Medición de los errores en las estimaciones realizadas a partir del panel del impuesto sobre la renta de las personas físicas (IRPF) del Instituto de Estudios Fiscales (IEF)

cesar.perez@ief.minhac.es, Instituto de Estudios Fiscales y Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid

1. Introducción

El presente trabajo tiene por objeto realizar un análisis exhaustivo de la problemática en la cuantificación del error de los estimadores obtenidos en paneles de datos con información procedente de registros administrativos a partir de muestras. Como aplicación práctica se presenta la cuantificación de los errores en los estimadores derivados del panel de datos del Impuesto sobre la Renta de las Personas Físicas del Instituto de Estudios Fiscales 1999-2007 (panel de IRPF del IEF). Cuando los estimadores a utilizar son de estructura matemática compleja (índices de Gini, Reynolds-Smolensky, Kakwani, etc.) se utilizarán métodos especiales de estimación de varianzas como el método de los grupos aleatorios, el método de las submuestras interpenetrantes y los métodos Bootstrap o de autogeneración. Estas estimaciones permitirán calcular los intervalos de confianza adecuados par cuantificar los efectos de las medidas de política fiscal sobre regiones a nivel de estrato, e incluso más pequeñas.

Palabras clave: Muestreo, Paneles, IRPF

2. Estructura del panel de IRPF 1999/2007 del IEF

El panel de IRPF del IEF contiene información de *rentas fiscales de personas y hogares* de una población representativa de los sujetos pasivos de IRPF en el Territorio de Régimen Fiscal Común a lo largo del tiempo. Este Panel responde al concepto de *Panel expandido*, es decir que anualmente se incluye una representación de las altas que se produzcan controlando también las bajas. Dado el objetivo perseguido y la información disponible, se considera que la opción más adecuada es la utilización de los *individuos* como *unidad muestral* al ser esta, y no las declaraciones, una unidad homogénea a lo largo del tiempo. Ello conduce a que, en el caso de las declaraciones conjuntas, se deban individualizar para cada uno de los cónyuges las rentas declaradas conjuntamente y así poder realizar la selección de la muestra; para ello se utilizará la información de la que la AEAT dispone y que de hecho utiliza en la elaboración de los Borradores que envía a los contribuyentes.

Para los individuos que resulten seleccionados según los criterios establecidos, se suministrará la información sobre las imputaciones individuales realizadas así como toda la información de sus declaraciones presentadas por ellos (sean obligados o no y realicen declaración conjunta o declaración individual). Así mismo, para poder llevar a cabo análisis referido a hogares, se suministrará el mismo conjunto de información referida a sus cónyuges, siempre que se disponga de la información que permita su identificación como tales cónyuges. Se considera como *año base del panel* 2003 porque es el primer año en que la Agencia Tributaria graba los datos hasta tres dígitos incluidos en los modelos de declaración que se usan como marco muestral.

Para *seleccionar la muestra*, del marco de lista que incluye las declaraciones de todos los individuos del territorio fiscal común (*ámbito geográfico*) se realiza la extracción en base a un *muestreo estratificado aleatorio*, siendo las *variables de estratificación y subestratificación* la Comunidad Autónoma de residencia (15 CCAA del Territorio de Régimen Fiscal Común, además de Ceuta y Melilla que se considerarán como una única comunidad autónoma), los niveles de ***renta bruta*** agrupados en 11 tramos (Negativas y 0, inferiores a 3.000 euros, superiores a 3.000 euros e inferior o igual a 6.000 euros, superiores a 6.000 euros e inferior o igual a 12.000 euros, superiores a 12.000 euros e inferior o igual a 18.000 euros, superiores a 18.000 euros e inferior o igual a 30.000 euros, superiores a 30.000 euros e inferior o igual a 60.000 euros, superiores a 60.000 euros e inferior o igual a 120.000 euros, superiores a 120.000 euros e inferior a 240.000 euros y superiores a 240.000 euros) y la fuente de renta (con dos valores posibles: proporción de ingresos del trabajo >50 por ciento y proporción de ingresos del trabajo menor o igual que el 50 por ciento). En cada CCAA (16 estratos) los individuos se agruparán según el nivel de renta bruta que le corresponda (16*11=176 estratos de segundo nivel) y, a continuación, en cada uno de los subestratos definidos según el tramo de renta, se separaran en dos grupos los individuos según el origen de dichas rentas (más del 50% de los ingresos son del trabajo y el 50% o menos provienen del trabajo). El resultado serán 176*2 = 352 estratos de último nivel en cada uno de los cuales se realiza la extracción aleatoria simple.

