

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES**  
**Departamento de Fundamentos del Análisis Económico II**



**INTERNET, BRECHA TECNOLÓGICA Y COMERCIO  
ELECTRÓNICO EN ESPAÑA**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**  
**PRESENTADA POR**

**Leonel Cerno**

Bajo la dirección del doctor:  
Teodosio Pérez Amaral

**Madrid, 2007**

- **ISBN: 978-84-669-3096-3**

**INTERNET, BRECHA TECNOLÓGICA Y COMERCIO  
ELECTRÓNICO EN ESPAÑA**

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO II

TESIS DOCTORAL

AUTOR:

**LEONEL CERNO**

DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO II  
UNIVERSIDAD COMPLUTENSE

DIRECTOR:

**TEODOSIO PÉREZ AMARAL**

DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO II  
UNIVERSIDAD COMPLUTENSE

NOVIEMBRE DE 2006

Al Pueblo del Chaco

## AGRADECIMIENTOS

Me gustaría agradecer primeramente a mi director Teodosio Pérez Amaral, a quién admiro y aprecio profundamente, por su constante apoyo y ayuda. Sus sugerencias han sido fundamentales tanto en las ideas originales así como en la calidad de la tesis.

También estoy agradecido a un gran número de personas que han aportado sus comentarios a lo largo de este período. En particular, me gustaría destacar los comentarios de los profesores Gary Madden y Russel Cooper, con quienes he coincidido en diversos congresos de la ITS tanto en España como en el extranjero. He de expresar igualmente mi gratitud al grupo de trabajo de la CMT, con quienes he discutido algunos de los temas abordados en esta investigación. No quiero olvidarme de agradecer profundamente a la profesora Francisca Cea D'Ancona, de la Universidad Autónoma de Madrid, con quien he compartido y discutido cuestiones de análisis multivariante abordados aquí. No puedo olvidar tampoco al profesor Alfonso Novales por sus valiosas sugerencias realizadas en mi seminario de tesis, las cuales he incorporado en este trabajo. A su vez, quiero agradecer al resto de los compañeros, amigos y profesores del departamento de Fundamentos del Análisis económico II de la Universidad Complutense, con los que he convivido los dos primeros años de cursos de doctorado y con quienes sigo en contacto permanente. Igualmente, deseo agradecer también a mis amigos y compañeros de trabajo de la Universidad Europea de Madrid, por su constante ánimo y por ofrecerme unas excelentes condiciones de trabajo.

El comienzo del camino que culmina con esta tesis fue posible gracias a la beca Fomec promovida por la Universidad Nacional del Nordeste (Argentina). El departamento de Economía auspició mi candidatura. Por eso no puedo dejar de mostrar mi agradecimiento a los profesores Antonio Besil, Susana Gelman, Mario Garber y Rubén Marcón, por haberme alentado y apoyado en este proyecto. Quiero agradecer en especial al profesor Miguel Angel Cantini, a quién siempre le deberé mi iniciación en el gusto por la estadística y la econometría.

Finalmente, y no por ello menos importante, en la elaboración de esta tesis doctoral ha resultado crucial el apoyo y comprensión de mi esposa Cecilia. Por ese motivo, he de mostrarle siempre mi más sincera gratitud.

# ÍNDICE

<i>Introducción</i> .....	1
---------------------------	---

## Capítulo I.

### **La Influencia de las Nuevas Tecnologías en la Economía**

1. Introducción.....	7
2. Revisión Bibliográfica.....	8
3. Fundamentos Teóricos.....	17
3.1. Introducción.....	17
3.2. Decisión en un Contexto de Alternativas Discretas.....	18
3.3. Teoría de la Utilidad Aleatoria.....	20
4. Conclusiones.....	23
5. Bibliografía.....	24

## Capítulo II.

### **Acceso y Uso de Internet en España**

1. Introducción.....	31
2. Análisis Gráfico – Descriptivo.....	34
2.1. Acceso a Internet desde el Hogar.....	35
2.2. Uso de Internet.....	38
3. Modelización.....	41
3.1. Funciones de Utilidad.....	42
3.2. Fundamentos para la Aproximación Econométrica.....	43
3.3. Especificación.....	48

4. Estimación y Discusión.....	50
4.1. Acceso a Internet desde el Hogar.....	51
4.2. Uso de Internet.....	55
5. Conclusiones.....	64
6. Referencias.....	65
Apéndice 1: Construcción de una Variable Proxy para la Renta Individual.....	70
Apéndice 2: Efectos Marginales de los Determinantes de la Probabilidad de Acceso a Internet por Banda Ancha desde el Hogar.....	75
Apéndice 3: Tablas Estadísticas.....	78
 Capítulo III.	
<b>Brecha Tecnológica en España</b>	
1. Introducción.....	83
2. Índice Sintético de Digitalización Geográfica para España.....	86
2.1. Un Índice de Digitalización por Comunidad Autónoma.....	88
2.1.1. Ponderaciones.....	89
2.2. Comparación de Índices y Predicción de la Brecha Tecnológica por Comunidad Autónoma.....	92
3. La Brecha Tecnológica y el Uso de Internet.....	100
3.1. Especificación: Uso de Internet por CCAA.....	102
3.2. Estimación y Resultados.....	103
3.3. Heterogeneidad.....	106
4. Determinantes del Grado de Digitalización en España.....	108
4.1. Especificación y Estimación MCO de una Función para el Grado de Digitalización Individual.....	109

4.2. Impacto de Determinantes a Distintos Niveles de Digitalización...	114
4.2.1. Regresión Cuantílica.....	115
4.3. Predicción del Nivel de Digitalización.....	123
4.3.1. El Algoritmo RETINA.....	123
4.4. Elección del Modelo Predictivo.....	127
4.4.1. Sensibilidad del Grado de Digitalización en las CCAA.....	128
4.6. Resumen de los Resultados.....	134
5. Conclusiones.....	135
6. Referencias.....	137
Apéndice 1: Definición de Variables.....	142
Apéndice 2: Resultados del Modelo de Poisson por Comunidad Autónoma.....	144
Apéndice 3: Conglomerados Jerárquicos de Comunidades Autónomas para Elasticidades de Uso de Internet con Respecto al Gasto en Equipamiento, la Edad y el Nivel de Estudios.....	149
Apéndice 4: Resultados de las Estimaciones RETINA.....	150

## Capítulo IV.

### **Comercio Electrónico en España**

1. Introducción.....	153
2. Comercio Electrónico y Uso de Internet.....	155
2.1. Análisis Descriptivo y Definición de Variables.....	157
2.1.1. Características de los Datos.....	158
2.2. El Modelo Logit.....	159
2.3. Compra en Internet: Especificación y Resultados.....	160
3. Modelos de Demanda de Comercio Electrónico.....	167

3.1. Descripción de Variables.....	170
3.2. Modelo de Demanda de Bienes Adquiridos en la Red.....	176
3.2.1. Resultados y discusión.....	177
3.3. Modelo de Previsión del Gasto.....	180
4. Conclusiones.....	184
5. Referencias.....	185
Apéndice 1: Gráficos de Efectos marginales (Modelo Logit).....	190
Apéndice 2: Función Lineal de Gasto con y sin Constantes Específicas.....	192
Apéndice 3: Función Lineal de Gasto Recomendada por RETINA.....	193

## Capítulo V.

### Conclusiones

1. Introducción.....	197
2. Conclusiones del Análisis del Acceso y Uso de Internet.....	198
2.1. Modelo para el Acceso a Internet desde el Hogar.....	198
2.2. Modelo para el Uso de Internet.....	199
3. Conclusiones de la Medición y Análisis de la Brecha Tecnológica.....	200
3.1. Modelo de Regresión Múltiple de los Determinantes del Nivel de Digitalización Individual en España.....	202
3.2. Modelo de Regresión Cuantílica de los Determinantes del Nivel de Digitalización Individual en España.....	203
3.3. Modelo de Previsión para el Nivel de Digitalización Individual en España.....	204
4. Conclusiones del Análisis del Comercio electrónico.....	206
4.1. Modelo de Impactos Sociodemográficos en la Compra de Bienes	

y Servicios en Internet.....	206
4.2. Modelos para la Cantidad de Transacciones a Través de Internet	
y Previsión del Gasto.....	208
5. Futuras Líneas de Investigación.....	209
<i>Cuadro Resumen de Ecuaciones Estimadas.....</i>	<i>211</i>
<i>Anexo: Base de Datos.....</i>	<i>213</i>

## INTRODUCCIÓN

Desde hace una década se viene dando un *boom* de la Economía Digital o Economía de la Información. Asimismo los años 90 se pueden considerar como la década de Internet. El incremento de la productividad en los países desarrollados durante estos años se debe en gran parte a la inversión en tecnologías de la información. Esto ha ido llevando a los agentes que forman parte de la economía a tomar conciencia de la importancia del mismo. Actualmente su estudio representa uno de los desafíos de la ciencia económica.

La potencialidad del fenómeno de Internet se observa en que es capaz de cambiar el tipo de relaciones sociales y la sociedad en su conjunto. En las próximas décadas seguramente guiará la economía y la sociedad aún con mayor contundencia.

En esta Tesis Doctoral se ha realizado un análisis del fenómeno enfocándolo por el lado de la demanda. Nos acercamos a la economía de la información desde tres perspectivas que son: los determinantes del acceso y el uso de Internet, la medición de la Brecha Digital entre Comunidades Autónomas y los determinantes del comercio electrónico. Para ello hemos utilizado una base de datos de las encuestas TIC-H (2003) que realiza anualmente el Instituto Nacional de Estadística (INE) seleccionando las variables a partir de los criterios teóricos en la literatura.

El primer paso de la presente investigación ha sido el estudio de los determinantes del acceso y uso de Internet en España. Los resultados obtenidos reflejan tanto similitudes como diferencias con estudios similares de otros países. Luego, y dada la necesidad de contar con un criterio de medición de los efectos de las diferencias en el acceso y uso de la tecnología digital, se propone un criterio de medición de la Brecha Tecnológica para las Comunidades Autónomas (CCAA). A partir de aquí se realizan

comparaciones, se miden las distancias y se investigan los determinantes que influyen en esta brecha entre CCAA.

Debido a las limitaciones del aparato teórico necesario para elaborar hipótesis sobre la forma funcional del proceso generador de las observaciones, se utiliza una herramienta flexible denominada RETINA (Relevant Transformations of the Input Network Approach) que selecciona la mejor especificación siguiendo un criterio de predicción fuera de la muestra. La utilidad de este algoritmo ha sido contrastado ampliamente en datos simulados. Actualmente ya se lo empieza a aplicar a datos reales. Tanto la elaboración de un indicador de Brecha Tecnológica como la obtención de un modelo para predicción de la misma suponen una de las principales aportaciones de esta Tesis Doctoral al permitir que tanto una como otra herramienta y su manera de utilizarla permitirá extender el análisis a otros ámbitos y objetivos distintos de los que se plantean aquí.

Por último se realiza un estudio de los determinantes de la demanda de bienes y servicios a través de la Red. La importancia de estudiar los determinantes de este fenómeno radica en la magnitud del crecimiento del gasto a través de Internet. En 2000 España gastó 213 millones de euros en comercio-e, mientras que en 2004 el gasto fue de 1.700 millones. Y este crecimiento no cesa. A partir de esta idea se obtiene una función de gasto en donde se observan los principales determinantes y se miden sus efectos.

La organización de los capítulos siguientes es la usual: en el capítulo primero se revisan algunos estudios precedentes que analizan el papel de Internet, la brecha tecnológica y el comercio electrónico en la economía, como también el marco teórico con el que luego se abordarán los análisis en los capítulos sucesivos. En el capítulo segundo se estudian los determinantes de acceso y uso de Internet en España, haciéndose un análisis exhaustivo y comparando los resultados entre sí y con los de la

literatura analizada previamente. En el capítulo tercero se propone un indicador que mide la brecha tecnológica entre CCAA y luego se estudian sus determinantes, eligiendo un modelo para predicción que sirva de herramienta de política. En el capítulo cuarto se estudian los determinantes que llevan al agente a elegir mercados virtuales en vez de los mercados convencionales, y luego se estima una función de gasto a través de la Red para el usuario de Internet en España. Por último, en el capítulo quinto se presentan las principales conclusiones y se hace un resumen de todos los modelos estimados en este trabajo.

# CAPÍTULO I

## La Influencia de las Tecnologías de la Información en la Economía

## **1. INTRODUCCIÓN**

En los últimos años, la economía del capital humano es más importante que la economía del capital físico. En países como Estados Unidos o España, la economía basada en el conocimiento representa respectivamente el 55,3% y el 38% del PIB anual. Los respectivos sectores de tecnologías de la información explican de forma directa el 8% y el 5% del PIB respectivamente. Si a esto le sumamos los impactos inducidos en forma indirecta, su influencia es aún mayor.

Debido al énfasis del papel jugado en estos impactos por las Tecnologías de la Información y de las Comunicaciones (TIC), y los cambios económicos producidos en cuanto al acceso a la información, los incrementos de productividad, el desplazamiento de lo tangible a lo intangible, etc., en las últimas dos décadas han proliferado trabajos que tienen como objetivo desarrollar modelos económicos de este fenómeno estudiándolo tanto por el lado de la oferta como por el lado de la demanda.

En los años 90, la tecnología que más incrementó su desarrollo y que se extendió hasta hacerse prácticamente imprescindible es Internet. En el año 2000, Internet comunicaba a 200 millones de personas. Hoy ya comunica a casi 1.000 millones. Este fenómeno impulsa además de los cambios económicos mencionados, cambios sociales tales como la ampliación del alcance y la potencialidad de las relaciones personales, cambio en los objetivos e instrumentos de la educación y de la formación, etc. Este fenómeno que emergió hace poco más de 20 años, hoy podemos decir que ha cambiado profundamente los hábitos cotidianos.

Así pues, en el presente capítulo se realiza una revisión de algunos de los principales trabajos que se centran en el estudio de los cambios económicos y sociales provocados por Internet y las nuevas tecnologías en general. Luego se exponen las metodologías

alternativas existentes a la hora de estimar funciones para los determinantes del acceso y uso de Internet, determinantes de la compra de bienes y servicios en la Red y funciones de gasto. De manera separada, se pueden obtener indicadores de grupos homogéneos de observaciones a distintos niveles de agregación a partir de indicadores o elasticidades calculadas reagrupándolas de acuerdo a su mayor similitud. Para ello utilizamos técnicas multivariantes de clasificación.

El presente capítulo se organiza de la siguiente manera: en el siguiente apartado se realiza una revisión bibliográfica de los principales estudios tanto a nivel teórico como empírico, que analizan el surgimiento de Internet y otras tecnologías de la información y los impactos provocados a nivel regional dentro de España. También se revisan trabajos teóricos y empíricos referidos al comercio-e, considerándolo como consecuencia directa de la evolución de Internet. En el tercer apartado se resume lo principal de la literatura econométrica referida a las técnicas que aplicaremos en los posteriores capítulos. Por último, se presentan las conclusiones y las referencias bibliográficas.

## **2. REVISIÓN BIBLIOGRÁFICA**

Desde que la información fue percibida como fuente de renta y riqueza, muchos han sido los intentos de desarrollar modelos teóricos que expliquen este fenómeno. Robert Solow (1956) fue el primer teórico en incorporar la tecnología en la función de producción. A partir de él surgieron muchos trabajos orientados a la Teoría del Crecimiento Económico en donde se trata de interaccionar la tecnología, el conocimiento en general, la formación y la información.

Por otro lado, la discusión acerca de la transmisión de la información en un principio se deriva en cuestiones de política de regulación de las telecomunicaciones, con trabajos

pioneros como Artle y Averous (1973), Rohlfs (1974), Von Rabenau y Stahl (1974) y Littlechild (1975) donde la demanda de acceso a la red telefónica está relacionada con el beneficio neto de utilizar el servicio telefónico como medida del excedente del consumidor. En todos estos trabajos, considerados seminales en cuanto a metodología de estudio, se observa al excedente del consumidor como una variable relacionada con los precios, el ingreso, y otros atributos sociodemográficos relevantes. También aquí, a partir de estos estudios surgió muchísima literatura tanto teórica como empírica en lo que hoy se considera una rama dentro de la economía, denominada Economía de las Telecomunicaciones. Dentro de esta rama tenemos la parte que estudia a las TIC como fenómeno económico. Las discusiones que sobresalen en este tipo de cuestión son principalmente discusiones de política. Así tenemos que Owen (2002) aborda la cuestión del debate de política sobre la difusión de Internet. En su artículo sostiene que el acceso residencial está garantizado para el futuro tal como ocurrió con la televisión por cable o los ordenadores personales. Según este autor, las barreras de acceso y uso de Internet tienen más que ver con la demanda (perfil del demandante) que con las políticas de regulación por el lado de la oferta.

Desde el punto de vista econométrico, un trabajo pionero que estudia la adopción de Internet de banda ancha en los hogares es el de Madden, Savage y Simpson (1996). Estos autores analizan una base de datos de 5.000 respuestas de una encuesta realizada en hogares australianos. Fueron los primeros en detectar que las características demográficas representan una de las influencias principales en la decisión individual de contratar el servicio de Internet de banda ancha en el hogar. Encuentran entre otras cosas, que las personas que no han terminado la escuela secundaria se muestran menos interesadas en contratar el servicio; personas que viven con por lo menos un miembro

nativo de Europa o Asia estarán más interesadas; la edad también influye en el interés aunque este va a variar según si el individuo pasó o no la barrera de los 65 años.

Cassel (1999) usa una encuesta llevada a cabo en Estados Unidos en 1997 de 30.000 observaciones. Se centra principalmente en las características de los individuos con más de una línea del teléfono fijo en el hogar. Esta autora demuestra que tener una línea de teléfono fijo adicional está positivamente asociado con el tamaño del hogar, la renta anual y la existencia de niños pequeños, suponiendo el acceso de Internet a través de la línea de teléfono convencional. La investigación además demuestra que las personas que poseen una línea de teléfono adicional y supuestamente se conectan a Internet en 56 kbps por un precio mensual de \$10 a \$30 pudieran estar interesadas en el futuro en una conexión de gran velocidad para \$30 a \$75. Sin embargo, también descubre que el interés va disminuyendo paralelamente con los precios. Este resultado sugiere que los potenciales usuarios requieren otras características además de un precio asequible (por ejemplo, 24 horas de conexión).

Goolsbee (2000) también estima una demanda de Internet de banda ancha con los datos de 100.000 respuestas de una encuesta realizada en 1999 en varias ciudades de Estados Unidos. Estima un modelo Probit que relaciona la probabilidad de adoptar Internet de banda ancha con cable-modem, el precio del servicio y algunas características particulares tales como el número de años que el individuo lleva utilizando Internet, la edad, el nivel de renta y la educación. Este modelo demuestra que la intención de acceder a Internet vía cable-modem aumentará a medida que los precios bajen. La elasticidad-precio de la demanda para distintos rangos va de  $-2,8$  a  $-3,5$ .

Duffy-Deno (2001) utiliza una muestra de hogares estadounidenses del cuarto trimestre de 1998 con 11.458 observaciones en donde 22% de la muestra posee dos o más líneas de teléfono. Estima un modelo logit de demanda residencial para las líneas

de teléfono adicionales para Estados Unidos. La elasticidad-precio de la demanda es mayor que la elasticidad-precio del teléfono primario.

Rappoport, Kridel y Taylor (2002) estiman la demanda de Internet en un contexto diferente del anterior. Usan datos con la información demográfica sobre hogares en 10 ciudades de Estados Unidos durante el mes de agosto en 2001. Todos los encuestados tienen algún tipo de acceso de Internet en el hogar y la atención se centra en qué tipo de acceso prefieren, de baja, o de alta velocidad. Los autores separan los datos en dos grupos –aquellos que escogen Internet de baja velocidad y los que escogen alta velocidad– y enfoca el estudio en las características que distinguen un grupo del otro. En otras palabras, van más allá de distinguir entre las personas que usan o no usan algún tipo de acceso a Internet. Usando un modelo logit anidado encuentran que el tamaño del hogar afecta positivamente la posibilidad de escoger Internet de banda ancha. Sorprendentemente descubren una correlación positiva entre los individuos de más edad y la adopción de banda ancha, más que los individuos más jóvenes. Además encuentran que los grandes usuarios (los que prefieren banda ancha) aunque visitan más sitios, gastan menos tiempo en cada sitio que los que prefieren la banda estrecha. La elasticidad-precio es aproximadamente  $-1,0$ .

Existe otra línea de estudios recientes que se centran en determinar la disposición a pagar de los usuarios (*Willingness to Pay*). Varian (2002) provee una evidencia interesante utilizando encuestas realizadas para un proyecto de la Universidad de Berkeley denominado Internet Demand Experiment (INDEX) en donde demuestra que la demanda de “velocidad” del servicio de Internet es bastante sensible al precio; las elasticidades-precio que estima están dentro del rango  $-2,0$  a  $-3,1$  para Internet de Banda Ancha. Otro descubrimiento interesante es que la mayoría de los participantes en la investigación de Berkeley estarían dispuestos a pagar no más de \$0,02 por minuto

(\$1,20 la hora) por el tiempo que se ahorra con una conexión a Internet de alta velocidad. Además, halla que los usuarios con profesiones técnicas y administrativas valoran más su tiempo y que están dispuestas a pagar más que otro tipo de usuario. Su conclusión es que aún queda tiempo para que se dé una gran oleada de demanda de Internet de Banda Ancha, por lo menos en un futuro cercano, aunque hay que considerar también que sus resultados son con experimentos de varios años atrás (entre 1998 y 1999)<sup>1</sup>.

Además de trabajos de investigación como los mencionados hasta el momento, existe una gran gama de informes oficiales como el realizado por el Departamento americano de Comercio (2002) en donde también se consideran la influencia sociodemográfica en la adopción de Internet de banda ancha y donde utilizan el mismo modelo propuesto por Madden et al. (1996) pero para datos de Estados Unidos y considerando distintos tipos de adopción. No es de sorprender que hayan llegado a conclusiones similares.

También se debe mencionar el informe emitido por la OCDE (2001) que analiza la adopción de Internet de banda ancha en 30 países con distintos niveles de desarrollo. Se observa que los países más avanzados en tecnología y número de usuarios de Internet deben su posición al apoyo y a la regulación de la competencia. También otro informe interesante es el que llevó a cabo el Telecommunication Research Group de la Universidad de Colorado (2003) en donde dentro de un marco teórico se plantean cuestiones tales como la definición del perfil de los consumidores adoptadores de Internet de banda ancha. Lo atractivo de este informe es que examina los problemas del mercado de Internet de banda ancha, los precios netos, la regulación gubernamental y el fenómeno del tele-trabajo. Excepto este último informe, el resto de los existentes se

---

<sup>1</sup> El proyecto INDEX dirigido por la Universidad de Berkeley se llevó a cabo durante los años 1998 y 1999. Los resultados se publicaron en 2002.

limitan a describir sin usar un soporte teórico. Dentro de esta línea, Gabel y Kwan (2000) reconoce una cierta escasez en la literatura econométrica referida a estos temas, y decide dirigir la investigación examinando el coste de la demanda residencial de Internet de banda ancha. Utilizando un método de muestreo estratificado selecciona una muestra de casi 300 observaciones y estiman un modelo logit restringido para analizar el servicio en donde la restricción es que los factores considerados influyen en la probabilidad de que la conexión por ADSL o por cable-módem pero en un contexto de buena instalación de Red. Los autores concluyen que a medida de que haya adultos entre 30 y 34 años de edad con ingresos que se acercan a la media en un contexto de buena instalación de Red, entonces la probabilidad de adopción del servicio aumentará. La variable que utilizan para cuantificar el coste de la demanda, y que es insignificante dentro del modelo que plantean, es la tele-densidad (el número de líneas de teléfono por millas cuadradas) y un coste de transmisión estimado de tráfico de Internet.

En lo que respecta al estudio de la brecha digital originada por las desigualdades en el acceso y uso de Internet y otras tecnologías, tenemos dos líneas de investigación incipientes. Una primera es la que estudia los patrones de difusión de la tecnología digital tanto a nivel macro como a nivel micro. Esta es la línea que siguen algunos reportes tales como el realizado por The United States Department of Commerce (2000), que aunque es un trabajo netamente descriptivo, es la primera contribución acerca del fenómeno. Este trabajo básicamente mide las diferencias en el acceso a Internet y otras tecnologías digitales en la población estadounidense, distinguiendo por consumidores, empresas y administración pública. A partir de este reporte han ido surgiendo otros como el de la OCDE (2001) en donde diferentes países son comparados en base a sus estadísticas relacionadas con las condiciones de acceso y uso de Internet y otras TIC. Dentro de la segunda línea, referida a los modelos de medición de la brecha

tecnológica más relacionados con Internet podemos mencionar el planteado en el McConnell International's E-readiness Survey (2001) en donde combinan linealmente variables elementales con el fin de generar un índice sintético de disposición al uso de Internet. Actualmente el Banco Mundial ([www.worldbank.org](http://www.worldbank.org)) como también organizaciones virtuales tales como [www.pewinternet.org](http://www.pewinternet.org), [www.digitaldivide.org](http://www.digitaldivide.org) o [www.digitaldividenetwork.org](http://www.digitaldividenetwork.org) publican informes periódicos comparando la brecha tecnológica entre países. Hay que mencionar que todos estos últimos reportes mencionados adoptan una aproximación netamente informativa y descriptiva sin concentrarse demasiado en los aspectos metodológicos relacionados con la medición y análisis del fenómeno.

Dentro de los enfoques metodológicos desarrollados recientemente tenemos al de Ricci (2000) que propone una escala de adopción para las tecnologías digitales en donde dicha escala resulta de la agregación de indicadores elementales. Luego Barbet y Coutinet (2001) analizan los vínculos existentes entre los determinantes de la digitalización y los efectos que provoca. Para ello proponen un modelo para medir las relaciones entre las características de las infraestructuras de las comunicaciones y la evolución de la economía digital. Selhofer y Mayringer (2001) proponen una metodología para medir el desarrollo de la sociedad de la información en los países europeos. Provee una visión de la identificación de las dimensiones del nivel de digitalización y proponen algunos indicadores. Corrocher y Ordanini (2002) plantean un modelo para medir la brecha digital a nivel geográfico. Empleando componentes principales agregan distintos factores en un índice sintético y consideran la dispersión en la distribución de un indicador que calculan como una medida que referencia a las diferencias entre regiones geográficas. Entre las conclusiones plantean algunas

implicaciones de política y exponen oportunidades y riesgos para los gobiernos que operan en un contexto económico digital.

Una de las principales consecuencias de la aparición de Internet es el Comercio Electrónico, también llamado Comercio-e. En la literatura existe toda una gama de autores que estudian las características de la competencia en los mercados electrónicos tales como la formación de precios, las elasticidades, los costes y la discriminación. El primer estudio acerca de la formación de precios en los mercados electrónicos la realizó Lee (1997) estudia el comportamiento de los precios en subastas de coches usados virtuales y convencionales entre 1986 y 1995, y encuentra que los precios virtuales son mayores que los precios de los mercados convencionales, y que dicha diferencia se agrandaba a lo largo del tiempo. Bailey (1998) examina los precios de libros y discos compactos también en ambos mercados y llega a conclusiones similares a las de Lee. Por otro lado, Brynjolfsson y Smith (1999) examinan también precios de libros y discos compactos vendidos en Internet y mercados convencionales durante 1998 y 1999 y encuentran que los precios son entre 9% y 16% menores en Internet aún incluyendo todos los costes adicionales al adquirir un producto (envío, impuestos locales, etc). La explicación que dan Brynjolfsson et al. dichas diferencias se deben a que los mercados virtuales se han vuelto mucho más eficientes entre 1996 y 1999. En cuanto a elasticidad de precios, Goolsbee (1998), utilizando los datos de una encuesta para Estados Unidos, encuentra que los consumidores online son mucho más sensibles a las políticas impositivas locales. La conclusión a la que llega es que los consumidores sujetos a pagar muchos impuestos al consumo tiene mayor probabilidad de comprar en Internet.

En cuanto a otros trabajos empíricos podríamos mencionar a Degeratu, Rangaswamy y Wu (1998) en donde comparan precios de comestibles vendidos en ambos tipos de mercados, y en donde encuentran que la sensibilidad de los precios será

menor para los mercados de comestibles virtuales que para los mercados convencionales. Muy relacionado está el estudio de Lynch y Ariely (1998) en donde analizan la sensibilidad de los precios en un mercado virtual simulado de vino. Los autores encuentran que los consumidores se centran en los precios sólo cuando hay algún otro tipo de información que haga diferenciar a los productos.

A continuación exponemos un cuadro resumen de la bibliografía comentada en este apartado:

### **Cuadro 1: Trabajos sobre Internet, Brecha Tecnológica y Comercio Electrónico**

Referencia	Datos – Enfoque	Conclusiones
<b>Acceso y Uso de Internet</b>		
Solow (1956)	Teoría del Crecimiento	<i>Incorporación de la tecnología en la función de producción</i>
Artle y Averous (1973) Rohlf's (1974) Von Rabenau y Stahl (1974) Littlechild (1975)	Economía del Bienestar	<i>La demanda telefónica está relacionada con el excedente del consumidor.</i>
Owen (2002)	Políticas de regulación	<i>El acceso doméstico a Internet está garantizado en el futuro tal como sucedió con otro tipo de tecnologías</i>
Madden, Savage y Simpson (1996)	Modelo econométrico aplicado a 5.000 hogares australianos	<i>Características demográficas como principal atributo a tener en cuenta como determinante de la adopción del servicio de Internet en el hogar</i>
Cassel (1999)	Modelo econométrico aplicado a 30.000 hogares estadounidenses	<i>Los usuarios potenciales requieren otras características del servicio además de tener un precio asequible</i>
Goolsbee (2000)	Modelo Econométrico aplicado a una muestra de 100.000 individuos en Estados Unidos	<i>La elasticidad-precio de la demanda de Internet de banda ancha varía dentro de un rango de -2,8 y -3,5</i>
Duffy-Deno (2001)	Modelo econométrico aplicado a una muestra de 11.458 observaciones estadounidenses	<i>La elasticidad-precio de la demanda de una segunda línea de teléfono es mucho más elástica que la línea de teléfono primaria</i>
Rappoport, Kridel y Taylor (2002)	Modelo Econométrico con información demográfica de 10 ciudades estadounidenses	<i>El tiempo se valora más con Internet de banda estrecha. Se visitan menos lugares que con banda ancha. La elasticidad precio del servicio es de -1,0</i>
Varian (2002)	Economía experimental	<i>Los usuarios con profesiones técnicas valoran más su tiempo. No se dará una gran explosión de demanda de Internet de banda ancha en un futuro cercano.</i>
The United States Department of Commerce (2002)	Informe descriptivo	<i>Aplicación del modelo de Madden (1996) para datos de Estados Unidos llegando a conclusiones similares</i>
OCDE (2001)	Informe descriptivo	<i>Los países más avanzados en tecnología de acceso y uso de Internet deben su posición a políticas de regulación acordes</i>

Telecommunication Research Group – University of Colorado (2003)	Compilación de trabajos – Informe	<i>Problemas del mercado de Internet de banda ancha, precio del servicio, políticas de regulación, fenómeno del tele-trabajo</i>
Gabel y Kwan (2000)	Modelo Econométrico aplicado a una muestra condicionada	<i>La probabilidad de demandar Internet de banda ancha aumenta para edades medias e ingresos medios condicionado a la existencia de buenas instalaciones de red de cableado</i>
<b>Brecha Tecnológica</b>		
The United States Department of Commerce (2002)	Informe descriptivo	<i>Medición de las diferencias en acceso y uso de Internet y otras tecnologías en EEUU.</i>
OCDE (2001)	Informe descriptivo	<i>Comparación del acceso y uso de Internet entre países con distintos niveles de desarrollo</i>
McConnell International's E-Readiness Survey (2001)	Informe descriptivo	<i>Combinación lineal de atributos logrando un indicador sintético</i>
<a href="http://www.worldbank.org">www.worldbank.org</a> <a href="http://www.digitaldivide.org">www.digitaldivide.org</a> <a href="http://www.digitaldividenetwork.org">www.digitaldividenetwork.org</a>	Informes descriptivos permanentes	<i>Comparación descriptiva de la brecha tecnológica entre países y propuestas de política</i>
Ricci (2002)	Econometría – Estadística Aplicada	<i>Escala subjetiva de adopción para las tecnologías digitales</i>
Barbet y Coutinet (2001)	Econometría – Estadística Aplicada	<i>Medición de los vínculos existentes entre los determinantes del nivel de digitalización entre los países y sus efectos</i>
Selhofer y Mayringer (2001)	Econometría – Estadística Aplicada	<i>Medición del desarrollo de la sociedad de la información en los países europeos</i>
Corrocher y Ordanini (2002)	Econometría – Estadística Aplicada	<i>Medición de la brecha tecnológica a través de la dispersión de los distintos niveles de digitalización</i>
<b>Comercio Electrónico</b>		
Lee (1997)	Microeconomía – Formación de precios	<i>Los precios de Internet son mayores que los precios de los mercados convencionales (mercado de coches usados)</i>
Bailey (1998)	Microeconomía – Formación de precios	<i>Las mismas conclusiones que Lee (1997) para el mercado de libros y discos compactos</i>
Brynjolfsson y Smith (1999)	Microeconomía – Formación de precios	<i>Conclusiones contrarias a las de Bailey (1998). La justificación está en el incremento de la eficiencia de los mercados virtuales</i>
Goolsbee (1998)	Modelo econométrico aplicado a datos demográficos	<i>Los consumidores online son mucho más sensibles a las políticas impositivas locales</i>
Degeratu, Rangaswamy y Wu (1998)	Modelo econométrico aplicado a datos demográficos	<i>La sensibilidad de los precios es menor en los mercados online que en los mercados convencionales</i>
Lynch y Ariely (1998)	Economía Experimental	<i>Se descubre la importancia de la información adicional que logra diferenciar los productos además de lograr un precio competitivo</i>

### 3. FUNDAMENTOS TEÓRICOS

#### 3.1. INTRODUCCIÓN

Dada la relevancia que actualmente presentan las TIC en la economía, surge la necesidad de explicar sus determinantes y el modo que tienen estos de influir en el

desarrollo del acceso y utilización de las mismas. La formalización matemática que se aplica en estos casos está basada en la que se utiliza generalmente en economía de las telecomunicaciones. Una buena compilación se encuentra en Taylor (1993).

