

VARIABLES EXPLICATIVAS DE LA PERCEPCIÓN IMPOSITIVA DE LOS CONTRIBUYENTES ESPAÑOLES: UN ANÁLISIS MEDIANTE ALGORITMOS GMDH

Javier Salinas Jiménez
(Instituto de Estudios Fiscales)
Daniel Santín González
(Universidad Complutense de Madrid)
José Félix Sanz Sanz
(Instituto de Estudios Fiscales)

Resumen:

Este trabajo estudia cuáles son las variables explicativas de la percepción impositiva que los contribuyentes españoles tienen de los tipos medios y marginales que pagan por el Impuesto de la Renta de las Personas Físicas (IRPF). Para ello se utiliza la encuesta que realiza anualmente el Instituto de Estudios Fiscales sobre las actitudes fiscales de los españoles. Los resultados obtenidos nos permiten concluir que, en media, los contribuyentes españoles sobreestiman el tipo medio de IRPF subestimando el marginal. Además, determinadas variables socioeconómicas individuales influyen en la cuantía de sobreestimación, acierto o subestimación de los tipos impositivos pero de forma no lineal. Este resultado ha podido ser contrastado comparando la bondad de ajuste obtenida por un modelo de regresión lineal múltiple y un algoritmo inductivo de redes neuronales cuyo objetivo es encontrar el modelo que mejor ajusta los datos mediante la construcción de polinomios (*Polynomial Neural Networks*). Esta metodología, también conocida como *The Group Method of Data Handling* (GMDH), [Farlow, 1984], es óptima para aquellos problemas económicos en los cuáles carecemos de un modelo teórico que respalde la teoría económica y la econometría clásica no encuentra una solución satisfactoria. En este caso, el algoritmo GMDH permitió descubrir relaciones no lineales entre variables socioeconómicas, tales como ingresos, edad, sexo y tamaño de la población de residencia y la percepción de tipos impositivos medio y marginal en IRPF, que no hubieran sido encontradas por un modelo lineal.

Códigos JEL: C45, D70, H24.

Palabras clave: percepción impositiva, GMDH, tipo medio IRPF, tipo marginal IRPF.

1. INTRODUCCIÓN.

En la mayoría del trabajo aplicado en donde se analiza la influencia de los impuestos se suele asumir, sin discusión, que los contribuyentes perciben correctamente el gravamen efectivo que soportan. Sin embargo, desde hace algún tiempo existe una creciente literatura que sugiere que las percepciones impositivas de los contribuyentes no tienen porque ajustarse necesariamente a la realidad [Rosen (1976), Fujii y Hawly (1988), Bartolome (1995), Arrazola et al (2000) y Gemmall et al (1999, 2000, 2001), Hettich y Winer (1988)]. Esta nueva hipótesis sugiere que para evaluar de manera correcta el impacto económico de los impuestos se hace necesario distinguir entre tipos impositivos formales y percibidos, siendo estos últimos los relevantes para explicar el comportamiento de los agentes económicos ante un shock impositivo. No tener en cuenta este desajuste entre tipos formales y percibidos puede conducir, por tanto, a evaluar erróneamente las implicaciones de la política impositiva y su impacto sobre el bienestar individual de los agentes económicos y sobre el grado de eficiencia de la economía.

La relevancia de esta cuestión exige contrastar empíricamente la existencia o ausencia de divergencias entre tipos reales y tipos impositivos percibidos. Este trabajo pretende aportar esa evidencia para el caso español a partir del análisis de una encuesta sobre actitudes fiscales. Además, se analizarán cuales son las variables socioeconómicas que influyen en la sobreestimación, subestimación o acierto de los tipos medio y marginal en el IRPF. En particular se estudiará si las diferencias entre los tipos impositivos reales y percibidos se relacionan de forma no lineal con determinadas variables tales como el nivel de ingresos, el sexo o el tamaño del municipio entre otras.

La estructura del artículo responde al siguiente esquema. Después de esta breve introducción, en el segundo epígrafe se exponen las principales características de los algoritmos GMDH. En el tercer epígrafe se describe la base de datos utilizada y se exponen los criterios adoptados para estimar el grado de corrección de la percepción impositiva. En el último epígrafe se presentan los principales resultados obtenidos y se presentan las principales conclusiones del trabajo.

2. LOS ALGORITMOS GMDH.

A la hora de contrastar empíricamente un modelo económico, normalmente se utilizan una serie de supuestos teóricos microeconómicos acerca del problema en cuestión. En muchas ocasiones estos supuestos se corresponden con modelos sólidos y ampliamente contrastados en la teoría económica. Sin embargo, como señalan Blundell et al. (1998, pp, 62), la obsesión con la linealidad de las variables en el análisis económico empírico, no está en ocasiones justificada por ninguna teoría económica fuerte que la respalde. En estos casos, los métodos denominados no paramétricos y semiparamétricos como *kernel regression* [Haerdle, 1990], *splines regression* [Eubank, 1999] o *projection pursuit regression*. [Friedman y Stuetzle, 1981] entre muchos otros son una solución para la búsqueda de posibles no linealidades.