Se observa que las variables de estratificación son adecuadas ya que dan a lugar a estratros muy homogéneos dentro de sí y muy heterogéneos entre sí. Los declarantes de cada Comunidad Autónoma se parecen porque la riqueza es una variable muy correlada con la Comunidad Autónoma y además diferencia muy acusadamente las Comunidades entre sí. Por otro lado, cada Comunidad tiene su legislación propia sobre el IRPF lo que también hace que los individuos de cada Comunidad se parezcan y se diferencien del resto de las Comunidades. El nivel renta también es una variable que produce estratos homogéneos dentro y heterogéneos entre ya que los ricos se parecen entre sí, los pobres también y además se diferencian mucho de los ricos. Lo mismo ocurre con la fuente de renta, ya que los individuos con sólo rentas del trabajo suelen parecerse entre sí y se diferencian bastante de los que tienen además rentas de capital y otras rentas adicionales. Se utilizará *afijación de mínima varianza* dada la gran desigualdad de variabilidades de la renta en los distintos estratos.

3. Estimadores

El *estimador de cualquier total poblacional* X en muestreo estratificado aleatorio es la suma de los estimadores del total en cada uno de los L estratos. Se tiene:

$$\hat{X}_{st} = \sum_{h=1}^{L} \hat{X}_h = \sum_{h=1}^{L} N_h \bar{x}_h = \sum_{h=1}^{L} \frac{N_h}{n_h} x_h = \sum_{h=1}^{L} fe_h x_h \qquad \begin{cases} \bar{x}_h = \textit{media muestral en el estrato h} \\ x_h = \textit{total muestral en el estrato h} \\ N_h = \textit{tamaño poblacional del estrato h} \\ n_h = \textit{tamaño muestral del estrato h} \\ fe_h = \textit{factor de elevación del estrato h} \end{cases}$$

Por lo tanto, para estimar cualquier total poblacional se suman los productos de los factores de elevación fe_h por los totales muestrales en cada estrato x_h. El estimador de cualquier media en muestreo estratificado aleatorio es la media ponderada de los estimadores de la media en cada estrato, siendo los coeficientes de ponderación $W_h = N_h/N$ de suma unitaria (N_h es el tamaño poblacional del estrato y N es el tamaño de la población).

$$\hat{\bar{X}}_{st} = \bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h = \sum_{h=1}^{L} \underbrace{\frac{N_h}{N}}_{W_h} \frac{1}{n_h} x_h = \frac{1}{N} \sum_{h=1}^{L} \frac{N_h}{n_h} x_h = \frac{1}{N} \sum_{h=1}^{L} fe_h x_h$$

Por lo tanto, *para estimar cualquier media poblacional* se suman los productos de los factores de elevación por los totales muestrales en cada estrato y se divide por el tamaño poblacional.

4. Tamaño de muestra y error de muestreo

El tamaño de muestra viene definido por un error relativo de muestreo menor del 1,5 por ciento, con un nivel de confianza adicional del 3 por mil (entre 300.000 y 400.000 individuos por año). Para afijación de mínima varianza, el tamaño de muestra para cometer un error relativo de muestreo dado $e_{r\alpha}$ en el año base viene dado por:

$$n = \frac{\lambda_\alpha^2 \left(\sum_{h=1}^{L} N_h S_h \right)^2}{e_{r\alpha}{}^2 N^2 \bar{X}^2 + \lambda_\alpha^2 \sum_{h=1}^{L} N_h S_h^2} \cong 400000$$

S_h^2 = cuasivarianza poblacional en el estrato h

Sabemos que el tamaño de muestra necesario para cometer un error de muestreo dado no depende del tipo de estimador que se utilice, por lo tanto, la expresión anterior es la misma para todos los estimadores posibles.