El presente apartado se centra fundamentalmente en el desarrollo de la modelización teórica que se empleará en las estimaciones de los posteriores capítulos. Para ello se exponen distintos modelos que permitirán la aproximación empírica al acceso y uso de Internet, brecha digital en el uso de las TIC, y utilización del comercio electrónico.

### **3.2. DECISIÓN EN UN CONTEXTO DE ALTERNATIVAS DISCRETAS**

Al estudiar el advenimiento de las TIC dentro de la vida cotidiana, la mayoría de las variables que se analizarán en la presente Tesis presentan valores que están limitados, o bien porque deben ser positivos, o porque sólo toman un reducido número de valores. Entonces es preciso emplear conceptos y tener en cuenta aspectos de la toma de decisiones cotidiana que lleva al agente a decidir entre una u otra tecnología. Se considera como agente tomador de decisiones al individuo, al grupo (familia, hogar), a una empresa o alguna institución pública. Dependiendo del mercado concreto, cada uno de estos se puede considerar como una unidad de decisión en el acceso y uso de las TIC en donde hay que tener en cuenta las diferencias entre las mismas unidades decisorias debido a la existencia de distintos patrones (preferencias, gustos, etc).

Sea un agente (individuo, familia, etc) que realiza una opción, o una serie de opciones en el tiempo, entre un conjunto de opciones. Se considera a la variable  $y$  como el resultado de esa decisión:

$y$ : Opción o secuencia de opciones elegidas en el tiempo

Nuestra meta aquí será entender el proceso de conducta que lleva a los agentes a tomar determinada opción dentro del conjunto de opciones posibles. Para ello se considera que existen *causas* que hacen que los agentes tomen la decisión del modo que la tomaron. A estas causas se las denomina determinantes, factores o atributos. Algunas de estos son observables y otros no lo son. Entonces, el proceso de conducta del agente se lo puede expresar de la siguiente manera:

$$y = h(x, \varepsilon) \quad (1.1)$$

Esta función incluye a los factores observables  $x$  y a los no observables  $\varepsilon$ , que al incluirse este último factor no observable no será una toma de decisión determinista y no se podrá predecir exactamente y solo se podrán predecir las *probabilidades* de opción de cada agente. La posibilidad de que el agente elija o tenga un resultado particular será entonces la probabilidad de que los atributos inobservables sean tales de que el proceso de conducta resulte como tal, es decir:

$$\Pr(y|x) = \Pr[\varepsilon : h(x, \varepsilon) = y] \quad (1.2)$$

Para expresar esta probabilidad de manera más manejable, se define una función *indicador*  $I[h(x, \varepsilon) = y]$  tal que:

- $I[h(x, \varepsilon) = y] = 1$  si la expresión de dentro del corchete es verdadera (si el valor de  $x$  combinado con  $\varepsilon$  induce al agente a optar por  $y$ ).
- $I[h(x, \varepsilon) = y] = 0$  si la expresión de dentro del corchete es falsa (si el valor de  $x$  combinado con  $\varepsilon$  induce al agente a no optar por  $y$ ).

Entonces la probabilidad de que el agente opte por  $y$  será el valor esperado de esa función indicador:

$$\Pr(y|x) = \Pr[I(h(x, \varepsilon) = y) = 1] = \int I(h(x, \varepsilon) = y) f(\varepsilon) d\varepsilon \quad (1.3)$$

Para calcular esta probabilidad es necesario calcular previamente la integral. Para ello se requieren ciertas especificaciones para  $h$  y  $f$  de manera tal de que la integral pueda ser expresada en forma cerrada (es decir, que la probabilidad pueda ser calculada exactamente). La especificación apropiada dependerá principalmente del objetivo y de los datos con los que se trabaje.

### 3.3. TEORÍA DE LA UTILIDAD ALEATORIA

La Teoría de la Utilidad Aleatoria (Mc Fadden, 1974) constituye la base teórica para modelizar las decisiones de los agentes. Como se comenta más arriba, a la hora de analizar estas decisiones nunca se dispone de información perfecta acerca de los factores que han influido en dicho proceso de elección. Para resolver esta cuestión se acude a la teoría de las probabilidades con la idea de incluir el efecto de los factores de naturaleza aleatoria de los cuales no se cuenta con información. Así pues, se define una función de utilidad como la siguiente:

$$U = \beta'x + \varepsilon \quad (1.4)$$

Esta función, que puede ser tanto positiva como negativa, consta de dos componentes: un componente observable,  $\beta'x$ , en donde  $x$  es el vector de variables y  $\beta$  es el vector de parámetros, y un componente no observable,  $\varepsilon$ . El agente tomará una decisión solo si la utilidad que le reporte dicha decisión es positiva, es decir, solo si le provee un beneficio neto. La probabilidad, dada la parte observable, será:

$$\Pr(y|x) = \int I(\beta'x + \varepsilon > 0) f(\varepsilon) d\varepsilon = \int I(\varepsilon > -\beta'x) f(\varepsilon) d\varepsilon$$

$$\Pr(y|x) = \int_{-\beta'x}^{\infty} f(\varepsilon) d\varepsilon = 1 - F(-\beta'x)$$

El modelo matemático utilizado para calcular  $\Pr(y|x)$  dependerá de la forma funcional de  $f(\varepsilon)$ . En esta Tesis utilizamos varios de ellos y los iremos describiendo según los usemos en cada capítulo. Aquí nos centraremos en las características comunes de todos ellos.

Los supuestos comunes de los modelos de utilidad aleatoria planteados primeramente por Marschak (1960) y de los cuales derivan los Modelos de Opción Discreta que utilizaremos en la modelización econométrica en los capítulos sucesivos, serán:

- I. Los agentes maximizan su utilidad.
- II. Los agentes responden a diferentes niveles de estímulo.
- III. Es posible representar al agente aún cuando no maximiza su utilidad, aunque en este caso habrá problemas de consistencia.

Bajo estos supuestos, el agente  $n$  toma una decisión entre un conjunto  $J$  de alternativas, y obtendrá un cierto nivel de utilidad para cada alternativa,  $U_{nj}$ , en donde  $j = 1, 2, \dots, J$ . Es decir que el agente elige la alternativa  $i$  sí y sólo si  $U_{ni} > U_{nj} \quad \forall j \neq i$ .

A esta utilidad se la puede observar desde dos puntos de vista; desde la óptica del agente tomador de decisiones, y desde la del investigador que intenta identificar el proceso de decisión de todos los agentes a la vez. Desde el punto de vista del agente tomador de decisiones, esta utilidad es conocida, pero desde el punto de vista del investigador no lo es. Sin embargo, lo que el investigador sí observará serán algunos atributos de las alternativas que tenía el agente decisor ( $x_{nj} \quad \forall j$ ) y algunos atributos propios del agente ( $s_n$ ). Entonces podrá especificar una función denominada *Utilidad Representativa* que relacione esos factores observados:

$$V_{nj} = V(x_{nj}, s_n) \quad \forall j = 1, \dots, J \quad (1.5)$$

Generalmente  $V$  dependerá de parámetros desconocidos para el investigador. Entonces, la utilidad estará compuesta de la siguiente manera:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad (1.6)$$

El componente  $\varepsilon_{nj}$  captura atributos que afectan a la utilidad pero que no están incluidos en  $V_{nj}$ . La distribución de probabilidad de  $\varepsilon_{nj}$  dependerá de cómo se especifique  $V_{nj}$ . Como no se conoce  $\varepsilon_{nj}$ , entonces se lo trata como un componente aleatorio.

La probabilidad de que el agente decisor  $n$  elija la alternativa  $i$  será:

$$\begin{aligned} \Pr_{ni} &= \Pr(U_{ni} > U_{nj}) \quad \forall j \neq i \\ \Pr_{ni} &= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \quad \forall j \neq i \\ \Pr_{ni} &= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) \quad \forall j \neq i \end{aligned} \quad (1.7)$$

Se observa que esta probabilidad se puede describir usando una función de distribución que utilizando la función de densidad  $f(\varepsilon_n)$  podría reescribirse como:

$$\Pr_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (1.8)$$

donde la función  $I(\cdot)$  es, como se dijo más arriba, la función indicador que tomará solo dos valores: 1 si el término entre paréntesis es verdadero, y 0 en caso contrario.

Se pueden obtener diferentes modelos a partir de diferentes especificaciones para la función de densidad  $f(\varepsilon_n)$ , es decir, para diferentes supuestos acerca de la forma funcional de la parte inobservable de la función de utilidad. La interpretación de los resultados dependerá de esta última.

#### **4. CONCLUSIONES**

Si bien hoy son prácticamente indispensables, las Tecnologías de la Información y de las Comunicaciones (TIC) han hecho su aparición en la vida cotidiana hace relativamente poco tiempo. El servicio de Internet por banda ancha tanto para hogares como para empresas pequeñas y medianas se comenzó a adquirir de manera cotidiana hace pocos años. Como consecuencia de este advenimiento, se hizo indispensable dirigir la investigación hacia este campo de manera de evaluar la influencia de las TIC en el incremento de la productividad, el valor potencial que tienen para los consumidores que acceden a las mismas más fácilmente que otros, o tal vez las posibilidades que se presentan para las empresas el hecho de poder ampliar el mercado utilizando los nuevos mercados virtuales, entre otras cuestiones. A su vez se buscan potenciales políticas para fomentar su incorporación en todos los ámbitos y poder aprovechar su enorme impacto. Desde hace aproximadamente 10 años en la literatura referida a economía de las telecomunicaciones se centró en el estudio de este fenómeno dando inicio a varias líneas de investigación tanto teóricas como empíricas analizando el fenómeno tanto desde el lado de la oferta (Servicio Universal, formación de precios, etc) como por el lado de la demanda (Willigness to Pay, determinantes de la demanda de Banda Ancha en diferentes contextos, etc). La finalidad de esta Tesis Doctoral es contribuir al debate realizando un estudio desde el lado de la demanda para España y utilizando técnicas de análisis empírico basadas en la Teoría de la Utilidad Aleatoria.

## **5. REFERENCIAS**

Artle, R. y Averous, C. "The Telephone System as a Public Good: Static and Dynamic Aspects" *Bell Journal of Economics and Management Science*, Vol. 4, No. 1, 89-100, spring 1973.

Barbet, P. y Coutinet, N. "Measuring the Digital Economy: State-of-the-Art Developments and Future Prospects" *Communications and Strategies*, 42, 153-184, 2001.

Bailey, J. "Electronic Commerce: Prices and consumer Issues for Three Products: Books, Compact Discs and Software" *Organization for Economic Co-Operation and Development, OCDE/GD(98)4*, 1998b.

Brynjolfsson, E. y Smith, M. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers" *Working Paper, MIT Sloan School of Management*, 1999.

Cassel, C., "Demand for and Use of Additional Lines by Residential customers", en Loomis and Taylor (eds.) *The Future of the Telecommunications Industry: Forecasting and Demand Analysis*, Kluwer Academic Publishers, Boston, 1999.

Corrocher, N. y Ordanini, A. (2002) "Measuring the digital divide: a framework for the analysis of cross-country differences" *Journal of Information Technology*, 17, 9-19.

Davidson, R. y MacKinnon, J., “Econometric Theory and Methods” Oxford University Press, 1993.

Degeratu, A., Rangaswamy, A. y Wu, J. “Consumer Choice Behavior in Online and Regular Stores: The Effects of Brand Name, Price and Other Search Attributes” Presentado en Marketing Science and the Internet, INFORM, College on Marketing Mini-Conference. Cambridge, MA, 6-8 March, 1998.

Duffy-Deno, K. T. “Demand for Additional Telephone Lines: An Empirical Note”, Information Economics and Policy, 13, 301-309, 2001.

Gabel, D. y Kwan, F. “Accessibility of Broadband Telecommunications Services by Various Segments of the American Population”, Telecommunications Policy Research Conference, August 2002.

Goolsbee, A. ”In A World Without Borders: The Impact of Taxes on Internet Commerce”, Working Paper, University of Chicago, 1999.

Goolsbee, A. “The Value of Broadband and the Deadweight Loss of Taxing New Technology”, Working Paper, University of Chicago, 2000.

Greene, W. “Econometric Analysis” 4<sup>th</sup>. Edition, Prentice Hall, 2000.

Jackson, M., Lookabaugh, T., Savage, S., Sicker, D. y Waldman, D. "Broadband Demand Study: Final Report" Telecommunications Research Group, University of Colorado, 2003.

Lee, H. "Do Electronic Marketplaces Lower the Prices of Goods?" Communications of the ACM, Vol. 41, No 12, Jan. 1997.

Littlechild, S. "Two-Part Tariffs and Consumption Externalities" Bell Journal of Economics, Vol. 5, No. 2, Autumn 1975, pp. 661-670.

Lynch, J. (Jr). y Ariely, D. "Interactive Home Shopping: Effects of Search Cost for Price and Quality Information On Consumer Price Sensivity, Satisfaction With Merchandise and Retention" Presentado en Marketing Science and the Internet, INFORM, College on Marketing Mini-Conference. Cambridge, MA, 6-8 March, 1998.

Maddala, G. "Limited Dependent and Qualitative Variables in Econometrics" Cambridge University Press, 1983.

Madden, G., Savage, S. y Simpson, M. "Information Inequality and Broadband Network Access: An Analysis of Australian Household Survey Data" Industry and Corporate Change, Oxford University Press, 1049-1056, 1996.

Marschak, J. "Binary Choice Constraints on Random Utility Indications" en K. Arrow, ed., Stanford Symposium on Mathematical Methods in the Social Science, Standford University Press, Standford, California, pp. 312-329, 1960.

McConnell International E-Readiness Survey “Ready? Net. Go! Partnership Leading the Global Economy”, Washington D.C., 2001.

Mc Fadden, D. “Conditional Logit Analysis of Qualitative Choice Behavior” in Zarembka (ed.) *Frontiers in Econometrics*, Academic Press, pp. 1117-1156, 1974.

Mc Fadden, D. “Econometric Analysis of Qualitative Response Models” in Z. Grilliches and M. Intriligator (eds.) *Handbook of Econometrics*, Amsterdam, North Holland, pp. 1376-1425, 1984.

Organization for Economic Cooperation and Development, “The Development of Broadband in OECD Countries”, October 29, 2001.

OECD Report (2001) “Understanding the Digital Divide” in [www.oecd.org/dataoecd/38/57/](http://www.oecd.org/dataoecd/38/57/)

Owen, B. “Broadband Mysteries” in Crandall, R. and Alleman, J. (ed.) *Broadband: Should we Regulate High-speed Internet Access?* AEI – Brookings Joint Center for Regulatory Studies, 2002.

Rappoport P., Taylor L., y Kridel D. “The Demand of Broadband: Access, Content, and the Value of Time”, in *Broadband: Should We Regulate High-Speed Internet Access?*, ed. by R. W. Crandall and J. H. Alleman, AEI-Brookings Joint Centre for Regulatory Studies, Washington, D.C., 2002.

Ricci, A. (2000) "Measuring information society dynamics of European data on usage of information and communication technologies in Europe since 1995" *Telematics and Informatics*, 17, 141-67.

Rolfhs, J. "A Theory of Interdependent Demand for a Communications Service" *Bell Journal of Economics and Management Science*, Vol. 5, No 1, Spring 1974, pp. 16-37.

Taylor, L. D. "Telecommunications Demand in Theory and Practice" Kluwer Academic Publishers, 1993.

Smith, M., Bailey, J. y Brynjolfsson, E., "Understanding the Digital Markets: Review and Assessment" en Brynjolfsson, E. y Kahin, B. "Understanding the Digital Economy", MIT Press, 1999.

The World Bank (2005) "Financing Information and Communication Infrastructure Needs in the Developing World: Public and Private Roles" Borrador para Discusión en <http://lnweb18.worldbank.org/ict/resources.nsf/>

Train, K. "Discrete Choice Methods with Simulation" Cambridge University Press, 2002.

Selhofer, H. y Mayringer, H. "Benchmarking the Information Society. Development in European Countries" *Communication and Strategies*, 43, 17-55, 2001.

Solow, R., "A Contribution to the Theory of Economic Growth" *Quarterly Journal of Economics*, 70, 1, pp. 65-94, February 1956.

U.S. Department of Commerce, and National Telecommunications & Information Administration, "A Nation Online: How Americans Are Expanding Their Use of the Internet", February 2002.

Varian, H. "The Demand for Bandwidth: Evidence from the INDEX Project", in *Broadband: Should We Regulate High-Speed Internet Access?*, ed. by R. W. Crandall and J. H. Alleman, AEI-Brookings Joint Centre for Regulatory Studies, Washington, D.C., 2002.

Von Rabenau, B. y Stahl, K. "Dynamic Aspects of Public Goods: A Further Analysis of the Telephone System" *Bell Journal of Economics and Management Science*, Vol. 5, No 2, Autumn 1974, pp. 651-669.

## CAPÍTULO II

### Acceso y Uso de Internet en España

## **1. INTRODUCCIÓN<sup>1</sup>**

En la actualidad se están desarrollando muchos estudios socio-económicos tanto de naturaleza teórica como empírica referidas al acceso y uso de Internet de alta velocidad (denominado Internet de banda ancha) en España así como en otros países.

Analizando el fenómeno desde una perspectiva socioeconómica en lo que se refiere a la adopción de un nuevo producto, el advenimiento de Internet ha evolucionado como algunos otros nuevos productos. Es decir, su demanda ha ido creciendo gradualmente hasta que en un tiempo muy corto se ha vuelto casi indispensable.

Desde una perspectiva histórica, el concepto de conectar sistemas y su empleo en una red compartida que permitiese la conexión entre dos ordenadores comenzó a desarrollarse a principios de la década de los 60. Posteriormente esta idea se ha ido desarrollando hasta que en 1969 se puso en marcha una red experimental que hiciera posible el intercambio de información entre ordenadores remotos. Esta red fue financiada en Estados Unidos por la Agencia de Proyectos de Investigación Avanzada (ARPA) –por esta razón al fenómeno inicialmente se lo denominó ARPANET- y se la considera el embrión de lo que hoy es Internet.

Hoy prácticamente ninguna actividad es concebible sin tener en cuenta la información disponible en esta grandísima red mundial. Nunca antes se ha visto una tecnología que se haya adoptado tan rápidamente y que haya tenido tanta influencia en prácticamente todos los ámbitos de la vida cotidiana. Hoy se puede decir que el fenómeno Internet ha transformado la tecnología y hasta la misma base cultural de los países desarrollados y en desarrollo.

---

<sup>1</sup> Una versión de este capítulo resumida y en inglés está publicada en Preissl, B. y Müller, J. eds, "Governance of Communication Networks" Physica-Verlag (Springer Company) 2006, pp. 333-353.

A partir de estas premisas iniciales, el enfoque de este capítulo es responder a las preguntas de qué impulsa a una familia a adquirir una conexión de Internet de banda ancha para su hogar, y qué atributos del individuo influyen en el mayor o menor uso del servicio. El estudio se realiza para España.

En el caso concreto de España en la actualidad el acceso a Internet desde los hogares no supera el 30%. En cuanto al uso del servicio, el porcentaje es un poco mayor, pero no demasiado. A diferencia de otros países europeos o Estados Unidos, no hace demasiado tiempo que se empezó a recopilar información y realizar estudios sobre Internet. Existen algunos estudios de tipo descriptivo-informativo llevados a cabo por fundaciones o empresas privadas<sup>2</sup>. Estos se centran en la descripción de la adopción de Tecnologías de la Información y de las Comunicaciones en el país. El Instituto Nacional de Estadística (INE) empezó a compilar esta información a partir del año 2001 por un encargo específico de la Comisión del Mercado de las Telecomunicaciones (CMT). Este capítulo fue elaborado a partir de los datos de la Encuesta TIC-H 2003 referida al Equipamiento y Uso de las Tecnologías de la Información en los Hogares. El ámbito geográfico es el conjunto del territorio español y el procedimiento es mediante entrevista personal en el domicilio del entrevistado. El fichero de microdatos contiene información de más de 18.900 individuos y más de 200 variables de interés.

Como primera aproximación, se puede decir que existe una gran variabilidad en el acceso y uso de Internet. De acuerdo a lo que podemos observar en el Gráfico 1 de la página siguiente, y de acuerdo a la encuesta que se maneja, el 25,2% de los hogares españoles accede a Internet, pero dicho acceso no se distribuye de manera homogénea. Este acceso está más extendido en Cataluña (32,7% de los hogares), el País vasco

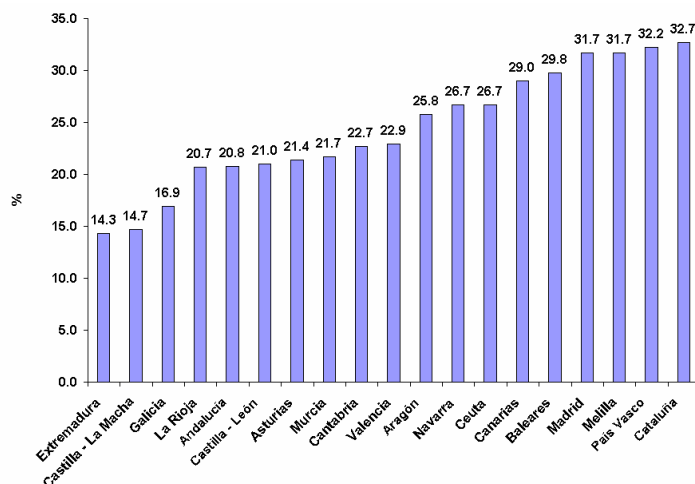
---

<sup>2</sup> [www.fundacionauna.com](http://www.fundacionauna.com), [www.red.es](http://www.red.es) entre otros.

(32,2%), Madrid (31,7%) y Melilla (31,7%). Esto contrasta con Extremadura (14,3%), Castilla-la Mancha (14,7%) y Galicia (16,9%).

El debate político principal en esta área se centra en como incentivar a los consumidores para que adopten el acceso a Internet y, en particular, adopten la tecnología de banda ancha. La discusión empieza definiendo la conexión a Internet y la tecnología de banda ancha. A modo de simplificación, diremos que *Banda Ancha* e *Internet de Alta Velocidad* son términos sinónimos.

**Gráfico 1:** Acceso a Internet desde el Hogar en España en el año 2003



Rappoport, Kridel y Taylor (2002) establecen claramente las similitudes y diferencias entre las características de la demanda de telefonía convencional y la adopción de algún tipo de conexión a Internet dentro del ámbito de la demanda de los servicios de telecomunicación. Lo que estos dos servicios tienen en común es que como cualquier otro perteneciente a las telecomunicaciones, no se consumen aisladamente. Hay toda una red de unidades productivas involucradas. Aparte de la interdependencia y externalidades vinculadas, también *Acceso* y *Uso* son conceptos diferentes. En cuanto a

las diferencias, la primera se refiere a la medida del output: mientras que en la telefonía convencional dicho output es medido en minutos, en el caso de Internet este será medido en velocidad de transmisión de archivos. Esta primera diferencia lleva a la segunda y mayor diferencia entre estos dos tipos de servicios de comunicación. Para el caso de algunas formas de acceso a Internet por banda ancha, como por ejemplo a través módem de cable, la velocidad se verá afectada por el número de líneas de acceso individual conectadas en cada instante, mientras que eso no ocurre si dichas conexiones se realizan a través de ADSL.

En el presente capítulo se estimarán funciones que explican los determinantes del acceso y uso de Internet por banda ancha en España basadas en las características de los consumidores utilizando los datos de la encuesta del INE descritos anteriormente. Las características de los consumidores se las considera desde tres atributos: económicos (nivel de renta permanente), tecnológicos (equipamiento en el hogar, uso de otras tecnologías) y sociodemográficos (tamaño familiar, nivel de educación alcanzado, edad, etc). El capítulo se organiza como sigue: en el próximo apartado se realiza un estudio gráfico-descriptivo del acceso y uso de Internet en España. En el apartado 3 se modeliza el fenómeno desde el punto de vista teórico de las funciones de utilidad y la especificación econométrica. En el apartado 4 se realiza la estimación y se discuten los resultados para el acceso y uso de Internet desde el hogar y otros lugares. Por último, en el apartado 5 se recogen las principales conclusiones de esta investigación y la bibliografía utilizada.

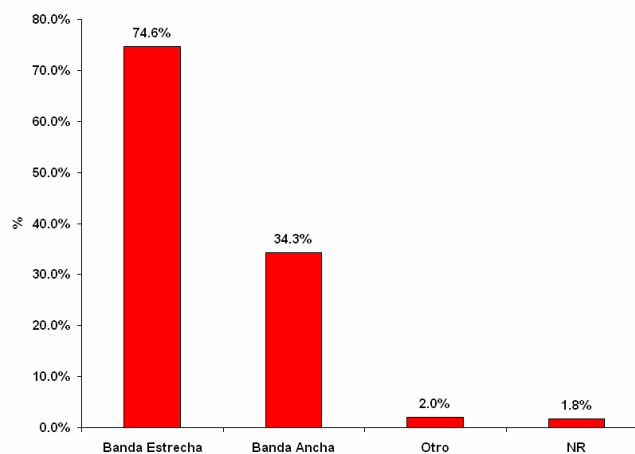
## 2. ANÁLISIS GRÁFICO-DESCRIPTIVO

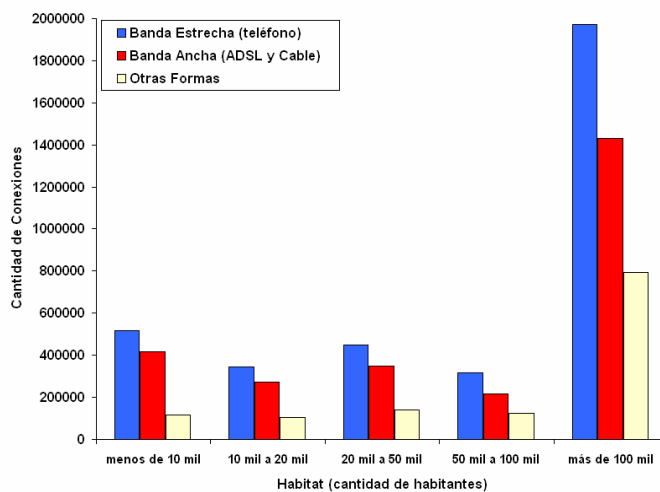
El objetivo del siguiente apartado es proveer una primera evidencia al debate sobre el acceso y uso de Internet, enfocando el análisis en la forma de conexión y el perfil del usuario. La adopción y naturaleza de algún tipo de conexión y la intensidad del uso del servicio serán endógenas en el modelo propuesto en esta investigación.

### 2.1. ACCESO A INTERNET DESDE EL HOGAR

A continuación se refleja la respuesta afirmativa a las preguntas “¿accede usted a algún tipo de conexión a Internet en su hogar?” y si es afirmativa, “¿con qué tipo de conexión accede a Internet?”. Se observa que en el Gráfico 2 los porcentajes no suman 100. Esto es debido a que las respuestas no son excluyentes:

**Gráfico 2:** Hogares Españoles con Conexión a Internet

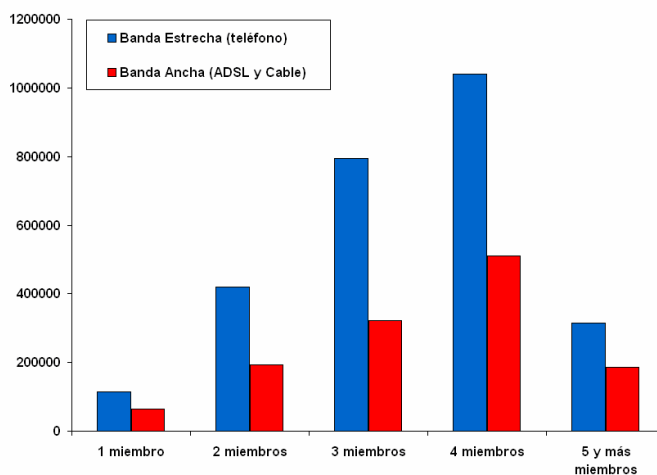


**Gráfico 3:** Acceso a Internet en el Hogar Teniendo en Cuenta el Hábitat

En el Gráfico 2 vemos la manera en que está distribuido el tipo de conexión dentro de los individuos que han manifestado que se conectan a Internet desde el hogar. En el Gráfico 3 vemos lo mismo pero teniendo en cuenta además el hábitat del individuo. Lo que se destaca claramente en ambos gráficos es que, para todos los hogares españoles con acceso a Internet (25,2% del total en el año 2003) la mayoría utiliza una conexión por banda estrecha a través de la línea telefónica convencional (74,6%). Solamente el 34,3% opta por una conexión por banda ancha u otras formas de conexión<sup>3</sup> (2,0%). A su vez se observa de que habría familias con más de un tipo de conexión debido a que la suma de estos porcentajes no es 100.

En cuanto a la relación existente entre el tipo de conexión y el tamaño del hogar, a continuación observamos el Gráfico 4:

<sup>3</sup> Por ejemplo la conexión WAP.

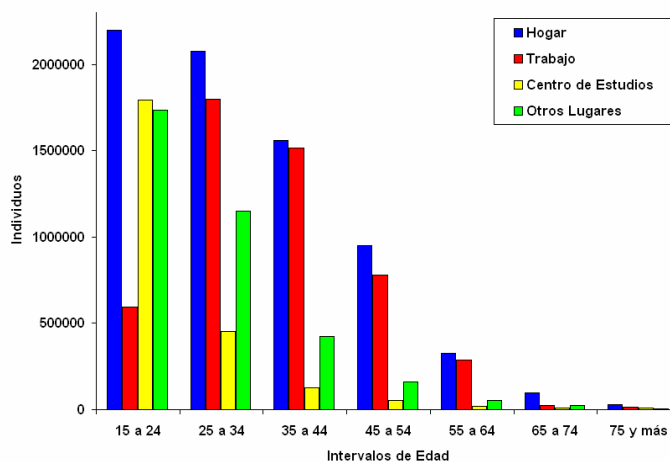
**Gráfico 4:** Acceso a Internet en el Hogar teniendo en cuenta su Tamaño

Se puede observar claramente que la conexión telefónica de banda estrecha persiste en todas las categorías consideradas, con los hogares de cuatro miembros prevaleciendo sobre todos. Sin embargo, se nota que habría considerables diferencias entre los porcentajes de acceso a través de un canal y otro dependiendo del tamaño del hogar. Por ejemplo, en los hogares con 4 miembros la conexión de banda estrecha es 40,2% más que la conexión de banda ancha. En cambio en familias con menos miembros esta diferencia es bastante menor (4,3% y 15,6% respectivamente). Esto podría explicarse por el hecho de que familias con uno o dos miembros presentan tipologías diferenciadas: una corresponde a jóvenes solteros o parejas sin hijos en edad laboral. La otra serían personas jubiladas. Salvo que su profesión requiera lo contrario, por lo general la primer tipología preferirá el acceso a Internet desde otros lugares (trabajo, bibliotecas, etc). En el caso de la segunda tipología, por lo general son personas con poco o ningún interés en acceder a Internet desde su hogar.

## 2.2. USO DE INTERNET

Con el Gráfico 5 comenzamos una descripción del uso de Internet relajando la restricción de que el acceso sea solo desde el hogar. En este caso relacionamos el uso de Internet con intervalos de edad. Aquí las respuestas son individuales (anteriormente se referían a hogares). Lo primero que observamos es que del total de las personas que manifiestan haber utilizado el servicio de Internet en los últimos tres meses, la mayoría lo hace desde el hogar a través de uno de los dos tipos de conexión ya comentados. En otras palabras, la mayoría de la población que usa Internet lo hace desde el hogar (59,7%), seguidos por aquellas personas que se conectan desde el trabajo (41,3%) desde otros lugares tales como cybercafés o bibliotecas (29,3%) y desde el centro de estudios (20,4%). De nuevo, aquí los porcentajes no suman 100 debido a que se consideran tablas de respuesta múltiple.

**Gráfico 5:** Uso de Internet por Edad y Lugar de Uso

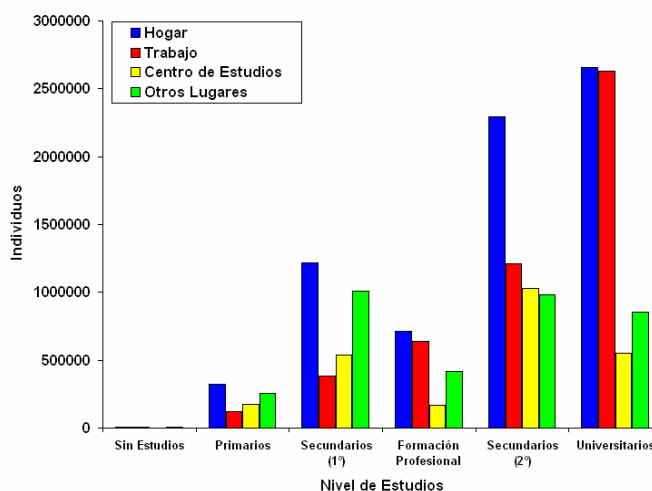


El Gráfico 5 ilustra la segmentación de la población española en términos de uso de Internet. Se observa que a pesar de que el uso principal de Internet en prácticamente

todas las edades será desde el hogar, sin embargo vemos que para individuos jóvenes (entre 15 y 24 años de edad) Internet se utiliza considerablemente en los centros de estudio, y gente de mayor edad utilizará más Internet desde el lugar de trabajo. Otra cuestión que se observa claramente en el gráfico es la relación inversa entre edad y uso de Internet, ya que el número de personas que se conecta a Internet decrece a medida que la edad va aumentando.

