilizando posteriormente estas observaciones, tomaremos la renta de arrendamiento PHOGUE como tal, y la renta de capital, por simplificación, como renta de capital mobiliario.

(17) Estos criterios genéricos se concretan en un proceso de emparejamiento en cuatro fases: (1) se buscan emparejamientos que cumplan simultáneamente los dos criterios: la no repetición y un determinado intervalo de proximidad de la renta de fusión; (2) si alguna observación queda sin emparejar, se elimina el criterio de no-repetición; (3) si sigue habiendo observaciones sin emparejar, se recupera el criterio de no-repetición, pero se elimina el intervalo de proximidad; (4) si aún quedan observaciones sin emparejar, se buscan parejas sin aplicar ninguno de los criterios. En la práctica, la gran mayoría de los emparejamientos se producen en la primera fase, alguno en la segunda, y prácticamente ninguno en las siguientes.

individuos. Dado que la desagregación individual es una de las ventajas que presenta PHOGUE, hemos optado por repartir las rentas de capital de manera proporcional a la renta bruta de cada miembro de la unidad familiar<sup>18</sup>.

Un tercer problema surge porque la fusión no ha tenido en cuenta las rentas de arrendamientos y las rentas de capital de cada observación. Por tanto, es posible que a una observación PHOGUE sin ese tipo de rentas le haya correspondido una observación Panel que sí las tenga. En el caso de las rentas mixtas de capital esto es algo plenamente normal, dado que precisamente uno de los problemas de PHOGUE es la no declaración de rentas de capital por parte de muchos contribuyentes. Lo contrario (que una observación PHOGUE con rentas de capital se quede sin rentas de capital Panel) es mucho más difícil que ocurra. Sin embargo, en el caso de los arrendamientos puede ocurrir lo contrario, ya que hay muchos individuos PHOGUE que tienen asignadas rentas de este tipo correspondientes a otros miembros del hogar al que pertenecen.

El proceso completo de fusión se resume en el cuadro 7.

### 3. PRINCIPALES RESULTADOS

Realizada la fusión, debemos comprobar si la muestra fusionada subsana los problemas planteados en el apartado 2. Para ello, procedemos a continuación a analizar los cambios globales que presenta la muestra fusionada en relación con la muestra original (PHOGUE), y la comparamos además con la muestra donante (Panel). En el cuadro 8 se recogen los valores agregados para cada muestra.

Tal y como se había previsto, las rentas de capital de la muestra fusionada se han visto incrementadas respecto a PHOGUE de manera muy importante, resultando las nuevas casi el triple que las originales. Esto lleva a una renta bruta más de un tercio superior, lo que a su vez provoca una cuota líquida del IRPF un 18% mayor.

Si la comparamos con Panel, la muestra fusionada presenta unas cifras absolutas mayores, ya que incluye a no declarantes que no están presentes en Panel. No obstante, la fusión ha incrementado la proporción de rentas de capital, situándose en la muestra fusionada en valores incluso superiores a los de Panel. Por último, la cuota líquida es casi idéntica en ambos casos, lo que demuestra que la fusión ha corregido, al menos, las rentas agregadas de los declarantes de la muestra original<sup>19</sup>.

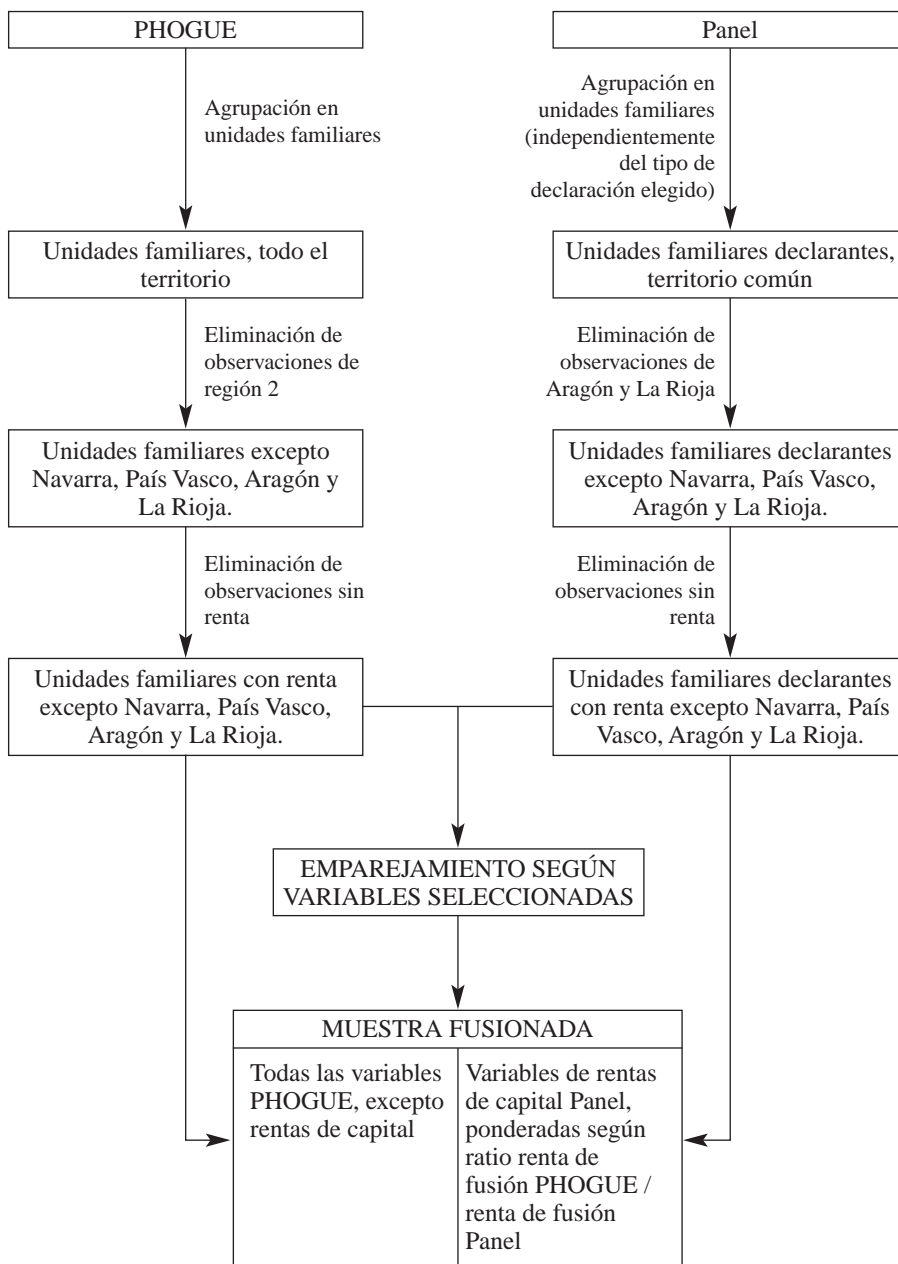
Para ver con más detalle cuáles han sido los cambios, analizamos separadamente cada uno de los dos tipos de renta de capital definidos. En primer lugar, el cuadro 9 las cifras de la renta de arrendamientos.