Una vez seleccionada la muestra con el tamaño y error anteriormente especificados, cualquier ***estimación a nivel de estrato para cualquier variable correlada con las variables de estratificación*** tendrá el error para el cual se calculó el tamaño de muestra, es decir, un 1,5 por ciento, con un nivel de confianza adicional del 3 por mil. Para ***estimaciones a niveles inferiores al de estrato***, habrá que usar ***áreas pequeñas, subpoblaciones*** o calcular los errores a través de las fórmulas de la estimación de las varianzas para muestreo estratificado y afijación proporcional que se especifican a continuación:

$$\hat{V}(\bar{x}_{st})=\frac{1}{n}\left(\sum_{h=1}^{L}W_h\hat{S}_h\right)^2-\frac{1}{N}\sum_{h=1}^{L}W_h\hat{S}_h^2, \quad \hat{V}\left(\hat{X}_{st}\right)=\frac{1}{n}\left(\sum_{h=1}^{L}N_h\hat{S}_h\right)^2-\sum_{h=1}^{L}N_h\hat{S}_h^2$$

$\hat{S}_h^2$ = cuasivarianza muestral en el estrato h

Los errores relativos estimados se calculan mediante las expresiones:

$$\hat{C}_v\left(\hat{X}_{st}\right)=\frac{\sqrt{\hat{V}(\hat{X}_{st})}}{\hat{X}_{st}} \qquad \hat{C}_v\left(\bar{x}_{st}\right)=\frac{\sqrt{\hat{V}(\bar{x}_{st})}}{\bar{x}_{st}}$$

5. Estimación del error para estimadores con estructura matemática compleja

Para realizar el análisis redistributivo del impuesto se suelen llevar a cabo análisis de desigualdad. En concreto, lo habitual es calcular los índices de Gini (IG) antes y después de la aplicación del impuesto; el índice de Reynolds-Smolensky (IRS), que expresa el grado de redistribución del impuesto así como la diferencia de los dos IG mencionados; y el índice de Kakwani (1977) (IK), que mide la progresividad del impuesto mediante la diferencia entre el IG de la renta antes de impuestos y un índice de concentración de las cuotas líquidas ordenadas según renta.

El índice de Gini es una medida de concentración relativa definida como la mitad de la diferencia media paara cada par de observaciones de renta, dividida por el valor media de la variable cuya distribución se evalúa, tradicionalmente expresado como:

$$G(y)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i-y_j\right|}{2n^2\bar{y}}$$

Siguiendo a Glasser (1962) y Dixon (1987), alternativamente la fórmula del coeficiente de Gini puede escribirse como sigue:

$$G(y)=\frac{1}{n(n-1)\bar{y}}\sum_{i=1}^{n}(2i-n-1)y_i$$

Adicionalmente, cuando se dispone de una muestra de tamaño n extraída de una población de tamaño N, el índice de Gini poblacional puede estimarse insesgadamente mediante el estimador siguiente:

$$\hat{G}(y)=\frac{1}{N\bar{y}}\sum_{i=1}^{n}K_iy_i\left(2\sum_{j=1}^{n}K_j-K_i-N\right)$$

donde y es la variable renta, n es el tamaño de la muestra, Ki es el factor de elevación y N es el tamaño poblacional.

Como el resto de los índices de desigualdad y progresividad dependen del índice de Gini, en este trabajo se cuantificará el error cometido.

Habitualmente el error absoluto de un estimador insesgado suele medirse, a partir de los datos de una muestra, mediante la estimación de su varianza. Pero el problema aparece al intentar estimar la varianza cuando la expresión del estimador es complicada, tal y como ocurre en el caso del estimador del Índice de Gini. En estas situaciones se acude a los métodos específicos de estimación de varianzas utilizados en la teoría del muestreo. Entre estos métodos tenemos el método de las muestras interpenetrantes, el método de los grupos aleatorios, el método de las semimuestras reiteradas, el método de Jacknife y el método Bootstrap

Método de las muestras interpenetrantes

El método de las muestras interpenetrantes se utiliza cuando tenemos un conjunto de dos o más muestras, elegidas con el mismo esquema de muestreo (independientes o no) y tales que cada una proporcione una estimación válida del parámetro que se pretenda estimar con el mismo error de muestreo. Si las muestras son independientes es fácil obtener un estimador insesgado de la varianza del estimador, tal y como se muestra a continuación.