Actualmente, existe una amplia gama de problemas en los que carecemos de un modelo económico teórico, siquiera aproximado, que relaciones variables y resultados. Este sería el caso que nos ocupa, acerca de las variables que explican que los individuos perciban tipos diferentes a los reales. Las percepciones podrían variar en función de distintas relaciones entre las variables: sexo, edad, nivel de estudios, población de residencia, ingresos y un largo etcétera. La intuición nos indica que todas estas variables pueden estar detrás de esta explicación aunque se desconoce la forma en que se relacionan. Si en un modelo de este tipo existieran relaciones no lineales más fuertes que las relaciones lineales éstas no serían encontradas usando modelos lineales, [Cohn et al., 1990, pp. 166]. Ello significaría una mala especificación del modelo que llevaría en su caso a una toma de decisiones incorrecta.

En estos casos podemos recurrir a buscar el modelo mediante métodos inductivos que nos ayuden a comprender mejor la complejidad de un determinado problema económico. La técnica GMDH (*Group Method of Data Handling*) sería una solución para enfrentarnos al análisis de estos problemas como posible alternativa a las técnicas tradicionales. A este grupo de técnicas también pertenecerían los árboles de decisión [Haerdle, 1990] o las redes neuronales artificiales [Bishop, 1996].

Los algoritmos inductivos de tipo GMDH, fueron propuestos por Ivakhnenko (1971) con el objetivo de identificar relaciones no lineales entre variables input y output. Este algoritmo consiste en construir, a partir de un conjunto de variables explicativas y una variable respuesta, aquel polinomio que mejor ajusta los datos del problema input-output cumpliendo

una serie de criterios previamente especificados. El polinomio final es normalmente implementado a través de una especificación de tipo Kolmogorov-Gabor. Esto es:

$$y = \beta_0 + \sum_{i=1}^M \beta_i x_i + \sum_{i=1}^M \sum_{j=1}^M \beta_{ij} x_i x_j + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M \beta_{ijk} x_i x_j x_k$$

Donde y es la respuesta deseada, $X(x_1, x_2, \dots, x_m)$ es el vector de variables explicativas o inputs y $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ es el conjunto de coeficientes o pesos.

Para la obtención del resultado final, se crea un conjunto de polinomios inicial que tiene en cuenta distintas interacciones entre las variables input. Aquellos que mejor ajustan los datos evolucionan y pasan a una siguiente etapa o capa, de tal manera que los que peor ajustan desaparecen. Estos modelos son reemplazados en sucesivas capas por nuevas combinaciones híbridas de polinomios hasta que no sea posible encontrar un modelo que ajuste mejor los datos¹. Con el fin de evitar una excesiva complejidad en la estructura del modelo ésta es controlada por el número de términos e interacciones que componen el polinomio de acuerdo con algún criterio de penalización. Aunque en principio podría pensarse que un polinomio más complejo supone automáticamente un modelo más predictivo, Farlow (1984) demuestra que la relación entre complejidad y capacidad predictiva tiene forma de “V”.

Por tanto, para evitar que el modelo final tenga un número excesivo de parámetros existen dos estrategias posibles. Por un lado se puede dividir la muestra en dos submuestras de forma que el modelo ajustado sea posteriormente validado con los datos reservados. Como alternativa, se puede emplear un término de penalización en el ajuste si la muestra disponible no es demasiado grande². Además, de forma previa a su aplicación las variables explicativas son escaladas linealmente³.

En economía el algoritmo GMDH ha sido aplicado para predecir variables macroeconómicas en Maddala et al. (1994, cap. 6); predicción de cifras de gasto público en

¹ No es el objetivo de este artículo realizar una revisión exhaustiva de los métodos de GMDH. Para este fin puede acudir a Farlow (1984), Oh et al. (2002) y Madala et al. (1994)

² En este trabajo hemos usado esta última técnica. El término de control fue el error de la predicción al cuadrado definido como $PSE = MSE + h \times var(a) \times k/n$. Donde MSE es el error en la respuesta promedio al cuadrado, N es el número de datos muestrales, k el número de coeficientes en el modelo, var(a) es la varianza de la respuesta real y h un coeficiente para penalizar el sobreajuste del modelo.

³ $X_i = (X_i - Min) / (Max - Min) \times 1$

Baker et al. (1999) o para modelar la función de producción educativa en Baker (2001). Liao (1992) afirma que los modelos neuronales de tipo GMDH pueden ser útiles a la hora de proporcionar nuevo conocimiento acerca de relaciones no lineales en las variables de problemas sociales.

3. ANALISIS EMPÍRICO DE LA PERCEPCIÓN IMPOSITIVA

3.1. DATOS EMPLEADOS

Para realizar el análisis empírico, hemos utilizado los datos correspondientes al año 2001 de la encuesta anual “Opiniones y actitudes fiscales de los españoles”, elaborada por el Instituto de Estudios Fiscales.