---

(18) En el caso de matrimonios con bienes gananciales sería más apropiado la asignación a partes iguales. Sin embargo, al carecer de información al respecto se ha optado por mantener la proporcionalidad con el resto de las rentas de cada individuo. En todo caso, en las simulaciones posteriores se ha comprobado que ambas opciones conducen a resultados agregados prácticamente idénticos.

(19) De hecho, la mayor proporción agregada de rentas de capital puede deberse a las altas proporciones de la decila inferior, compuesta en su mayoría por no declarantes (Picos, 2004a).

Cuadro 7: ESQUEMA DEL PROCESO DE FUSIÓN PHOGUE-PANEL



Fuente: Elaboración propia.

**Cuadro 8: CAMBIOS EN LAS MAGNITUDES GLOBALES<sup>(1)</sup>**

	PHOGUE	Panel	Muestra fusionada
Renta bruta (RB)	182.502.777.653	187.572.527.892	195.823.020.429
Rentas de capital (RK)	6.219.729.240	11.667.115.792	16.225.521.034
Proporción (RK/RB)	3,41%	6,22%	8,28%
Cuota líquida IRPF 1995 <sup>(2)</sup>	23.284.125.691	27.048.994.000	27.486.263.144

Fuente: Elaboración propia.

(1) Datos elevados a población de territorio común excepto Aragón y la Rioja.

(2) Las cuotas líquidas de PHOGUE y la muestra fusionada han sido calculadas con el simulador que se describe en el epígrafe siguiente. La de Panel corresponde a la cuota resultante de la autoliquidación de las estadísticas tributarias proporcionadas por Badespe (<http://www.estadief.minhac.es/>).

**Cuadro 9: RENTA DE ARRENDAMIENTOS: PHOGUE FRENTE A MUESTRA FUSIONADA <sup>(1)</sup>**

	PHOGUE	Muestra fusionada	Diferencia
Masa (EUR)	1.574.984.926	3.468.850.281	+120.25%
Perceptores	1.261.466	874.979	-30.64%
Renta media (EUR)	1.249	3.964	+217.37%

Fuente: Elaboración propia.

(1) Datos elevados a población.

Lo primero que llama la atención es el acusado crecimiento de la masa total (que más que se duplica) y de la renta media de los perceptores (que en PHOGUE era de poco más de 1.200 euros y en la muestra fusionada llega casi a 4.000). Sin embargo, el número de perceptores individuales disminuye casi un 30%, como cabía esperar del hecho de que PHOGUE reparta las rentas de alquileres entre todos los miembros del hogar. Estos resultados son coherentes con los análisis realizados en el apartado 1, de los cuales se deducía que el problema principal de PHOGUE en este campo era fundamentalmente de infravaloración individual, y no de ocultación.

Para conocer los efectos individuales clasificamos a los individuos de la muestra fusionada según los cambios que hayan sufrido sus rentas de arrendamientos en el proceso de fusión. Los resultados son los que muestra el cuadro 10.

Los datos del cuadro parecen no corroborar los resultados agregados, ya que la escasa diferencia entre individuos que aumentan su renta e individuos que la ven reducida no hace pensar en que se produzca un incremento de la percepción media. La información debe completarse con cifras monetarias. Así, mientras que

**Cuadro 10: RENTA DE ARRENDAMIENTOS: CAMBIOS INDIVIDUALES**

Cambia su situación	Aumentan su renta	0,57%
	Reducen su renta	0,50%
	Pasan a tener renta	3,67%
	Pierden su renta	5,76%
Mantienen su situación	Siguen sin tener renta	89,50%

Fuente: Elaboración propia.

las pérdidas medias son de 1.800 euros, las ganancias medias llegan a 4.088 euros. Asimismo, los individuos que pierden sus rentas en la fusión (fundamentalmente miembros de hogares en los que hay algún individuo que percibe rentas de este tipo) pierden de media 1.156 euros, mientras que a los que pasan a tener renta (probables ocultaciones) les corresponden de media 4.234 euros.

A continuación analizamos los cambios producidos en las rentas mixtas de capital, recogidos en términos agregados en el cuadro 11.