En cuanto al uso de Internet y el nivel de estudios alcanzado observamos en el gráfico de la página siguiente que existe una clara relación directa entre ambos fenómenos. Esto tendrá sentido debido a que generalmente Internet como servicio es muy utilizado en puestos de trabajo en donde se requiere cierto grado de cualificación.

**Gráfico 6:** Uso de Internet por Nivel de Estudios y Lugar de Uso

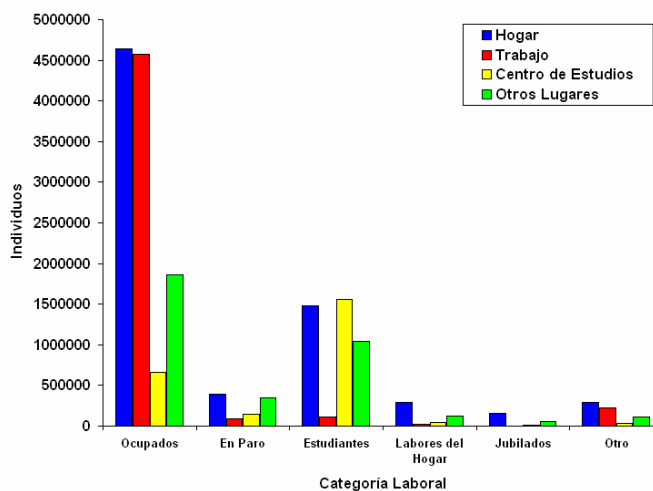


En el Gráfico 6 se ve que el uso de Internet se incrementa con el nivel de cualificación del individuo. En todos los niveles de estudios el uso que prima es desde el hogar, tal como veíamos en el Gráfico 5. Vemos además que para los individuos con nivel universitario el acceso desde el lugar de trabajo es casi el mismo. Lo mismo

observamos, aunque en una proporción menor, para los individuos con Formación Profesional de nivel terciario. El uso de Internet desde otros lugares será lo que le sigue en importancia al uso desde el hogar en el caso de la primera etapa de los estudios secundarios y en los estudios primarios. Y, como es de esperar, las personas que no han llegado a completar ningún nivel de estudios representan un escaso 0,5% de los usuarios de Internet y casi no se perciben en el gráfico. Después de lo que observamos, encontramos evidencia clara de que el nivel de estudios alcanzado estará muy asociado con el uso de Internet.

Por último y considerando la categoría laboral del individuo, a continuación presentamos el Gráfico 7:

**Gráfico 7:** Uso de Internet por Categoría Laboral y Lugar de Uso



Tal y como es de esperar, los individuos que utilizan más el servicio de Internet serán los que pertenecen a la población ocupada y en donde la proporción de uso desde el hogar o desde el lugar de trabajo es casi la misma. Le siguen los estudiantes, en donde como es lógico mayormente se conectan a Internet desde el centro de estudios. El resto de las categorías laborales representa una escasa proporción, aunque destacamos a los individuos en paro, en donde se percibe el uso desde el hogar y desde otros lugares

(cybercafés, bibliotecas, etc) desde donde se supone están buscando trabajo, y a las personas dedicadas a las labores del hogar, en donde de conectarse a Internet lo hacen principalmente desde el hogar.

En general podríamos decir, como una primera aproximación a los datos, de que el demandante del acceso a Internet es el tipo de individuo que habita en grandes centros urbanos (con más de 100.000 habitantes) y tiene una familia con tres o más miembros, mientras que el usuario en general es el individuo con nivel de cualificación medio o superior, de edad media y se encuentra empleado o estudiando.

Luego de esta descripción básica de los datos, se busca descubrir y medir la dependencia entre las variables descriptas. A continuación se utiliza un modelo económico y su aproximación econométrica para explicar dichas relaciones.

### **3. MODELIZACIÓN**

Como ya se dijo anteriormente, una de las características básicas de la modelización económica en telecomunicaciones es la distinción entre el acceso y el uso. Para ello se utilizan las ideas teóricas propuestas originalmente para la demanda de líneas telefónicas. Es obvio que el uso de un servicio dado será solo posible si el usuario accede al servicio. Es decir, *el uso será condicional en el acceso*. Al mismo tiempo, si alguien opta por el acceso del servicio (por ejemplo desde el hogar) será porque planea darle algún tipo de uso. O sea que *el acceso es condicional al uso*. Esta observación es central en el modelo pionero propuesto por Artle y Averous (1973) y constituye la piedra angular en la modelización de la demanda de las telecomunicaciones (ver Taylor, 1994).

### 3.1. FUNCIONES DE UTILIDAD

La evidencia teórica en la demanda de Internet sugiere que el servicio de Internet es utilizado para ahorrar dinero y tiempo. Siguiendo el marco teórico planteado en Taylor<sup>4</sup> (1994), existen dos tipos de agentes:

$G_0$  : Subconjunto de agentes sin acceso a la Red.

$G_1$  : Subconjunto de agentes con acceso a la Red.

La función de utilidad para el individuo  $i$ -ésimo se la expresa como:

$$U^i = U^i(x^i, \delta^i q^i) \quad (2.1)$$

donde  $x^i$  es el vector de los otros bienes consumidos por el  $i$ -ésimo agente, y las variables dicotómicas determinan el estado de acceso del agente, es decir:

$$q^i \begin{cases} q & \forall i \in G_1 \\ 0 & \forall i \in G_0 \end{cases} \quad \delta^i \begin{cases} 1 & \text{si el agente accede} \\ 0 & \text{otra alternativa} \end{cases}$$

El planteamiento entonces se presenta como un problema de maximización de las siguientes funciones de utilidad individual:

---

<sup>4</sup> Lester Taylor resume las contribuciones a la teoría de la demanda telefónica desde mediados de los años 70. El acceso-no acceso al servicio juega un importante rol en el análisis.

$$\begin{aligned} U^1 &= U^1(x^1, q) \text{ si } \delta=1 \\ U^0 &= U^0(x^0) \text{ si } \delta=0 \end{aligned} \quad (2.2)$$

### 3.2. FUNDAMENTOS PARA LA APROXIMACIÓN ECONOMÉTRICA

La literatura sugiere que el uso de Internet difiere de otros usuarios de telecomunicaciones en el tipo de atributo relevante para uno y otro servicio. Esto es sostenido por Rappoport *et al.* (2002) en donde exponen las diferencias entre la demanda telefónica y la demanda de Internet. En la misma línea Jackson *et al.* (2002) utiliza un modelo de opción entre trabajo-ocio asumiendo que el agente quiere ganar dinero, tener momentos de ocio y además actividad online.

La teoría sugiere que la maximización de la utilidad del agente que utiliza el servicio de Internet está condicionada por el consumo de otros bienes y su distribución de dinero y tiempo. Una aproximación lineal a la función de utilidad condicional se presenta como sigue:

$$\begin{aligned} U_i^* &= x_i^T \beta + \varepsilon_i \\ i &= 1, 2, \dots, n. \end{aligned} \quad (2.3)$$

en donde  $U_i^*$  es la utilidad latente que el agente  $i$  experimenta al acceder y/o usar Internet,  $\beta$  es el vector de parámetros desconocidos a estimar y representa el vector de utilidades marginales de cada uno de los regresores que se encuentran en el vector  $x_i^T$ , y  $\varepsilon_i$  es el término de error.

Los parámetros de la función de utilidad latente individual (las utilidades marginales) son estimados en base a la información obtenida y con la cual hemos hecho el análisis descriptivo del apartado anterior. En este capítulo se utiliza el enfoque de la función de utilidad aleatoria para analizar las características de los usuarios de Internet en España y observar los impactos que las características sociodemográficas tales como el hábitat, el género sexual, la edad o el nivel de estudios entre otras provocan en la probabilidad de que los agentes opten por acceder a Internet desde el hogar o usar el servicio tanto desde el hogar como desde otros sitios tales como el trabajo, el centro de estudios, etc.

Teniendo en cuenta las relaciones entre el acceso y el uso de Internet comentado más arriba, se puede decir que la división de la muestra de tamaño  $N$  es tal como sigue:

$P_1$  : Agentes con acceso a Internet en el hogar.

$P_2$  : Agentes sin acceso a Internet en el hogar, pero usan el servicio en otros lugares.

$P_3$  : Agentes sin acceso a Internet en el hogar, y que no usan el servicio.

$$N = P_1 + P_2 + P_3$$

Es decir que no se puede medir directamente la utilidad que proporciona Internet debido a que existe información no disponible. Esto último se lo tendrá en cuenta al especificar el modelo. Siguiendo a Meng y Schmidt (1985), para obtener eficiencia y consistencia, estimamos conjuntamente una ecuación de selección para determinar quién tiene acceso y quién no (ecuación de selección), y una segunda ecuación (de comportamiento) para explicar qué tipo de ancho de banda se tiene en el hogar. Las estimaciones se obtienen aplicando el método de máxima verosimilitud.

Para evaluar esta función de verosimilitud es necesario conocer las probabilidades de cada uno de los eventos observados. En el modelo que se especifica se está haciendo referencia a cuatro eventos, dependiendo de si el individuo es “elegido” para participar en la muestra (por ejemplo, que el individuo pertenezca al grupo  $G_1$ ), y la elección que se haga del tipo de conexión a Internet dentro de los elegidos primeramente.

Dados los cuatro eventos comentados, solo tres son observados debido a que no sabemos qué elección del tipo de conexión habrían hecho los agentes pertenecientes a  $G_0$ :

**Tabla 1: Probabilidades de Acceso a Internet y Tipo de Conexión en el Hogar**

		Tipo de Conexión	
		$q^i = q$	$q^i = 0$
Acceso	$\delta^i = 1$	$P(q^i = q   \delta^i = 1)$	$P(q^i = 0   \delta^i = 1)$
	$\delta^i = 0$	$P(q^i = q   \delta^i = 0)$	$P(q^i = 0   \delta^i = 0)$

Mientras que no se tengan observaciones para calcular las probabilidades  $P(q^i = 0 | \delta^i = 1)$  y  $P(q^i = 0 | \delta^i = 0)$ , sin embargo a través de  $\delta^i$  tenemos que:

$$P(q^i = 0 | \delta^i = 1) + P(q^i = 0 | \delta^i = 0) = P(q^i = 0)$$

Entonces, para poder considerar la función de probabilidad como tal:

$$P(q^i = q | \delta^i = 1) + P(q^i = q | \delta^i = 0) + P(q^i = 0) = 1$$

De ahí que la función de verosimilitud quedará expresada:

$$L = \prod_{i=1}^N \left[ P(q^i = q | \delta^i = 1) \right]^{q^i \delta^i} \left[ P(q^i = q | \delta^i = 0) \right]^{q^i (1-\delta^i)} \left[ P(q^i = 0) \right]^{(1-q^i)} \quad (2.4)$$

Esta relación bivalente puede ser considerada a través de un procedimiento en dos etapas que permita obtener una estimación consistente de los parámetros y controlar el sesgo debido a la selección no aleatoria de la muestra (Heckman, 1979). Siguiendo a Sigelman y Zeng (1999), este método es utilizado cuando el interés se centra fundamentalmente en la relación que existe entre  $x$  e  $y$  pero la información se encontrará disponible sólo en el caso en que otra variable,  $z^*$ , exceda cierto valor concreto. Específicamente:

$$\begin{bmatrix} y_i^* \\ z_i^* \end{bmatrix} = \begin{bmatrix} x_i \beta \\ w_i \gamma \end{bmatrix} + \begin{bmatrix} u_i \\ v_i \end{bmatrix} \quad (2.5)$$

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim NID \left[ 0, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right]$$

La primera ecuación está centrada en la conducta del individuo acerca de la elección, mientras que la segunda se centra en el interés de la selección dentro de la muestra, y  $\rho$  es el coeficiente de correlación entre los errores<sup>5</sup>. El estimador está basado en la siguiente esperanza condicional de la variable  $y$  :

---

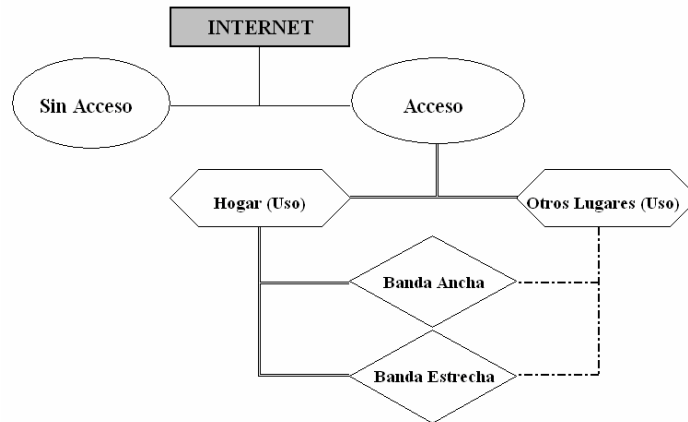
<sup>5</sup> La restricción de que la varianza de  $v_i$  es igual a 1 es impuesta porque de  $z_i^*$  sólo se observa esto. De hecho las variables que actualmente serán observadas son  $y_i$  y  $z_i$  que están relacionadas con  $y_i^*$  y  $z_i^*$  tal que  $y_i = y_i^*$  y  $z_i = 1$  sólo si  $z_i^* > 0$ . Si  $z_i = 0$  entonces  $y_i$  será inobservada. Véase Davidson y MacKinnon (1993).

$$E[y_i | z_i^* > 0] = x\beta + \sigma^2 \rho \lambda_i (-w_i \gamma) \quad (2.6)$$

en donde  $\lambda_i$  es la inversa del Ratio de Mills. Tal y como podemos ver, la ecuación (2.6) implica que la esperanza condicional de  $y_i$  será  $x_i\beta$  sólo cuando la correlación entre ambos errores sea no significativa. En caso contrario, dicha esperanza siempre se verá afectada por la ecuación de selección de la muestra.

A continuación se representan las decisiones de acceso y uso de Internet en el Gráfico 8:

**Gráfico 8:** Relaciones entre el Acceso y el Uso de Internet



Tal y como podemos ver en el Gráfico 8, la demanda de algún tipo de conexión a Internet para el hogar sólo puede darse si antes se cuenta con dicho acceso. A continuación se especifica un modelo para el acceso y uso de Internet manteniendo esta premisa y a partir de los atributos económicos, tecnológicos y sociodemográficos del

individuo. A su vez, el modelo utilizado está basado en la maximización de la utilidad del agente que accede y usa Internet en un determinado escenario. Es decir:

$$\text{Acceso a Internet} = f [\text{atr. económicos, atr. tecnológicos, atr. sociodemográficos}]$$

$$\text{Uso de Internet} = g [\text{atr. económicos, atr. tecnológicos, atr. sociodemográficos} | \text{Acceso}]$$

### 3.3. ESPECIFICACIÓN

De acuerdo al tipo de datos disponibles, utilizaremos un modelo Probit binario corregido por sesgo de selectividad. Así, en la primera especificación referida al acceso de Internet en el hogar del  $i$ -ésimo individuo se ajustan dos ecuaciones con ambas variables endógenas ficticias. La primera ecuación distinguirá entre los individuos que acceden y no acceden a Internet en el hogar, y la segunda lo hará entre los que acceden por banda ancha y los que no. Es decir que el modelo tiene la estructura

$$\begin{aligned} y_{1i}^* &= x_{1i}\beta_1 + u_{1i} \\ y_{2i}^* &= x_{2i}\beta_2 + u_{2i} \end{aligned} \quad (2.7)$$

$$\begin{aligned} y_{1i} &= 1 & \text{si } y_{1i}^* > 0 & & y_{2i} &= y_{2i}^* & \text{si } y_{1i}^* > 0 \\ y_{1i} &= 0 & \text{si } y_{1i}^* < 0 & & y_{2i} &= 0 & \text{si } y_{1i}^* > 0 \end{aligned}$$

que está referida al tipo de conexión en el hogar ( $y_{2i} = 1$  si posee conexión por banda ancha), pero que sólo tendrá una respuesta cuando la variable  $y_{1i} = 1$ , es decir, cuando el individuo tenga alguna conexión a Internet en el hogar. Esto significa que la variable  $y_{1i}$  elegirá no aleatoriamente la muestra para estimar un modelo para medir el efecto de los determinantes de la conexión a Internet por banda ancha en el hogar. Las variables

que componen el vector de regresores están explicadas en la tabla que se presenta a continuación:

**Tabla 2: Regresores**

<i>Nombre</i>	<i>Definición</i>
<u><i>Atributos Económicos</i></u>	
○ <i>income</i>	Índice de Renta Permanente Familiar (su elaboración se encuentra detallada en el apéndice).
<u><i>Atributos Tecnológicos</i></u>	
○ <i>pc</i>	= 1, si se conecta a Internet con un PC en el hogar; = 0 en otro caso
○ <i>laptop</i>	= 1, si se conecta a Internet con un ordenador portátil; = 0 en otro caso
○ <i>mobile</i>	= 1, si se conecta a Internet con el teléfono móvil; = 0 en otro caso
○ <i>frequer</i>	Intensidad de uso de Internet (cantidad de usos trimestrales = 70, 14, 3, 1, 0)
○ <i>usagecomp</i>	Intensidad de uso del ordenador (cantidad de usos trimestrales = 70, 14, 3, 1, 0)
○ <i>broadband</i>	= 1, si se tiene conexión de banda ancha en el hogar; = 0 en otro caso
<u><i>Atributos Sociodemográficos</i></u>	
○ <i>bestudying</i>	= 1, si el individuo está estudiando; = 0 en otro caso
○ <i>studylevel</i>	Nivel de estudios alcanzados (medido en años de estudio).
○ <i>housemembers</i>	Número de miembros en el hogar
○ <i>habitat</i>	Tamaño del hábitat del individuo (intervalo poblacional)
○ <i>male</i>	= 1, si el individuo es hombre; = 0 en otro caso
○ <i>age</i>	Edad del individuo

Aproximamos el modelo para los determinantes de la conexión por banda ancha en el hogar a nivel individual usando una muestra de toda España como:

$$y_{2i} = \beta_0 + \beta_1 \text{income}_i + \beta_2 \text{pc}_i + \beta_3 \text{laptop}_i + \beta_4 \text{mobile}_i + \beta_5 \text{frequser}_i + \beta_6 \text{bestudying}_i + \\ + \beta_7 \text{housemembers}_i + \beta_8 \text{habitat}_i + \beta_9 \text{male}_i + \beta_{10} \text{age}_i + \beta_{10} \text{agesq}_i + \beta_{11} \text{IMR}_i(x) + \varepsilon_i$$

$i = 1, \dots, 18.948$  individuos

(2.8)

donde  $\text{IMR}_i(x) = \frac{\phi(z^T \gamma)}{[1 - \Phi(z^T \gamma)]}$  es la inversa del Ratio de Mills, que como se explica

más arriba, se añade como regresor para corregir el sesgo de selectividad. La variable endógena  $y_{2i}$  es una variable ficticia que representa el output de la función indicador referida a si el individuo tiene una conexión a Internet por banda ancha en el hogar (es decir, igual a 1 si el individuo tiene conexión por banda ancha en el hogar y 0 en otro caso).

#### 4. ESTIMACIÓN Y DISCUSIÓN

La variable endógena en la ecuación de comportamiento se refiere a “tener o no Internet de Banda Ancha en el hogar” y es un indicador de la probabilidad de dicho suceso. Con la estimación del modelo especificado se tratará de medir cómo cada variable considerada afectará esa probabilidad. A su vez y como se comenta antes, el efecto marginal en  $y_{2i}$  está compuesto tanto por el efecto de la ecuación de selección como por la ecuación de comportamiento. En otras palabras, cada predictor en el modelo podría aparecer no solamente como una variable exógena en la ecuación de

comportamiento, sino también como un componente de  $IMR(x)$ . Una consecuencia de esto es que el efecto en  $k$  unidades de cambio de la variable exógena no será simplemente  $k$  veces el efecto de una unidad de cambio en este vector. El cambio en la variable endógena dependerá no solamente de la magnitud del cambio de la variable explicativa, sino también del punto de partida desde el que dicho cambio tiene lugar (Sigelman y Zeng, 1999).

#### **4.1. ACCESO A INTERNET EN EL HOGAR**

En la Tabla 3 se exponen los resultados de la estimación del modelo especificado en (2.8) referida a un modelo Probit binario con sesgo de selectividad corregido. Observamos que el signo de los coeficientes de las variables significativas son los esperados ya que los coeficientes de la renta, de tener un PC en el hogar y de la intensidad de uso de Internet tienen signos positivos. Lo contrario sucederá con los coeficientes de la edad y el tamaño del hábitat, cuyos coeficientes serán todos negativos. Observamos además que encontrarse estudiando, el número de miembros en el hogar y pertenencia al género masculino serán variables no significativas.

**Tabla 3: Resultados de la Ecuación (2.8)**

Variable Dependiente: $y_{2i}$ (Internet por banda ancha desde el hogar)			
Regresores	$\hat{\beta}$	$\epsilon$	$ z $
Constante	-9,030	---	15,98
<i>income</i>	0,837	0,99	13,19
<i>Pc</i>	1,883	1,58	33,45
<i>Laptop</i>	0,534	0,95	7,77
<i>Mobile</i>	0,261	0,08	3,01
<i>Frequer</i>	-0,002	-0,09	0,09
<i>bestudying</i>	0,031	0,12	0,59
<i>housemembers</i>	0,022	0,001	1,29
<i>Habitat</i>	0,060	0,02	7,23
<i>Male</i>	0,086	0,29	2,28
<i>Age</i>	-0,012	-0,66	2,03
<i>Agesq</i>	0,000084	0,44	1,28
<i>IMR(x)</i>	2,611	---	12,37

Muestra: 18.948 individuos, 4.470 con conexión a Internet en el hogar.  
 Razón de Verosimilitudes = -9229,138 (P-valor > chi-cuadrado = 0,00)  
 Pseudo  $R^2 = 0,4591$

En la columna  $\epsilon$  se incluyen las elasticidades en las medias calculadas a partir de los efectos marginales. Se han hecho diversos contrastes de especificación (tipo Wald) en los que no se rechazaba la hipótesis nula de ausencia de variables omitidas. Los valores de la columna  $z$  son los del estadístico de contraste para la significación individual del regresor dentro del modelo. Dichos valores son los equivalentes al ratio  $t$ , que al ser la muestra tan grande, estos valores tienden a distribuirse como una  $N[0,1]$ . Al observar los resultados del contraste de razón de verosimilitudes y el Pseudo  $R^2$  se puede afirmar que el modelo presenta buenos resultados de ajuste global

tanto en significatividad como de bondad de ajuste. A continuación se presenta la Tabla 4 con la validación cruzada entre valores observados y valores predichos<sup>6</sup>:

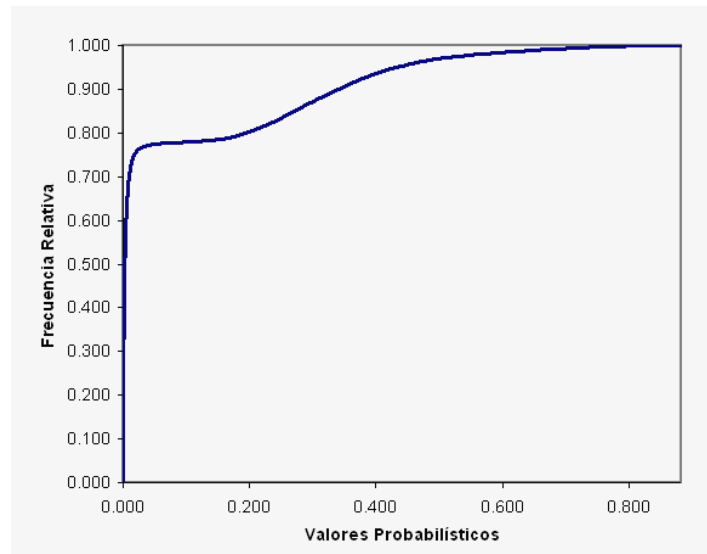
**Tabla 4: Validación del Modelo para el Acceso a Internet**

		Real		
		<i>BA=0</i>	<i>BA=1</i>	<i>Total</i>
Predicho	<i>BA=0</i>	15.286	620	15.906
	<i>BA=1</i>	2.131	911	3.042
	<i>Total</i>	17.417	1.531	18.948

El modelo tiene un 85,48% de aciertos totales, con un 59,50% de sensibilidad (aciertos iguales a uno), que es un porcentaje razonablemente alto dado el ajuste del modelo y el escaso porcentaje de poseedores de banda ancha en el hogar en la muestra total. Esto significa que los regresores propuestos tienen una importante influencia en la probabilidad de adquirir una conexión a Internet por banda ancha para el hogar.

A continuación presentamos el Gráfico 8 con las probabilidades predichas acumuladas a partir del modelo que acabamos de estimar. Aquí se observa que la probabilidad de adquirir una conexión por banda ancha en el hogar es mínima para más del 70% de la muestra. Esto se complementa con el análisis descriptivo anterior donde veíamos que en la muestra sólo un 25,2% poseía algún tipo de conexión a Internet en el hogar.

<sup>6</sup> Considerando que de la muestra total sólo un 8% posee una conexión a Internet de banda ancha en el hogar, el umbral utilizado en la validación cruzada es de 0,08.

**Gráfico 8:** Probabilidades Predichas Acumuladas

En cuanto a la evolución de los efectos marginales, los gráficos de los mismos se encuentran en el apéndice del final del capítulo. En general se observa que todos los atributos considerados contribuyen a explicar la probabilidad de adquisición del acceso a Internet por banda ancha en el hogar. En todos los casos la probabilidad aumenta y el impacto de cada uno de los regresores es medido a través de la diferencia entre la función de predicción estimada con y sin el mismo. Se observa por ejemplo que, después del impacto que provoca la renta, los atributos tecnológicos (*pc*, *laptop* y *mobile*) tendrán un considerable impacto en la probabilidad de adquisición de banda ancha, mientras que los atributos sociodemográficos tales como el hábitat y el género masculino también tendrán impactos considerables, aunque menores. Resumiendo:

- Una renta elevada implica una alta inclinación a demandar una conexión de banda ancha para el hogar. Esto se refleja en el signo positivo y la evolución del efecto marginal de la variable *income*.
- La variable *age* es una variable que tiene un comportamiento particular. En este modelo se la ha incluido dos veces: en niveles y al cuadrado. Esto significa que, *coeteris paribus*, el efecto marginal en la probabilidad de adquirir la conexión

por banda ancha será la función lineal  $\beta_9 + 2\beta_{10} \text{ age}$ . Como se observa en los resultados de la estimación, *age* tiene un coeficiente negativo y *agesqr* lo tiene positivo. Esto significa que el efecto en la probabilidad de adquirir Internet por banda ancha en edades tempranas es inverso, pero tenderá a aumentar hasta volverse positivo en edades mayores (aproximadamente sobre los 50 años).

- Los atributos tecnológicos tienen un rol positivo en la probabilidad de demandar banda ancha. Esto es lógico debido a que si la gente quiere un acceso a Internet para el hogar, será necesario tener previamente un ordenador de sobremesa o un portátil. Es más, se observa que también la variable *mobile* tiene un efecto positivo en la probabilidad de demandar banda ancha. Sin embargo, la intensidad de uso del servicio de Internet (*frequer*) es una variable no significativa.

#### 4.2. USO DE INTERNET

A continuación se aborda el estudio de los determinantes del uso de Internet. Este problema será diferente del estudiado anteriormente (suscripción al servicio de Internet desde el hogar es una decisión más “colectiva”) debido a que ahora se busca observar el determinado perfil de un individuo y saber si tiene más o menos probabilidad de ser usuario potencial de Internet, o no. ¿Qué es lo que busca el usuario al utilizar Internet? Según se comenta en la primera parte del artículo, el usuario de Internet buscará principalmente velocidad en la transmisión de datos. Esto lo puede conseguir conectándose a Internet tanto desde el hogar como también desde otros lugares.

La distinción entre usuarios y no usuarios de Internet se podría resolver además utilizando técnicas de análisis de distancias y otras técnicas multivariantes tales como el

análisis discriminante, siempre y cuando se puedan conocer los parámetros de las distribuciones. Cuando los datos disponibles para la clasificación no son normales, como las variables que se utilizan en este análisis, que en su mayoría son discretas, una posibilidad que se tiene es construir un modelo que explique los valores de la variable de clasificación. En el presente análisis entonces, se añadirá al análisis una variable que toma valor cero cuando la observación corresponde a un no usuario, y tomará valores 1, 2, 3 y 4 según el usuario utilice Internet en uno, dos tres o cuatro sitios distintos. Los sitios considerados son el hogar, el lugar de trabajo, el centro de estudios y otros lugares tales como bibliotecas, cybercafés, etc. Este planteamiento será así debido a que los valores no corresponden a un ordenamiento propiamente dicho ni tampoco son alternativas mutuamente excluyentes, pues nada quita de que se pueda usar Internet desde uno o más de los cuatro lugares considerados.

En principio se podría analizar estos datos utilizando el modelo de regresión lineal múltiple, pero la abundancia de ceros, los valores tan pequeños que toma la variable endógena y la naturaleza claramente discreta del fenómeno lleva a buscar una especificación que tenga en cuenta todas estas características (Greene, 1998). Se especifica entonces un modelo de regresión de Poisson para la variable  $y_i$  cuyo campo de variación es el intervalo discreto  $[0, 4]$ . Este tipo de especificación supone que  $y_i$  es una realización de una variable aleatoria con distribución de Poisson de parámetro  $\lambda_i$ , y que ese parámetro está relacionado con los regresores  $x_i$ . La función de probabilidad de este modelo será:

$$\Pr(y|x) = \frac{e^{-\lambda} \lambda^y}{y!}$$

La crítica principal que recibe el modelo de Poisson es que supone que la media y la varianza son iguales. El fallo de este supuesto (denominado *sobredispersión*) tiene unos efectos cualitativamente similares a la heteroscedasticidad en los mínimos cuadrados, aunque en este caso el efecto sobre la varianza del estimador es mayor. Un modelo similar que recoge la sobredispersión es el modelo Binomial Negativo, cuya función es similar a la de Poisson, aunque en este caso se introduce un parámetro  $\alpha$  que recoge la sobredispersión y que converge a 1 cuando la sobredispersión converge a cero (Cameron y Trivedi, 1998), lo que provoca que las estimaciones sean similares a las del modelo de Poisson. La función de probabilidad entonces tiene la forma:

$$\Pr(y|x, \alpha) = \frac{e^{-\alpha\lambda} \alpha \lambda^y}{\Gamma(y+1)}$$

La formulación logarítmico-lineal de este tipo de modelos es

$$\ln \lambda_i = \beta^T x_i \quad (2.9)$$

La esperanza condicional y los efectos marginales vendrán dados por

$$E[y_i|x_i] = V[y_i|x_i] = \lambda_i = e^{\beta^T x_i} \quad (2.10)$$

$$\frac{\partial E[y_i|x_i]}{\partial x_i} = \lambda_i \beta \quad (2.11)$$

Las variables explicativas no serán exactamente las mismas del modelo anterior, aunque se sigue la misma filosofía de considerar los atributos económicos, tecnológicos y sociodemográficos de los individuos. La especificación será entonces

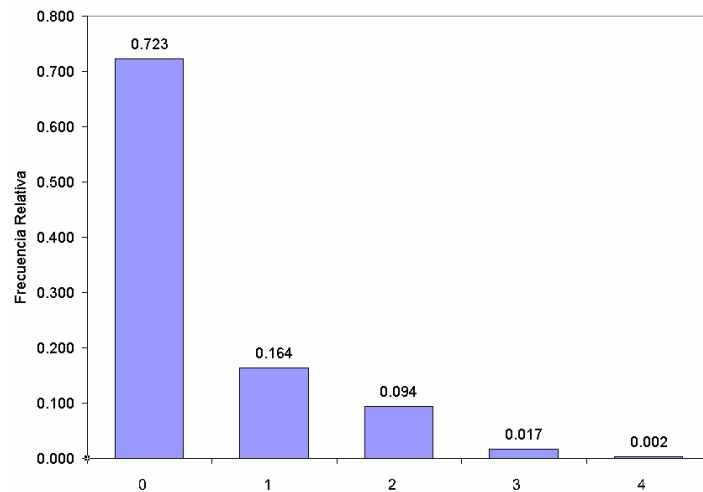
$$USE_i = \exp(\beta_0 + \beta_1 \text{income}_i + \beta_2 \text{broadband}_i + \beta_3 \text{usagecomp}_i + \beta_4 \text{bestuding}_i + \beta_5 \text{studylevel}_i + \beta_6 \text{housemembers}_i + \beta_7 \text{habitat}_i + \beta_8 \text{male}_i + \beta_9 \text{age}_i) + u_i$$

$$i = 1, \dots, 18.948$$

(2.12)

En particular se busca detectar las relaciones existentes entre distintas intensidades de uso y la relación entre algunos de los atributos tecnológicos y sociodemográficos. A continuación se observa el histograma de la variable *USE* :

**Gráfico 9:** Uso de Internet



En el gráfico se puede observar que la mayor parte de los individuos se declaran no usuarios de Internet (72,3%). A estos les siguen los usuarios de menos intensidad (en

uno y dos lugares, con 16,4% y 9,4% respectivamente). Los usuarios más intensos (en tres y cuatro lugares distintos) representan menos del 2% en total.