El tipo de muestreo utilizado para la selección de la muestra ha sido el de conglomerados, polietápico con estratificación de las unidades de primera etapa (secciones censales con representación en las diecisiete Comunidades Autónomas y en cinco estratos de hábitat por tamaño de población) y subestratificación de la última unidad muestral (individuos), según la ocupación o actividad principal de los miembros componentes de los hogares elegidos⁴. El tamaño muestral es de 1.506 entrevistados. Para optimizar esta muestra se ha realizado una afijación muestral no proporcional, asignando 1.155 unidades para los segmentos de los económicamente activos y 351 unidades para los no activos.

En la edición correspondiente al año 2001 se incluyeron, por primera vez, cinco cuestiones dirigidas a obtener información sobre la percepción de los españoles respecto a su carga tributaria percibida en los dos impuestos con mayor volumen de recaudación, IRPF e IVA. Asimismo, se incorporan preguntas relativas a las preferencias de los entrevistados sobre varias alternativas de reforma fiscal. Junto a estas nuevas cuestiones, la encuesta recoge también otras variables relacionadas con características personales del entrevistado y de su hogar⁵.

⁴ Se han considerado los siguientes segmentos: Empresarios no agrarios, empresarios agrarios, profesionales, trabajadores del sector público, trabajadores del sector privado y no activos económicamente (jubilados y pensionistas, estudiantes, amas de casa y parados) categoría que sirve de referencia en este trabajo.

⁵ En este conjunto de variables se encuentran, entre otras, el estado civil, la edad, los ingresos brutos mensuales del entrevistado y de su cónyuge, la existencia de minusválidos en el hogar y el número de hijos y de otros descendientes y ascendientes.

En relación con la percepción impositiva del IRPF, único impuesto que abordaremos en esta investigación, y el problema que nos planteamos en este trabajo, los entrevistados contestaron a las siguientes cuestiones en relación al año de referencia de la encuesta:

1.- El porcentaje de ingresos que cree que pagó a Hacienda en concepto de IRPF, tanto él como su cónyuge (tipo medio de IRPF).

2.-El porcentaje de sus ingresos que cree que tendría que pagar a Hacienda en concepto de IRPF, por ingresar 100.000 pesetas más que el año anterior (tipo marginal de IRPF).

Con el fin de determinar el grado de corrección de las percepciones de los entrevistados, se ha estimado el tipo medio y marginal de IRPF que soportó cada entrevistado, comparándolos posteriormente con los tipos percibidos.

En relación con el tipo medio de IRPF, se ha estimado para cada individuo la cantidad que debe ingresar por este impuesto, calculándose el tipo medio como el porcentaje de sus ingresos brutos que supone dicha cantidad. Dado que a partir de la Encuesta utilizada sólo conocemos el tramo de ingresos en el que se sitúa el entrevistado, se han calculado dos tipos medios, uno para el extremo inferior del tramo y otro para el superior. De esta forma, si el tipo medio percibido por el entrevistado se encuentra entre los dos tipos medios, se ha considerado que la percepción del entrevistado era correcta y la diferencia cero. Si la respuesta del entrevistado es inferior al menor de los tipos medios estimados, entonces el entrevistado subestima lo que paga, obteniéndose la diferencia mediante la resta entre su respuesta y este mínimo. De manera similar, si la respuesta es superior al mayor de los tipos medios calculados, entonces el entrevistado sobreestima lo que paga, obteniéndose la diferencia mediante la resta entre su respuesta y este máximo.

Del mismo modo, por lo que respecta al tipo marginal del IRPF, se ha considerado como intervalo aceptable el comprendido entre los tipos marginales correspondientes a los extremos máximo y mínimo del intervalo de renta en el que se encuentra el entrevistado más/menos un punto. Si el tipo percibido por el entrevistado está comprendido en el intervalo definido, se considera que la percepción es correcta. Si por el contrario, el tipo percibido es menor al mínimo, el contribuyente subestima su tipo marginal. En cambio, si es superior al máximo, sobreestima el tipo marginal que le corresponde. Al igual que ocurría con el tipo medio, el

error de percepción es cero para los que tienen una percepción correcta; para el resto se ha considerado el resultado de restar al tipo percibido el extremo inferior del intervalo para los que subestiman y el superior para los que sobreestiman.

Para estimar las diferencias entre los tipos percibidos y los reales sólo se han utilizado aquellos cuestionarios que contestaron a las tres preguntas relativas a las percepciones. Esto ha exigido realizar una reponderación de la muestra con objeto de ajustarla a la población. Así, en los resultados para el tipo medio de IRPF se han considerado 392 cuestionarios y 393 para el tipo marginal.

En la tabla 1 se ofrecen los resultados obtenidos sobre las percepciones impositivas de los contribuyentes analizados.

Tabla 1. Percepciones impositivas.