**Cuadro 11: RENTA MIXTA DE CAPITAL: PHOGUE FRENTE A MUESTRA FUSIONADA<sup>(1)</sup>**

	PHOGUE	Muestra fusionada	Diferencia
Masa (EUR)	4.644.744.314	12.756.670.753	+174,65%
Perceptores	5.017.983	13.687.759	+172,77%
Renta media (EUR)	926	932	+0,65%

Fuente: Elaboración propia.

(1) Datos elevados a población.

El comportamiento de las rentas de capital mixto es sustancialmente diferente al de las rentas de arrendamientos. Las cifras que más se incrementan son en este caso la masa total y el número de perceptores, que casi se triplican. Sin embargo, la práctica igualdad de ambos incrementos hace que la renta media se mantenga constante. Este resultado es coherente con el análisis realizado en el apartado 2, según el cual el principal problema de las rentas de capital mixto era su ocultación, y no una infravaloración de lo declarado por los encuestados.

Este resultado se observa también si analizamos los efectos individuales de la fusión, que quedan recogidos en el cuadro 12.



Cuadro 12: RENTA MIXTA DE CAPITAL: CAMBIOS INDIVIDUALES

Cambia su situación	Aumentan su renta	10,62%
	Reducen su renta	9,44%
	Pasan a tener renta	54,08%
	Pierden su renta	7,12%
Mantienen su situación	Siguen sin tener renta	18,73%

Fuente: Elaboración propia.

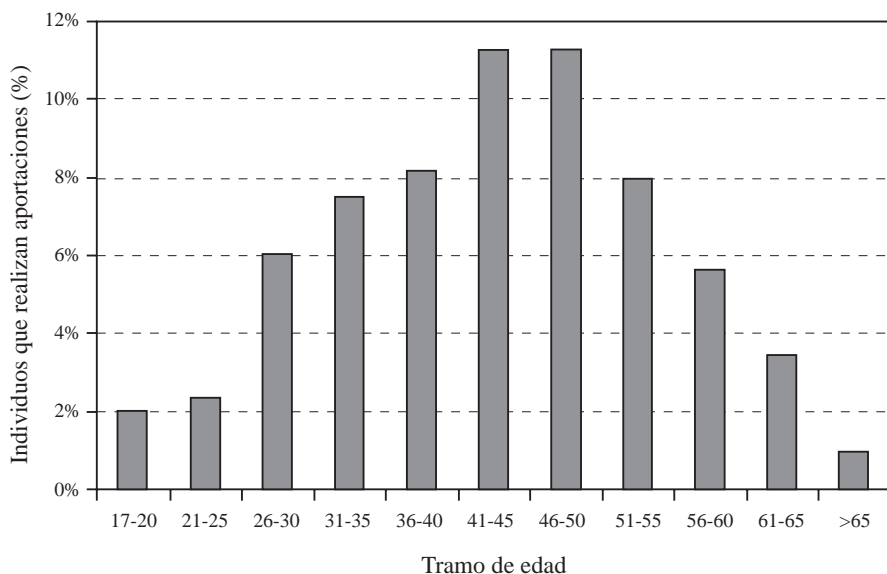
En este caso nos encontramos con dos grupos con comportamientos diferenciados: el formado por los individuos cuya situación cambia cuantitativamente (su renta de capital se incrementa o disminuye con la fusión), y el formado por aquellos que cambian también cualitativamente (pasan de no tener renta de capital a tenerla, o viceversa). En el primero el porcentaje de individuos que ven incrementadas sus rentas de capital (10,62%) es muy similar al de los que las ven reducidas (9,44%), lo que podríamos interpretar que en PHOGUE existen tantos casos de infravaloración como de sobrevaloración. Sin embargo, es más razonable pensar que la similitud de las cifras se debe al criterio utilizado para el emparejamiento de observaciones, ya que al no tener en cuenta los valores de las rentas de capital originales de PHOGUE, posibilita que existan “cruces” de rentas entre individuos de iguales características, de forma que los incrementos de unos se correspondan con reducciones de otros. De hecho, la similitud de cifras medias de ganancias (1.517 euros) y pérdidas (1.415) así parece corroborarlo.

Por el contrario, en el segundo se observa una gran diferencia entre el porcentaje de individuos que pasa a tener renta de capital mixto (más de la mitad de la población) y el porcentaje que pierde sus rentas (poco más de un 7%). Puesto que no tiene sentido que individuos que tenían rentas de capital en PHOGUE no las tengan tras realizar la fusión, la interpretación también debe realizarse en términos de intercambio de rentas entre iguales, de manera que por cada individuo que pierde sus rentas de capital mixto con la fusión, es de esperar que haya otro que las gane. De esta forma la diferencia entre las dos porcentajes (46,96%) puede entenderse como la cifra neta de individuos que no tenían rentas de capital en PHOGUE y que pasan a tenerlas después de la fusión. Podemos interpretar, por tanto, que la fusión ha sacado a la luz las rentas de capital mixto ocultas por buena parte de la población encuestada en PHOGUE.

Una vez comprobado que mediante la fusión se corrigen los problemas de las rentas de capital diagnosticados en el apartado 1, contrastamos la robustez de la fusión mediante las aportaciones a planes de pensiones. Esta variable está muy relacionada con la edad del contribuyente, pero fue asignada de Panel a PHOGUE sin tener ésta en cuenta, ya que no formaba parte de las variables de fusión. La distribución que podemos esperar a priori es que sean los trabajadores de mediana edad los que realicen más aportaciones, siendo éstas menores en los jóvenes y los

próximos a la jubilación, y prácticamente inexistentes en los jubilados. Tal y como muestra el gráfico 1, esta pauta se cumple en la muestra fusionada, lo que representa un buen indicador de que las variables de fusión fueron correctamente elegidas.