Cuando existe un número excesivo de respuestas nulas, como en este caso, se podría pensar en la existencia no de uno sino de dos procesos generadores de los datos: uno para los valores nulos y otro para los que no lo son. Entonces, a los modelos de Poisson y Binomial Negativo se les agrega un filtro  $\omega$  que puede modelizarse como una función logit o probit, ambos con resultados similares. Este filtro  $\omega$  representa la probabilidad de un cero extraordinario (por no pasar el filtro). Este tipo de modelos se denominan *Poisson inflado de ceros* cuya especificación es:

$$\Pr(y|x, \omega) \begin{cases} \omega + (1-\omega) \frac{e^{-\lambda} \lambda^y}{y!} & \forall y = 0 \\ (1-\omega) \frac{e^{-\lambda} \lambda^y}{y!} & \forall y \geq 1 \end{cases}$$

y el *Binomial Negativo inflado de ceros* con función igual a:

$$\Pr(y|x, \omega, \alpha) \begin{cases} \omega + (1-\omega) \frac{e^{-\alpha\lambda} \alpha \lambda^y}{\Gamma(y+1)} & \forall y = 0 \\ (1-\omega) \frac{e^{-\alpha\lambda} \alpha \lambda^y}{\Gamma(y+1)} & \forall y \geq 1 \end{cases}$$

Entonces, como se comentaba más arriba, de acuerdo a las características de la variable endógena (datos de recuento con gran cantidad de ceros), el modelo más idóneo será alguno de los mencionados. Para evaluar dicha idoneidad, se practica un doble contraste. El primero basado en la significatividad del parámetro  $\alpha$ , que si es significativo implica la presencia de sobredispersión y viceversa. El segundo lugar

también puede contrastarse en función de la bondad de ajuste y la función de verosimilitud. En el caso de la existencia de sobredispersión, la función de verosimilitud binomial negativa ajustará mejor que la de Poisson. Por otra parte, un contraste para poder comparar y elegir un modelo con filtro contra la alternativa de un modelo tradicional fue propuesto por Vuong (1989). Básicamente consiste en comparar la bondad de ajuste de un modelo de Poisson (Binomial Negativo) respecto a un modelo de Poisson inflado de ceros (Binomial Negativo inflado de ceros). El contraste calcula, para cada observación, el logaritmo de la relación entre el valor previsto en un modelo inflado de ceros y el valor previsto en un modelo tradicional. A esta nueva variable (denominada  $m$ ) se le calcula el estadístico:

$$V = \frac{\sqrt{n}(\bar{m})}{S_m} \succ Z = N(0,1)$$

donde  $n$  es el número de observaciones,  $\bar{m}$  es la media y  $S_m$  la desviación típica. Dicho estadístico se distribuye como una normal tipificada. Para un nivel de significación del 5%, un valor  $V$  superior a 1,96 favorecerá al modelo inflado de ceros.

A continuación se ven los resultados del modelo estimado para el uso de Internet, en donde presentan además de los coeficientes, los ratios de incidencia y los efectos marginales.

Tabla 5: Resultados del Modelo del Uso Individual de Internet

Variable Endógena ( <i>USE</i> )	Modelos			$e^{\hat{\beta}_i}$	Efecto Marginal ( $dy/dx$ )	$z$
	Poisson	Binomial Negativo	Poisson inflado de ceros			
Constante	-2.701	-2.701	-0,940	---	---	32.45
<i>income</i>	0.086	0.086	0,320	1.089	0.126	27.52
<i>broadband</i>	0.357	0.357	0,219	1.429	0.061	12.57
<i>usagecomp</i>	0.327	0.327	0,009	1.387	0.048	27.69
<i>bestudying</i>	0.394	0.394	0,269	1.484	0.067	12.57
<i>studylevel</i>	-2.156	-2.156	-0,833	0.116	-0.316	12.24
<i>housemembers</i>	-0.056	-0.056	-0,019	0.946	-0.008	5.48
<i>habitat</i>	0.007	0.007	0,003	1.007	0.001	1.37*
<i>male</i>	0.221	0.221	0,068	1.247	0.033	9.66
<i>age</i>	-0.031	-0.031	-0,011	0.969	-0.005	30.68
$\alpha$	---	1,22E-07	---	---	---	0,001*
Vuong	---	---	39,69	---	---	---
Probabilidad $N(0,1)$			0,00			
Pseudo $R^2$	0,3912	0,3707	0,3912	---	---	---
LR Chi-cuadrado	12.986,61	11.905,47	1.140,72	---	---	---
Probabilidad $\chi^2_{(1)}$	0,00	0,00	0,00			
Media de <i>USE</i> : 0.4123 Desviación Típica de <i>USE</i> : 0.7500 Observaciones = 18.816						

(\*) no significativo al 95%

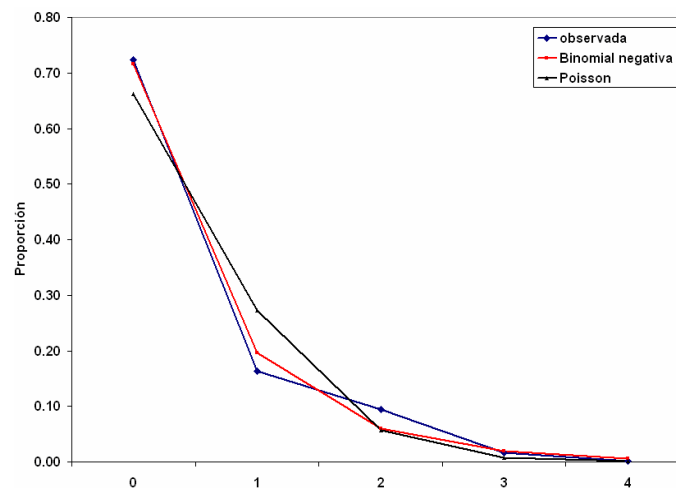
Se observa que la evidencia favorece al modelo de Poisson, debido a que el estimador del parámetro  $\alpha$  no es significativo. También se ve que la bondad de ajuste y los valores del ratio de verosimilitud son superiores (0,3912 es mayor que 0,3707 y 12.986,61 lo es con respecto a 11.905,47). Por otro lado, la evidencia también favorece a la existencia de un filtro, al observar que el estadístico de Vuong es significativo.

Como contribución adicional a favor del modelo de Poisson, aunque vemos que la media y la varianza de la variable endógena no son iguales, al comparar los ajustes de ambos modelos no habría evidencia de sobredispersión dañina:

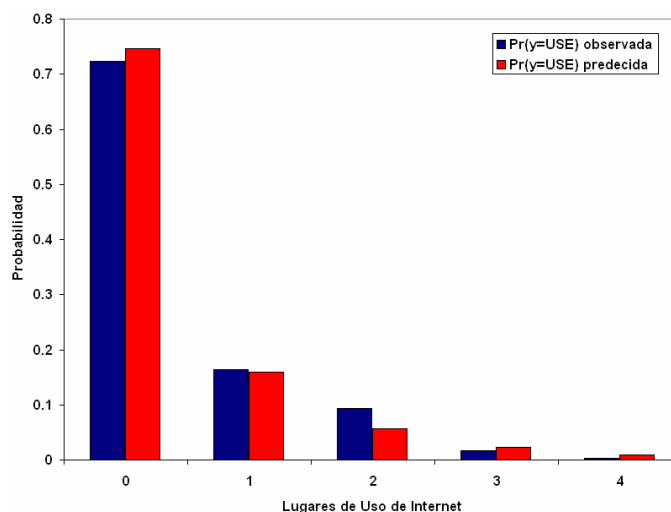
**Tabla 6: Comparación de Ajustes de los Modelos Poisson y Binomial Negativo**

Rango de Variación de USE	Probabilidades binomiales negativas Media = 4123 Sobredispersión= 1,2162	Probabilidades de Poisson Lambda 0,4123
0	0,7159	0,6621
1	0,1966	0,2730
2	0,0598	0,0563
3	0,0188	0,0077
4	0,0060	0,0008

**Gráfico 10: Comparación de Ajustes de los Modelos Poisson y Binomial Negativo**



A continuación, en el Gráfico 11 se observan las predicciones con el modelo de Poisson seleccionado:

**Gráfico 11:** Predicciones del Modelo Estimado

Se observa que la mayoría de las variables tienen efectos positivos en la cantidad esperada de sitios de uso de Internet. En el caso del ingreso, un incremento unitario del índice *income* provoca un incremento de 8,9% en el número de sitios esperados de uso (porcentaje denominado *incidencia* y obtenido a partir del cálculo de  $e^{\hat{\beta}}$ ). Esto equivale a un efecto marginal de 0,126. Podría deberse a que el incremento del ingreso incrementa a su vez la adquisición de bienes y servicios que sólo se adquieren a un determinado nivel de ingresos, Internet entre estos. Es decir que Internet se comporta no como un bien de primera necesidad sino más bien como un bien normal o un bien de lujo.

El tener acceso de banda ancha en el hogar tiene una incidencia de 42,9% en los sitios esperados. Esto podría deberse al hecho de que tener el servicio disponible en el hogar crea un hábito de uso que lleva a utilizar el servicio fuera del hogar. El efecto marginal es de 0,061. Algo parecido sucede con el otro atributo tecnológico considerado que es la frecuencia de uso del ordenador, herramienta casi fundamental para poder

acceder a la Red. Su incidencia en los sitios esperados de uso es positiva de 38,7% y su efecto marginal es de 0,048.

El hecho de encontrarse estudiando tiene la incidencia positiva más grande en los lugares esperados de uso (48,4%). Esto tiene sentido y demuestra una vez más la importancia que Internet ha adquirido en la educación y la enseñanza.

Otro efecto positivo en el uso de Internet, aunque de escasa incidencia, es el género masculino. Pertenecer a este género incrementa los lugares de uso esperados en 24,7%. Su efecto marginal es de 0,033.

Por otro lado, la tasa de incidencia nos indica que la edad, el número de miembros en el hogar y el nivel de estudios tienen una relación inversa con el número de sitios esperados en donde utilizar Internet. Aumentar un año el nivel en estudios acabados disminuye el número esperado en 88,4% (porcentaje obtenido a partir de  $0,116 - 1$ ). Recordemos que esta variable se refiere al nivel de estudios alcanzado hasta el momento. Esto podría deberse a que al aumentar el nivel de estudios habría una concentración de sitios para utilizar Internet (hogar y centro de estudios primero, y hogar y trabajo después). Arriba veíamos que el efecto de encontrarse estudiando aún era exactamente lo contrario. El efecto marginal del nivel de estudios es de  $-0,316$ .

Por último, tenemos el caso de los miembros del hogar la disminución del número esperado de sitios de uso es de 5,4% equivalente a un efecto marginal de  $-0,008$ . El aumentar un año más de edad la disminución de sitios esperados de uso es de 3,1%. Esto equivale a un efecto marginal de  $-0,005$ .

## **5. CONCLUSIONES**

EL objetivo de este capítulo fue detectar y analizar los determinantes de la demanda de Internet en España, tanto del acceso en el hogar con conexiones de alta velocidad, como del uso del servicio como tal.

En primer lugar se describieron las características que podrían afectar al acceso y el uso de Internet utilizando un análisis comparativo entre las distintas tipologías de usuario, el tipo de conexión desde el hogar, y la cantidad de sitios de uso del servicio. Se encontró que el 25,2% de la muestra accede a Internet con algún tipo de conexión desde el hogar. También se encontró que este porcentaje es menor que los que se declaran usuarios de Internet (35,5% de la muestra). Estos porcentajes varían considerablemente según de qué región se trate. Además se han observado relaciones entre las características sociodemográficas y el uso de Internet similares a las de otros países desarrollados.

Luego se proponen dos modelos econométricos que explican los determinantes de ambas demandas. Primero se especifica y estima un modelo para el acceso a Internet en el hogar a través de un Probit binario con corrección del sesgo de selectividad (Heckman, 1979) y se mide la importancia de determinantes como el ingreso o los atributos tecnológicos y su impacto positivo en la probabilidad de adquisición de Internet de banda ancha en el hogar, o de determinantes sociodemográficos tales como el hábitat y la edad con impactos negativos.

El segundo modelo propuesto se refiere al uso de Internet teniendo en cuenta el número de sitios de uso (cuatro en total). Para esta estimación y por el tipo de información de la variable endógena se acaba utilizando un modelo de regresión de Poisson inflado de ceros, donde se observa que los efectos de los tres atributos

mencionados arriba son en general directamente proporcionales al uso de Internet, con excepciones tales como la edad o la cantidad de miembros en el hogar. Observamos que los mayores efectos sobre el uso de Internet los provocan el contar con una conexión a Internet por banda ancha en el hogar, encontrarse estudiando en ese momento y la pertenencia al género masculino.

En el próximo capítulo se continúa con lo visto hasta aquí del acceso y uso de Internet y se analizan las diferencias entre las distintas regiones españolas en cuanto a grado de digitalización alcanzado y también se miden y detectan sus determinantes.

## **6. REFERENCIAS**

Artle, R. y Averous, C. "The Telephone System as a Public Good: Static and Dynamic Aspects" *Bell Journal of Economics and Management Science*, Vol. 4, No. 1, 89-100, spring 1973.

Cameron, A. y Trivedi, P. "Regression Analysis of Count Data". Cambridge: Cambridge University Press (1998).

Cassel, C., "Demand for and Use of Additional Lines by Residential customers", in Loomis and Taylor (eds.) *The Future of the Telecommunications Industry: Forecasting and Demand Analysis*, Kluwer Academic Publishers, Boston, 1999.

Davidson, R. y MacKinnon, J., "Econometric Theory and Methods" Oxford University Press, 1993.

Duffy-Deno, K. T. "Demand for Additional Telephone Lines: An Empirical Note", *Information Economics and Policy*, 13, 301-309, 2001.

Friedman, M. "A Theory of the Consumption Function", Princeton University Press for the National Bureau of Economic Research, 1957.

Gabel, D. y Kwan, F. "Accessibility of Broadband Telecommunications Services by Various Segments of the American Population", Telecommunications Policy Research Conference, August 2002.

Goodman, A. y Kawai, M. "Permanent Income, Hedonic Price and Demand for Housing: New Evidence", *Journal of Urban Economics*, 12, 214-237, 1982.

Goolsbee, A. "The Value of Broadband and the Deadweight Loss of Taxing New Technology", Mimeo, University of Chicago, 2000.

Greene, W. "Econometric Analysis" Prentice Hall Ed.,1998.

Heckman, J. "Sample Selection Bias as a Specification Error" *Econometrica*, Vol. 47, No. 1, January 1979.

Jackson, M., Lookabaugh, T., Savage, S., Sicker, D. y Waldman, D. "Broadband Demand Study: Final Report" Telecommunications Research Group, University of Colorado, 2003.

Madden, G. y Coble-Neal, G., “Australian Residential Telecommunications Consumption and Substitution Patterns” preliminary draft, 15<sup>th</sup> International Telecommunications Society Meeting, Berlin, Germany, 4-7 September 2004.

Madden, G., Savage, S. y Simpson, M. “Information Inequality and Broadband Network Access: An Analysis of Australian Household Survey Data” *Industry and Corporate Change*, Oxford University Press, 1049-1056, 1996.

Mc Fadden, D. “Conditional Logit Analysis of Qualitative Choice Behavior” in Zarembka (ed.) *Frontiers in Econometrics*, Academic Press, pp. 1117-1156, 1974.

Mc Fadden, D. “Econometric Analysis of Qualitative Response Models” in Z. Grilliches and M. Intriligator (eds.) *Handbook of Econometrics*, Amsterdam, North Holland, pp. 1376-1425, 1984.

Meng, C. y Schmidt, P. “On the Cost of Partial Observability in the Bivariate Probit Model” *International Economic Review*, Vol. 26, No. 1, february 1985.

Organization for Economic Cooperation and Development, “The Development of Broadband in OECD Countries”, October 29, 2001.

Owen, B. “Broadband Mysteries” in Crandall, R. and Alleman, J. (ed.) *Broadband: Should we Regulate High-speed Internet Access?* AEI – Brookings Joint Center for Regulatory Studies, 2002.

Pérez Amaral, T., F. Alvarez y B. Moreno “Business Telephone Traffic Demand in Spain 1980-1991: an Econometric Approach”, *Information Economics and Policy*, 7, 115-134, 1995.

Rappoport P., Taylor L., y Kridel D. “The Demand of Broadband: Access, Content, and the Value of Time”, in *Broadband: Should We Regulate High-Speed Internet Access?*, ed. por R. W. Crandall y J. H. Alleman, AEI-Brookings Joint Centre for Regulatory Studies, Washington, D.C., 2002.

Sigelman, L. y Zeng, L. “Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models” *Political Analysis*, 8:2, The George Washington University Working Papers, December 16, 1999.

Taylor, L. D. “Towards a framework for analyzing internet demand”, Manuscript, U. of Arizona, 2000.

Taylor, L. D. “Telecommunications Demand in Theory and Practice” Kluwer Academic Publishers, 1993.

U.S. Department of Commerce, and National Telecommunications & Information Administration, “A Nation Online: How Americans Are Expanding Their Use of the Internet”, February 2002.

Varian, H. “The Demand for Bandwidth: Evidence from the INDEX Project”, Mimeo, University of California, Berkeley, 2002.

Vuong, Q. "Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses"

Econometrica 57: 307-333

## APÉNDICE 1: Construcción de una Medida Aproximada de la Renta Individual

En la literatura econométrica, cuando se estiman tanto funciones de demanda como cualquier otro tipo de ecuación que tenga en cuenta los determinantes del acceso y uso de Internet, se hace necesario utilizar variables que hagan referencia a los atributos económicos, singularmente la renta.

A pesar de lo necesario de esta información para los modelos planteados tanto en este como en los modelos que se plantearán en los capítulos siguientes, este tipo de información concreta no está disponible en la muestra que se utiliza. Por ejemplo, teniendo en cuenta que la decisión de contratar un servicio de Internet para el hogar dependerá no sólo del precio, sino también de la restricción presupuestaria familiar, será necesario encontrar alguna variable que represente estos tópicos mencionados en el estricto sentido de lo que representa, es decir *poder de mercado* para el caso del precio, y *poder adquisitivo* en los distintos tipos de renta.

Para el caso del caso del precio, este corresponderá a un atributo del producto dadas las distintas alternativas de conexión a Internet (conexión con teléfono convencional, ADSL, LDSI y por cable). Sin embargo, se opta por omitirlo al considerar el tipo de muestra utilizada (datos de sección cruzada) y además debido al proceso de regulación de precios de las telecomunicaciones existente en España. Estas dos causas harían que dicha variable tenga poca variabilidad.

En el caso de la renta se utilizará el concepto de *Renta Permanente* (Friedman, 1957), considerado más conveniente que el concepto de renta corriente para analizar la elección de este tipo de producto.

Para obtener un indicador que sirva de variable *proxy* de la renta permanente familiar, y dada la disponibilidad de información contenida en la muestra, se optó por

basarse en la propuesta de Goodman y Kawai (1982) aplicada al mercado de la vivienda. Estos autores utilizan un modelo de capital humano donde la renta está determinada por la inversión en capital humano y capital no humano.

En nuestro caso consideramos todas las observaciones en la muestra (18.948 entrevistas realizadas en toda España), en donde a partir de estas se plantea un índice de tipo promedio ponderado a través de la siguiente especificación:

$$income_i = 0,625HK_i + 0,375NHK_i \quad (2.13)$$

donde:

$income_i$ : Índice de Renta permanente.

$HK_i$ : Inversión en Capital Humano del agente  $i$  –ésimo.

$NHK_i$ : Inversión en Capital no Humano del agente  $i$  –ésimo.

1) Variables consideradas en la construcción de  $HK_i$ : estas variables incluyen:

- a) Encontrarse estudiando en ese momento.
- b) Nivel de educación alcanzado.

Estos dos conceptos son cuantificados a través de siete variables ficticias expuestas en el siguiente cuadro:

**Tabla A. Variables para el Capital Humano**

<i>Variable</i>	<i>Definición</i>
<i>Ecursa</i>	=1 si el agente entrevistado está estudiando en ese momento; =0 en otro caso.
<i>1<sup>er</sup> nivel</i>	=1 si el agente entrevistado no completó ningún nivel de estudios; =0 en otro caso.
<i>2<sup>do</sup> nivel</i>	=1 si el agente entrevistado ha completado la educación

	primaria; =0 en otro caso.
3 <sup>er</sup> nivel	=1 si el agente entrevistado ha completado la primera parte de la educación secundaria; =0 en otro caso.
4 <sup>to</sup> nivel	=1 si el agente entrevistado ha completado la segunda parte de la educación secundaria; =0 en otro caso.
5 <sup>to</sup> nivel	=1 Si el agente entrevistado ha completado cursos de Formación Profesional; =0 en otro caso.
6 <sup>to</sup> nivel	=1 Si el agente entrevistado ha completado alguna carrera universitaria de más de tres años; =0 en otro caso.

Se considera que en promedio lleva 16 años desde el ingreso en el primer grado de la escuela primaria hasta la obtención del grado universitario. Así, teniendo en cuenta seis niveles de estudios, y dándoles a todos el mismo peso, se supondrá que se utilizan 2,66 años por cada nivel condicionado al anterior. De acuerdo con esto, cada nivel será ponderado con el siguiente coeficiente:

$$\pi_i = \frac{2,66}{16} j$$

$$j = 1, 2, \dots, 6.$$

Entonces la proporción de capital humano vendrá dado por la siguiente definición

$$HK_i = 0,5 ecursa + 0,5 \pi_i.$$

- 2) Variables consideradas en la construcción de  $NHK_i$ : Dados los datos de la muestra, los modelos propuestos en este capítulo y los que se propondrán en capítulos posteriores, se considerará a la inversión en capital no humano a la inversión en tecnología. Esta inversión puede ser cuantificada en nueve variables ficticias tales como las que se muestran en la siguiente Tabla B:

**Tabla B. Variables para el Capital no Humano**

<i>Variable</i>	<i>Definición</i>
$x_1$	=1 si el agente entrevistado recibe canales de televisión a través de antena parabólica; =0 en otro caso
$x_2$	=1 si el agente entrevistado recibe canales de televisión a través de cable; =0 en otro caso
$x_3$	=1 si el agente entrevistado posee un ordenador personal de sobremesa; =0 en otro caso
$x_4$	=1 si el agente entrevistado posee un ordenador portátil; =0 en otro caso
$x_5$	=1 si el agente entrevistado posee un número de teléfono fijo; =0 en otro caso
$x_6$	=1 si el agente entrevistado posee un teléfono móvil; =0 en otro caso
$x_7$	=1 si el agente entrevistado tiene contratada una cadena musical de alta fidelidad; =0 en otro caso
$x_8$	=1 si el agente entrevistado posee una vídeo casetera; =0 en otro caso
$x_9$	=1 si el agente entrevistado posee un dvd; =0 en otro caso

Así, y siguiendo la misma aproximación, la medida del capital no humano será:

$$NHK_i = \frac{\sum_{h=1}^9 x_h}{9}$$

$$h = 1, 2, \dots, 9$$

Las ponderaciones utilizadas pueden variar de acuerdo a distintos puntos de vista. Desde lo planteado en este apéndice se prefiere dejar todas las magnitudes consideradas con el mismo peso, aunque si bien se le da mayor importancia al capital humano en conjunto. Esto es así debido a que el capital humano se puede considerar como más *permanente* que el capital no humano debido a que este último está sujeto a depreciaciones mientras que los años de estudio acabados quedan con el individuo para

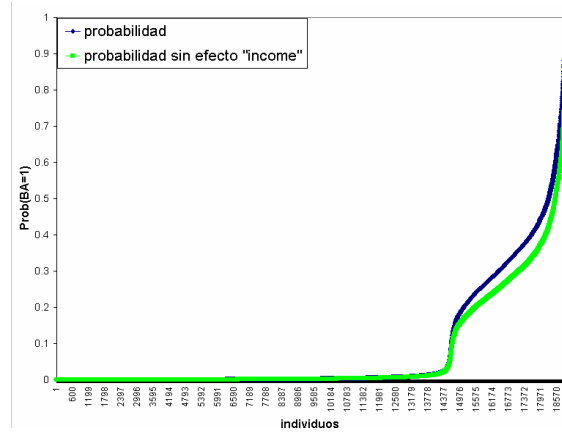
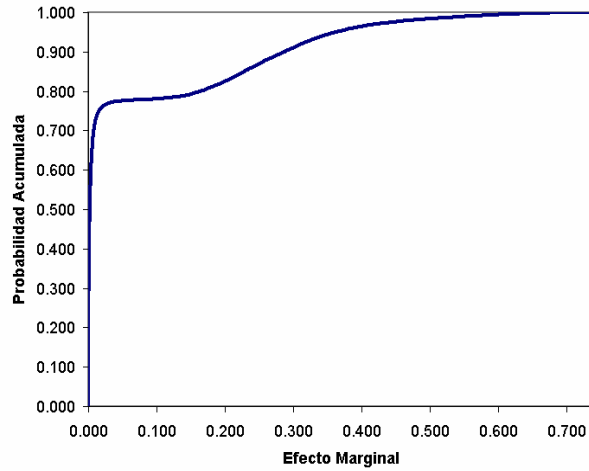
siempre y tienen una relación directa con sus ingresos obtenidos en el mercado laboral. A su vez, las variables consideradas aquí para medir el capital no humano están asociadas a bienes de consumo tecnológico que se supone tienen una tasa de depreciación muy elevada.

## **APÉNDICE 2: Efectos Marginales de los Determinantes de la Probabilidad de Acceso a Internet por Banda Ancha desde el Hogar<sup>7</sup>**

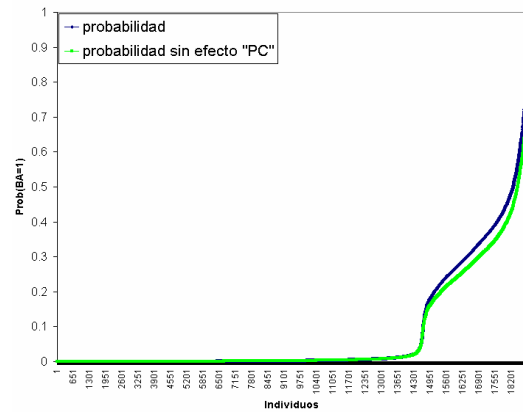
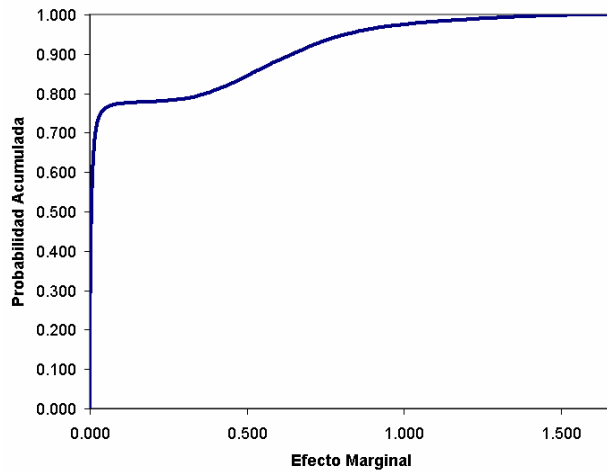
A diferencia de lo que sucede en los modelos de regresión lineal, los efectos marginales de los modelos no lineales no serán necesariamente los parámetros del modelo. En el caso de los modelos logit, probit o cualquier otro modelo de regresión no lineal, los efectos marginales variarán con los valores del vector  $X$  de variables explicativas. Para interpretar el modelo estimado resulta útil calcular los efectos marginales para varios valores de  $X$ , y para representarlos se pueden utilizar dos maneras:

- a. Observando la variación de la probabilidad acumulada a medida que el efecto marginal varía (gráficos de la izquierda). En este caso se pueden evaluar las distintas intensidades en la probabilidad acumulada a medida que el efecto marginal va aumentando.
- b. Utilizando los coeficientes estimados para obtener las probabilidades en función de todas las variables explicativas fijando primero su valor en la media, y luego el mismo cálculo pero con la variable estudiada igual a cero (gráficos de la derecha). Aquí el efecto marginal se puede ver como la diferencia entre estas dos funciones.

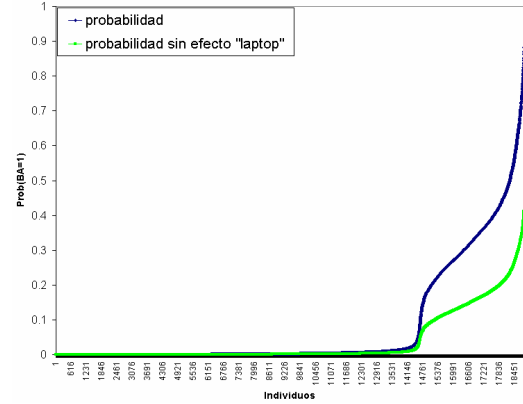
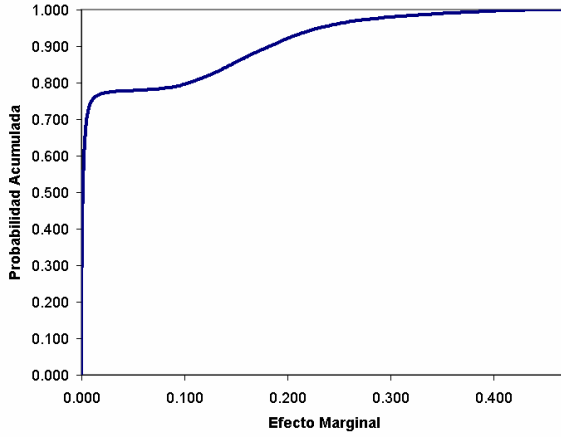
A continuación observamos los efectos marginales del modelo para el acceso a Internet.



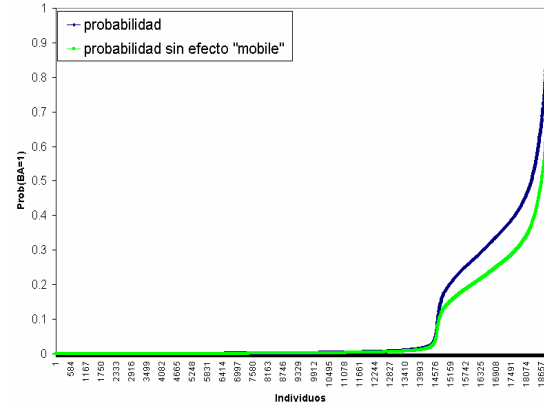
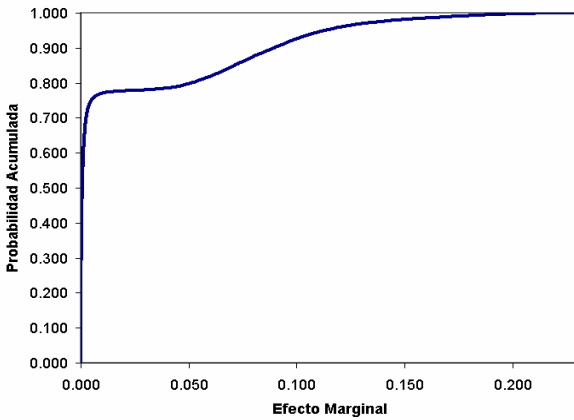
### Efecto Marginal de la Variable *income*



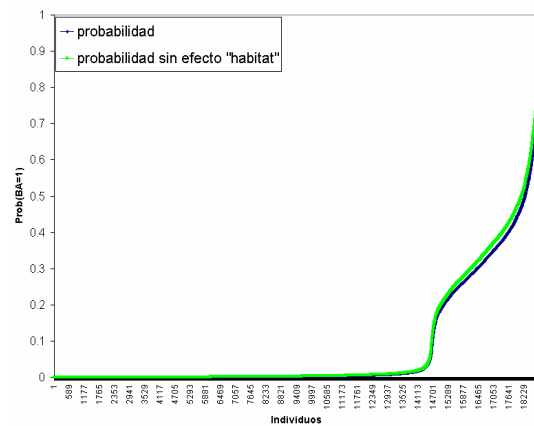
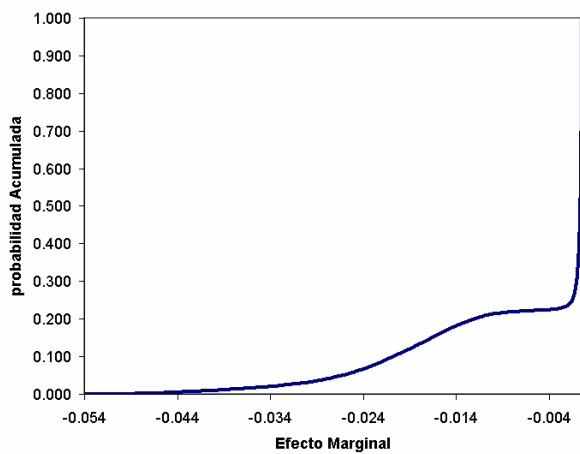
### Efecto Marginal de la Variable *PC*



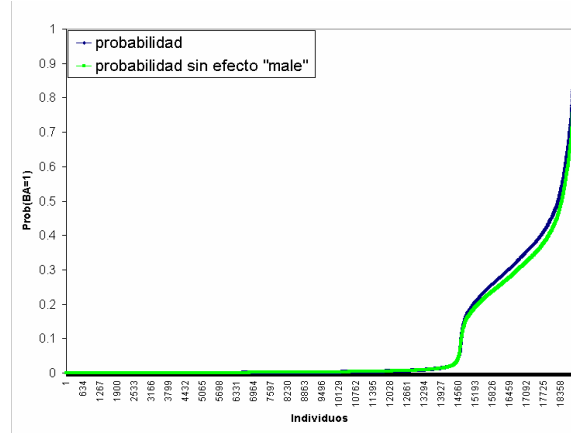
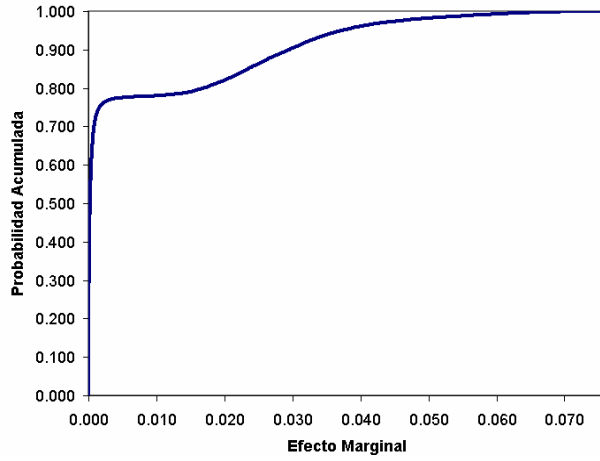
Efecto Marginal de la Variable *laptop*



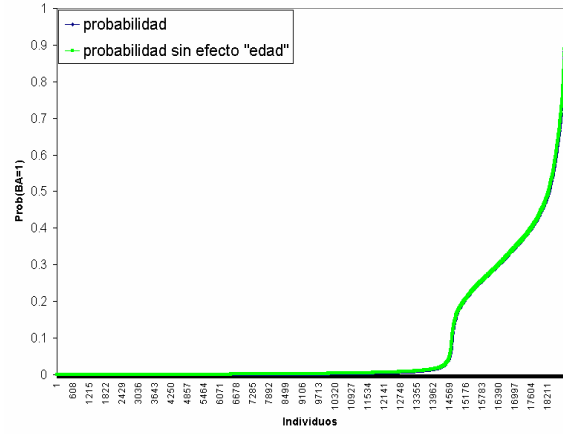
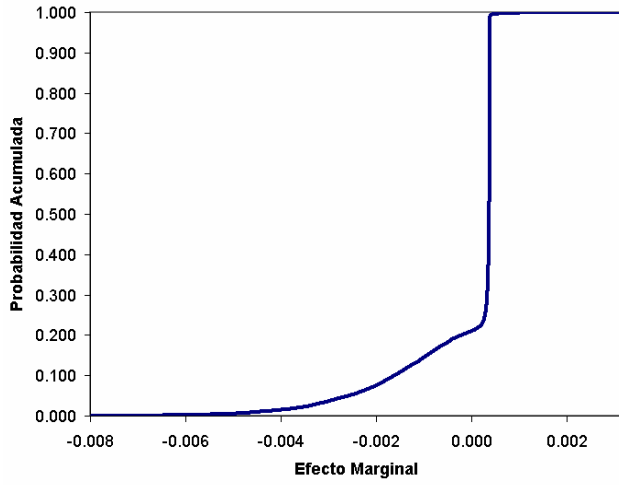
Efecto Marginal de la Variable *mobile*



Efecto Marginal de la Variable *habitat*



**Efecto Marginal de la Variable *male***



**Efecto Marginal de la Variable *age***

### APÉNDICE 3: Tablas Estadísticas

**Tabla A: Hogares con Acceso a Internet en el Hogar, por Tipo de Acceso, Hábitat y Tamaño del Hogar**

	HOGARES CON ACCESO A INTERNET	Línea telefónica convencional	Banda Ancha (ADSL, RDSI, Red Cable)	Otras formas de conexión	NS / NC
Total viviendas	3599054	2683193	1275790	70818	63968
Más de 100.000 habitantes y capitales de provincia	1971135	1432824	791896	35229	30932
De 50.000 a 100.000 habitantes	318494	215815	124037	4620	5929
De 20.000 a 50.000 habitantes	449671	346816	139323	11020	6459
De 10.000 a 20.000 habitantes	344367	272629	104080	3460	8387
Menos de 10.000 habitantes	515388	415109	116454	16488	12262
Hogares de 1 miembro	170929	114978	63340	6161	1917
Hogares de 2 miembros	550044	419119	194215	5775	3118
Hogares de 3 miembros	1017838	794450	321855	24802	23179
Hogares de 4 miembros	1407514	1039018	511005	18903	31104
Hogares de 5 o más miembros	452729	315628	185375	15177	4651

NS / NC = no sabe – no contesta.