	Porcentaje horizontal			Descriptivos	
	Subestiman	Correcta	Sobreestiman	Media	Desviación típica
Tipo medio IRPF	16,3	23,0	60,7	4,23	8,38
Tipo marginal IRPF	86,8	8,4	4,8	-10,2	8,735

Como puede observarse, el 23% de los contribuyentes tienen una percepción acertada de su tipo medio de IRPF, el 16,3% subestiman dicho tipo mientras que el 60,7% lo sobreestiman. La diferencia media entre el tipo medio percibido y el real es de 4,23 puntos con una desviación típica de 8,38. Por lo que respecta al tipo marginal de IRPF, solamente es percibido de forma correcta por un 8,4% de los contribuyentes, mientras que el 86,8% subestiman lo pagado por este concepto y el 4,8% lo sobreestima. La media de la diferencia entre el tipo percibido y el real es de -10,2 puntos con una desviación típica de 8,735.

3.2. VARIABLES EXPLICATIVAS Y METODOLOGÍA

Con objeto de ahondar en las posibles causas que explican la percepción impositiva de los tipos medio y marginal de los contribuyentes se ha estimado, por mínimos cuadrados ordinarios (MCO) y con algoritmos GMDH, un modelo en el que la variable dependiente es la diferencia entre el tipo percibido y el real. Además, hemos utilizado como metodología para encontrar el modelo que mejor aproxime los datos, la regresión por pasos sucesivos o

stepwise regression, [Darlington, 1990]. A partir de un modelo que se hace explícito se va alterando el mismo iterativamente, introduciendo y extrayendo variables, en base a algún criterio de entrada o salida definido previamente. La búsqueda se detiene cuando los criterios de entrada-salida se detienen o tras un número de pasos definido.

El criterio de entrada-salida seguido es el conocido como *pasos sucesivos*. Mediante este método en cada paso se introduce la variable independiente que no se encuentre ya en la ecuación y que tenga la probabilidad para F más pequeña, si esa probabilidad es suficientemente pequeña. Las variables ya introducidas en la ecuación de regresión se eliminan de ella si su probabilidad para F llega a ser suficientemente grande. El método termina cuando ya no hay más variables candidatas a ser incluidas o eliminadas⁶. La elección de un modelo de este tipo es especialmente útil como mecanismo para eliminar determinadas relaciones no lineales encontradas por el algoritmo GMDH que pudieran ser reiterativas y sólo contribuyan a dotar de mayor complejidad el modelo. En aras de comparar los resultados también se ha aplicado, además de introducir todas las variables y aplicar MCO, regresión por pasos sucesivos en el modelo lineal.

Como variables explicativas se han considerado la ocupación, el nivel de estudios, la edad, el tamaño del hogar, el estado civil, el logaritmo neperiano de los ingresos netos y el tamaño del municipio de residencia. Para ambos modelos se usaron exactamente el mismo número de variables.

⁶ Los valores de entrada y de salida utilizados fueron 0,05 y 0,1 respectivamente.

4. RESULTADOS.

4.1. LA PERCEPCIÓN DEL TIPO MEDIO.

Los resultados de las estimaciones para el tipo medio se ofrecen en la tabla 2.

Tabla 2. Estimación MCO de la percepción impositiva del tipo medio en IRPF.

	Tipo medio IRPF (Introducir)	Tipo Medio IRPF Stp, pasos suc	Tipo Medio IRPF (stepwise GMDH)
Constante	23,468 (0,00)	23,636 (0,00)	21,504 (0,00)
Sexo (Hombre = 0; Mujer=1)			
Edad	0,615 (0,02)		
Tamaño hog(nº miembros)	1,059 (0,02)	1,277 (0,00)	1,252 (0,00)
Habitat1 (Menos 10.000)	2,778 (0,03)		1,983 (0,00)
Habitat2 (Entre 10.000 y 50.000)			
Habitat3 (Entre 50.001 y 200.000 y cp)	3,014 (0,04)		
Habitat4 (Entre 200001 y 500000)*			
Empresarios no agrarios		2,637 (0,01)	4,361 (0,00)
Empresarios agrarios			3,541 (0,00)
Profesionales			
Asalariados			
Asalariados en sector público			
Soltero			
Casado			
Edad al cuadrado	-0,006 (0,03)		
Est. básicos (EGB)			
Estudios medios (BUP, FP)			
Est. Universitarios 1 ^{er} ciclo**	-2,839 (0,05)		
LogN (ingresos netos)***	-4,924 (0,00)	-3,281 (0,00)	-3,308 (0,00)
INTERACCIONES GMDH****			
Asal. públicos x habitat1			
Asalariados x Est. medios			
sexo x habitat3			132,92 (0,00)
sexo x tamaño hogar			
univers (ciclo corto) x LN(ing. netos)			
Edad x LN(ing. netos)			
soltero x LN(ing. netos)			
habitat3 x sexo x [LN(ing.netos)] ²			
profesional x edad al cuadrado			0,001709 (0,00)
sexo x habitat3 x LN(ing. netos)			-18,444 (0,00)
Contraste F	3,076 (0,00)	10,858 (0,00)	12,974 (0,00)
R ² corregido	0,092	0,071	0,198

Nota: Entre paréntesis los p-valores. Para p-valor menor o igual a 0,05 el parámetro es significativo al 5%. Para p-valor menor a 0,1 el parámetro es significativo al 90%.