**Gráfico 1: DISTRIBUCIÓN POR EDADES DE LAS APORTACIONES A PLANES DE PENSIONES**



Fuente: Elaboración propia.

#### 4. EJERCICIO DE MICROSIMULACIÓN

Con el fin de ilustrar la utilidad de la fusión estadística llevada a cabo en el epígrafe anterior, se realizan a continuación varias simulaciones a partir de la muestra fusionada. Para ello se utiliza un microsimulador fiscal creado al efecto, que tiene como principales características la simulación de dos bases imponibles (trabajo y capital) con tratamiento diferenciado, y la posibilidad de elegir independientemente en cuál de ellas tributa cada renta de capital<sup>20</sup>.

(20) Este microsimulador está basado en el desarrollo del simulador de IRPF de la Subdirección General de Estudios Tributarios del Instituto de Estudios Fiscales (SIRPIEF), que trabaja con PHOGUE y simula sólo impuestos sintéticos. Para más información sobre las características del simulador utilizado en este trabajo puede consultarse Picos (2004b).

Dado que la ventaja de la muestra fusionada es la de ofrecer rentas de capital más fiables y detalladas que PHOGUE, se proponen, partiendo del actual IRPF, dos reformas fiscales sencillas que incidan únicamente en este tipo de rentas, y que lo hagan de manera opuesta. Así, una primera reforma, a la que llamaremos Reforma Dual, consiste en extender el tratamiento especial de las ganancias patrimoniales de más de un año (tipo único del 15%) a todas las rentas de capital. La segunda reforma va en sentido opuesto, eliminando dicho tratamiento especial, y gravando todas las rentas según la tarifa progresiva vigente en 2003. Las principales características del IRPF de 2003 y de las dos reformas propuestas quedan resumidas en el cuadro 13.

Cuadro 13: IMPUESTOS SIMULADOS

		Rentas incluidas				
Base liquidable	Tipo	IRPF 2003	Reforma Dual	Reforma Progresiva	Mínimos personales y familiares	
General	0-4.000	15%	Todas las rentas	Rentas de trabajo y actividades económicas	Todas las rentas	Personal general: 3.400
	4.000-13.800	24%	excepto GP>1 año			Familias monopar.: 5.550
	13.800-25.800	28%				Adicional >65 años: 800
	25.800-45.000	37%				Adicional >75 años: 1.000
	Más de 45.000	45%				Adicional primer hijo: 1.400
Especial	15%	GP>1 año	Rentas de capital	-	Adicional segundo hijo: 1.500 Adicional tercer hijo: 2.300 Adicional hijos <3 años: 1.200 Adicional asc. >65 años: 800	

Fuente: Elaboración propia.

Una vez definidos los impuestos, nuestro objetivo es simularlos y analizar los resultados obtenidos en términos recaudatorios y redistributivos. En cuanto a la primer cuestión, el cuadro 14 recoge los principales datos de cada uno de ellos<sup>21</sup>.

Como cabría esperar, las dos reformas simuladas tienen efectos opuestos. Mientras que la Reforma Dual reduce la recaudación y todas sus variables asociadas, la Reforma Progresiva las aumenta. Sin embargo, ambos efectos no tienen la misma magnitud, siendo la reducción de recaudación debida a la Reforma Dual mucho más elevada que la ganancia asociada a la Reforma Progresiva. Lo primero es consecuencia directa del peso significativo de las rentas de capital en la muestra fusionada (6,59%), cifra que estaba muy infravalorada en PHOGUE (3,49%) y que habría impedido llegar a resultados como el que se presenta. El segundo resultado tiene que ver con la poca significatividad de las ganancias patrimoniales

(21) Los datos monetarios de la muestra fusionada han sido actualizados a 2003 según IPC.

Cuadro 14: MEDIDAS DE RECAUDACIÓN<sup>(1)</sup>

Concepto	IRPF 2003	Reforma Dual	Reforma Progresiva
Renta bruta (millones de EUR)	246.230	246.230	246.230
Suma de cuotas líquidas (mill. EUR)	28.226	25.452	28.448
Tipo medio efectivo	13,01%	11,74%	13,11%
Tipo marginal medio	23,47%	22,34%	23,62%
Tipo marginal medio ponderado	28,64%	26,36%	28,88%
Elasticidad recaudación-renta	1,6182	1,5498	1,6266

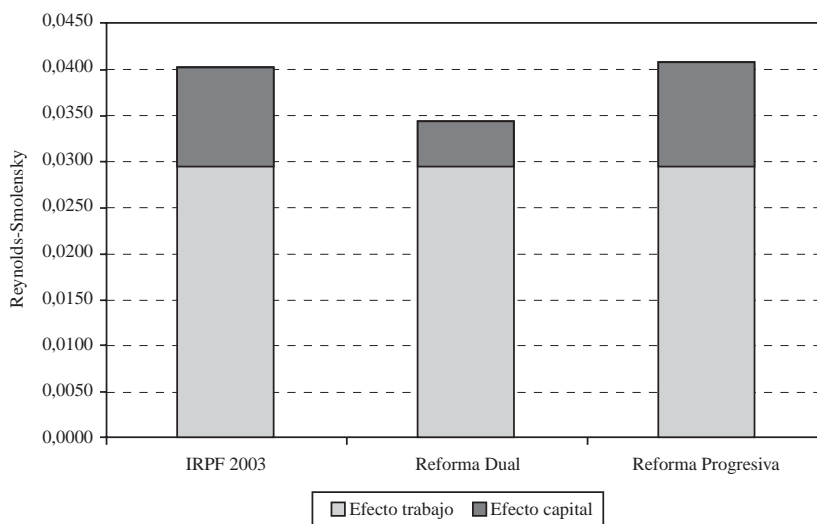
Fuente: Elaboración propia.