Fuente: INE (2003)

**Tabla B: Uso de Internet en los Últimos 3 Meses por Lugar de Uso, Sexo y Edad**

	HAN UTILIZADO INTERNET EN LOS ÚLTIMOS 3 MESES	Desde la Vivienda	Desde el centro de trabajo	Desde el centro de estudios	Desde otros lugares	NS / NR
Total Personas	12130100	7236572	5013926	2471758	3548983	71242
Sexo: Hombre	6533251	4067344	2831831	1155415	1952407	46466
Sexo: Mujer	5596850	3169228	2182094	1316343	1596576	24776
Edad: De 15 a 24 años	3882397	2196185	595930	1794897	1733285	24852
Edad: De 25 a 34 años	3726407	2077879	1800401	454561	1150003	26479
Edad: De 35 a 44 años	2490057	1557464	1513678	126683	424040	9401
Edad: De 45 a 54 años	1359696	951357	778809	53462	159011	2531
Edad: De 55 a 64 años	505463	327448	287312	20690	53581	2620
Edad: De 65 a 74 años	123142	95335	25305	12078	22295	149
Edad: De 75 y más años	42939	30904	12491	9387	6769	5211

NS / NC = no sabe – no contesta.

Fuente: INE (2003)

**Tabla C: Uso de Internet en los Últimos 3 Meses por Lugar de Uso y Nivel de Estudios Terminados**

	HAN UTILIZADO INTERNET EN LOS ÚLTIMOS 3 MESES	Desde la Vivienda	Desde el centro de trabajo	Desde el centro de estudios	Desde otros lugares	NS / NC
Sin Nivel de Estudios Terminados	5076	5076	5076	0	5076	0
Educación Primaria	643677	320428	122600	178006	255502	7546
Primera etapa de Educación Secundaria	2347652	1218957	386531	538936	1011951	25408
Segunda etapa de Educación Secundaria	3681069	2292043	1213735	1026828	980574	13540
Formación Profesional de Grado Superior	1281563	712290	639317	167675	417186	4942
Educación Superior Universitaria	4114286	2656406	2628970	551727	855207	19806
Estudios terminados: Otros	56779	31372	17697	8585	23488	0

NS / NC = no sabe – no contesta.

Fuente: INE (2003)

**Tabla C: Uso de Internet en los Últimos 3 Meses por Lugar de Uso y Categoría Laboral y Situación Profesional**

	HAN UTILIZADO INTERNET EN LOS ÚLTIMOS 3 MESES	Desde la Vivienda	Desde el centro de trabajo	Desde el centro de estudios	Desde otros lugares	NS / NC
Activos ocupados	7740944	4635858	4569428	665472	1862498	31995
Activos parados	736229	388399	84908	147775	347333	12601
Inactivos: Estudiantes	2570675	1477916	108935	1561948	1039710	14430
Inactivos: Labores del hogar	448341	290098	23528	49167	128233	7004
Inactivos: Pensionistas	210348	155502	4538	15508	56962	5211
Otra situación laboral	420897	286130	219920	31888	111580	0
Situación Profesional (en trabajo principal): Trabajador por cuenta ajena	6473552	3779741	3865177	549182	1614280	24184
Situación Profesional (en trabajo principal): Trabajador por cuenta propia	1267391	856117	704252	116290	248218	7811

NS / NC = no sabe – no contesta.

Fuente: INE (2003)

## CAPÍTULO III

### Brecha Tecnológica en España

## 1. INTRODUCCIÓN

*Brecha tecnológica o digital* es el acceso desigual de las personas a la tecnología digital. Esta desigualdad puede tener distintas dimensiones: geográfica, económica, étnica, por edades, etc. Puede ser medida tanto a nivel agregado (continentes, países, etc.) como a nivel desagregado (regiones, sectores, consumidores, etc.). Actualmente la discusión de la existencia de una brecha tecnológica entre países o entre regiones de cada país se lleva a cabo tanto a nivel descriptivo como a nivel cuantitativo. Preguntas acerca de los motivos de dichas diferencias vienen siendo formuladas hace relativamente poco tiempo, y en cuanto a su medición la mayoría de las contribuciones no pasan de ser meramente descriptivas, sin adentrarse en la determinación y medición de sus determinantes.

A tales enfoques responderían estudios como los de Randstad, Red.es, Fundación AUNA, Banco Mundial y OCDE entre otros. También existen una serie de contribuciones previas a la elaboración de ciertas medidas (Corrocher y Ordanini, 2002; Martin, 1995; Mansell y When, 1998; Ricci, 2000). Estos trabajos intentan cuantificar la difusión del fenómeno digital utilizando metodologías cuantitativas.

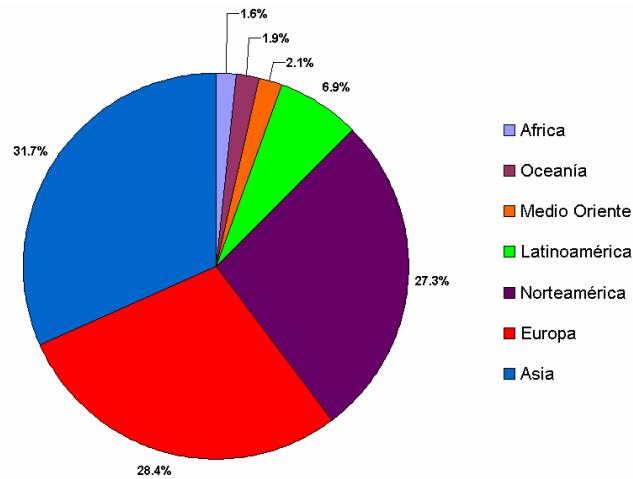
Generalmente el uso de Internet se considera un buen indicador del uso del resto de la tecnología digital. Informes recientes<sup>1</sup> revelan que existen actualmente cerca de 1.000 millones de usuarios de Internet en todo el mundo. En el Gráfico 1 siguiente se observa cómo están distribuidos dichos usuarios, donde vemos que Asia concentra el 31,7% de la población on-line mundial, con más de 257 millones de usuarios. Le siguen Europa, con 230 millones de usuarios (28,4%) y Norteamérica con más de 222 millones de usuarios (27,3%). Luego, muy por debajo en cantidad le siguen Latinoamérica con casi

---

<sup>1</sup> Disponibles en <http://n-economia.com/frindex.html>

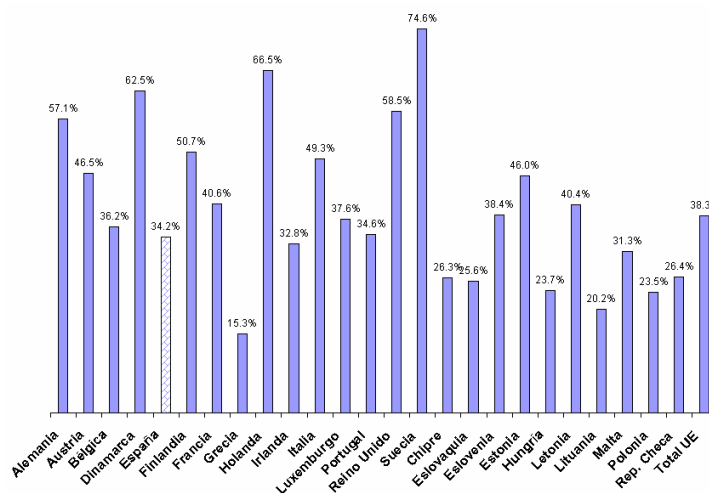
56 millones de usuarios (6,9%), Medio Oriente con 17 millones (2,1%), Oceanía<sup>2</sup> con más de 15 millones (1,9%) y por último África con el menor número de usuarios (1,6%) equivalente a casi 13 millones de personas.

**Gráfico 1:** Distribución Mundial del Total de Usuarios de Internet (año 2004)



Centrándonos en el continente europeo podemos apreciar que la distribución de los cerca de 250 millones de usuarios es bastante desigual. Existen países sumamente avanzados en cuanto al uso de Internet y otras TIC y otros que apenas están surgiendo, entre los que se encontraría España. En el Gráfico 2 vemos que países como Suecia u Holanda están a la cabeza de la UE con un 74,6% y 66,5% de población usuaria de Internet. España se encuentra entre los países con menor uso (sólo el 34,2% de la población en 2004 declara utilizar Internet), aunque es un porcentaje superior al de países como Grecia (15,3%), u otros recién ingresados en la Unión (Hungría 23,7%, Polonia 23,5% o Lituania 20,2%).

<sup>2</sup> En términos relativos este continente estaría a la altura de Asia, Europa o Norteamérica si consideramos que entre Australia y Nueva Zelanda en conjunto no llegan a 26 millones de habitantes.

**Gráfico 2:** Distribución de Usuarios de Internet en la UE

España es un país donde existe una diferencia considerable en cuanto al uso de las nuevas tecnologías (TIC). En un informe de Red.es, de diciembre de 2004, en donde se comparan datos de uso de TIC entre regiones del país y con el resto de Europa, se destacan datos tales como que sólo el 23,5% de las empresas españolas utiliza el comercio electrónico, que no todas las empresas acceden a Internet (el 76,1% tiene acceso a Internet por Banda Ancha), o que la aportación al PIB del mercado de las TIC en España no supera el 1,68% en 2004. Sin embargo dicha brecha tecnológica no se ve tan claramente al realizar la comparación por regiones. Cataluña, Madrid y el País Vasco tienen niveles de convergencia tecnológica con las regiones más desarrolladas de Europa (más del 89% de sus empresas compran y venden a través de la Red), mientras que en Galicia y Extremadura no se alcanzaba el 21 % de sus empresas. En cuanto al acceso a Internet desde los hogares, las diferencias entre regiones también son bastante evidentes como ya se lo observa en el capítulo anterior en el Gráfico 1. Si bien en los hogares españoles se experimentó una tasa de crecimiento medio anual de 36,4% durante el período 2000-2004 (Red.es, 2004), en 2003 el 25,2% de los hogares ya

cuenta con acceso a Internet (Cerno y Pérez-Amaral, 2005), dicho crecimiento fue dándose de manera desigual<sup>3</sup>.

A continuación se amplía el análisis al resto de las tecnologías digitales. Esto pondrá en evidencia las diferencias existentes a nivel geográfico. A su vez, revela la necesidad de estudiar dichas diferencias, cuantificarlas con un índice sintético y medir el efecto de los posibles determinantes.

El presente Capítulo trata el problema de la brecha digital en España y analiza sus determinantes. Comenzamos en la siguiente sección con la elaboración de un índice de brecha tecnológica para España desde dos puntos de vista: primero a partir del consumo en tecnología y luego desde el punto de vista del grado de digitalización. El primero (índice de consumo de tecnología) es para utilizarlo en un modelo de regresión para el uso de Internet por CCAA. El segundo (nivel de digitalización) para utilizarlo en otro modelo esta vez para explicar sus determinantes. En las secciones 3 y 4 se utilizan modelos econométricos de comportamiento para medir el efecto de los factores que influyen en el nivel de digitalización, a nivel individual para cada CA de España. Por las características que presenta la variable endógena, se utiliza primero un modelo de regresión de Poisson para estimar las elasticidades que nos ayudarán a separar la muestra en grupos homogéneos para tratar la heterogeneidad y obtener otro modelo para el grado de digitalización que realice la mejor predicción por fuera de la muestra utilizando el algoritmo RETINA (Relevant Transformations of the Inputs Network Approach) propuesto recientemente por Pérez-Amaral, Gallo y White (2003).

---

<sup>3</sup> En el último informe de Red.es (XI Oleada, julio de 2006), se observa una tendencia creciente en todos los puntos mencionados aquí. Por ejemplo, actualmente el porcentaje de hogares con conexión permanente a Internet es de 33,9%; el 89,95% de las empresas se conectan a Internet y el 48,46% posee sitio Web propio. A su vez, el 48,9% de los mayores de 15 años han accedido a Internet alguna vez, y el 31,6% lo hizo en la última semana.

## 2. ÍNDICE SINTÉTICO DE DIGITALIZACIÓN GEOGRÁFICA PARA ESPAÑA

Hablar de *digitalización* se refiere a la evolución que permite que los entes (individuos o empresas) estén conectados y sus sistemas de información se entiendan y puedan interactuar entre sí<sup>4</sup>.

Al iniciar la presente investigación se ha encontrado que la medición de la brecha tecnológica es poco homogénea dentro de la literatura. Las contribuciones realizadas hasta ahora en su mayoría usan diferentes medidas. En este capítulo se planteará cómo medir el grado de digitalización en las zonas geográficas españolas y cómo dicha medición puede ser utilizada como herramienta para orientar políticas.

Existen algunos puntos importantes a tener en cuenta relacionados con la agregación de elementos y con las ponderaciones para construir un índice sintético. El proceso podría seguir el criterio de ponderar de manera objetiva o basarse en ponderaciones subjetivas de acuerdo a criterios de importancia considerados de manera predeterminada. Para identificar los niveles de agregación y la provisión de las ponderaciones tenemos que pensar que la brecha tecnológica se está refiriendo a una *diferencia* entre las distintas unidades (en nuestro caso Comunidades Autónomas de España), entonces la medición a realizar deberá reflejar esta diferencia. Esto nos lleva al planteamiento de qué elementos se considerarán y qué ponderaciones se incluirán. El detalle de la construcción del índice se encuentra a continuación.

---

<sup>4</sup> [www.dmconsulting.es/digitalizacion.htm](http://www.dmconsulting.es/digitalizacion.htm)

## 2.1. UN INDICE DE DIGITALIZACIÓN POR COMUNIDAD AUTÓNOMA

Siguiendo la idea planteada en Corrocher y Ordanini (2002) donde se desarrolla una medida para el nivel de digitalización dentro de un conjunto de países a través de indicadores agrupados por un factor de digitalización, la dispersión del valor final calculada para cada zona geográfica constituye una medida de brecha tecnológica dentro del sistema. En este apartado se pretende elaborar una medida que además sirva para determinar las diferencias por zona geográfica (Comunidad Autónoma en este caso), y a su vez pueda ser utilizada al nivel de desagregación tal que alcance al mismo individuo.

Como punto de partida se consideran todas las observaciones de la muestra (los 18.948 individuos entrevistados en 2003 en toda España<sup>5</sup>) y la siguiente definición:

$$I_{TIC_i} = \left\{ \sum_{j=1}^n \left[ q_{ij} D_{ij} \prod_{j=1}^k (1 + q_{ijk} D_{ijk}) \right] \right\} \quad (3.1)$$

$i=1, \dots, 18.948$

en donde:

- $I_{TIC_i}$  : Índice TIC que cuantifica el grado de digitalización para cada individuo de la muestra.
- $D_{ij}$  : Variable ficticia intergrupar (=1 si el agente  $i$  posee algún elemento entre los distintos conjuntos de tecnología  $j$ ).
- $q_{ij}$  : Peso de importancia del conjunto de tecnología  $j$ .

<sup>5</sup> Para mayor detalle, véase el Apéndice Estadístico del Capítulo II.

- $D_{ijk}$  : Variable ficticia intragrupal (=1 si el agente  $i$  dentro del conjunto de tecnología  $j$  posee algún elemento  $k$ ).
- $q_{ij}$  : Peso de importancia del elemento dentro del conjunto  $j$ .

Los grupos de elementos que se han considerado son los siguientes planteados en la Tabla 1:

**Tabla 1:** Equipamientos de Tecnología Considerados para cada Individuo

Grupos $j$	Subgrupos $k$
<i>Televisión</i>	Antena convencional Antena parabólica Cable (incluye satélite) Televisión digital Terrestre
<i>Ordenador</i>	Ordenador de sobremesa Portátil Otro tipo(incluye agenda electrónica)
<i>Telefonía</i>	Teléfono fijo Teléfono móvil
<i>Internet</i>	Conexión por banda ancha
<i>Otros Equipamientos</i>	Cadena musical Radio Vídeo (incluye videoconsola) DVD Fax Automóvil con sistema de navegación

### 2.1.1. PONDERACIONES

En una segunda parte y una vez que ya se identificaron los factores de digitalización a tener en cuenta, la tarea es decidir qué grado de importancia se le dará a cada uno dentro del índice sintético. A partir de esta cuestión se presenta una primera posibilidad: la utilización de un peso *objetivo*, es decir una ponderación dentro de cada subgrupo, y

de cada grupo dentro del total. En este caso se estaría haciendo referencia al *consumo* que se supone realiza el individuo que está orientado a las nuevas tecnologías. Dentro de esta primera posibilidad, las ponderaciones se realizan entonces de acuerdo a la proporción de cada uno de los elementos de los subgrupos  $k$  dentro del total de cada grupo  $j$ .

Recordando lo definido más arriba, que  $D_{ij}$  es una variable ficticia intergrupal (=1 si el agente  $i$  posee algún elemento entre los distintos conjuntos de tecnología  $j$ ),  $q_{ij}$  es el peso de importancia del conjunto de tecnología  $j$  y  $D_{ijk}$  es una variable ficticia intragrupal (=1 si el agente  $i$  dentro del conjunto de tecnología  $j$  posee algún elemento  $k$ ), entonces queda:

$$q_{ijk} = \frac{\text{cantidad de elementos } k}{\text{cantidad total de elementos en } j} = \frac{n_{\gg k}}{n_{> j}} \quad q_{ij} = \frac{\text{cantidad de elementos } j}{\text{cantidad total de elementos}} = \frac{n_{> j}}{n_{\gg}}$$

$i = 1, 2, \dots, 18.948$  individuos.

$j = 1, 2, \dots, 5$  grupos de equipamientos.

$k = 1, 2, \dots, n_k$  subgrupos de equipamientos.

$$n_{\gg k} = \sum_{i=1}^{18.948} D_{ijk}$$

$$n_{> j} = n_{\gg k_1} + n_{\gg k_2} + \dots = \sum_{j=1}^5 \sum_{i=1}^{18.948} D_{ijk} = \sum_{i=1}^{18.948} D_{ij}$$

$$n_{\gg} = n_{> j_1} + n_{> j_2} + \dots = n_{\gg k_1} + n_{\gg k_2} + \dots = \sum_{k=1}^{16} \sum_{j=1}^5 \sum_{i=1}^{18.948} D_{ijk}$$

Utilizando estas ponderaciones planteadas aquí, lo que se obtendrá es un índice del consumo individual en tecnología digital, que correspondería a la parte del poder adquisitivo dedicado a adquirir y mantener los elementos que el individuo dice poseer.

En la segunda posibilidad se varían los pesos de importancia anteriormente considerados. Por ejemplo, si se consideran pesos objetivos, el tener televisión o línea de teléfono fijo en el hogar será más importante que tener conexión de Banda Ancha, debido a que es más probable que el individuo posea más de lo primero que de lo segundo<sup>6</sup>. A partir de esta idea se han rediseñado las ponderaciones de acuerdo a los siguientes criterios expuestos en la siguiente la Tabla 2:

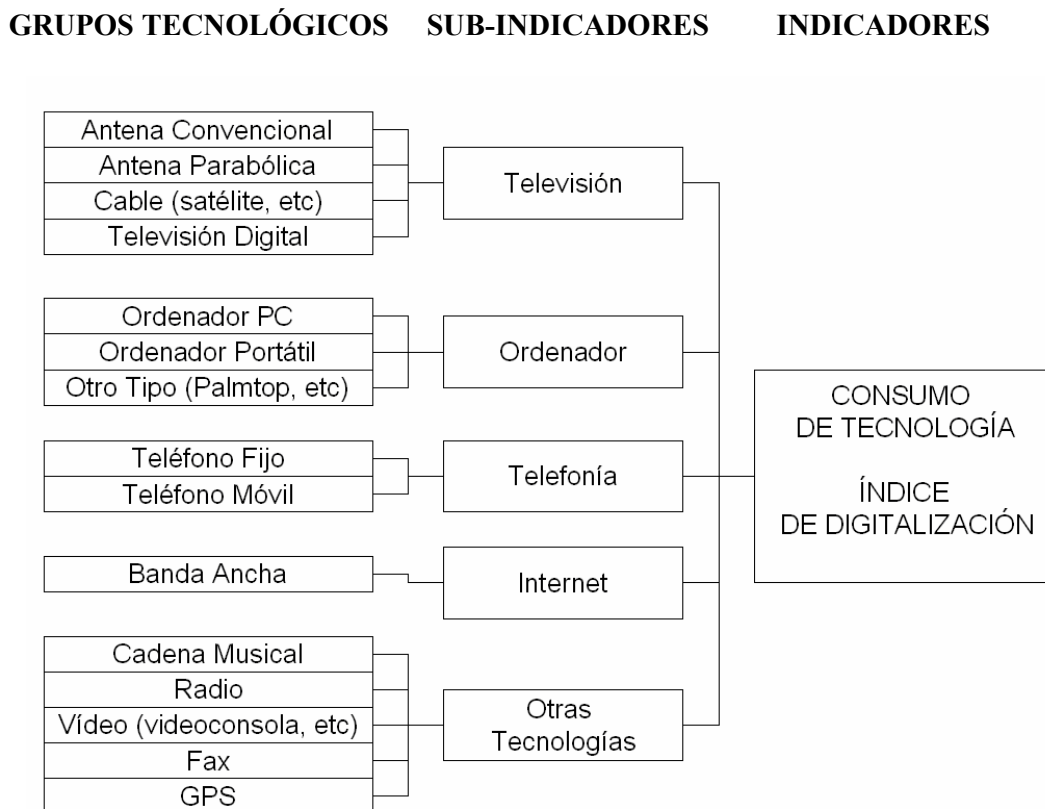
**Tabla 2:** Nuevos Pesos de cada Equipamiento en el Índice

Grupos <i>j</i>	PESOS del GRUPO	Subgrupos <i>k</i>	PESOS del SUBGRUPO
<i>Televisión</i>	15%	Antena convencional Antena parabólica Cable Televisión digital	30% 15% 20% 35%
<i>Ordenador</i>	15%	PC Portátil Otros tipos	35% 45% 20%
<i>Telefonía</i>	15%	Teléfono fijo Teléfono móvil	50% 50%
<i>Internet</i>	35%	banda ancha	100%
<i>Otros Equipamientos</i>	10%	Cadena musical Radio Vídeo DVD Fax Automóvil con GPS	20% 15% 15% 15% 15% 20%

<sup>6</sup> Es posible que casi todos tengan televisión y teléfono fijo en casa o un móvil, aunque la conexión a Internet en el hogar, según hemos observado en el capítulo anterior, no llega al 30% de los hogares españoles. Es decir que tener televisión y/o teléfono fijo *pesará más* que tener Internet porque el agente *suele consumir más en media* en estas TIC.

El índice ITIC da pesos objetivos a las alternativas dentro de cada grupo. Al variarlas a partir del criterio de darle mayor peso a determinado tipo de tecnologías es posible obtener un *Índice Sintético del Grado de Digitalización* para cada individuo, que luego se puede agregar para cada Comunidad Autónoma. Esto nos permitirá evaluar la magnitud de la brecha tecnológica a nivel de Comunidad Autónoma (CCAA), o por zonas geográficas según se determinen. En el siguiente apartado se construyen y se comparan los índices de digitalización por Comunidad Autónoma. A partir de estos valores se determina la brecha tecnológica para España. En el siguiente esquema se resume la idea planteada en este apartado:

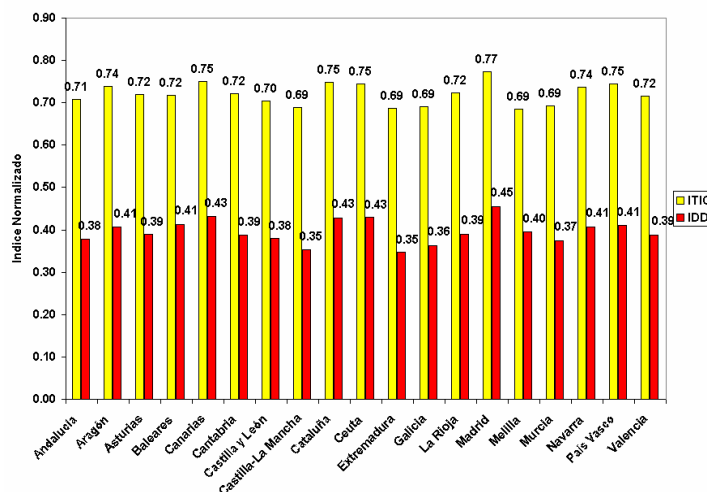
**Gráfico 3:** Esquema de Elaboración de los Índices



## 2.2. COMPARACIÓN DE ÍNDICES Y PREDICCIÓN DE LA BRECHA TECNOLÓGICA POR CCAA

A continuación se presentan los valores de ambos índices por CCAA. Los puntos de vista planteados en los pesos de importancia se refieren al consumo en equipamiento y tecnología (ITIC), y al nivel de digitalización (IDD). Este último aproxima la predisposición del individuo hacia las nuevas tecnologías. Lo hace a partir de la asignación de pesos de importancia a los distintos grupos de tecnologías considerados en la construcción del índice dándole el mayor peso a Internet, tal como se detalla en el apartado anterior:

**Gráfico 4:** Índice Normalizado del Consumo de Tecnología (ITIC) y Nivel de Digitalización (IDD)



En una primera aproximación se puede apreciar la diferencia en el nivel medio de ambos índices. Esto será así por la diferencia existente en el peso del componente “Internet de banda ancha”. También se observa que hay diferencias entre los valores calculados para cada CA de cada índice, a pesar de que el rango de variación

aparentemente no es importante tanto en uno como en otro enfoque. Sabiendo que valor 1 es el valor máximo, se observa que Madrid es la CA con mayor consumo de tecnología y con mayor nivel de digitalización, aunque en este último caso con salvedades, ya que vemos que ninguna CCAA llega siquiera al 50% del nivel considerado. Esto estaría relacionado con lo que se ve más arriba al comparar España con otras naciones europeas donde se aprecian diferencias importantes en cuanto al uso de Internet, teniendo en cuenta de que es el factor más importante en la construcción del índice IDD.

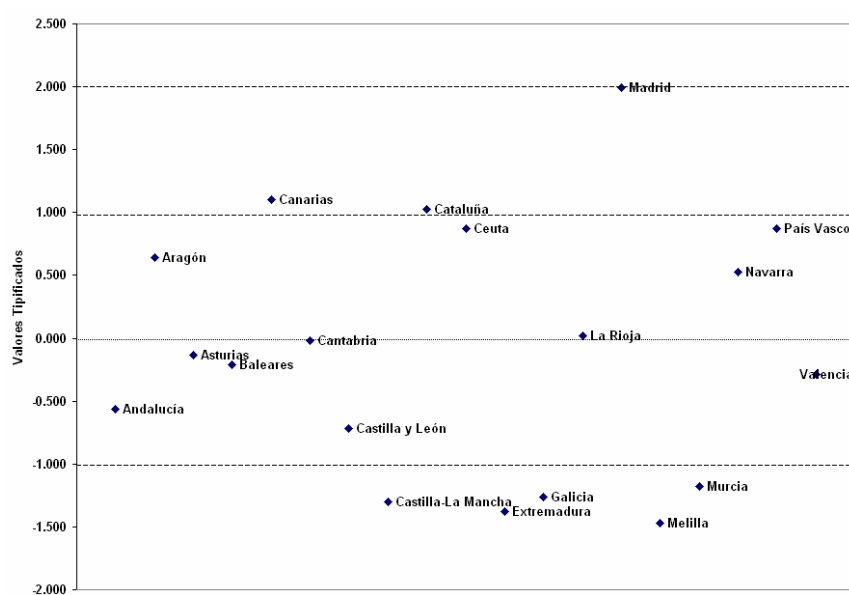
Centrando el análisis en las CCAA, vemos que Melilla consume menos tecnología que Extremadura, pero sin embargo la supera bastante en el ranking de nivel de digitalización. Esto se debe a que Melilla (como Ceuta) es sólo una *ciudad* dentro de las CCAA, en donde obviamente el consumo en tecnología será mucho menor que el resto por tener una población mucho menor, pero el nivel de digitalización de sus habitantes estará muy por encima. Otras CCAA como Cataluña, País Vasco, Extremadura o Castilla-La Mancha se mantienen en niveles altos y bajos en nivel de consumo en tecnología y nivel de digitalización respectivamente.

El paso siguiente es medir la brecha digital entre CCAA, definiendo esta como la diferencia existente entre las mismas en cuanto a consumo de tecnología y nivel de digitalización. Después de calcular niveles de consumo y digitalización para cada CA, se utiliza la media del índice como valor de referencia, obtenida a partir de los valores normalizados observados anteriormente. A partir de ahí se mide la distancia que cada Comunidad tiene respecto de esa media y con respecto a otras. Esto tendrá las siguientes implicaciones:

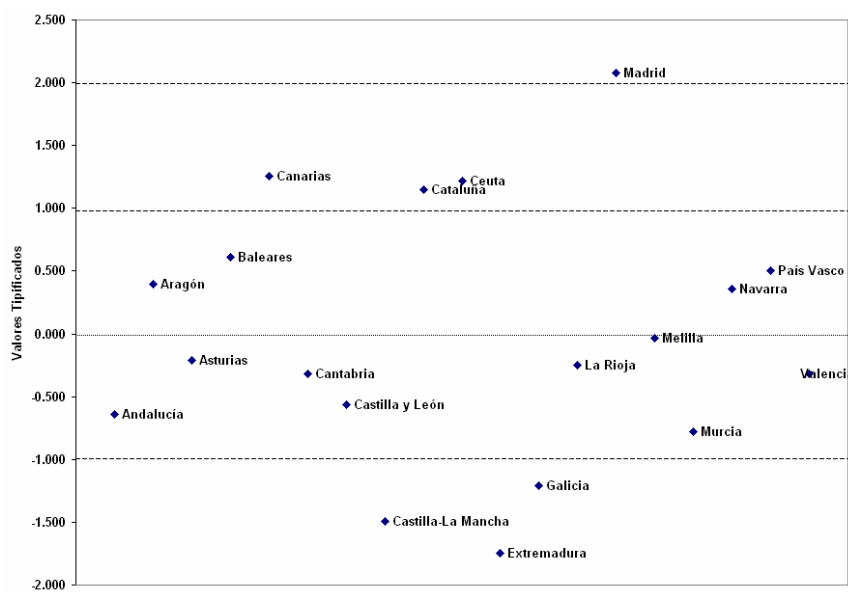
- Un valor de brecha tecnológica “total” (es decir con respecto a todos los individuos) puede ser calculada a partir de una medida de dispersión relativa tal como el coeficiente de variación (cociente entre la desviación típica y la media aritmética).
- La brecha tecnológica por CCAA puede ser medida a partir de las distancias de su valor normalizado del índice con el resto de los valores o con la media total. Esta distancia representa la “similitud” o “diferencia” para cada CA, ya sea en cuanto al consumo como también a nivel de digitalización.
- Por el tipo de muestra que se utiliza en el análisis no se planteará en esta Tesis, pero la evolución de la brecha tecnológica se puede medir a través del tiempo, tanto a nivel de CCAA como agregado.