* La categoría de referencia es el hábitat 5? Más de 500.000 habitantes.

** La categoría de referencia es poseer estudios universitarios superiores.

*** Para ambos tipos impositivos los ingresos se refieren a los del entrevistado.

**** Interacciones descubiertas por el modelo GMDH.

De los tres modelos considerados el que logró un mayor ajuste fue el que tuvo en cuenta las interacciones descubiertas por el algoritmo GMDH. Se puede comprobar como el ajuste de este modelo en términos del coeficiente de determinación corregido supera en casi el doble al del modelo estimado mediante MCO.

La diferencia entre los tipos medios por IRPF percibidos y reales está positivamente relacionada con la edad, el tamaño del hogar y determinadas profesiones, y de forma negativa con el nivel de ingresos. Se observa además como los habitantes en municipios de entre 50.000 y 200.000 habitantes parecen tener un comportamiento distinto al resto. El algoritmo GMDH detecta además que dentro del hábitat 3 las mujeres presentan un comportamiento especial. Las tablas 3 y 4 muestran como el ajuste del modelo GMDH es mayor que en el resto de modelos para el hábitat 3 (tabla 3) y para las mujeres que residen en municipios de este tamaño.

Tabla 3: Coeficientes de correlación entre valores reales y estimaciones para individuos residentes en municipios de entre 50.000 y 200.000 habitantes.

		Tipo Medio IRPF	Tipo Medio IRPF Stepwise	Tipo Medio IRPF Introducir	Tipo Medio IRPF GMDH Stepwise
Tipo Medio IRPF	Correlación de Pearson	1	,415(**)	,439(**)	,645(**)
	Sig. (bilateral)	.	,000	,000	,000
	N	71	71	71	71

** La correlación es significativa al nivel 0,01 (bilateral).

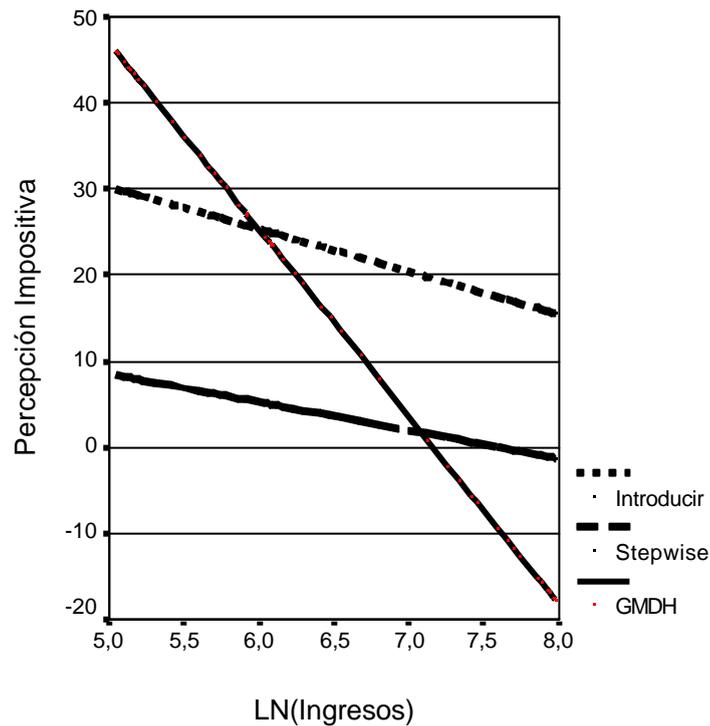
Tabla 4: Coeficientes de correlación entre valores reales y estimaciones para mujeres residentes en municipios de entre 50.000 y 200.000 habitantes.

		Tipo Medio IRPF	Tipo Medio IRPF Stepwise	Tipo Medio IRPF Introducir	Tipo Medio IRPF GMDH Stepwise
Tipo Medio IRPF	Correlación de Pearson	1	,506(**)	,572(**)	,764(**)
	Sig. (bilateral)	.	,010	,003	,000
	N	25	25	25	25

** La correlación es significativa al nivel 0,01 (bilateral).

En el gráfico 1 se presenta una simulación de cómo varía la percepción impositiva del tipo medio con el nivel de ingresos para las mujeres en el hábitat 3 teniendo en cuenta los coeficientes de los tres modelos.

Gráfico 1: Simulación* del cambio estimado en la percepción impositiva del tipo medio con los ingresos a partir de los distintos modelos construidos para mujeres residentes en municipios de entre 50.000 y 200.000 habitantes.



* Para individuos que viven solos con una edad media de 44 años.

En esta simulación se observa como el modelo GMDH estima una caída abrupta de la percepción del tipo medio a medida que aumenta el nivel de ingresos respecto a las estimaciones realizadas por los métodos lineales.

4.2. LA PERCEPCIÓN DEL TIPO MARGINAL

Los resultados de las estimaciones para el tipo medio se ofrecen en la tabla 5.

Tabla 5: Estimación MCO de la percepción impositiva del tipo marginal en IRPF.