Sean $\hat{\theta}_1, \hat{\theta}_2, \cdots \hat{\theta}_k$ estimadores insesgados de θ basados en k muestras independientes. Su media

$$\hat{\theta} = \frac{1}{k}\sum_i^k \hat{\theta}_i$$

es también un estimador insesgado de θ, ya que:

$$E(\hat{\theta}) = \frac{1}{k}\sum_i^k \underbrace{E(\hat{\theta}_i)}_{\theta} = \frac{k\theta}{\theta} = \theta$$

y su varianza puede calcularse fácilmente como:

$$V(\hat{\theta}) = V\left(\frac{1}{k}\sum_i^k \hat{\theta}_i\right) = \frac{1}{k^2}\sum_i^k V(\hat{\theta}_i) = \frac{kV(\hat{\theta}_i)}{k^2} = \frac{V(\hat{\theta}_i)}{k}.$$

Además, un estimador insesgado de esta varianza es :

$$\hat{V}(\hat{\theta}) = \frac{1}{k(k-1)}\left(\sum_i^k \hat{\theta}_i^2 - k\hat{\theta}^2\right)$$

En nuestro caso, para el año base del panel, utilizamos 20 muestras independientes de tamaño 20.000 declaraciones del IRPF. El estimador de la varianza para el índice de Gini será:

$$\hat{V}(\hat{G}) = \frac{1}{k(k-1)}\left(\sum_i^k \hat{G}_i^2 - k\hat{G}^2\right) \qquad \hat{G} = \frac{1}{k}\sum_i^k \hat{G}_i$$

Para nuestros datos en el año base del panel tenemos:

$$\hat{V}\left(\hat{G}\right)=\frac{1}{k\left(k-1\right)}\left(\sum_{i}^{k}\hat{G}_i^2-k\hat{G}^2\right)=0{,}00000293063$$

Este error absoluto lo relativizamos a partir del coeficiente de variación, lo que nos lleva al siguiente resultado:

$$\hat{C}_V\left(\hat{G}\right)=\frac{\sqrt{\hat{V}(\hat{G})}}{\hat{G}}=0{,}001295489$$

Podemos concluir por tanto que el error relativo para el estimador del índice de Gini es del 0,1295%, es decir, aproximadamente del uno por mil. Estamos ante un resultado óptimo derivado del elevado tamaño de las muestras, del elevado número de muestras y de la elevada precisión de las propias muestras.

El método de los grupos aleatorios

Se extrae una muestra de n unidades de una población de tamaño N. Dicha muestra se subdivide en K submuestras de igual tamaño m, de modo que n=K.m. Estas submuestras se denominan grupos aleatorios, y además de ser submuestras de la muestra, también son muestras de la población completa. La formación de los K grupos aleatorios de tamaño m dentro de una muestra W de tamaño n puede realizarse considerando una permutación aleatoria de los números 1,2,...,n y eligiendo el primer grupo aleatorio formado por los elementos de la muestra que ocupan los lugares definidos por los m primeros números de la permutación. El segundo grupo aleatorio se formará con los elementos de la muestra que ocupan los lugares definidos por el segundo conjunto de m números de la permutación. Así sucesivamente se formarán los K grupos aleatorios correspondientes a la muestra.

En estas condiciones si $\hat{\theta}$ es un estimador insesgado de la característica poblacional θ basado en la muestra completa W, y si $\hat{\theta}_r$ es un estimador insesgado de la característica poblacional θ basado en el r-ésimo grupo aleatorio, un estimador insesgado de la varianza de $\hat{\theta}$ es el siguiente:

$$\hat{V}(\hat{\theta})=\frac{1}{K(K-1)}\sum_{r=1}^{K}(\hat{\theta}_r-\hat{\theta})^2$$

Este método de los grupos aleatorios es igualmente válido si se subdivide la muestra completa W de tamaño n en K grupos aleatorios de distintos tamaños m_1, m_2,...,m_k cuya suma sea n. En este caso la condición $V(\hat{\theta}_r)=KV(\hat{\theta})$ se transforma en

$V(\hat{\theta}_r)=\frac{1}{\lambda_r}V(\hat{\theta})$ con $\lambda_r=\frac{m_r}{n}$. Tomando $\hat{\theta}=\sum_{r=1}^{K}\lambda_r\hat{\theta}_r$ tenemos que:

$$\frac{1}{K-1}\sum_{r=1}^{K}\lambda_r\left(\hat{\theta}_r-\hat{\theta}\right)^2 \; insesgado\ de\ \ V(\hat{\theta})$$

En nuestro caso utilizamos 20 submuestras independientes de tamaño 20.000 declaraciones del IRPF. El estimador de la varianza para el índice de Gini en el año base del panel será:

El estimador de la varianza para el índice de Gini será:

$$\hat{V}(\hat{G}) = \frac{1}{K(K-1)} \sum_{r=1}^{K} (\hat{G}_r - \hat{G})^2 \qquad \hat{G} = \text{Gini de la muestra inicial}$$

Para nuestros datos tenemos:

$$\hat{V}(\hat{G}) = \frac{1}{K(K-1)} \sum_{r=1}^{K} (\hat{G}_r - \hat{G})^2 = 0{,}00000297912$$

Este error absoluto lo relativizamos a partir del coeficiente de variación, lo que nos lleva al siguiente resultado:

$$\hat{C}_V\left(\hat{G}\right) = \frac{\sqrt{\hat{V}(\hat{G})}}{\hat{G}} = 0{,}001307335$$

Podemos concluir por tanto que el error relativo para el estimador del índice de Gini es del 0,13%, es decir, aproximadamente del uno por mil. Estamos ante un resultado equivalente al del método anterior.

Métodos Bootstrap o de autogeneración

El método de autogeneración (*bootstrap*) se emplea, entre otras cosas, para la estimación aproximada de varianzas para estimadores complejos.

Para llevarlo a cabo partimos de la muestra de tamaño 400000 declaraciones de IRPF extraída de una población de 16 millones de declaraciones en el año base. A continuación extraemos de la muestra inicial M=1000 muestras con reposición, también de tamaño 400000 y calculamos en cada una de ellas el estimador $\hat{\theta}_j^*$ para el cual estamos calculando el error (índice de Gini).

La precisión del estimador se obtiene por la expresión:

$$\hat{\sigma}_{BOOT} = \sqrt{\frac{\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)^2 - \left(\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)\right)^2 \Big/ M}{M-1}}$$

Para nuestros datos obtenemos:

$$\hat{\sigma}_{BOOT1000} = \sqrt{\frac{\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)^2 - \left(\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)\right)^2 \Big/ M}{M-1}} = 0{,}00052414$$

Para expresar el error anterior en términos relativos utilizamos el coeficiente de variación:

$$\hat{C}_V\left(\hat{G}\right) = \frac{\hat{\sigma}_{BOOT1000}}{\hat{G}} = 0{,}00125543$$

Si ahora consideramos M=5000 muestras, tenemos los siguientes resultados:

$$\hat{\sigma}_{BOOT5000} = \sqrt{\frac{\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)^2 - \left(\sum_{j=1}^{M}\left(\hat{\theta}_j^*\right)\right)^2 \Big/ M}{M-1}} = 0{,}00052414$$

Para expresar el error anterior en términos relativos utilizamos el coeficiente de variación:

$$\hat{C}_V\left(\hat{G}\right) = \frac{\hat{\sigma}_{BOOT5000}}{\hat{G}} = 0{,}00122569$$

Se observa que con el método Bootstrap nos movemos también en un error aproximado del uno por mil. Además, al elevar el número de muestras del método Bootstrap se gana precisión. La cuantificación de esta ganancia en precisión s del 2,334%.

Por otro lado, todo los métodos obtienen una precisión parecida para el estimador del ïndice de Gini.

Al repetir los cálculos para los diferentes años del panel, se obtienen resultados similares.

6. Implicaciones de política fiscal

Los modelos de microsimulación sobre datos de panel provenientes de registros administrativos no sólo permiten evaluar los efectos de las políticas públicas actuales, sino también las de sus posibles reformas, a través de la proyección de los probables cambios normativos sobre una base de datos representativa en el tiempo de la población afectada. Para comparar una misma población en dos situaciones distintas en el tiempo (la inicial y la que resultaría de la hipotética aplicación de los cambios previstos) es ineludible disponer de la dimensión temporal que ofrecen los paneles de datos y de la medición de las observaciones sobre los mismos elementos de la población en los distintos momentos del tiempo. Los errores relativos acotados en las estimaciones facilitan ese trabajo.

A través de los datos de panel de IRPF es posible evaluar con fiabilidad (errores mínimos) los factores que explican los movimientos en la recaudación a lo largo del tiempo, o la medición del impacto sobre el tipo medio efectivo del impuesto derivado de cambios legales que afecten a la tarifa y a las deducciones del IRPF, o el efecto redistributivo del IRPF y el grado de progresividad del gravamen a través de los índices habituales. La medición de los efectos recaudatorios y redistributivos de una reforma y sus implicaciones sobre el bienestar social se ven muy favorecidos cuando se dispone de datos de panel y se aplican las técnicas adecuadas de microsimulación sobre ellos.