A continuación se presentan los gráficos 4 y 5 en donde se observan las diferencias existentes entre la media del índice y los valores por CCAA (valores tipificados):

**Gráfico 5:** Comparación entre CCAA del Consumo de Tecnología



**Gráfico 6:** Comparación entre CCAA del Nivel de Digitalización



En la siguiente tabla están representados los valores de las diferencias observadas en ambos gráficos:

**Tabla 3.** Diferencia entre Valores del Índice con Respecto a la Media del Total

	<b>ITIC</b>	<b>IDD</b>
Media del Índice	<b>0,721</b>	<b>0,396</b>
Desviación Típica	<b>0,036</b>	<b>0,071</b>
CCAA	Var. ITIC	Var. IDD
Andalucía	-1,7%	-4,52%
Aragón	2,6%	2,81%
Asturias	-0,2%	-1,49%
Baleares	-0,5%	4,32%
Canarias	4,2%	8,87%
Cantabria	0,2%	-2,25%
Castilla-León	-2,3%	-4,01%
Castilla-La Mancha	-4,4%	-10,58%
Cataluña	4,0%	8,11%
Ceuta	3,4%	8,61%
Extremadura	-4,7%	-12,35%
Galicia	-4,2%	-8,56%
La Rioja	0,3%	-1,74%
Madrid	7,4%	14,68%
Melilla	-4,9%	-0,23%
Murcia	-4,0%	-5,53%

Navarra	2,1%	2,55%
País Vasco	3,4%	3,56%
Valencia	-0,8%	-2,25%
Total Porcentajes (valores absolutos)	<b>55,31%</b>	<b>107,02%</b>
Media de Porcentajes (valores absolutos)	1,94%	4,04%

De lo visto se puede decir:

- La media del índice de consumo en equipamiento ITIC (0,721) tiene un valor mayor que la del nivel de digitalización (0,396). Esta discrepancia se da en parte por las diferencias existentes en las ponderaciones de los dos índices. En el caso de IDD por ejemplo, al dársele un mayor peso a Internet de banda ancha y otros componentes más avanzados, los valores del indicador tienden a disminuir. Esto evidencia la diferencia existente entre el consumo de cualquier tipo de tecnología, y la de aquellas más avanzadas.
- Tal como se comenta más arriba, considerando como medida de brecha tecnológica como la dispersión entre los valores de los índices, si la misma se mide en cuanto al consumo destinado a equipamiento (ITIC) será menor que la existente en cuanto a nivel de digitalización y afinidad con las nuevas tecnologías con IDD (0,036 es menor que 0,071). Esto también se puede observar en la suma total de los porcentajes de variación con respecto a las medias (55,31% y 107,02%). Esto se debe a que un incremento en el nivel de renta per capita como el que experimentaron las regiones de España después de su ingreso en la UE se tradujo, como es lógico, en un mayor consumo en equipamiento. Sin embargo el nivel de digitalización tendrá otros determinantes además de la renta y tal como se estudia en las secciones posteriores de este capítulo.

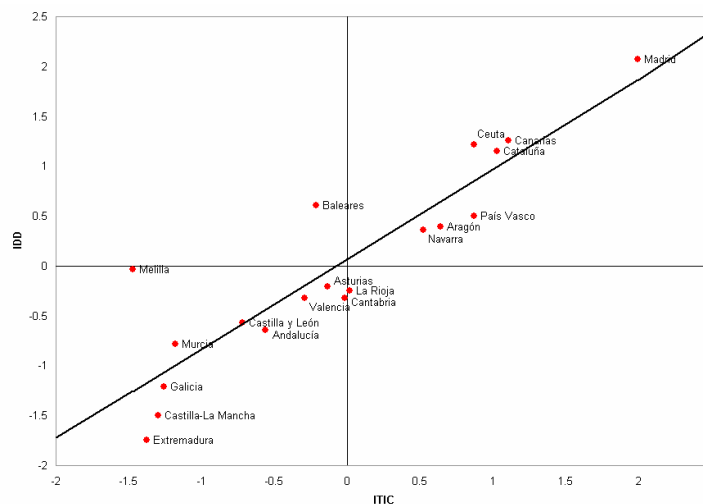
- Madrid es la CA que consume más equipamiento y que mayor nivel de digitalización tiene respecto al resto. La diferencia existente entre Madrid y el valor medio supera las dos desviaciones típicas positivas en ambos casos según podemos observar en los gráficos 4 y 5. La diferencia respecto a la media en ITIC es positiva de 7,4%. Respecto a IDD es también positiva y de casi el doble (14,68%). Es decir que Madrid es la CA más favorecida según ambos índices.
- Luego de Madrid, las CCAA más favorecidas en cuanto a renta destinada a equipamiento son Canarias y Cataluña, con diferencias positivas respecto a la media total de 4,2% y 4,0% respectivamente. Le siguen de cerca el País Vasco y Ceuta, ambos con 3,4%.
- Los valores de las diferencias en cuanto a nivel de digitalización serán también positivos de 8,87% para Canarias y de 8,11% para Cataluña. Para Ceuta la diferencia es de 8,61%. El País Vasco está más cercano a la media con una diferencia positiva de sólo el 3,56%.
- Como contrapartida, las CCAA más desfavorecidas en cuanto a consumo de equipamiento y nivel de digitalización son Castilla-La Mancha, Galicia y Extremadura. Las diferencias tanto para el consumo en equipamiento como en el nivel de digitalización serán por debajo de la media en -4,4% y -10,58% para Castilla-La Mancha, -4,7% y -12,35% para Extremadura y en -4,2% y -8,56% para Galicia.
- En cuanto al consumo de equipamiento del resto de las CCAA, se encuentran dentro de la banda de más-menos una desviación típica, algunas muy cercanas a la media (Asturias, Baleares, Cantabria y La Rioja), otras más o menos favorecidas (dentro de las que se encuentran por encima de la media las que más

destacan son Aragón y Navarra, y de las por debajo a Andalucía y Castilla y León).

- El cuanto al nivel de digitalización del resto de las CCAA, la dispersión es mayor. Exceptuando a Melilla con un valor muy cercano a la media, las Comunidades se sitúan claramente por encima o por debajo de la media tal como observamos en el Gráfico 5.

A continuación se comparan en el Gráfico 6 ambos índices, en donde se observa que no todas las CCAA tienen ambos valores igualmente altos o bajos. Por ejemplo aquellas comunidades que están por encima de la línea de tendencia son las relativamente más favorecidas en cuanto a nivel de digitalización, mientras que las que se encuentran por debajo serán aquellas comunidades con gasto en tecnología relativamente mayor que el nivel de digitalización alcanzado.

**Gráfico 6: Diferencias entre CCAA**



Las CCAA que se encuentran en el cuadrante superior derecho serán las más favorecidas en tecnología en general, mientras que las que se encuentran en el inferior

izquierdo serían las menos. Los casos “atípicos” de Melilla y Baleares indican que el consumo en tecnología es especialmente bajo, mientras que en nivel de digitalización están en o por encima de la media.

También es posible realizar un análisis similar a partir de las matrices de distancias de ambos enfoques del índice entre las distintas CCAA. Para ello se podrían aplicar técnicas de análisis multivariante tales como el escalamiento multidimensional o los biplots, pero el objetivo ahora es explicar los determinantes que hacen que existan estas diferencias. Para ello, en la sección siguiente se calcularán los impactos de los determinantes del uso de Internet haciendo el análisis por CCAA.

### **3. LA BRECHA TECNOLÓGICA Y EL USO DE INTERNET**

Para modelizar la brecha tecnológica o tomar decisiones de política, es conveniente comenzar estudiando los motivos del uso de Internet. De los servicios considerados, Internet es el que posee mayor potencial para reducir las distancias entre el grado de digitalización de distintas regiones. Esto es así debido a que el acceso y uso de Internet traen aparejados el establecimiento de una red para que sea posible tal conexión, además del uso de distintas forma de conexión ya sea a través de ordenadores, teléfonos móviles, etc. A partir de aquí se comenzará por modelizar los determinantes del uso de Internet, separando la muestra por CCAA.

Actualmente en la literatura empírica se analizan los determinantes de la demanda de acceso y uso de Internet de forma relativamente extensa, por ejemplo Madden, Savage y Simpson (1996), Cassel (1999), Rappoport, Taylor y Kridel (2002) y Cerno y Pérez-Amaral (2005) entre otros, destacan las implicaciones del fenómeno.

Para elaborar un modelo empírico que mida los determinantes de la brecha tecnológica a nivel regional en España, se utiliza la base de datos ya comentada en el capítulo anterior sobre acceso y uso de Nuevas Tecnologías en España, con la salvedad de que para este capítulo dicha información se organiza por CCAA, 18 en total, donde tomamos a Ceuta y Melilla como una sola.

En el análisis se empleará la misma variable  $USE_i$  del capítulo anterior, en donde se la empleó en un modelo para la muestra completa. Esta vez el análisis será condicional por CA. La variable  $USE_i$  se refería a los sitios de uso de Internet y tomaba valores en el intervalo discreto  $[0,4]$ , en donde una particularidad que se observaba para toda la muestra era que lo que más abunda son los valores nulos:

**Tabla 4:** Estadísticos del Número de Lugares de Uso de Internet por los Individuos

Frecuencias Absolutas					
0	1	2	3	4	Total
13.705	3.093	1.776	329	45	18.948
Media	Mediana	Desviación Típica	Coefficiente de Variación	Simetría	Curtosis
0,412	0,000	0,750	181,9	1,821	2,716

A primera vista observamos que casi tres cuartos de la población se declara no usuaria de Internet (13.705 equivale al 72,3% de la muestra). El 27,7% restante sí utiliza el servicio dentro del intervalo de 1 a 4 lugares distintos.

Como se comenta más arriba, en el Capítulo II se modelizan los determinantes del uso de Internet para la muestra completa. Dichos determinantes eran de tipo económico, tecnológico y sociodemográfico. Las conclusiones obtenidas son coherentes con la literatura y con lo esperado para el caso de España. Ahora, la finalidad en este capítulo es obtener un modelo que explique las diferencias en el grado de digitalización existente actualmente en las CCAA españolas. Para ello, y tal como se comenta más

arriba, dentro de los determinantes de dichas diferencias se incluye entre otros el uso de Internet. Sin embargo, por el análisis del apartado anterior referido a la medición de la brecha digital, y por el tipo de datos utilizados, no se puede descartar que exista heterogeneidad entre las CCAA. De no tenerla en cuenta en el modelo, se podrían llegar a resultados contradictorios y estimaciones inconsistentes. A partir de esta idea entonces, se utilizará un modelo análogo al del Capítulo II para el uso de Internet, pero con una particularidad distinta: la de recoger la heterogeneidad. Para ello se hará el análisis del uso de Internet condicional por CA e incluyendo como variable explicativa el consumo en equipamiento ( $ITIC_i$ ). Para recoger la heterogeneidad se calcularán y se agruparán las elasticidades del uso de Internet con respecto a las variables explicativas más significativas dentro del modelo especificado por CCAA.

### 3.1. ESPECIFICACIÓN: USO DE INTERNET POR CCAA

Al igual que en el capítulo anterior, y siempre teniendo en cuenta que la variable  $USE_i$  no se distribuye normalmente por ser una variable de recuento (la naturaleza discreta viola el principio de continuidad de la distribución normal), además de tomar valores pequeños y poseer una abundancia de ceros, a continuación se especifica un modelo de Poisson para el uso de Internet a nivel de CCAA. La variable endógena será el uso de Internet. Las variables explicativas utilizadas serán de cuatro tipos:

Variable endógena: Es la cantidad de lugares en donde la persona se conecta. La variable es  $USE$  y varía entre 0 y 4.

### Variables explicativas

Consumo en tecnología: Su efecto está cuantificado por la variable *ITIC*, y es una medida *proxy* del poder adquisitivo del individuo dedicado a la tecnología. Su elaboración está explicada en detalle en el apartado anterior, donde fue propuesta como un índice para reflejar el grado de participación en el mercado de equipamiento digital a nivel individual y la posición en el mismo. En efecto, los valores de este indicador serán mayores cuando mayor sea el consumo de tecnología.

Demanda de servicios complementarios: Estas se refieren a tener acceso a Internet en la vivienda y al equipamiento utilizado para conectarse. El acceso a Internet en la vivienda está recogido por la variable ficticia *HOUSINT* y el equipamiento para conectarse está expresado en la variable *EQ*.

Principales variables sociodemográficas: Estas son las referidas al sexo (*MALE*), la edad (*AGE*), el hábitat (*HABITAT*) y el nivel de estudios (*STUDYLEVEL* y *BESTUDYING*).

Para mayor detalle ver la tabla de Definición de Variables del apéndice.

### **3.2. ESTIMACIÓN Y RESULTADOS**

El modelo para el uso de Internet individual para cada una de las CCAA que se plantea es aquel que sigue la siguiente relación:

$$USO_i = f(\text{gasto en tecnología}_i, \text{atributos sociodemográficos}_i | \text{equipamiento}_i) + \text{error}_i$$

donde  $i = 1, 2, \dots, n_k$ , y  $n_k$  representa el tamaño de la muestra por cada CA. El consumo de tecnología se refiere a una de las formas planteadas de índice para medir la brecha tecnológica y en esta especificación representa el consumo en bienes complementarios. Nuestro análisis por CCAA recoge las diferencias entre las mismas permitiendo heterogeneidad. El modelo a estimar por máxima verosimilitud es:

$$USE_{ik} = \beta_{0k} + \beta_{1k}ITIC_{ik} + \beta_{2k}HOUSINT_{ik} + \beta_{3k}EQ_{ik} + \beta_{4k}MALE_{ik} + \beta_{5k}AGE_{ik} + \beta_{6k}HABITAT_{ik} + \beta_{7k}STUDYLEVEL_{ik} + \beta_{8k}BESTUDYING_{ik} + u_{ik} \quad (3.2)$$

donde  $USE_i = \ln \lambda_i$ , el subíndice  $i$  representa al individuo y el subíndice  $k = 1, 2, \dots, 18$  a la CA considerada.

El detalle de los resultados de cada CA se puede revisar en el apéndice, donde vemos que las 18 ecuaciones tienen un ajuste satisfactorio con pseudo  $R^2$  entre 0,32 y 0,42. A su vez, los coeficientes estimados serán significativos conjuntamente en todas las regresiones. A continuación presentamos una tabla de las elasticidades, refiriéndose estas al impacto porcentual provocado en el uso de Internet por variaciones porcentuales unitarias en el gasto en equipamiento, la edad y el nivel de estudios:

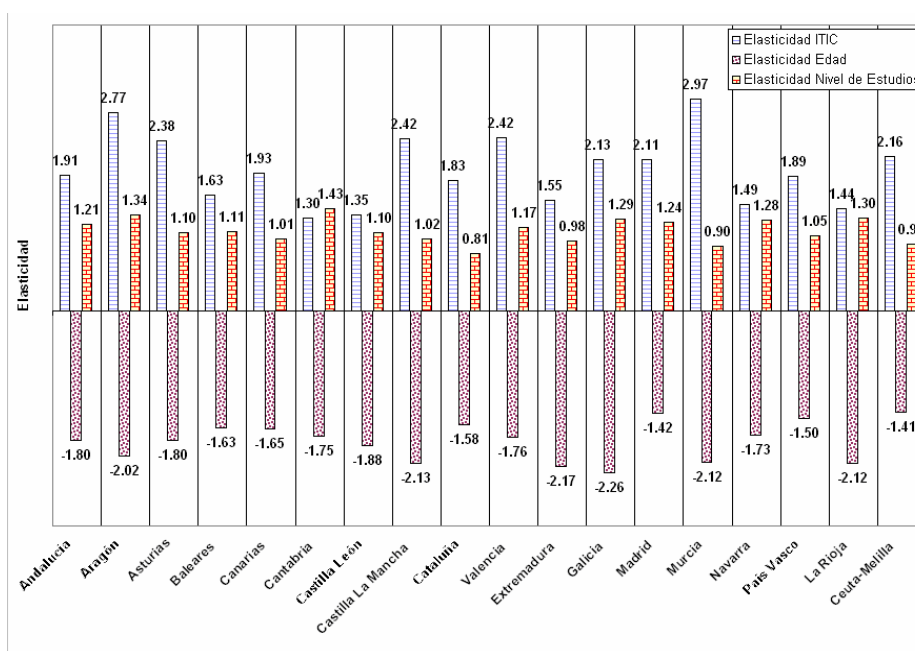
**Tabla 5:** Elasticidades del Uso de Internet por CCAA

COMUNIDAD AUTÓNOMA	Tamaño de la Muestra por CCAA	(% Variación del Uso de Internet		
		Gasto	Edad	Nivel de Estudios
Andalucía	2.147	1,91	-1,80	1,21
Aragón	996	2,77	-2,02	1,34
Asturias	893	2,38	-1,80	1,10
Baleares	579	1,63	-1,63	1,11

<b>Canarias</b>	787	1,93	-1,65	1,01
<b>Cantabria</b>	544	1,30	-1,75	1,43
<b>Castilla León</b>	1.304	1,35	-1,88	1,10
<b>Castilla La Mancha</b>	1.012	2,42	-2,13	1,02
<b>Cataluña</b>	1.833	1,83	-1,58	0,81
<b>Valencia</b>	1.377	2,42	-1,76	1,17
<b>Extremadura</b>	831	1,55	-2,17	0,98
<b>Galicia</b>	1.171	2,13	-2,26	1,29
<b>Madrid</b>	1.567	2,11	-1,42	1,24
<b>Murcia</b>	773	2,97	-2,12	0,90
<b>Navarra</b>	1.202	1,49	-1,73	1,28
<b>País Vasco</b>	1.181	1,89	-1,50	1,05
<b>La Rioja</b>	549	1,44	-2,12	1,30
<b>Ceuta-Melilla</b>	203	2,16	-1,41	0,94

En el cuadro de arriba se ve cómo varían las elasticidades por CCAA y que estas dependerán de según qué variable se esté considerando. A continuación se presenta un gráfico comparativo con los valores de la Tabla 5, en donde se puede ver que las similitudes o diferencias en elasticidades entre CCAA dependerán de las variables a considerar:

**Gráfico 7: Elasticidades del uso de Internet**



Aquí observamos que el uso de Internet en Murcia es más sensible al consumo de tecnología con un valor de 2,97, esto quiere decir que si en Murcia aumenta el consumo individual en equipamiento un 1%, el número de lugares de uso de Internet se incrementará un 2,97%. Para el caso de Comunidades como Madrid o Cataluña, en donde anteriormente se veía que el consumo en equipamiento era de los primeros del ranking, dicho impacto es de 2,11% y 1,83% respectivamente. Observamos también que en general el uso de Internet tendrá elasticidades mayores que uno respecto al consumo de tecnología.

Por otro lado, observamos que en todos los casos la edad tiene un impacto negativo en el uso de Internet. Las elasticidades serán todas negativas, encontrándose los mayores impactos en las CCAA más atrasadas en cuanto a nivel de digitalización como hemos visto que serían Galicia, Extremadura o Castilla-La Mancha con valores de menos 2,26%, 2,17% y 2,13% respectivamente. Ceuta-Melilla, Madrid y País Vasco serán las CCAA cuyas elasticidades con respecto a la edad estarían en el otro extremo, con valores de -1,41, -1,42 y -1,50 y respectivamente.

En cuanto al nivel de educación, es interesante observar que Cataluña, una de las CCAA con mayor nivel de digitalización, tenga el menor valor de elasticidad del uso de Internet con respecto al grado de educación adquirido (0,81). Esto quiere decir que de aumentar 1% el nivel de educación adquirido a nivel agregado, el incremento en el uso de Internet sería menos que proporcional, de 0,81%. Algo parecido sucedería con Ceuta-Melilla o Murcia, ambas regiones con valores por debajo de la unidad (0,94 y 0,90 respectivamente). Entre estas se encuentra Extremadura, la CCAA con menor grado de digitalización y con un valor de 0,98. Las regiones con elasticidades mayores son aquellas de nivel de digitalización medio como Cantabria, Aragón y La Rioja, con elasticidades de 1,43, 1,34 y 1,30 respectivamente.

### 3.3. HETEROGENEIDAD

Modelizar los determinantes del nivel de digitalización que lleva a que exista una brecha tecnológica en España requerirá tener en cuenta la heterogeneidad que hemos observado entre las distintas CCAA. En el apartado anterior hemos realizado regresiones por CCAA y hemos calculado elasticidades. Siguiendo este criterio, a continuación agruparemos a las CCAA según sus elasticidades, intentando de esta manera identificar en dichos grupos comportamientos similares en algún sentido<sup>7</sup>. Para lograr este objetivo utilizamos el método de Ward de agrupamiento jerárquico por variables, también denominado Análisis de Conglomerados. Con este método partimos desde las medias de los subgrupos definiendo una medida global de heterogeneidad que es la suma de las distancias euclídeas al cuadrado entre cada elemento y la media de su grupo:

$$W = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)^T \quad (3.3)$$

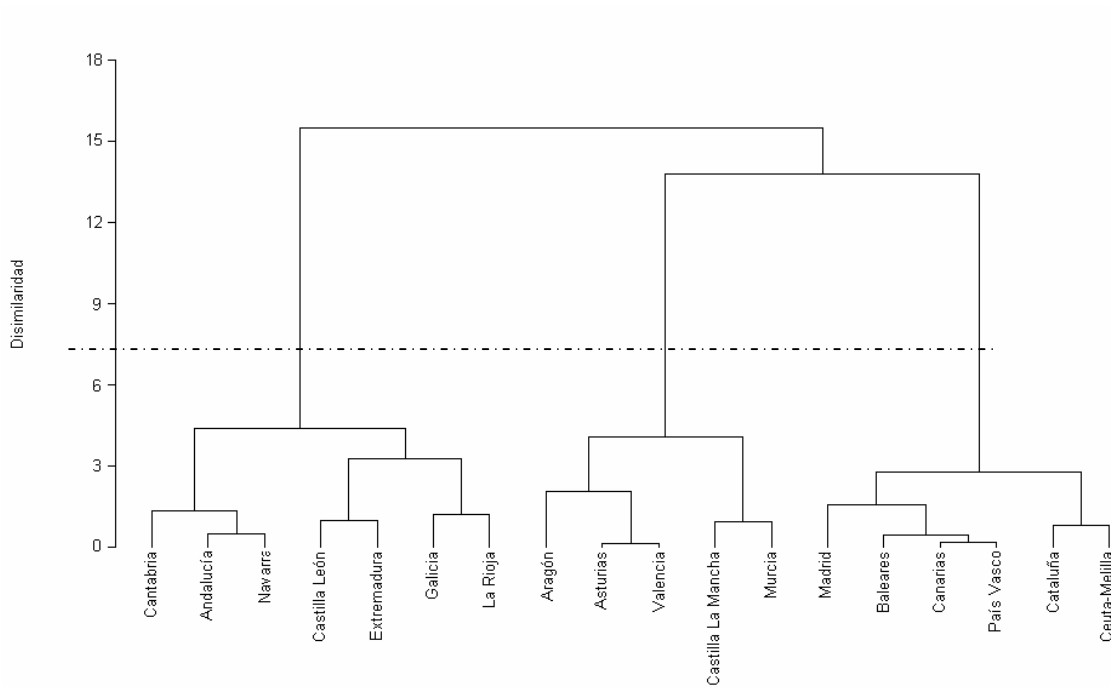
El procedimiento comienza suponiendo que cada dato forma un grupo, por tanto  $W = 0$ . Luego se unen los elementos más próximos de manera tal de que se produzca el mínimo incremento en  $W$ . Entonces en la siguiente etapa tendremos  $n - 2$  grupos de un elemento y uno de dos elementos. El algoritmo seguirá hasta unir todos los grupos en uno solo. En la representación gráfica (dendrograma) se determina el corte a un nivel de distancia dado y así obtenemos una clasificación del número de grupos existentes y los

---

<sup>7</sup>Agrupar por elasticidades es uno de los criterios existentes para agrupar las CCAA. Por ejemplo, también se podría agrupar por efecto marginal del regresor o por ratio de incidencia, aunque los resultados son similares.

elementos que lo forman. Este tipo de análisis es muy útil para explorar datos y detectar subgrupos homogéneos dentro del total de la muestra, determinando previamente el grado de similitud que se prefiere entre los elementos de cada grupo. En nuestro caso la evidencia es bien clara al observar el próximo Gráfico 8, en donde vemos que se forman tres grupos de CCAA (el resto de los resultados en el apéndice):

**Gráfico 8:** Árbol jerárquico de las CCAA a partir de las tres elasticidades calculadas



En este gráfico vemos que la partición se la puede realizar a un nivel bajo de agregación, observando la existencia clara de tres grupos dentro de los cuales existe una cierta similitud. A partir de aquí los grupos formados serán los siguientes:

**Tabla 6:** Conglomerados de CCAA

GRUPO 1: Menos Sensibles	Cantabria Andalucía Navarra Castilla y León	Extremadura Galicia La Rioja
GRUPO 2: Sensibilidad Media	Aragón Asturias Valencia	Castilla – La Mancha Murcia
GRUPO 3: Más Sensibles	Madrid Baleares Canarias	País Vasco Cataluña Ceuta – Melilla

Entonces vemos la existencia de tres grupos diferenciados por sus mayores o menores predisposiciones a las nuevas tecnologías según aumente la renta, el nivel de estudios o la edad, sea por el consumo, el uso o la misma afinidad. En la sección siguiente se considerará individualmente a cada uno de estos grupos en el modelo de los determinantes de la brecha tecnológica.

#### 4. DETERMINANTES DEL GRADO DE DIGITALIZACIÓN EN ESPAÑA

En el apartado anterior se detectaron tres grupos de CCAA caracterizados por su mayor o menor elasticidad de uso de Internet respecto a tres determinantes del mismo. Utilizando dichos grupos, a continuación se analizan los determinantes de la brecha tecnológica en España con un modelo de regresión lineal para explicar el índice de brecha tecnológica, medida por el grado de digitalización ( $IDD_i$ ).

La hipótesis que se quiere comprobar es que la brecha tecnológica es en realidad el resultado de otras brechas existentes tales como la brecha educativa, la brecha de edad o la brecha de la renta, entre otras. Las desigualdades estructurales existentes en otros ámbitos como el económico o el sociodemográfico contribuirán en gran manera a configurar la brecha tecnológica en España.

La comprobación de este supuesto se abordará desde tres líneas de análisis:

- Estimación de una función de regresión MCO de los determinantes del nivel de digitalización individual en España: Esto permitirá calcular los impactos de los atributos que se piensan influyen en el nivel de digitalización individual.
- Estimación de una función de regresión cuantílica de los determinantes del nivel de digitalización individual en España: Con este enfoque, complementario del anterior, se pretende establecer que las políticas de reducción de la brecha digital no deben orientarse a las mismas causas en toda la población.
- Estimación de un modelo de previsión del nivel de digitalización individual en España: Un modelo de previsión permitirá completar el análisis determinando si la brecha digital existente es debido a la existencia de otras “brechas” tales como la edad, el nivel de renta o el nivel de estudios.

Seguidamente se realizan las estimaciones mencionadas previa discusión de los aspectos metodológicos. Inmediatamente después se presentan e interpretan los resultados obtenidos a partir de cada enfoque planteado.

#### **4.1. ESPECIFICACIÓN Y ESTIMACIÓN DE UNA FUNCIÓN PARA EL GRADO DE DIGITALIZACIÓN INDIVIDUAL**

A continuación se especifica una función lineal del grado de digitalización individual en función de los que se consideran como sus determinantes. Esta especificación permitirá determinar de qué manera impactan los atributos que se cree que influyen en el nivel de digitalización individual. En la literatura teórica y empírica

existe un muy variado número de funciones de este tipo, pero sin duda la forma funcional lineal en dobles logaritmos es la forma tradicional debida tanto a su simplicidad como a su fácil forma de interpretar, ya que se puede demostrar que equivale a una función de elasticidad constante. He aquí el modelo:

$$\begin{aligned} \ln IDD_i = & \beta_0 + \beta_1 \ln IRPC_i + \beta_2 \ln CPRO_i + \beta_3 \ln POBLAC_i + \\ & + \beta_4 \ln AGE_i + \beta_5 \ln STUDYLEVEL_i + \beta_6 \ln USOR_i + \beta_7 \ln USOIN_i + \\ & + \beta_8 MALE_i + \beta_9 BESTUDYING_i + \sum_{h=1}^3 \gamma_h CLUSTER_{ih} + \sum_{m=1}^6 \phi_m WORK_{im} + u_i \end{aligned} \quad (3.4)$$

El detalle y la definición de cada una de las variables se pueden encontrar en el apéndice. La función especificada expresa que el nivel de digitalización individual está en función de variables económicas a nivel agregado como el nivel de renta per capita *IRPC* y de la capacidad productiva *CPRO*, ambas variables tomadas a nivel provincial. A su vez, está en función atributos sociodemográficos tales como el hábitat poblacional, medido por la variable *POBLAC*, la edad (*AGE*) y el nivel de estudios alcanzados (*STUDYLEVEL*). También se consideran atributos de uso de las nuevas tecnologías tales como la frecuencia de uso del ordenador y la frecuencia de la conexión a Internet (*USOR* y *USOIN* respectivamente). El resto de las variables consideradas corresponden a variables ficticias como la pertenencia al género masculino (*MALE*), encontrarse estudiando en ese momento (*BESTUDYING*), las ficticias *CLUSTER* correspondientes a los tres grupos de CCAA detectados en el apartado anterior, y las categorías laborales consideradas en las variables *WORK*.

Con los datos disponibles a continuación se procede a estimar por mínimos cuadrados ordinarios (MCO) la función lineal (3.4) especificada. Para corregir posibles

problemas de heteroscedasticidad con forma desconocida, se utiliza la matriz de varianzas y covarianzas sugerida por White (1980) denominada *Matriz de Varianzas y Covarianzas Consistente a Heteroscedasticidad de Forma Desconocida*.

Se estima el modelo a partir de la muestra completa de 18.948 observaciones. Los resultados de esta estimación serían aplicables a todos los individuos. De aquí en adelante, la referencia a este modelo será a través de las siglas de “Modelo Básico Incondicional” (MBI).

**Tabla 7:** Estimaciones Robustas de Funciones Lineales de Digitalización

Variable a Explicar	Todos los Individuos de la Muestra
Nivel de Digitalización (IDD)	MBI
Constante	-0,37 (1,02)*
ln <i>IRPC</i>	0,10 (2,14)
ln <i>CPRO</i>	-0,01 (0,17)*
ln <i>POBLAC</i>	0,02 (0,58)*
ln <i>AGE</i>	-0,07 (7,94)
ln <i>STUDYLEVEL</i>	0,19 (26,34)
ln <i>USOR</i>	0,06 (22,95)
ln <i>USOIN</i>	0,09 (30,80)
<i>MALE</i>	0,01 (2,57)
<i>BESTUDYING</i>	0,01 (1,33)*
<i>CLUSTER1</i>	-0,02 (2,44)
<i>CLUSTER2</i>	-0,02 (2,89)
<i>CLUSTER3</i>	---
<i>WORK1</i>	-0,06 (7,36)
<i>WORK2</i>	-0,10 (7,35)
<i>WORK3</i>	-0,04 (2,33)
<i>WORK4</i>	-0,05 (5,37)
<i>WORK5</i>	-0,18 (17,81)
<i>WORK6</i>	---
Observaciones	18.796
$R^2$	0,424
AIC	-2,058
RCMSPE	0,358
F (Pr > F <sub>c</sub> )	0,000

Los resultados de la estimación expuestos arriba en la Tabla 7 evidencian que:

- I. La elasticidad de la renta per capita es significativa en la muestra y tiene el signo positivo esperado. Es decir que un incremento del 1% en la misma provocará un incremento positivo de 0,10% en el índice de digitalización.
- II. El caso de la variable referida a la edad del individuo muestra que la elasticidad en la muestra completa tiene signo negativo y es igual a 0,07. Esto revela lo visto en la literatura, en donde la edad suele ser una variable con efecto inverso en el acceso y uso de nuevas tecnologías. Sin embargo en algunos trabajos (Martin, 1995) en donde se trabaja con muestras de individuos más familiarizados con las nuevas tecnologías (por ejemplo los usuarios de Internet) la edad tendrá un efecto positivo.
- III. En cuanto al nivel de estudios alcanzados, la variable será significativa con un efecto positivo de 0,19 para toda la muestra. Curiosamente sin embargo, la variable ficticia referida al hallarse estudiando, no será significativa.
- IV. El caso de las variables referidas a la frecuencia de uso del ordenador y conexión a Internet, los resultados revelan que la frecuencia de uso del ordenador tiene una elasticidad positiva de 0,06 para el caso de la muestra completa. El caso de la frecuencia de conexión a Internet tendrá un efecto positivo de 0,09. Para el caso de la pertenencia al género masculino, el efecto será significativo y positivo de 0,01.
- V. Para el caso de las variables de control referidas a los grupos de CCAA obtenidos en el apartado anterior, se observa que pertenecer a los dos grupos menos sensibles hará que el efecto sea negativo de -0,02 para ambos grupos respecto a la categoría base (en este caso el grupo más sensible *CLUSTER3*).

Se advierte además que se utiliza *CLUSTER3* como categoría base (es decir, omitiéndola de la estimación) para evitar la multicolinealidad exacta.