	Tipo marginal IRPF	Tipo marginal Stp. pasos suc.	Tipo Marginal IRPF (stepwise GMDH)
Constante		15,660 (0,00)	-44,966 (0,00)
Sexo			
Edad			
Tamaño hog(n° miembros)		0,879 (0,02)	
Habitat1			
Habitat2	2,744 (0,06)		
Habitat3			
Habitat4			
Empresarios no agrarios			
Empresarios agrarios			
Profesionales			
Asalariados			
Asal. en sector público			
Soltero			
Casado			
Edad al cuadrado			
Est. Básicos (EGB)			
Estudios medios (BUP, FP)			
Est. Universitarios 1 ^{er} ciclo			
LogN (ingresos netos)***	-4,594 (0,00)	-4,017 (0,00)	
INTERACCIONES GMDH			
sexo x habitat3			
edad x Est. basicos			
asalariados x Est. basicos			
Emp. no agrarios x LN(ing. netos)			
habitat3 x sexo x LN(ing.netos)			
sexo x asalariados x LN(ing.netos)			-0,649 (0,01)
sexo x asalariados x LN(ing.netos) x h3			1,675 (0,00)
sexo x [LN(ing.netos)] ³			
sexo x Est. univers x [LN(ing.netos)] ²			
Est. Univers x [LN(ing.netos)] ²			
edad al cubo			
edad x LN(ing.netos)			
edad al cubo x LN(ing.netos)			
sexo x edad al cuadrado			0,0009446 (0,04)
Asalariados en s. publico x basicos			6,870 (0,02)
habitat1 x est. medios			-3,094 (0,02)
Tamaño hogar x LN(ing.netos)] ²			0,01889 (0,00)
[LN(ing.netos)] ²			2,865 (0,00)
[LN(ing.netos)] ³			-0,312 (0,00)
Contraste F	2,290 (0,00)	15,593 (0,00)	10,508 (0,00)
R ² corregido	0,059	0,070	0,164

Nota: Entre paréntesis los p-valores. Para p-valor menor o igual a 0,05 el parámetro es significativo al 5%. Para p-valor menor a 0,1 el parámetro es significativo al 90%.

* Para ambos tipos impositivos los ingresos se refieren a los del entrevistado

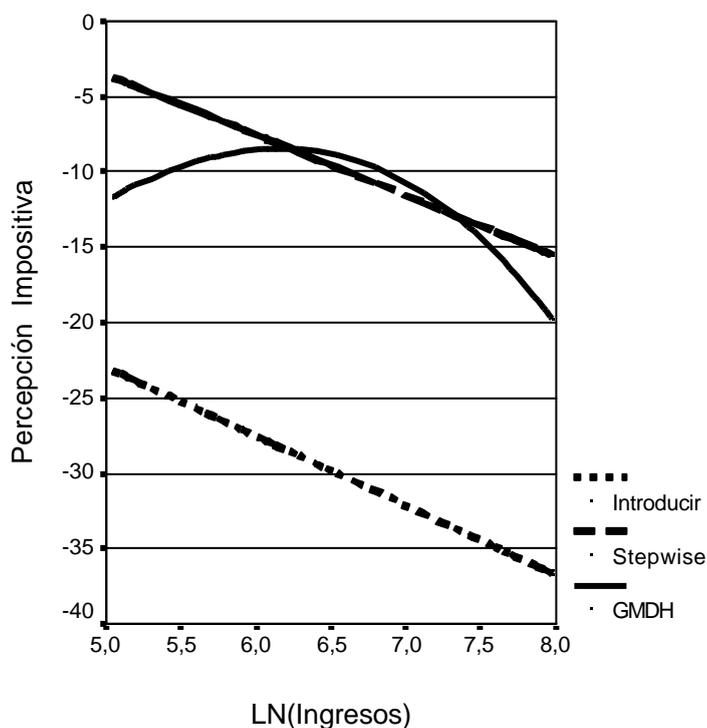
** Interacciones descubiertas por el modelo GMDH.

*** Para ambos tipos impositivos los ingresos se refieren a los del entrevistado.

De nuevo de los tres modelos considerados el que logró un mayor ajuste fue el algoritmo GMDH. Se puede comprobar como el ajuste de este modelo en términos del coeficiente de determinación corregido supera otra vez en casi el doble al del modelo estimado mediante MCO y por pasos sucesivos.

La diferencia entre los tipos marginales por IRPF percibidos y reales está positivamente relacionada con el sexo, la edad, el tamaño del hogar y los asalariados públicos con nivel de estudios básicos, y de forma negativa con ser asalariado y mujer o con individuos que viven en núcleos de población poco habitados y poseen estudios medios. En cuanto a los ingresos el efecto no es claro a primera vista dados los términos cuadrado y cubo encontrados. Para poder estudiar su efecto hemos simulado como variaría la percepción impositiva para un individuo varón no asalariado en el sector público que vive solo. En este análisis debemos tener en cuenta que la curva estimada podría quedar ligeramente desplazada por el lugar de residencia o por el nivel de estudios. El resultado final se muestra en el gráfico 2.