(1) Cifras calculadas sobre individuos.

en la muestra fusionada, lo que provoca que el paso a un tratamiento progresivo tenga escasos efectos sobre la redistribución<sup>22</sup>.

Los efectos redistributivos de los tres impuestos quedan recogidos en el gráfico 2.

Gráfico 2: ÍNDICE DE REYNOLDS-SMOLENSKY Y DESCOMPOSICIÓN TRABAJO-CAPITAL



Fuente: Elaboración propia.

(22) Es posible que hoy en día la significatividad de las ganancias patrimoniales sea mayor, pero en todo caso la simulación muestra el nivel de detalle con el que la muestra fusionada permite analizar los efectos recaudatorios y redistributivos.

Los resultados son paralelos a los obtenidos para la recaudación: mientras que la Reforma Dual reduce de manera significativa la redistribución, medida por el índice de Reynolds-Smolensky (altura total de las barras), la Reforma Progresiva la incrementa casi imperceptiblemente. Lo primero se debe a que los individuos más pobres serán indiferentes a la reforma, porque su renta de capital tributa igual antes de la reforma (al tipo mínimo del 15%) que después (al tipo único del 15%), mientras que según subimos en la escala de renta los ahorros son cada vez mayores<sup>23</sup>. El resultado de la Reforma Progresiva tiene la misma explicación, pero en sentido contrario, y su menor magnitud se debe a la ya mencionada poca relevancia de las ganancias patrimoniales en el total de rentas de capital.

El gráfico recoge también la descomposición del índice de Reynolds Smolensky entre efecto trabajo y efecto capital. El primero mide el efecto redistributivo que provoca el tratamiento fiscal de las rentas de trabajo, lo que se logra simulando el impuesto correspondiente como si las rentas de capital estuviesen exentas. El efecto capital recoge la diferencia del efecto redistributivo total con el efecto trabajo, midiendo así el efecto que tiene en sí mismo el tratamiento fiscal de las rentas de capital.

Como cabría esperar, el efecto trabajo es considerablemente mayor que el efecto capital y, al no modificarse el tratamiento fiscal de las rentas de trabajo, es igual en los tres impuestos simulados. Por tanto, los cambios redistributivos vienen dados exclusivamente por los cambios en el gravamen de las rentas de capital. En este sentido, el gráfico nos muestra como el tipo único de la Reforma Dual mantiene una capacidad redistributiva positiva, aunque ésta es menos de la mitad que en el impuesto actual<sup>24</sup>. Por su parte, el efecto capital de la Reforma Progresiva es sólo ligeramente mayor que el del IRPF actual.

Tras analizar los cambios en la redistribución global, podemos preguntarnos quién gana y quién pierde con cada una de las reformas. Para ello utilizaremos la Participación Relativa en la Cuota (PRC) por decilas, que se calcula como el ratio entre el tipo medio efectivo de una decila y el tipo medio efectivo global. El gráfico 3 muestra los resultados.

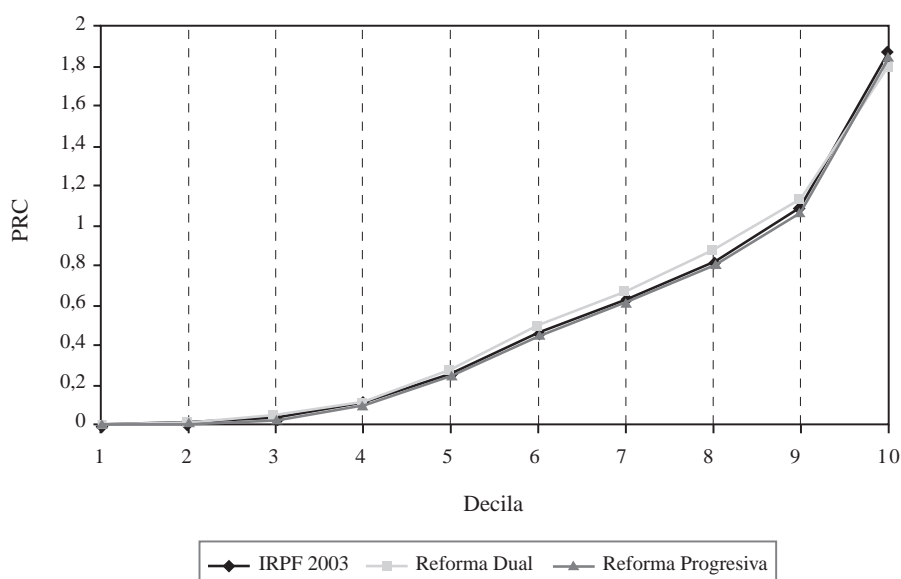
Lo primero que llama la atención es la coincidencia casi exacta de la Reforma Progresiva con el IRPF actual. Esto demuestra que los pequeños cambios de recaudación y de redistribución vistos anteriormente se distribuyen equitativamente a lo largo de la población, siendo el efecto imperceptible en todas las decilas. Por el contrario, la Reforma Dual presenta una pauta diferente. Como cabría esperar, esta reforma perjudica a las rentas bajas y beneficia a las altas; no obstan-

---

(23) Nótese que esto no ocurriría si el tipo elegido fuese mayor que el 15%, ya que entonces los resultados dependerían de las proporciones de renta de capital que tuviesen los individuos, siendo los resultados inciertos a priori (Picos, 2004a).