- VI. Por último, el caso de las variables de control referidas a las categorías laborales es particular ya que todas son significativas y en todos los casos tendrán efectos negativos. Hay que considerar que se trata de constantes específicas referidas a las calificaciones dentro del mercado laboral (ocupado por cuenta ajena, por cuenta propia, en el paro, labores del hogar, estudiantes y pensionados) y que no tienen en cuenta el grado de calificación del empleo en sí. Dicho efecto se mide en los parámetros de la variable referida al nivel de estudios alcanzado.

De este modo entonces la estimación de un modelo para explicar los determinantes del nivel de digitalización individual derivará en conclusiones similares o diferentes según se trate de una muestra completa o restringida sólo a los usuarios de Internet.

#### **4.2. IMPACTO DE DETERMINANTES A DISTINTOS NIVELES DE DIGITALIZACIÓN**

Hasta aquí se ha evidenciado que los cambios en el nivel de digitalización en España dependerán de factores tanto observables como no observables. La explicación hasta ahora de las diferencias existentes entre los distintos niveles de brecha tecnológica se ha concentrado en encontrar variables que se relacionan con características de los individuos tales como la renta per capita del entorno geográfico, la edad o la educación entre otras. En el apartado 3 se analizaron y detectaron grupos homogéneos a nivel de CA. Luego, dichos grupos han sido incluidos en el modelo anterior recogiendo la heterogeneidad. Sin embargo, y a pesar de contar con significatividad global en ambos

modelos, los valores bajos de bondad de ajuste revelan que el efecto de la variable explicativa en la variable endógena puede no ser el mismo en la media condicional de la variable endógena debido a un efecto *leverage* considerable provocado principalmente por la existencia de grupos aislados de observaciones anómalas.

#### 4.2.1. REGRESIÓN CUANTÍLICA

Una buena metodología para modelizar situaciones en donde existe heterogeneidad de comportamiento o un número elevado de observaciones anómalas, es la regresión cuantílica (Koenker y Basset, 1978). Esta forma de regresión es una técnica semiparamétrica que puede verse como una extensión de la regresión clásica MCO. La diferencia fundamental radica en que en los MCO se estima una función de *media* condicional, mientras que la regresión cuantílica equivale a un grupo de modelos para varias funciones de *cuantiles* condicionales. El caso principal es el estimador de la regresión mediana que minimiza la suma de errores absolutos. Otras funciones cuantílicas condicionales se estiman a partir de minimizar una suma de errores absolutos ponderada asimétricamente.

Se define como *cuantíl* a un valor  $Y_\tau$  de la muestra que deja a un porcentaje  $\tau$  de las observaciones por encima de  $Y_\tau$  y un porcentaje  $(1-\tau)$  por debajo de este. Entonces, la regresión cuantílica (MRC) permite estimar distintas regresiones para distintos valores de  $\tau$  ocupando *toda* la información muestral pero captando las diferencias entre el porcentaje superior e inferior separado por  $Y_\tau$ .

La justificación adicional del uso de esta metodología se centra en buscar responder las siguientes preguntas:

- ¿Responderán de manera distinta los cuantiles más bajos y más altos del nivel de digitalización frente a iguales cambios en algún determinante del mismo?
- ¿Deben las políticas de reducción de brecha digital centrarse en las mismas causas para toda la población?
- Los determinantes observables del nivel de digitalización individual, ¿son estos fenómenos estadísticamente significativos y tienen todos los mismos impactos entre distintas porciones de la distribución del nivel de digitalización?

El modelo de regresión clásico MCO estima los parámetros de la especificación  $y = x'\beta + u$  con el supuesto de  $u \succ N(0, \sigma^2)$ , de modo que a  $E(y|x) = x'\beta$  se la conoce como función de regresión y la identidad  $\partial E(y|x)/\partial x = \partial y/\partial x = \beta$  sólo se da si la perturbación  $u$  es independiente de  $x$ . En una situación en donde existen un elevado grupo de observaciones anómalas o una persistente heteroscedasticidad, el regresor  $x$  no afecta a  $E(y|x)$  pero sí afectará a  $y$ . Los MRC intentan modelizar el efecto del regresor  $x$  sobre toda la distribución de  $y$ .

De forma análoga a los MCO, el MRC para el  $\tau$ -ésimo cuantil de la distribución de la variable endógena  $y$  puede ser visto:

$$Q_{y|x}(\tau) = x'\beta(\tau)$$

Con:

$$\frac{\partial Q_{y|x}(\tau)}{\partial x} = \beta(\tau)$$

Es decir que se está permitiendo que el efecto de la variable  $x$  sobre  $y$  sea distinto en distintos lugares de la distribución de  $y$  dado  $x$ .

La estimación de  $\beta(\tau)$  (Koenker y Basset, 1978) se la obtiene resolviendo el siguiente programa lineal:

$$\hat{\beta}(\tau) = \arg \min \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta)$$

$$\rho_{\tau}(z) \equiv z(\tau - I(u < 0))$$

Las ventajas de esta metodología son, en primer lugar, su contundencia estadística a valores extremos, dado que el MRC ajusta hiperplanos entre las observaciones, de modo que el estimador MRC para un valor dado de  $\tau$  no varía ante la presencia de observaciones anómalas en la variable dependiente. La segunda ventaja de este método es que permite ajustar distintas regresiones para distintos cuantiles en la distribución utilizando toda la información muestral lo que permitirá responder a las preguntas realizadas más arriba. En el contexto de la función de los determinantes del nivel de digitalización, la estimación MRC dará coeficientes o elasticidades factoriales respecto del nivel de digitalización individual para los distintos cuantiles que se definan.

Antes de presentar y discutir los resultados es conveniente señalar tres aspectos metodológicos. Primero, el MRC aplicado a la estimación se realizó para valores de  $\tau$  iguales a 0,10, 0,25, 0,50, 0,75 y 0,90. Segundo, las regresiones se realizaron de forma simultánea, que tiene la ventaja de que las varianzas son menores en relación con la estimación por separado dado que se las estima por bootstrapping, obteniéndose estadísticos  $t$  de mayor potencia y porque permite hacer contrastes sobre coeficientes de

manera intra e intercuantílica (Martins y Pereira, 2003). Finalmente, en algunos casos es particularmente útil la regresión del modelo en diferencias de cuantiles. Por ejemplo, si el interés se centra en medir las diferencias de un mismo coeficiente entre el cuantil 0,75 y el cuantil 0,25 y determinar si esta diferencia es significativa o no, se debe establecer el siguiente modelo:

$$Q_{0,75}(y) = \beta_{0(0,75)} + \beta_{1(0,75)}X_1 + \beta_{2(0,75)}X_2$$

$$Q_{0,25}(y) = \beta_{0(0,25)} + \beta_{1(0,25)}X_1 + \beta_{2(0,25)}X_2$$

Luego, el modelo en diferencias de cuantiles está dado por:

$$Q_{0,75}(y) - Q_{0,25}(y) = (\beta_{0(0,75)} - \beta_{0(0,25)}) + (\beta_{1(0,75)} - \beta_{1(0,25)})X_1 + (\beta_{2(0,75)} - \beta_{2(0,25)})X_2$$

Esta última ecuación establece coeficientes que corresponden a las diferencias entre coeficientes de los modelos ajustados para los cuantiles 0,75 y 0,25. Los contrastes de dichas diferencias se realizan a partir de errores estándar calculados por medio de bootstrapping.

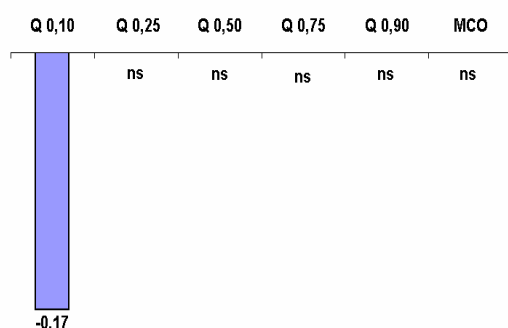
En la Tabla 8 más abajo se muestran los resultados de las estimaciones simultáneas para los distintos cuantiles considerados. Se puede señalar que las estimaciones tienen un buen ajuste al observar los valores del Pseudo  $R^2$  de cada una de las ecuaciones. A su vez, la gran mayoría de los coeficientes estimados son estadísticamente significativos. Se observa que el valor constante sólo será significativo en las partes inferiores de la distribución del nivel de digitalización individual (0,10 y 0,25), mientras que para la mediana y cuantiles superiores no será significativo.

**Tabla 8:** Resultados de la Regresión Cuantílica Simultánea para *IDD*

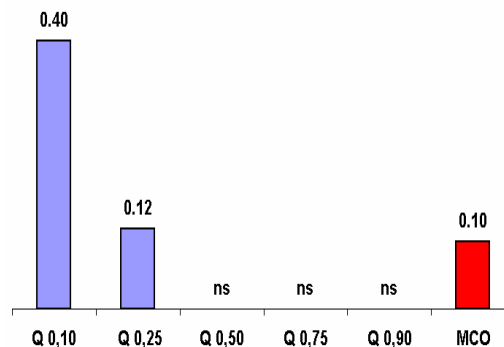
$\tau$	0,10	0,25	0,50	0,75	0,90
Variable Endógena: Nivel de Digitalización ( <i>IDD</i> )					
Constante	-2,71 (3,27)	-0,84 (2,36)	-0,21 (0,65)*	0,18 (0,26)*	-0,18 (0,25)*
ln <i>IRPC</i>	0,40 (3,93)	0,12 (2,36)	0,05 (1,09)*	0,06 (0,75)*	0,13 (1,49)*
ln <i>CPRO</i>	-0,17 (2,17)	-0,04 (1,17)*	-0,02 (0,08)*	0,02 (0,32)*	-0,01 (0,09)*
ln <i>POBLAC</i>	0,18 (2,27)	0,05 (1,57)*	0,02 (0,50)*	-0,01 (0,12)*	0,01 (0,19)*
ln <i>AGE</i>	-0,11 (6,93)	-0,07 (7,93)	-0,07 (10,55)	-0,07 (5,74)	-0,004(0,22)*
ln <i>STUDYLEVEL</i>	0,19 (10,89)	0,13 (8,49)	0,16 (14,04)	0,21 (17,09)	0,21 (11,35)
ln <i>USOR</i>	0,03 (11,74)	0,04 (12,83)	0,09 (35,86)	0,04 (13,09)	0,02 (4,58)
ln <i>USOIN</i>	0,03 (5,91)	0,11 (15,68)	0,08 (34,35)	0,09 (31,41)	0,07 (17,19)
<i>MALE</i>	0,02 (1,76)	0,01 (2,14)	0,02 (4,39)	0,01 (1,14)*	-0,02 (1,72)
<i>BESTUDYING</i>	0,02 (1,17)*	0,02 (1,23)*	0,02 (1,78)	0,01 (0,13)*	0,01 (0,98)*
<i>CLUSTER1</i>	-0,04 (3,63)	-0,01 (2,11)	-0,03 (4,75)	-0,01 (0,71)*	-0,04 (2,34)
<i>CLUSTER2</i>	-0,02 (2,09)	-0,01 (2,27)	-0,03 (4,44)	-0,02 (1,71)	-0,04 (2,41)
<i>CLUSTER3</i>	---	---	---	---	---
<i>WORK1</i>	-0,05 (5,29)	-0,06 (5,57)	-0,05 (6,66)	-0,06 (5,33)	-0,05 (4,68)
<i>WORK2</i>	-0,10 (3,90)	-0,09 (5,11)	-0,07 (6,42)	-0,12 (4,39)	-0,08 (2,73)
<i>WORK3</i>	0,04 (1,58)*	0,002 (0,08)*	-0,04 (2,79)	-0,05 (3,09)	-0,02 (1,06)*
<i>WORK4</i>	-0,01 (0,47)*	-0,05 (3,87)	-0,04 (4,71)	-0,06 (2,39)	0,001 (0,03)*
<i>WORK5</i>	-0,13 (6,82)	-0,10 (9,07)	-0,17 (30,00)	-0,30 (18,32)	-0,25 (13,77)
<i>WORK6</i>	---	---	---	---	---
Observaciones	18.796	18.796	18.796	18.796	18.796
Pseudos $R^2$	0,1438	0,2213	0,3229	0,3061	0,1750

En los gráficos siguientes se observan las comparaciones de las elasticidades del nivel de digitalización respecto a las variables continuas para los distintos cuantiles considerados y la estimación MCO como referencia:

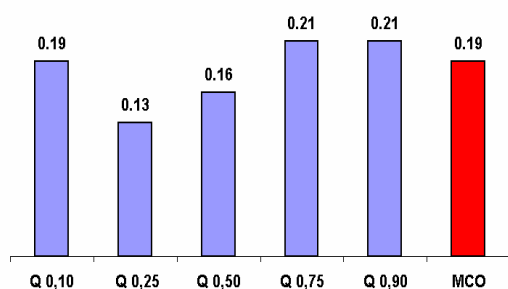
**Gráfico 9:** Capacidad Productiva



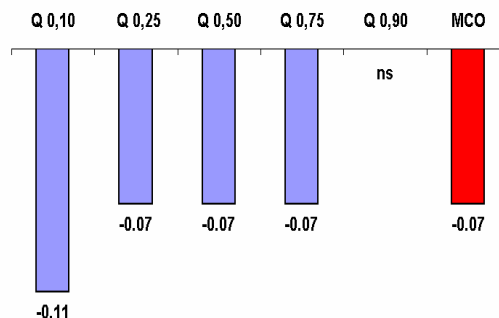
**Gráfico 10:** Renta per Capita



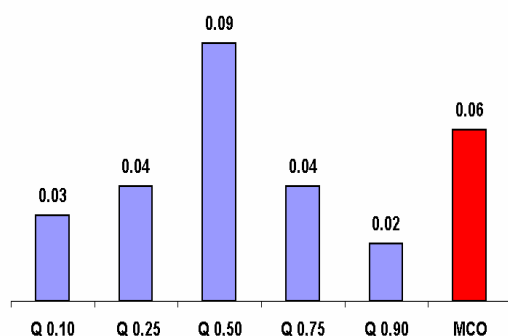
**Gráfico 11: Nivel de Estudios**



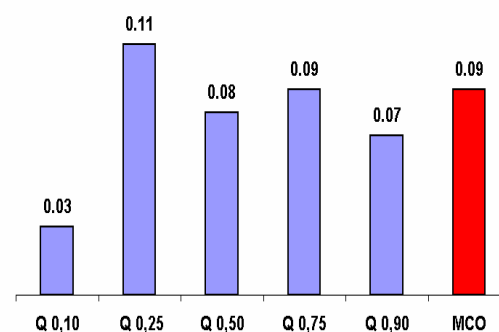
**Gráfico 12: Edad**



**Gráfico 13: Frecuencia de Uso del PC**



**Gráfico 14: Frecuencia de Uso de Internet**



De estos resultados se puede decir:

- Se observa que la situación es distinta del resto para los cuantiles 0,10 y 0,25 referidos a los valores más bajos de digitalización individual. Por ejemplo, la capacidad productiva provincial sólo será significativa en el cuantil 0,10, mientras que la renta per capita lo será además solo en el cuantil 0,25. Es decir que, en las provincias o regiones en donde el aporte al PIB nacional se incremente el 1% hará disminuir el nivel de digitalización un 0,17% *sólo* en los individuos con niveles más bajos del mismo. En el resto de la distribución esta variable no influye.
- El caso de la renta per capita tendrá elasticidades de 0,40 en el cuantil 0,10 y de 0,12 en el cuantil 0,25. En el resto de los cuantiles tampoco influirá. La curiosidad aquí es que a medida que se va aumentando el nivel de

digitalización individual (es decir, pasándose de cuantiles inferiores a superiores) estas variables pierden significatividad.

- Al observar las elasticidades con respecto al nivel de estudios alcanzado, se observa que ocurre lo opuesto, ya que los valores de elasticidad serán mayores en los cuantiles superiores de la distribución (0,21% tanto para el cuantil 0,75 y 0,90). La variable edad no será significativa en el cuantil más alto de la distribución. Lo contrario ocurre con el cuantil más bajo, en donde el impacto es el mayor de todos con un valor negativo de 0,11%.
- Luego, las variables referidas a la frecuencia de uso del ordenador y de conexión a Internet tendrán a lo largo de toda la distribución impactos positivos, con más o menos intensidad según qué cuantil se trate. Es el caso de la frecuencia del uso del PC en donde la mayor elasticidad se encuentra en la mediana (cuantil 0,50) con un valor de 0,09%. Para el caso de la frecuencia de conexión a Internet el mayor impacto será en el cuantil 0,25 con un valor de 0,11%.

Asimismo y para completar el análisis se observan a continuación (véase Tabla 9) los resultados del modelo en diferencias descrito más arriba para las porciones centrales de la distribución definidas por las diferencias de los cuantiles extremos (0,90 – 0,10) y de los cuantiles (0,75 – 0,25). Se puede observar que en este caso ambos términos constantes son significativos, y que en la porción central más grande de la distribución (definida por la diferencia de los cuantiles 0,10 y 0,90) la mayoría de las variables continuas serán significativas. Las únicas variables que no serán significativas pero que interesantemente sí lo son para la porción central más pequeña (definida por la diferencia de los cuantiles 0,25 y 0,75) serán la variable

referida al nivel de estudios y la variable ficticia referida a encontrarse estudiando en ese momento (*STUDYLEVEL* y *BESTUDYING*). La variable que será significativa en ambas porciones es la referida a la frecuencia de conexión a Internet con un valor positivo de 0,03 para el caso del rango mayor y negativa de 0,02 para el rango menor.

**Tabla 9:** Resultados de la Regresión Intercuantílica para *IDD*

<b>Modelo en Diferencias</b>	<b>Cuantiles 0,10 y 0,90</b>	<b>Cuantiles 0,25 y 0,75</b>
Constante	2,53 (2,99)	1,02 (2,07)
ln <i>IRPC</i>	-0,27 (2,54)	-0,5 (0,82)*
ln <i>CPRO</i>	0,16 (2,19)	0,06 (1,28)*
ln <i>POBLAC</i>	0,16 (2,19)	-0,06 (1,27)*
ln <i>AGE</i>	0,10 (3,65)	0,001 (0,12)*
ln <i>STUDYLEVEL</i>	0,02 (0,84)*	0,07 (5,55)
ln <i>USOR</i>	-0,01 (1,97)	0,003 (0,59)*
ln <i>USOIN</i>	0,03 (5,88)	-0,02 (4,07)
<i>MALE</i>	-0,04 (3,04)	-0,005 (0,81)*
<i>BESTUDYING</i>	-0,01 (0,34)*	-0,02 (4,07)
<i>CLUSTER1</i>	-0,001 (0,07)*	-0,005 (0,62)*
<i>CLUSTER2</i>	-0,02 (1,22)*	-0,003 (0,34)*
<i>CLUSTER3</i>	---	---
<i>WORK1</i>	0,001 (0,10)*	0,004 (0,23)*
<i>WORK2</i>	-0,02 (0,72)*	-0,03 (1,45)*
<i>WORK3</i>	-0,06 (1,84)	-0,05 (2,64)
<i>WORK4</i>	0,01 (0,34)*	-0,004 (0,20)*
<i>WORK5</i>	-0,12 (5,07)	-0,20 (9,30)
<i>WORK6</i>	---	---
Observaciones	18.976	18.976
Pseudo R <sup>2</sup> ( $\tau = 0,90$ )	0,1750	
Pseudo R <sup>2</sup> ( $\tau = 0,10$ )	0,1438	
Pseudo R <sup>2</sup> ( $\tau = 0,75$ )	0,3061	
Pseudo R <sup>2</sup> ( $\tau = 0,25$ )	0,2213	

Otro resultado importante será referido a las variables ficticias consideradas, donde las únicas que serán significativas en ambos modelos serán estar en el paro (*WORK3*) con un impacto negativo en el nivel de digitalización individual de 0,06 y 0,05 para la

porción mayor y menor respectivamente) y dedicarse a las labores del hogar (*WORK5* con un impacto negativo de 0,12 y 0,20 respectivamente).

Resumiendo entonces, en general se observa que los individuos con mayor nivel de digitalización responderán de manera distinta que los de menor nivel a cambios en los determinantes. Es decir que toda política orientada a reducir la brecha digital deberá centrarse primeramente en este tipo de diferencias de impacto y en los fenómenos que resultan ser estadísticamente significativos. Por ejemplo, percibir que en sectores de la población de nivel de digitalización bajo, el mayor peso lo tendrán variables de tipo estructural como el nivel de renta o la edad, mientras que en sectores de nivel de digitalización elevado estas no son motivos relevantes mientras que sí lo serán el nivel de estudios alcanzados, el uso del ordenador o la frecuencia de conexión a Internet entre otros.

### **4.3. PREDICCIÓN DEL NIVEL DE DIGITALIZACIÓN**

En el apartado anterior está evidenciada la posibilidad de que la brecha digital sea en realidad la consecuencia de brechas existentes de edad, de educación o de renta. Tener en cuenta esta posibilidad será importante al momento de tomar decisiones de política y para ello a continuación se plantea un modelo de predicción que permita prever la reducción de la brecha digital obtenida al considerar variaciones hipotéticas en las otras brechas consideradas.

Entonces, para obtener un buen modelo de predicción se utilizará el algoritmo RETINA (Relevant Transformations of the Inputs Network Approach) y se seguirán los pasos metodológicos de Pérez-Amaral, Gallo y White (2005).

#### 4.3.1. ALGORITMO RETINA

Propuesto por Pérez-Amaral, Gallo y White (2003), este algoritmo fue concebido como una herramienta para la construcción y selección de modelos. Tiene la flexibilidad de los modelos de redes neuronales (White, 1989) al permitir no linealidades e interacciones entre las variables explicativas a través de transformaciones de las variables originales. A su vez asegura la concavidad en los parámetros de la función de verosimilitud como en los modelos lineales comunes (es decir que evita complejidades numéricas en la estimación) y posee la capacidad de identificar el conjunto de atributos con mayor capacidad predictiva (principio de parsimonia). Este procedimiento es particularmente útil en aquellas situaciones en las cuales el investigador no posee hipótesis a priori acerca de la forma funcional que debe adoptar el fenómeno que desea explicar, y necesita una valoración sobre qué variables o transformaciones de estas serán más útiles para la predicción. También la efectividad de RETINA fue contrastada con simulaciones en Pérez-Amaral *et al.* (2003) en donde se reporta un gran porcentaje de aciertos en reconocer además el proceso generador de los datos.

El algoritmo funciona como sigue:

- Paso 0: Ordenamiento y Construcción de Predictores
  - Generación de un conjunto de variables transformadas a partir de las variables originales  $\zeta(X) = \{W_1, \dots, W_m\}$ .
  - Partición de la muestra en tres submuestras.
- Paso 1: Construcción del Modelo “Candidato”
  - Utilizando los datos de la primer submuestra:

- Ordenamiento de las variables de  $\zeta(X)$  de acuerdo a su correlación simple en valor absoluto con la variable endógena (de tal manera que  $W_{(1)}$  será la transformación con mayor correlación con la variable endógena,  $W_{(2)}$  será la segunda más correlacionada, y así sucesivamente).
- Consideración de varios conjuntos de regresores potenciales  $\zeta_\lambda(X)$  en donde cada uno está indexado con un “umbral de tolerancia” de colinealidad  $\lambda$  entre 0 y 1.
- El número máximo de regresores está controlado por el número de valores de  $\lambda$  elegidos (siendo  $W_{(1)}$  el único regresor en el caso de que  $\lambda = 0$ )
- Utilizando los datos de la primera y segunda submuestras:
  - Estimación de regresiones de la variable endógena con cada uno de los conjuntos de  $\zeta_\lambda(X)$  utilizando sólo los datos de la primera submuestra y computando un criterio de predicción fuera de la muestra utilizando sólo los datos de la segunda submuestra. Este criterio de predicción es el Error Cuadrático Medio Robusto de Predicción fuera de la Muestra (RCMSPE en sus siglas en inglés).
  - Selección del modelo que posea el mejor RCMSPE,  $\zeta_{\lambda^*}(X)$ .
- Paso 2: Estrategia de Búsqueda
  - Utilizando los datos de la segunda y tercera submuestras
    - Elección del modelo más parsimonioso estimando todos los modelos incluyendo una constante y todos los regresores de

$\zeta_{\lambda^*}(X)$  en el orden en que originalmente fueron incluidos, y además en el orden producido en base a la correlación en la segunda submuestra.

- Utilizar una medida de evaluación de los modelos por fuera de la muestra (RCMSPE utilizando los datos de la tercera submuestra).
- Paso 3: Selección del Modelo
  - Repetición de los pasos 1 y 2 cambiando el orden de las submuestras, produciendo modelos potencialmente candidatos para cada ordenación de las tres submuestras.
  - Elección del modelo que de mejor desempeño resulta en toda la muestra.

Resumiendo, la idea general que hay detrás del procedimiento planteado en RETINA es la obtención de un modelo predictivo guiándose con la estructura que se halla en los datos, no se requiere una teoría como guía en la búsqueda y obtención de información potencial. Los principios de la construcción de este algoritmo son la flexibilidad, la búsqueda selectiva, el control de la multicolinealidad y la capacidad predictiva por fuera de la muestra. A continuación se utilizará RETINA en la obtención de un modelo de predicción para el nivel de digitalización individual para todas las CCAA de España.

Se comienza con la especificación de un modelo lineal de referencia denominado modelo *Básico*, que este será el ya especificado en (3.5). Se trataba de un modelo en el que no se utilizaba ninguna estrategia particular de selección. Este será el punto de partida. Ahora se utilizará RETINA para encontrar la mejor especificación que se denominará modelo *Predictivo*. Este permitirá la inclusión de no linealidades en los

regresores controlando siempre el nivel de multicolinealidad. La selección posterior se la utilizará para realizar predicciones.

A continuación se presentan los resultados más destacados. Se buscará el mejor modelo predictivo fuera de la muestra. Para ello se seguirá el criterio del modelo con el menor Error Cuadrático Medio de Previsión fuera de la Muestra (CMSPE).

#### 4.4. ELECCIÓN DEL MODELO PREDICTIVO

A continuación se presenta el modelo predictivo para la muestra completa seleccionado por RETINA, en donde se evidencia que tener en cuenta la heterogeneidad de la muestra y la inclusión de no linealidades en las especificaciones mejora el ajuste y la capacidad predictiva del modelo.

**Tabla 10:** Comparación de los resultados entre Modelo Básico y Predictivo

	Modelo Básico	Modelo RETINA
Parámetros	17	17
Transformaciones (W)	--	10
Constantes Específicas	9	--
Pendientes Específicas	--	6
$\bar{R}^2$	0,426	0,704
RCMSPE(1000)	0,358	0,257
AIC	-2,058	-2,718

Tener en cuenta la heterogeneidad y la inclusión de no linealidades en las especificaciones puede mejorar tanto el ajuste como la capacidad predictiva del modelo RETINA al compararlo con el modelo básico. Se destaca que el RCMSPE disminuye de 0,358 a 0,257. También mejoran los valores de bondad de ajuste y el AIC (de 0,426 a 0,704 y de -2,058 a -2,718 respectivamente, todos valores más que aceptables en una muestra de corte transversal). Además se observa que RETINA no selecciona los

regresores continuos originales, aunque sí transformaciones de estos (ver apéndice). Así tenemos regresores tales como  $\ln(STUDYLEVEL)*\ln(EDAD)$ , como  $\ln(EDAD)^2$  y como  $\ln(STUDYLEVEL)^2$ , todos estos significativos.

Si bien nada impide realizar una interpretación de los resultados del modelo RETINA, en este caso dicho modelo sólo se utilizará para realizar predicciones. Para aclarar este último comentario se sostiene de que existe un gran paralelismo entre RETINA y el enfoque ARIMA para la previsión de series temporales. Primero porque los modelos estimados con el enfoque ARIMA son utilizados únicamente para realizar predicciones y no para interpretarlos económicamente. Además se tiene que en ambos enfoques el orden inicial de los potenciales regresores se realiza a partir de correlaciones entre la variable dependiente y cada regresor candidato en el caso de RETINA y el valor retardado en el caso de ARIMA, la identificación está basada en funciones de correlación simple y parcial, y la elección del modelo para predecir se realiza a partir del RCMSPE.

#### **4.4.1. SENSIBILIDAD DEL NIVEL DE DIGITALIZACIÓN Y DE LA BRECHA TECNOLÓGICA EN LAS CCAA**

Más arriba hemos visto que lo que más influirá en la existencia de una brecha tecnológica a nivel de CCAA y a nivel individual en España serán, además de factores tecnológicos tales como la adquisición y uso de equipamiento o el acceso a Internet desde el hogar, también lo serán factores demográficos tales como el nivel de estudios, la edad y el género sexual. Otra característica es que estos factores impactarán posiblemente de manera diferente en cada CCAA.

En los últimos años en España se ha avanzado en el grado de sofisticación on line de la Administración tanto de las CCAA como del Estado central, aunque no todas las CCAA han avanzado por igual. En 2004 los presupuestos dedicados a las TIC por parte de las autonomías suponen ya el 73% del total de partidas asignadas, pero la inversión por habitante oscila entre los 87 euros en el País Vasco y los 10 euros que dedica la Región de Murcia (Fundación AUNA, 2005). También hay grandes diferencias en el número de personas que se han incorporado a Internet en el último año, en donde la comunidad con mayor crecimiento fue Cataluña (con un 22,65% del total), seguidas por Castilla y León, Andalucía y el País Vasco. Como contrapartida en el último puesto se encuentra Castilla-La Mancha, que incluso ha perdido usuarios.

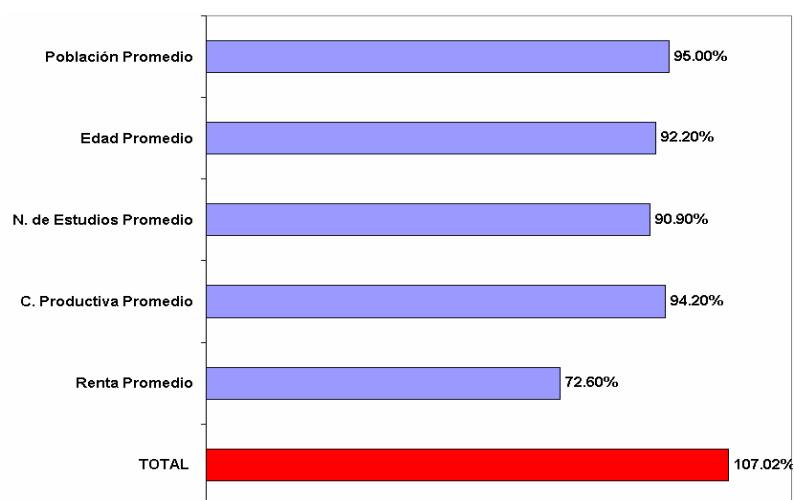
Después de seleccionar un modelo para la predicción en el apartado anterior, a continuación se estudiarán los impactos de los distintos factores del grado de digitalización que influyen directamente en la brecha tecnológica. El análisis se hará diferenciando por CA y se utilizará el modelo RETINA debido a que como se ha visto arriba, es el mejor de los dos que se han estimado por MCO desde el punto de vista estadístico.

Para evaluar dicha sensibilidad se ha fijado cada una de las variables una a una a su valor promedio y observando la variación provocada en las predicciones. Es decir, se observa el cambio porcentual del resultado obtenido en la variable dependiente para cada una de las CCAA calculando directamente la variación del nivel de digitalización y a través de esta el ensanchamiento o disminución de la brecha digital entre las CCAA.

A continuación la Tabla 11 y el Gráfico 15 muestran la sensibilidad del nivel de digitalización respecto a sus principales determinantes. Estos son los logaritmos de la renta ( $\ln IRPCPRO$ ), la capacidad productiva ( $\ln CPRO$ ), el nivel de estudios alcanzados ( $\ln STUDYLEVEL$ ), la edad ( $\ln AGE$ ) y la población ( $\ln POBLAC$ ).

**Tabla 11:** Sensibilidad del Nivel de Digitalización Respecto a sus Determinantes<sup>8</sup>

Comunidad Autónoma	Renta	Capacidad Productiva	Nivel de Estudios	Edad	Población
Andalucía	-2.0%	-5.4%	-4.1%	-5.4%	-5.1%
Aragón	1.5%	3.0%	2.6%	3.1%	2.9%
Asturias	-1.0%	-2.0%	-1.9%	-1.0%	-1.4%
Baleares	2.6%	3.7%	4.6%	4.0%	4.6%
Canarias	5.7%	4.1%	4.0%	3.8%	4.6%
Cantabria	-1.5%	-0.9%	-1.7%	-1.2%	-1.1%
Castilla – León	-4.0%	-3.0%	-4.2%	-3.6%	-4.1%
Castilla La Mancha	-8.0%	-8.8%	-8.0%	-9.8%	-10.0%
Cataluña	6.3%	5.6%	9.2%	8.5%	8.4%
Valencia	0.6%	-1.4%	0.9%	0.1%	0.0%
Extremadura	-8.4%	-11.0%	-10.2%	-11.8%	-12.0%
Galicia	-5.4%	-7.1%	-6.6%	-6.8%	-7.1%
Madrid	12.8%	12.4%	16.2%	15.6%	15.8%
Murcia	-4.3%	-7.4%	-5.6%	-6.2%	-6.7%
Navarra	1.4%	4.1%	3.4%	4.7%	4.4%
País Vasco	-0.3%	2.2%	4.6%	3.6%	2.8%
La Rioja	-1.6%	1.0%	-1.0%	0.0%	0.0%
Ceuta	4.3%	6.9%	-0.3%	2.7%	3.6%
Melilla	1.1%	4.0%	-2.0%	-0.3%	0.3%
Media de Valores Absolutos	<b>2.6%</b>	<b>3.9%</b>	<b>3.3%</b>	<b>2.4%</b>	<b>2.3%</b>
Total Porcentajes en Valor Absoluto	<b>72.6%</b>	<b>94.2%</b>	<b>90.9%</b>	<b>92.2%</b>	<b>95.0%</b>

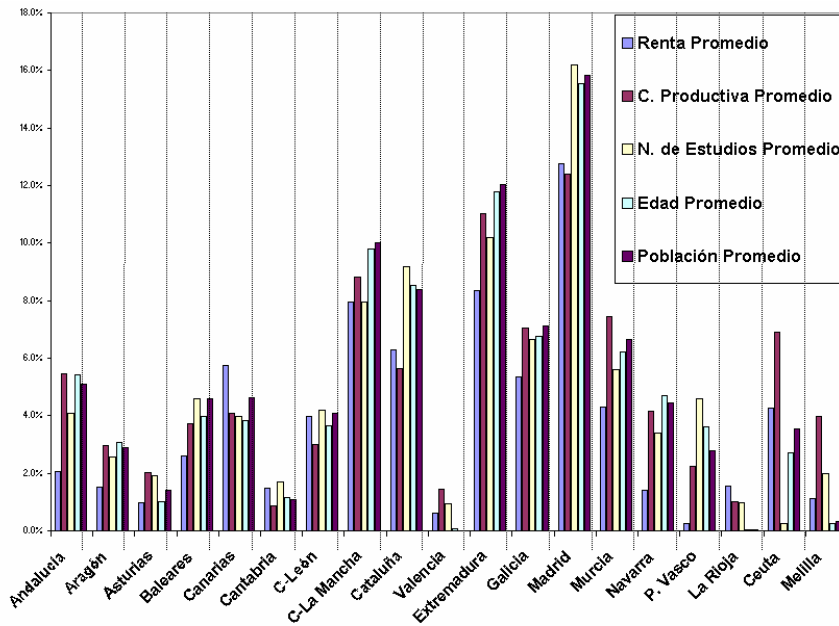
**Gráfico 15:** Total de Porcentajes de Variación con Respecto al Valor Medio

<sup>8</sup> Variación del IDD medio de cada CA. Porcentajes obtenidos como los expuestos en la Tabla 3.