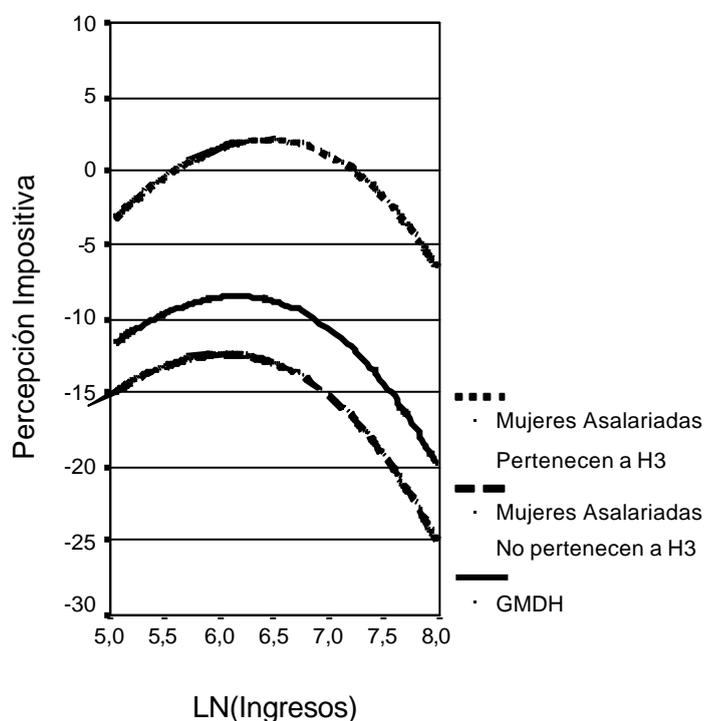
Gráfico 2: Simulación del cambio estimado en la percepción impositiva del tipo marginal con los ingresos a partir de los distintos modelos construidos.



Dentro de la tendencia general a la subestimación, el gráfico 2 muestra como hasta los 475 euros aproximadamente los varones tienden a mantener la percepción de los tipos marginales pagados con una tendencia positiva hacia el tipo marginal pagado. A partir de este nivel de ingresos la percepción caería hacia una mayor subestimación. En el gráfico 2 se muestra además como este comportamiento no lineal no es detectado por los modelos lineales estimados.

A continuación hemos desagregado en el modelo GMDH el efecto que tiene de nuevo la variable sexo en núcleos poblacionales de entre 50.000 y 200.000 habitantes. Los resultados se muestran en el gráfico 3.

Gráfico 3: Simulación del cambio estimado en la percepción impositiva del tipo marginal con los ingresos para mujeres asalariadas que pertenecen y no pertenecen al hábitat 3 y comparación con el modelo principal.

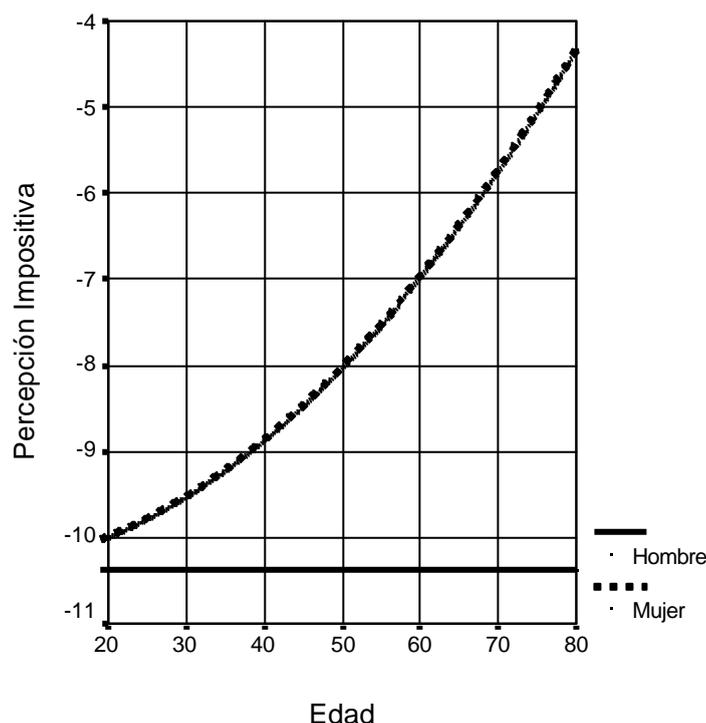


El gráfico muestra como las mujeres asalariadas en el hábitat 3 tienden en relación al resto a sobreestimar el tipo marginal pagado. El máximo lo alcanzarían las mujeres que perciben unos 640 €. A partir de esta cantidad de ingresos los resultados de percepción vuelven a decrecer. El comportamiento para las mujeres asalariadas que no viven en el hábitat 3 tiene

una tendencia parecida si bien el máximo se alcanza antes, en torno a los 430 € para posteriormente decrecer de forma más pronunciada que en el caso anterior. De nuevo en este gráfico queda reflejada la complejidad del problema a tratar. Siguiendo una tendencia parecida, la percepción varía a ritmos distintos a medida que incorporamos determinados efectos no lineales en el modelo.