(24) Nótese que este resultado no tendría por qué darse. Por definición, un gravamen proporcional es redistributivamente neutral en términos de Reynolds-Smolensky, porque no provoca cambios en la concentración de renta antes y después de impuestos. La clave en este caso está en que la proporcionalidad se aplica sólo sobre una parte de la renta; al ser ésta variable entre individuos, los efectos (progresivos o regresivos) dependerán de cómo se repartan dichas proporciones a lo largo de la escala de renta.

Gráfico 3: PRC POR DECILAS



Fuente: Elaboración propia.

te, llama la atención comprobar que los beneficiados están situados en la última decila, mientras que el resto de las decilas salen perjudicadas<sup>25</sup>.

## 5. CONCLUSIONES

En los últimos años se ha producido en España un importante crecimiento de la utilización de muestras de microdatos para las investigaciones de economía pública, debido tanto a los avances de la informática como a las mejoras en los datos disponibles. Sin embargo, las fuentes existentes siguen presentado limitaciones para ciertos objetivos de investigación. Este es el caso del PHOGUE, cuyos rentas de capital presentan una deficiente calidad tanto en términos cuantitativos como cualitativos, lo que cuestiona los resultados de las simulaciones que se lleven a cabo con ellas.

Este trabajo proporciona una solución a este problema, mediante la realización de una fusión estadística de PHOGUE con el Panel de Declarantes de IRPF

(25) Esto no implica que estos individuos paguen más que antes, ya que por definición del impuesto será imposible. Lo que significa es que, con el nuevo impuesto, su tipo medio se incrementa respecto al tipo medio global, lo que implica que su situación empeora en términos relativos.

del Instituto de Estudios Fiscales. Tras un complejo proceso de homogeneización y emparejamiento de observaciones, se obtiene una nueva muestra con las características básicas de PHOGUE, pero cuyas rentas de capital proceden de Panel. De los análisis efectuados a posteriori se deduce que el proceso de fusión logra sus objetivos, puesto que las nuevas variables mejoran la calidad cuantitativa y cualitativa de las variables originales de PHOGUE.

La muestra resultante es así un buen sustituto de PHOGUE, particularmente para llevar a cabo simulaciones que afecten con especial intensidad a las rentas de capital. Así lo corroboran las simulaciones realizadas en el apartado 4, en las que se constatan las ventajas de utilizar la muestra fusionada tanto en lo referente al detalle de las rentas de capital (permite múltiples variantes en el tratamiento fiscal de las mismas) como a su magnitud (los cambios fiscales tienen efectos significativos sobre los resultados finales).

En definitiva, los resultados obtenidos tienen una especial relevancia para un periodo como el actual en el que se suceden propuestas de reforma fiscal que casi siempre afectan a las soluciones aplicadas a este tipo de rentas. Además, aunque la fusión se ha limitado a la tercera ola de PHOGUE, la técnica descrita podría extenderse a otras, proporcionando un panel corregido que mejoraría la simulación de políticas públicas cuyos efectos quisieran analizarse a lo largo de varios años.

#### ANEXO: COMPARACIÓN DE RENTAS PARA LA CREACIÓN DE LA VARIABLE RENTA DE FUSIÓN

Renta de trabajo – pese a que PHOGUE distingue varios tipos de renta, Panel ofrece una desagregación cuantitativa que no ofrece PHOGUE. Así, para cada receptor proporciona los ingresos íntegros (tanto dinerarios como en especie), los gastos deducibles (en los que están incluidos, entre otros, las cotizaciones a la Seguridad Social y una cantidad a tanto alzado por gastos generales), los ingresos netos y las retenciones practicadas. Por el contrario, PHOGUE ofrece sólo la renta disponible (monetaria y en especie) en el momento del cobro, es decir, los rendimientos una vez restadas las cotizaciones de la Seguridad Social y las retenciones del IRPF<sup>26</sup>.

En el cuadro 15 aparecen las variables que nos proporcionan los distintos conceptos de renta de una y otra fuente<sup>27</sup>.

Como se ve, ninguno de los tres conceptos de renta aparece reflejado simultáneamente en las dos muestras. Por tanto, es necesario realizar cálculos en una de las bases de datos para obtener un concepto comparable al de la otra. Lo más razonable es adaptar Panel a PHOGUE restando en Panel las retenciones al rendi-

---

(26) PHOGUE proporciona una variable de elevación a bruto de los ingresos, pero se ha decidido no hacer uso de ella, debido a que está calculada como media del ratio neto/bruto para todos los ingresos en los que se solicita y para todos los miembros de un hogar.

(27) De PHOGUE se han excluido aquellas rentas que, según la legislación del IRPF de 1995, no eran gravables, ya que en ningún caso serán comparables con las rentas que aparecen en Panel.

**Cuadro 15: RENTA DE TRABAJO: COMPARACIÓN DE DEFINICIONES**

Concepto		PHOGUE	Panel	
Rendimientos íntegros	A	–	V1, V2	
Gastos deducibles	Cotizaciones sociales Cuotas a sindicatos Otros gastos deducibles	B	– – –	V3
Contribuciones imputadas por promotores a planes de pensiones	C	–	V4	
Rendimiento neto	A-B+C	–	V5	
Retenciones	D	–	V88, V89	
Renta disponible	A-B-C-D	pi111, pi131, pi1321, pi1322, pi133, pi134, pi136	–	

Fuente: Elaboración propia.

miento neto (V5 - V88 - V89), para así obtener la renta disponible comparable con PHOGUE<sup>28</sup>. Adicionalmente, deberemos dar valor nulo a las rentas negativas de Panel derivadas de unas retenciones mayores que los ingresos netos.