Si además el panel se diseña con estratificación geográfica, es posible realizar estudios de ámbito territorial. De esta forma será posible, por ejemplo, estudiar las distintas alternativas de descentralización del IRPF como instrumento de financiación autonómica y los efectos redistributivos de la cesión del impuesto a las Comunidades Autónomas. Se pueden así medir los cambios distributivos de la renta en las CC AA derivados de sucesivas reformas, así como simular escenarios de descentralización de la imposición sobre la renta personal. Juega un papel clave en todo ello la estimación de las magnitudes con errores conocidos y óptimos.

7. Bibliografía

[1]. **PROGRESIVIDAD Y REDISTRIBUCIÓN EN EL IRPF ESPAÑOL: PANEL 82-98. ONRUBIA, RODADO, SARRALDE, PÉREZ (PAPEL DE TRABAJO IEF 23/2006)**

[2]. PANEL DE RENTA DEL INSTITUTO DE ESTUDIOS FISCALES 1999/2007: CÉSAR PÉREZ, FIDEL PICOS Y JORGE ONRUBIA (IEF-2010).

[3]. **MICROSIMULACIÓN MEDIANTE FUSIÓN DE PHOGUE Y PANEL DE DECLARENTES PARA EVALUAR REFORMAS FISCALES: FIDEL PICOS. REVISTA DE ECONOMÍA APLICADA Nº 41 VOLUMEN 14 (2006)**

[4]. **MODELOS DE MICROSIMULACIÓN: APLICCACIONES A PARTIR DEL PANEL DE DECLARANTES POR IRPF DEL INSTITUTO DE ESTUDIOS FISCALES. AYALA, ONRUBIA Y RUIZ HUERTA. ICE Nº 68 (2004)**

[5]. LA MUESTRA DE DECLARANTES DE IRPF 2002/2003: DESCRIPCIÓN GENERAL Y PRINCIPALES MAGNITUDES. DOCUMENTOS DE TRABAJO DEL IEF 15/05 Y 20/06 FIDEL PICOS SÁNCHEZ, MARÍA ANTIQUEIRA PÉREZ, CÉSAR PÉREZ LÓPEZ, ALFREDO MORENO SÁEZ, CARMEN MARCOS GARCÍA Y SANTIAGO DÍAZ DE SARRALDE MIGUEZ

[6]. LA MUESTRA DE DECLARANTES DE IRPF 2004: DESCRIPCIÓN GENERAL Y PRINCIPALES MAGNITUDES. DOCUMENTOS DE TRABAJO DEL IEF 25/07. FIDEL PICOS SÁNCHEZ, CÉSAR PÉREZ LÓPEZ, ALFREDO MORENO SÁEZ Y SANTIAGO DÍAZ DE SARRALDE MIGUEZ .

[7]. LA MUESTRA DE DECLARANTES DE IRPF 2005: DESCRIPCIÓN GENERAL Y PRINCIPALES MAGNITUDES. DOCUMENTOS DE TRABAJO DEL IEF 9/09. FIDEL PICOS SÁNCHEZ, CÉSAR PÉREZ LÓPEZ, ALFREDO MORENO SÁEZ, SANTIAGO DÍAZ DE SARRALDE MIGUEZ Y CARMEN GONZALEZ QUEIJA.

[8]. LA MUESTRA DE DECLARANTES DE IRPF 2006: DESCRIPCIÓN GENERAL Y PRINCIPALES MAGNITUDES. DOCUMENTO DE TRABAJO DEL IEF 28/09. FIDEL PICOS SÁNCHEZ, CÉSAR PÉREZ LÓPEZ, Y MARÍA DEL CARMEN GONZÁLEZ QUEIJA.

[9]. TÉCNICAS DE MUESTREO ESTADÍSTICO. CÉSAR PÉREZ LÓPEZ. GARCETA (2010)

[10]. **MUESTREO ESTADÍSTICO. CONCEPTOS Y PROBLEMAS RESUELTOS. CÉSAR PÉREZ LÓPEZ. PEARSON EDUCACIÓN – PRENTICE HALL (2005)**

[11]. TÉCNICAS DE MUESTREO ESTADÍSTICO. TEORÍA, PRÁCTICA Y APLICACIONES INFORMÁTICAS. CÉSAR PÉREZ LÓPEZ. RAMA (1999)