Por último hemos simulado en el modelo GMDH, en esta ocasión fijando el nivel de ingresos en la media muestral, como varía la percepción del tipo marginal con la edad y el sexo. El resultado de este ejercicio se muestra en el gráfico 4.

Gráfico 4: Simulación del cambio estimado en la percepción impositiva del tipo marginal con la edad para mujeres y hombres no asalariados en el modelo GMDH.



* Para unos ingresos medios de 1034,32 €

El resultado muestra como las mujeres asalariadas tienden a aumentar la percepción del tipo marginal pagado a medida que cumplen años mientras que en el caso de los hombres esta percepción permanece invariable con la edad.

5. CONCLUSIONES

En este trabajo se ha podido comprobar como los algoritmos GMDH pueden ser una posible alternativa a los modelos tradicionales para descubrir no linealidades contenidas en los datos. Los beneficios de este tipo de técnica son más evidentes en aquellos problemas en los que no exista una teoría económica sólida explicativa del mismo.

Entre los principales resultados obtenidos en el análisis llevado a cabo, cabe destacar que la diferencia entre los tipos medios por IRPF percibidos y reales está positivamente relacionada con la edad, el tamaño del hogar y determinadas profesiones, y de forma negativa con el nivel de ingresos. En relación con la diferencia entre los tipos marginales por IRPF percibidos y reales, hay que indicar que parece estar positivamente relacionada con el sexo, la edad, el tamaño del hogar y los asalariados públicos con nivel de estudios básicos, y de forma negativa con ser asalariado y mujer o con individuos que viven en núcleos de población poco habitados y poseen estudios medios.

Creemos, en definitiva, que la riqueza de la información ofrecida por este tipo de análisis puede ayudar a evitar resultados erróneos que se producirían utilizando modelos convencionales. Así, en determinadas ocasiones el supuesto de linealidad supone una mala especificación del modelo. Por ello otra utilidad de este tipo de modelos sería contrastar la no existencia de este problema proporcionando mayor robustez a la hipótesis de linealidad. Las implicaciones que se derivan de este tipo de estudios en diferentes áreas del conocimiento pueden ser sumamente útiles en términos de política económica.

REFERENCIAS

- Arrazola, M., de Hevia J. y Sanz J. (2000): "More on tax perception and labour supply: the Spanish case", *Economics Letters*, 67, pp. 15-21.
- Baker, B. D. y Richards, C. E. (1999): "A comparison of conventional linear regression methods and neural networks for forecasting educational spending". *Economics of Education Review*, Vol.18, pp. 405-415.
- Baker, B. D. (2001): "Can flexible non-linear modeling tell us anything new about educational productivity?". *Economics of Education Review* 20 (2001) 81-92.
- Bartolome, C. A.M. (1995): "Which tax rate do people use: Average or marginal?". *Journal of Public Economics*, 56, pp. 79-96.
- Bishop, C. M. (1996): *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press.
- Blundell, R. y Duncan, A. (1998): "Kernel Regression in Empirical Microeconomics". *The Journal of Human Resources*. Winter 1998, Volume 33, Issue 1, pp. 62-87.
- Cohn, E. y Geske, T. (1990): Production and cost functions in education. In E. Cohn & T. Geske. *The economics of education* (3^a ed.). pp. 159-210. New York: Pergamon reprinted by Academic Advantage, Columbia, SC, 1998.
- Darlington, R. B. (1990): *Regression and Linear Models*. New York. Mc Graw-Hill Publishing Company.
- Eubank, R. L. (1999): *Nonparametric Regression and Spline Smoothing*, 2nd ed., Marcel Decker. New York.
- Farlow, S. J. (1984): *Self-organizing method of modelling: GMDH type algorithms*. Marcel Dekker, New York.
- Friedman, J. H. y Stuetzle, W. (1981): "Projection Pursuit Regression", *Journal of the American Statistical Association*, 76, 817-823.
- Fujii, E.T. y Hawley, C. B. (1988): "On the accuracy of tax perceptions", *The Review of Economics and Statistics*, 70 (2), pp. 344-347
- Greene, W. H. (1998): *Econometric analysis*, Prentice Hall.
- Haerdle, W. (1990): *Applied Nonparametric Regression*, Cambridge University Press. Cambridge.
- Instituto de Estudios Fiscales (2002) "Opiniones y actitudes fiscales de los españoles". Documento de Trabajo 14/02.

- Ivakhnenko, A. G. (1971): "Polynomial theory of complex systems", *IEEE Transactions on Systems, Man and Cybernetics*, SMC-1, pp. 364-378.
- Liao, T. F. (1992): "A modified GMDH approach for social science research: exploring patterns of relationships in the data". *Quality and Quantity*, 26, 19-38.
- Maddala, H. R. e Ivakhnenko, A. G. (1994): Inductive learning algorithms for complex systems modelling. Boca Raton, FL: CRC Press, Inc.
- Oh, S. y Pedrycz, W. (2002): "The design of self-organizing Polynomial Neural Networks". *Information Sciences*, 141, pp. 237-258.