Para comprobar si las magnitudes son comparables debemos tener en cuenta que Panel recoge sólo a los perceptores de rentas de trabajo que declaran. Si comparamos Panel directamente con PHOGUE, observamos que en esta última base de datos hay un 35% más de individuos de este tipo, que perciben una renta media un 17% menor (ya que los perceptores más pobres no declaran). Por tanto, para constatar la comparabilidad de los datos recurrimos a las estadísticas de retenciones de Agencia Estatal de la Administración Tributaria e Instituto de Estudios Fiscales (1997) (en adelante AEAT) que, teniendo el mismo origen que Panel, proporcionan datos agregados de perceptores de salarios, pensiones y prestaciones por desempleo de toda la población, y no sólo de los declarantes del IRPF. El cuadro 16 recoge el resultado de la comparación<sup>29</sup>.

De los tres tipos de renta, los datos de PHOGUE más ajustados a AEAT son los de los salarios. La percepción salarial media de una observación individual PHOGUE es sólo un 2,48% menor que la percepción salarial media de AEAT. Sin embargo, esta práctica igualdad en la renta media procede de un número de perceptores y una masa de rentas menores, en torno a un 9%.

(28) No obstante, a la hora de obtener el rendimiento neto de Panel, a los rendimientos íntegros se les ha restado el concepto “Otros gastos deducibles”, que no recoge necesariamente gastos reales al tratarse de una reducción fiscal, por lo que para lograr tener en Panel una renta disponible comparable con PHOGUE, restaremos al rendimiento neto las retenciones, pero le sumaremos el concepto “Otros gastos deducibles”.

(29) En este caso tomamos de PHOGUE la renta elevada a íntegro, ya que este es el dato que proporciona AEAT. La elevación a íntegro ha sido realizada por la Unidad de Microsimulación del Instituto de Estudios Fiscales.



Cuadro 16: RENTA DE TRABAJO: COMPARACIÓN DE MAGNITUDES (PHOGUE-AEAT)

Concepto		PHOGUE <sup>(1)</sup>	AEAT	Diferencia
Salarios	Masa (EUR)	117.179.228.232	132.762.206.528	-11,74%
	Perceptores	9.753.390	10.776.418	-9,49%
	Renta media (EUR)	12.014	12.320	-2,48%
Pensiones	Masa (EUR)	32.111.214.157	39.771.375.365	-19,26%
	Perceptores	4.957.318	6.300.876	-21,32%
	Renta media (EUR)	6.478	6.312	2,62%
Prestaciones desempleo <sup>(2)</sup>	Masa (EUR)	4.330.821.257	6.728.543.125	-35,64%
	Perceptores	1.658.496	2.885.262	-42,52%
	Renta media (EUR)	2.611	2.332	11,97%

Fuente: Elaboración propia.

(1) Datos elevados a población.

(2) En PHOGUE se han eliminado dos observaciones con rentas mayores de 10.000.000 pesetas (60.101,21 EUR) que casi con seguridad estaban computadas por error como prestaciones por desempleo.

Estos datos empeoran cuando observamos los perceptores de pensiones: el número de perceptores y la masa total son bastante menores en PHOGUE, en torno a un 20%. Por el contrario, la renta media se mantiene prácticamente igual, en este caso por encima en un 2,62%. Los datos que presentan mayores diferencias son los de percepciones de desempleo: el número de perceptores de este tipo de rentas en PHOGUE es menor en más de un 40%, mientras que la masa total de rentas es más un tercio más pequeña. Como resultado, la percepción media por desempleo es un caso 12% mayor en PHOGUE que en AEAT.

En resumen, las diferencias de renta global y las de número de perceptores son elevadas, pero las diferencias de renta individual no lo son.

Renta de autónomos – estas rentas presentan un problema similar al de las rentas de trabajo. Mientras que Panel ofrece los rendimientos íntegros, el neto y las retenciones (aunque en este caso no ofrece separadamente los gastos deducibles), PHOGUE ofrece de nuevo sólo un dato, bajo la denominación de “rendimiento neto de trabajo por cuenta propia”. Según la descripción de esta variable que aparece en Instituto Nacional de Estadística (1996), el rendimiento neto se corresponde de nuevo con lo que hemos denominado renta disponible (rendimiento neto menos las retenciones y los pagos fraccionados).

Al igual que en el caso de las rentas de trabajo, adaptaremos Panel a PHOGUE, calculando la renta disponible de la primera muestra mediante la resta de las retenciones y pagos fraccionados al rendimiento neto. Asimismo, daremos valor nulo a las rentas disponibles negativas de Panel, ya que las posibles rentas negativas de PHOGUE son consignadas como nulas.

El cuadro 17 ofrece un resumen de la cuestión, con indicación de las variables que en cada caso recogen los conceptos mencionados.

Cuadro 17: RENTA DE AUTÓNOMOS: COMPARACIÓN DE DEFINICIONES				
Concepto		Panel		
		PHOGUE	Declaración por módulos	Resto
Rendimientos íntegros	A	–	–	V17, V19, V21, V23
Gastos deducibles	B	–	–	–
Rendimiento neto	A-B	–	V25, V26	V18, V20, V22, V24
Retenciones y pagos fraccionados	C	–	V91, V92, V96	
Renta disponible	A-B-C	pi112		–

Fuente: Elaboración propia.

Para la comparabilidad de magnitudes no es necesario acudir a fuentes externas, ya que todos los perceptores de rentas de autónomos deben aparecer en Panel por estar en 1995 obligados a declarar fuese cual fuese la cuantía de sus ingresos. En el cuadro 18 se muestra la comparación PHOGUE-Panel de rentas de autónomos.

Cuadro 18: RENTA DE AUTÓNOMOS: COMPARACIÓN DE MAGNITUDES <sup>(1)</sup>			
Concepto	PHOGUE	Panel	Diferencia
Masa (EUR)	15.516.703.006	18.742.799.712	-17,21%
Perceptores	1.913.006	2.433.360	-21,38%
Renta media (EUR)	8.111	7.702	5,31%

Fuente: Elaboración propia.

(1) Datos elevados a población.

De nuevo nos encontramos con una situación en la que PHOGUE presenta un número considerablemente menor de perceptores (un 21% menos), y una masa de rentas también menor (casi un 17%). Como resultado, la renta media es mayor en PHOGUE en un 5,31%.

El cuadro 19 resume las dos variables definidas, así como las diferencias de renta media encontradas.

Cuadro 19: VARIABLES DE FUSIÓN TIPO RENTA

Variable	PHOGUE	Panel <sup>(1)</sup>	Diferencia <sup>(2)</sup>
Renta de trabajo	pi111 + pi131 + pi1321 + pi1322 + pi133 + pi134 + pi136	V5+OGDT-V88-V89	Salarios: -2,51% Pensiones: 2,54% Desempleo: 12,11%
Renta de autónomos	pi112	V18+V20+V22+V24+V25 +V26-V91-V92-V96	5,31%

Fuente: Elaboración propia.

(1) Si los valores son negativos, se computan como nulos.

(2) En renta de trabajo, las diferencias son entre PHOGUE y AEAT; en renta de autónomos, diferencia entre PHOGUE y Panel.



#### REFERENCIAS BIBLIOGRÁFICAS

- Agencia Estatal de la Administración Tributaria e Instituto de Estudios Fiscales (1997): *Empleos, salarios y pensiones en las fuentes tributarias 1995*, Madrid: Ministerio de Economía y Hacienda.
- Alegre, J., J. Arcarons, S. Calonge y A. Manresa (2001): “El método de fusión estadística: una aplicación entre la encuesta de presupuestos familiares (EPF90) y una muestra de declarantes del Impuesto sobre la Renta de las Personas Físicas (IRPF90)”, en Labeaga y Mercader (coord.) (2001): *Desigualdad, redistribución y bienestar: una aproximación a partir de la microsimulación de reformas fiscales*, Madrid: Instituto de Estudios Fiscales.
- Analistas Financieros Internacionales, DFC (2000): *El ahorro y los mercados financieros: situación actual y perspectivas*, Madrid: Bolsa de Madrid.
- 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*, vol. 43, n.º 148, págs. 241-280.
- Argandoña, A. (1986): “Los determinantes del ahorro de las familias” *Papeles de Economía Española*, n.º 28, págs. 317-361.
- Calonge, S. y A. Manresa (2001): “Fusión estadística de bases de datos”, cap. 4 de Calonge, S.; Manresa, A. (2001): *Incidencia fiscal y del gasto público social sobre la distribución de la renta en España y sus CCAA*, Bilbao: Fundación BBVA.
- Cohen, M.L. (1991): “Statistical matching and microsimulation models”, en Citro, C.F. and Hanushek, E.A. (eds.) (1991): *Improving Information for social policy decisions. The uses of microsimulation modelling*, vol. II, Washington D.C.: National Academy Press. Cit por Sutherland *et al.* (2001).
- Dayal, N., J. Gomulka, L. Mitton, H. Sutherland y R. Taylor (2000): “Enhancing Family Resources Survey income data with expenditure data from the Family Expenditure Survey: data comparisons”, *University of Cambridge Department of Applied Economics Microsimulation Unit Research Notes*, n.º MU/RN/40.
- Instituto Nacional de Estadística (1996): *Panel de Hogares de la Unión Europea (PHOGUE): metodología*, Madrid: Instituto Nacional de Estadística.

- Picos, F. (2004a): "Modelo Dual de IRPF y equidad: un nuevo enfoque teórico y su aplicación al caso español", *Papeles de Trabajo del Instituto de Estudios Fiscales*, n.º 08/04.
- Picos, F. (2004b): *Modelo Dual de IRPF y equidad: un estudio de la viabilidad y los efectos de su aplicación en España*, Madrid: Instituto de Estudios Fiscales (Colección Investigaciones, n.º 04/04).
- Sutherland, H., R. Taylor y J. Gomulka (2001): "Combining household income and expenditure data in policy simulations", *University of Cambridge Department of Applied Economics Working Papers / Microsimulation Unit Discussion Papers*, n.º MU0101.

*Fecha de recepción del original: marzo, 2004*

*Versión final: abril, 2005*

#### ABSTRACT

Since the early nineties, microsimulation techniques have been frequently used on research in Public Economics in Spain, typically for measuring the effects of tax changes. For this purpose, one of the most frequently used data sources has been the European Community Household Panel (ECHP). Despite providing very rich information, this source supplies undervalued capital income data with little disaggregation. This problem must be solved if we want to give certainty to microsimulation exercises. Our proposal consists of assigning values of capital income that come from a sample of Spanish taxpayers (Panel de Declarantes del IRPF) to ECHP by means of statistical matching techniques. The result is a more complete and robust database and, therefore, more reliable when evaluating the effects of tax reforms on revenue and redistribution. These conclusions are illustrated by performing several tax simulations.

*Key words:* personal income tax, microsimulation, European Community Household Panel, statistical matching.

*JEL classification:* H24 C81.