En cuanto a la sensibilidad del grado de digitalización respecto a la renta, cabe notar que si esta fuera la misma para todos los usuarios de Internet (renta promedio), el total de las diferencias de los valores del índice de cada CA respecto de la media total disminuiría desde 107,02% a 72,60%. Es decir que si todas las CCAA tuviesen la renta promedio, se disminuiría la brecha tecnológica en un 34,42%. Para calcular este porcentaje de variación, igualamos el nivel de renta a su media dentro del modelo RETINA estimado, y a partir de ahí se observan las variaciones en las predicciones. O sea que igualar a su media todas las observaciones del nivel de renta provoca un descenso en la brecha digital y sugiere que es una variable con influencia sobre dicha brecha. También se observa que las mayores variaciones porcentuales positivas son provocadas en las comunidades de mayor nivel de digitalización tales como Madrid (12,8%) y Cataluña (6,3%), y las negativas en las comunidades de menor digitalización tales como Extremadura (-8,4%) y Castilla-La Mancha (-8,0%). Esto podría deberse a que regiones con menor tendencia a lo digital sean a su vez regiones con bajo nivel de renta per capita y que obviamente serán menos propensos al uso de las TIC que los usuarios de regiones de mayor nivel de renta.

Algo parecido ocurre con el resto de determinantes considerados tales como la Capacidad Productiva, el Nivel de Estudios, la Edad y la Población, que al reemplazar los valores observados por sus valores medios provocan reducciones en la brecha digital de 12,82%, 16,12%, 14,82% y 12,02% respectivamente, en todos los casos los mayores impactos por encima de la media serán Madrid y Cataluña, mientras que los más bajos serán Extremadura y Castilla-La Mancha. Sin embargo los efectos serán distintos y variarán de acuerdo a la Comunidad considerada. A continuación se observa el Gráfico 15:

**Gráfico 15:** Sensibilidad del Nivel de Digitalización Respecto a sus Determinantes



Lo primero que se ve es que las CCAA más sensibles a los determinantes serán en primer lugar Madrid, que a su vez es la Comunidad más digitalizada de todas y donde los porcentajes de variación son mayores. Curiosamente le siguen en sensibilidad las CCAA menos digitalizadas como Extremadura y Castilla-La Mancha, mientras que Cataluña (con un nivel de digitalización también elevado) viene seguidamente en el cuarto puesto. En el otro extremo se encuentran CCAA que a pesar de tener un nivel de digitalización medio, presentan un bajo porcentaje de variación, es decir un bajo nivel de sensibilidad a los determinantes. Estas son Valencia, La Rioja y Cantabria. Esto se da porque tienen valores de las variables explicativas más próximos a la media de todas las CCAA.

Además, se observa por ejemplo que la Renta será el determinante principal para reducir la brecha tecnológica sólo en Canarias. En las otras CCAA dicho impacto, aunque sea también importante, no será el del principal impacto. En Madrid por ejemplo, el determinante principal lo será el nivel de estudios alcanzados, lo mismo que en Castilla y León y Cataluña. La capacidad productiva lo será en Andalucía, Ceuta y

Melilla. En las Comunidades más desfavorecidas en digitalización como Extremadura y Castilla-La Mancha, el principal impacto lo ejerce la población, que parece ser responsable de gran parte de la brecha digital.

Entonces se puede concluir que

- Con respecto a la media de la Comunidad Europea, las políticas de reducción de la brecha digital en CCAA con mayor nivel de digitalización tales como Madrid, País Vasco o Cataluña no deberían ser las mismas que en otras comunidades con un menor nivel como es el caso de Extremadura, Galicia o Castilla-La Mancha, ya que hemos visto que la sensibilidad del nivel de digitalización a sus determinantes en las distintas CCAA no son similares. Esto sugiere un estudio detallado a la hora de implementar una política de reducción de la brecha digital teniendo en cuenta otros factores tales como el nivel de estudios o la edad media de la población, el uso cotidiano de algún tipo específico de equipamiento tales como ordenadores, etc, además de sólo promover el acceso y uso de las TIC.
- Vemos que el grado de digitalización será muy sensible en comunidades poco digitalizadas. Esto sugiere que una política de fomento de adquisición y uso de equipamientos podría ser adecuada. La liberalización de los mercados de telecomunicaciones ha estimulado nuevas demandas de equipamiento en los hogares y una política de reducción de la brecha digital podría apuntar a este aspecto.
- La renta es el atributo más importante de los considerados debido al impacto que provoca en el nivel de digitalización y la forma en que reduce la brecha digital.

- Una política de reducción de la brecha digital entre CCAA podría considerar favorecer el acceso a Internet en todas sus formas (ADSL o cable módem) incluyéndolo en el Servicio Universal.
- El nivel de estudios afecta a la digitalización de manera dispar. El nivel de digitalización será más sensible a este factor en Comunidades más industrializadas. Esto podría estar correlacionado con la necesidad de capital humano y alto grado de especialización para determinadas tareas.

#### **4.5. RESUMEN DE LOS RESULTADOS**

Los resultados en general sugieren:

- I. La renta, la capacidad productiva, la edad, el hábitat poblacional y el nivel de estudios son buenos predictores de la brecha tecnológica, y su importancia para el modelo será grande tanto de manera lineal como interactuando entre sí o con otras variables.
- II. RETINA mejora la capacidad predictiva al incluir no linealidades y pendientes específicas.
- III. Los impactos de los factores que provocan la brecha tecnológica entre las CCAA de España no serán de la misma magnitud según se hable de un grupo con buena predisposición a la digitalización o de un grupo con menor predisposición. En el análisis de sensibilidad se ha visto cómo los atributos que influyen en la digitalización lo harán de manera distinta según se trate de una Comunidad con menor o mayor nivel de digitalización. Esto debería ser tenido en cuenta en la formulación de políticas de reducción de la brecha digital tales como determinadas campañas de precios o descuentos.

IV. Podría pensarse de que el problema “coyuntural” de la existencia de una brecha digital es el resultado de la existencia de otras brechas “estructurales” tales como brecha de renta, brecha poblacional, brecha de edad o brecha de nivel de estudios alcanzados. Esta idea anima a futuras investigaciones orientadas en este aspecto.

## **5. CONCLUSIONES**

La importancia de la brecha digital entre países y regiones contrasta con el relativamente escaso esfuerzo dedicado a su estudio. Este capítulo se centra en el análisis y la medición de la brecha tecnológica dentro de las Comunidades Autónomas de España. El planteamiento del problema de cómo medir adecuadamente dicha brecha se comienza intentando responder las siguientes preguntas:

- a. ¿Qué es la Brecha Digital?
- b. ¿Cómo se puede medir?
- c. ¿Cuáles son sus causas más relevantes?
- d. ¿Cuáles son algunos de sus efectos?
- e. ¿Cómo se la puede disminuir?

A partir de estas preguntas se propuso un indicador del nivel de digitalización y otro del consumo en tecnología. A partir de las dos formas que toma el indicador, se comparan los resultados para las CCAA considerando que un valor de brecha tecnológica “total” puede ser medida a partir de una medida tal como el coeficiente de variación, por ser esta una medida de dispersión relativa. Una vez calculado este

coeficiente para toda la muestra, la brecha por CCAA se mide a partir de las distancias del valor normalizado del índice con la media total.

Una vez calculado el índice propuesto para cada Comunidad y demostradas las diferencias existentes en los niveles de digitalización entre CCAA, y siempre teniendo por finalidad explicar los determinantes de la existencia de estas diferencias, el siguiente paso será agrupar las CCAA para recoger la heterogeneidad existente en la muestra. Para ello se estima un modelo econométrico de Poisson para el uso de Internet estimado con datos individuales para cada CCAA. Los resultados sugieren que los determinantes que provocan mayor variación porcentual son el gasto en equipamiento, la edad y el nivel de educación, aunque no con el mismo impacto para todas las Comunidades. A partir de aquí se formuló una pregunta adicional:

- f. ¿Las diferencias entre las elasticidades se deben a diferencias en el comportamiento entre CCAA, o las brechas observadas se pueden explicar por diferencias en el nivel de renta y otros factores demográficos?

Tener en cuenta las diferencias entre las elasticidades es importante debido a que equivale a decir que existen distintos impactos porcentuales en la variable a explicar por parte de los regresores. Entonces, para clasificar a las CCAA de acuerdo a su similitud se utiliza un análisis de conglomerados donde se distinguen tres grupos bien diferenciados: el de menor predisposición a la digitalización (Grupo 1), el de predisposición media (Grupo 2) y el de predisposición mayor (Grupo 3).

Una vez obtenidos dichos grupos, a continuación se los tiene en cuenta en la modelización del nivel de digitalización considerando el elevado número de observaciones anómalas utilizando primero una regresión cuantílica y posteriormente

regresiones MCO en donde se especifica un modelo básico que se mejora sustancialmente en habilidad predictiva fuera de la muestra a partir de la inclusión de interacciones y no linealidades sugeridas por el algoritmo RETINA. A partir de él se estiman las diferencias que existen en los impactos de los atributos para los distintos grupos. Se observa claramente que no serán de la misma magnitud según se trate de uno o de otro.

En futuras investigaciones sería recomendable analizar las interacciones recomendadas por RETINA para la elaboración de un fundamento de teoría económica en la modelización de la brecha tecnológica, tanto para España como a nivel más general. Debido a que es un fenómeno que evoluciona rápidamente, sería interesante contar con paneles de datos para estudiar la dinámica del grado de digitalización. En particular, el impacto del posible efecto de las políticas públicas, como se ensaya en el apartado 4.5. Finalmente, estos resultados animan a utilizar la metodología propuesta para otros contextos (a otro nivel de agregación o para otros servicios).

## **6. REFERENCIAS**

Cassel C (1999) “Demand for and Use of Additional Lines by Residential Customers”. En: Loomis, Taylor (eds.) *The Future of the Telecommunications Industry: Forecasting and Demand Analysis*. Kluwer Academic Publisher, Boston

Cerno, L. y Pérez-Amaral, T. (2005) “Demand of Internet Access and Use in Spain” mimeo presentado en la ITS Conference on Regional Economic Development, Pontevedra, España.

Corrocher, N. y Ordanini, A. (2002) “Measuring the digital divide: a framework for the analysis of cross-country differences” *Journal of Information Technology*, 17, 9-19.

Davidson R. y MacKinnon J. (2003) “Econometric Theory and Methods”. Oxford University Press.

Fundación AUNA (2005) “Informe España 2005” en [http://www.fundacionauna.com/areas/25\\_publicaciones/](http://www.fundacionauna.com/areas/25_publicaciones/)

Koenker, R. y Basset, G. (1978) “Regression Quantiles” *Econometrica*, vol 46(1), pp. 33-50.

Lebart, L., Morineau, A. y Warwick, K. (1984) “Multivariate Descriptive Statistical Analysis”, Wiley.

Madden G, Savage S y Simpson M (1996) “Information Inequality and Broadband Network Access: An Analysis of Australian Household Survey Data”. *Industry and Corporate Change*, Oxford University Press, pp 1049–1056.

Mansell, R. y When, U. (1998) “Knowledge Societies Information Technology for Suitable Development” Oxford University Press, Oxford and New York.

Marinucci, M. (2005) "RETINA Winpack for Real Data: A quick guide for Automatic Model Selection" Universidad Complutense de Madrid, Somosaguas, Madrid.

Marinucci, M. y Pérez-Amaral, T. (2005) "Econometric Modeling of Business Telecommunications Demand Using RETINA and Finite Mixtures" mimeo presentado en la ITS Conference on Regional Economic Development, Pontevedra, España.

Martin, W. (1995) "The Global Information Society" Bowker Saur, London.

Martins, P. y Pereira, P. (2004) "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries" Labour Economics 11, pp. 355-371.

OECD Report (2001) "Understanding the Digital Divide" en [www.oecd.org/dataoecd/38/57/](http://www.oecd.org/dataoecd/38/57/)

Pelletiere, D. y Rodrigo, G. (2000) "An Empirical Investigation of the Digital Divide in the United States" mimeo presentado en la Western Economics Association annual conference.

Peña, D. (2002) "Análisis Multivariante de Datos" Mc Graw Hill Eds., Madrid.

Pérez-Amaral, T., Gallo, G. y White, H. (2003) “A Flexible Tool for Model Building: The Relevant Transformations of the Inputs Network Approach (RETINA)” *Oxford Bulletin of Economics and Statistics* 65, supplement 1, 821-838.

Pérez-Amaral, T., Gallo, G. y White, H. (2005) “A Comparison of Complementary Automatic Modeling Methods: RETINA and PcGets” *Econometric Theory*, 21, 262-277.

Randstad Report (2004) “Quality at Work in the EU-15: The Knowledge Society” en [www.eurofound.eu.int/ewco/reports/](http://www.eurofound.eu.int/ewco/reports/)

Rappoport P, Taylor L y Kridel D (2002) “The Demand of Broadband: Access, Content, and the Value of Time”. En: Crandall RW, Alleman JH (eds.) *Broadband: Should We Regulate High-Speed Internet Access?* AEI-Brookings Joint Centre for Regulatory Studies, Washington, D.C.

Red.es – Ministerio de Industria, Turismo y Comercio (2004) “Las TIC en los Hogares Españoles” en <http://observatorio.red.es/estudios/consumo/index.html>

Ricci, A. (2000) “Measuring information society dynamics of European data on usage of information and communication technologies in Europe since 1995” *Telematics and Informatics*, 17, 141-67.

Sneath, P. y Sokal, R. (1973) “Numerical Taxonomy” W.H. Freeman, San Francisco.

Taylor LD (1994) "Telecommunications Demand in Theory and Practice". Kluwer Academic Publishers.

The World Bank (2005) "Financing Information and Communication Infrastructure Needs in the Developing World: Public and Private Roles" Draft for Discussion in <http://lnweb18.worldbank.org/ict/resources.nsf/>

White, H. (1980b) "A Heteroscedasticity-Consistent Covariance Matrix Estimator and Direct Test of Heteroscedasticity" *Econometrica*, 48, pp. 817-838.

White, H. (1989) "Learning in Artificial Neural Network: A Statistical Perspective" reimpresso en White, H. (1992) "Artificial Neural Networks: Approximation and Learning Theory" pp. 90-131. Blackwell.

**APÉNDICE 1: Definición de variables**

<i>Variable</i>	<i>Definición</i>
<i>USE</i>	Cantidad de sitios donde el individuo usa Internet (hasta 4)
(*) <i>IDD</i>	Índice de digitalización
(*) <i>ITIC</i>	Consumo Individual en tecnología
<i>HOUSINT</i>	Ficticia =1 si el individuo accede a Internet en el hogar
<i>EQ</i>	Cantidad de equipamiento informático del individuo en el hogar
<i>IRPC</i>	Índice de Renta per Capita a nivel provincial
<i>CPRO</i>	Capacidad Productiva a nivel provincial (porcentaje de aporte provincial al PBI nacional)
<i>USOR</i>	Intensidad de uso del ordenador (cantidad de usos trimestrales igual a 70, 14, 3, 1, 0)
<i>USOIN</i>	Intensidad de uso de Internet (cantidad de conexiones trimestrales igual a 70, 14, 3, 1, 0)
<i>MALE</i>	Ficticia =1 si el individuo es de sexo masculino
<i>AGE</i>	Edad del individuo
<i>POBLAC</i>	Población provincial
<i>STUDYLEVEL</i>	Grado alcanzado en nivel de estudios (en años)
<i>BESTUDYING</i>	Ficticia =1 si el individuo aún se encuentra estudiando
<i>SIZE1</i>	Ficticia =1 si es de una capital de provincia con más de 500.000 habitantes
<i>SIZE2</i>	Ficticia =1 si es de una capital de provincia
<i>SIZE3</i>	Ficticia =1 si es de una municipio con más de 100.000 habitantes
<i>SIZE4</i>	Ficticia =1 si es de un municipio entre 50.000 y 100.000 habitantes
<i>SIZE5</i>	Ficticia =1 si es de un municipio entre 20.000 y 50.000 habitantes
<i>SIZE6</i>	Ficticia =1 si es de un municipio entre 10.000 y 20.000 habitantes
<i>SIZE7</i>	Ficticia =1 si es de un municipio con menos de 10.000 habitantes
<i>COM1</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Andalucía
<i>COM2</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Aragón
<i>COM3</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Asturias
<i>COM4</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Baleares
<i>COM5</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Canarias
<i>COM6</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Cantabria
<i>COM7</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Castilla y León
<i>COM8</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Castilla – La Mancha
<i>COM9</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Cataluña
<i>COM10</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Valencia
<i>COM11</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Extremadura
<i>COM12</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Galicia
<i>COM13</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Madrid
<i>COM14</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Murcia
<i>COM15</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de Navarra
<i>COM16</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de País Vasco

<i>COM17</i>	Ficticia =1 si pertenece a la Comunidad Autónoma de La Rioja
<i>COM18</i>	Ficticia =1 si pertenece a las Comunidades de Ceuta o Melilla
<i>WORK1</i>	Ficticia =1 si se encuentra trabajando por cuenta propia
<i>WORK2</i>	Ficticia =1 si se encuentra trabajando por cuenta ajena
<i>WORK3</i>	Ficticia =1 si se encuentra parado
<i>WORK4</i>	Ficticia =1 si es estudiante
<i>WORK5</i>	Ficticia =1 si se dedica a las labores del hogar
<i>WORK6</i>	Ficticia =1 si es pensionista
<i>CLUSTER1</i>	Ficticia =1 si pertenece a las Comunidades Autónomas de Madrid, Baleares, Canarias, País Vasco, Cataluña, Ceuta o Melilla.
<i>CLUSTER2</i>	Ficticia =1 si pertenece a las Comunidades Autónomas de Aragón, Asturias, Valencia, Castilla-La Mancha o Murcia.
<i>CLUSTER3</i>	Ficticia =1 si pertenece a las Comunidades Autónomas de Cantabria, Andalucía, Navarra, Castilla y León, Extremadura, Galicia o La Rioja.

---

(\*) Nota: La construcción de estas dos variables fue explicada en el apartado 2 de este capítulo.

## APÉNDICE 2: Resultados de la regresión de Poisson por CA

En el siguiente apéndice se detallan las regresiones de Poisson para el uso de Internet en cada CA. Todas están referidas a la especificación (3.2) del apartado 3. Como se explica allí, la variable dependiente se refiere a la cantidad de sitios de uso de Internet, y las variables explicativas están referidas al consumo en tecnología, la demanda de servicios complementarios y atributos sociodemográficos. En la columna  $\hat{\beta}$  se encuentran los coeficientes calculados (semielasticidades), en la columna  $\frac{dy}{dx}$  los efectos marginales, y en la  $\varepsilon$  las elasticidades, donde parte de las cuales se exponen en la Tabla 5 y el Gráfico 7. En todos los casos se exponen los ratios  $t$  entre paréntesis. Las observaciones varían de acuerdo a la CA que se considere, siendo la muestra proporcional al tamaño poblacional de cada una. La sigla LR es el ratio de verosimilitudes igual a  $2\ln(L(H_1)-L(H_0))$  que se distribuye asintóticamente como una  $\chi^2$ , con un número de grados de libertad igual a la diferencia entre los valores hipotéticos de los parámetros en  $H_0$  y  $H_1$ . El pseudo  $R^2$  es una medida global de ajuste para este tipo de modelos igual a  $1 - \frac{L(\hat{\beta})}{L(\hat{\beta}_0)}$ , donde el numerador es la función de verosimilitud en el punto máximo para el modelo con los parámetros estimados, y el denominador es la misma función de verosimilitud pero que sólo incluye al término constante. Tal como se podrá observar a continuación, todas las regresiones tienen buenas medidas de bondad de ajuste y serán globalmente significativas. Como se explica en el apartado 3, aunque la mayoría de los regresores son significativos, sólo se han utilizado los tres que resultan significativos en todas las regresiones realizadas (*ITIC*, *AGE* y *STUDYLEVEL*).

Regresores	Andalucía			Aragón		
	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$
Constante	-3.776 (8.58)	---	---	-4.263 (5.73)	---	---
<i>ITIC</i>	1.189 (5.07)	.144 (5.44)	1.905 (5.07)	1.655 (4.28)	.218 (4.65)	2.770 (4.28)
<i>HOUSINT</i>	.515 (3.54)	.074 (2.90)	.105 (3.54)	.620 (2.91)	.096 (2.37)	.162 (2.91)
<i>EQ</i>	.152 (1.79)	.028 (1.76)	.035 (1.79)	-.109 (.81)	-.014 (.81)	-.032 (.81)
<i>MALE</i>	.207 (2.83)	.026 (2.74)	.087 (2.83)	.255 (2.58)	.034 (2.50)	.115 (2.58)
<i>AGE</i>	-.038 (11.44)	-.005 (12.13)	-1.803 (11.44)	-.039 (8.71)	-.005 (8.75)	-2.017 (8.71)
<i>HABITAT</i>	.017 (1.06)	.002 (1.06)	.077 (1.06)	-.212 (.67)	-.003 (.67)	-.129 (0.67)
<i>STUDYLEVEL</i>	2.306 (13.41)	.279 (11.57)	1.211 (13.41)	2.332 (9.61)	.306 (8.12)	1.342 (9.61)
<i>BESTUDYING</i>	.389 (4.63)	.055 (3.74)	.048 (4.63)	.211 (1.75)	.030 (1.54)	.022 (1.75)
<b>Observaciones</b>	2147			996		
<b>LR</b>	1,280.74 (p-valor=0.00)			662.03 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3724			.3809		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Asturias			Balears		
	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$
Constante	-3.828 (5.18)	---	---	-3.253 (4.26)	---	---
<i>ITIC</i>	1.459 (3.83)	.201 (4.12)	2.377 (3.83)	1.002 (2.44)	.168 (2.57)	1.625 (2.44)
<i>HOUSINT</i>	.491 (2.00)	.079 (1.67)	.098 (2.00)	.776 (2.79)	.157 (2.26)	.230 (2.79)
<i>EQ</i>	.202 (1.23)	.028 (1.22)	.044 (1.23)	.073 (.44)	.012 (.43)	.024 (.44)
<i>MALE</i>	.062 (.55)	.008 (.55)	.024 (.55)	.274 (2.14)	.047 (2.08)	.124 (2.14)
<i>AGE</i>	-.034 (7.62)	-.005 (7.79)	-1.799 (7.62)	-.034 (6.13)	-.006 (6.40)	-1.633 (6.13)
<i>HABITAT</i>	-.006 (.18)	-.0008 (.18)	-.019 (.18)	-.013 (.42)	-.002 (.42)	-.043 (.42)
<i>STUDYLEVEL</i>	1.919 (7.36)	.264 (6.69)	1.096 (7.36)	2.003 (6.75)	.336 (6.16)	1.113 (6.75)
<i>BESTUDYING</i>	.230 (1.74)	.035 (1.54)	.025 (1.74)	.234 (1.61)	.043 (1.43)	.029 (1.61)
<b>Observaciones</b>	893			579		
<b>LR</b>	536.80 (p-valor=0.00)			378.19 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3628			.3598		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Canarias			Cantabria		
	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$
Constante	-3.077 (4.93)	---	---	-3.350 (3.61)	---	---
<i>ITIC</i>	1.135 (3.38)	.273 (3.57)	1.930 (3.38)	.793 (1.70)	.094 (1.76)	1.30 (1.70)
<i>HOUSINT</i>	.365 (1.68)	.095 (1.52)	.107 (1.68)	1.139 (3.80)	.196 (2.56)	.248 (3.80)
<i>EQ</i>	.151 (1.02)	.037 (1.02)	.049 (1.02)	-.069 (.39)	-.008 (.40)	-.017 (.39)
<i>MALE</i>	.267 (2.63)	.065 (2.56)	.121 (2.63)	-.342 (2.30)	.042 (2.17)	.141 (2.30)
<i>AGE</i>	-.037 (8.25)	-.009 (9.21)	-1.646 (8.25)	-.034 (5.15)	-.004 (5.48)	-1.75 (5.15)
<i>HABITAT</i>	-.024 (.83)	-.006 (.83)	-.075 (.83)	-.046 (1.35)	-.006 (1.33)	-.164 (1.35)
<i>STUDYLEVEL</i>	1.928 (7.92)	.465 (7.78)	1.010 (7.92)	2.537 (7.00)	.300 (5.99)	1.429 (7.00)
<i>BESTUDYING</i>	.429 (3.80)	.120 (3.13)	.072 (3.80)	.267 (1.45)	.035 (1.26)	.025 (1.45)
<b>Observaciones</b>	787			544		
<b>LR</b>	513.98 (p-valor=0.00)			366.43 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3257			.4131		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Castilla - León			Castilla - La Mancha		
	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$
Constante	-2.994 (5.16)	---	---	-3.951 (4.90)	---	---
<i>ITIC</i>	.845 (2.82)	.102 (3.00)	1.350 (2.82)	1.550 (3.86)	.125 (4.24)	2.417 (3.86)
<i>HOUSINT</i>	.773 (4.11)	.120 (3.05)	.164 (4.11)	.695 (2.08)	.073 (1.56)	.100 (2.08)
<i>EQ</i>	.150 (1.33)	.018 (1.31)	.036 (1.33)	-.058 (.20)	-.005 (.20)	-.009 (.20)
<i>MALE</i>	.336 (3.67)	.041 (3.46)	.152 (3.67)	.038 (.33)	.003 (.33)	.017 (.33)
<i>AGE</i>	-.036 (8.70)	-.004 (8.97)	-1.877 (8.70)	-.041 (7.53)	-.003 (7.14)	-2.128 (7.53)
<i>HABITAT</i>	-.022 (1.01)	-.003 (1.01)	-.791 (1.01)	-.008 (.26)	-.0006 (.26)	-.034 (.26)
<i>STUDYLEVEL</i>	2.002 (9.02)	.243 (7.79)	1.098 (9.02)	2.093 (7.71)	.170 (6.17)	1.017 (7.71)
<i>BESTUDYING</i>	.320 (2.80)	.044 (2.34)	.040 (2.80)	.497 (3.52)	.049 (2.58)	.053 (3.52)
<b>Observaciones</b>	1304			1012		
<b>LR</b>	916.90 (p-valor=0.00)			590.69 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.4145			.4110		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Cataluña			Valencia		
	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$
Constante	-2.894 (6.23)	---	---	-4.347 (7.34)	---	---
<i>ITIC</i>	1.081 (4.53)	.243 (4.74)	1.832 (4.53)	1.494 (4.76)	.176 (5.26)	2.420 (4.76)
<i>HOUSINT</i>	.733 (5.27)	.1941 (4.34)	.223 (5.27)	.703 (3.82)	.103 (2.92)	.157 (3.82)
<i>EQ</i>	.073 (0.92)	.016 (.91)	.025 (.92)	.050 (.44)	.006 (.43)	.013 (.44)
<i>MALE</i>	.103 (1.57)	.024 (1.55)	.042 (1.57)	.233 (2.68)	.028 (2.58)	.100 (2.68)
<i>AGE</i>	-.319 (11.29)	-.007 (11.81)	-1.58 (11.29)	-.036 (8.97)	-.004 (9.13)	-1.763 (8.97)
<i>HABITAT</i>	.005 (.32)	.001 (.32)	.029 (.32)	.015 (.70)	.001 (.70)	.071 (.70)
<i>STUDYLEVEL</i>	1.42 (9.37)	.320 (9.34)	.805 (9.37)	2.177 (10.31)	.257 (8.38)	1.172 (10.31)
<i>BESTUDYING</i>	.277 (3.43)	.070 (2.99)	.028 (3.43)	.331 (3.27)	.045 (2.71)	.038 (3.27)
<b>Observaciones</b>	1833			1377		
<b>LR</b>	1223.31 (p-valor=0.00)			962.93 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3339			.4141		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Extremadura			Galicia		
	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\varepsilon$
Constante	-3.17 (3.91)	---	---	-3.662 (5.74)	---	---
<i>ITIC</i>	.996 (2.36)	.084 (2.52)	1.553 (2.36)	1.361 (4.09)	.122 (4.41)	2.126 (4.41)
<i>HOUSINT</i>	.743 (1.92)	.083 (1.44)	.109 (1.92)	.909 (3.92)	.114 (2.67)	.150 (3.92)
<i>EQ</i>	.054 (.18)	.005 (.18)	.009 (.18)	-.203 (1.22)	-.018 (1.21)	-.388 (1.21)
<i>MALE</i>	.170 (1.21)	.015 (1.18)	.071 (1.21)	.099 (.97)	.009 (.95)	.042 (.95)
<i>AGE</i>	-.043 (6.79)	-.004 (7.07)	-2.174 (6.79)	-.044 (9.45)	-.004 (9.11)	-2.260 (9.11)
<i>HABITAT</i>	.022 (.66)	.002 (.66)	.103 (.66)	-.020 (.72)	-.002 (.72)	-.084 (.72)
<i>STUDYLEVEL</i>	1.976 (6.34)	.166 (5.36)	.982 (6.34)	2.428 (10.02)	.217 (7.58)	1.285 (10.02)
<i>BESTUDYING</i>	.492 (3.02)	.051 (2.22)	.046 (3.02)	.240 (1.93)	.024 (1.66)	.026 (1.66)
<b>Observaciones</b>	831			1171		
<b>LR</b>	385.54 (p-valor=0.00)			780.07 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3557			.4252		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Madrid			Murcia		
	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$
Constante	-3.748 (7.89)	---	---	-4.515 (5.75)	---	---
<i>ITIC</i>	1.221 (4.89)	.337 (5.20)	2.114 (4.89)	1.899 (4.60)	.163 (5.08)	2.97 (4.60)
<i>HOUSINT</i>	.568 (4.44)	.176 (3.82)	.181 (4.44)	.563 (2.09)	.058 (1.66)	.113 (2.09)
<i>EQ</i>	.086 (1.36)	.024 (1.35)	.034 (1.36)	.0145 (.08)	.001 (.08)	.003 (.08)
<i>MALE</i>	.152 (2.31)	.042 (2.28)	.067 (2.31)	.342 (2.66)	.030 (2.46)	.149 (2.66)
<i>AGE</i>	-.030 (10.99)	-.008 (11.83)	-1.418 (10.99)	-.042 (7.38)	-.004 (7.03)	-2.117 (7.38)
<i>HABITAT</i>	.033 (1.99)	.009 (1.98)	.172 (1.99)	-.010 (.24)	-.001 (.24)	-.028 (.24)
<i>STUDYLEVEL</i>	1.956 (11.86)	.540 (11.83)	1.238 (11.86)	1.733 (5.79)	.148 (4.92)	.904 (5.79)
<i>BESTUDYING</i>	.277 (3.77)	.084 (3.35)	.048 (3.77)	.420 (2.86)	.042 (2.19)	.055 (2.86)
<b>Observaciones</b>	1567			773		
<b>LR</b>	1135.60 (p-valor=0.00)			491.66 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3360			.4092		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	Navarra			País Vasco		
	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$	$\hat{\beta}$	$dy/dx$ (*)	$\epsilon$
Constante	-3.015 (5.36)	---	---	-4.029 (5.38)	---	---
<i>ITIC</i>	.892 (3.06)	.149 (3.21)	1.488 (3.06)	1.120 (2.78)	.141 (2.93)	1.889 (2.78)
<i>HOUSINT</i>	.423 (2.09)	.079 (1.84)	.111 (2.09)	1.110 (3.88)	.183 (2.87)	.335 (3.88)
<i>EQ</i>	.220 (1.53)	.037 (1.52)	.062 (1.53)	.190 (.85)	.024 (.84)	.060 (.85)
<i>MALE</i>	.258 (2.83)	.044 (2.75)	.121 (2.83)	.257 (2.73)	.033 (2.69)	.121 (2.73)
<i>AGE</i>	-.035 (9.30)	-.006 (9.95)	-1.731 (9.30)	-.029 (6.84)	-.004 (6.97)	-1.504 (6.84)
<i>HABITAT</i>	-.040 (2.00)	-.007 (1.98)	-.164 (2.00)	-.003 (.10)	-.0003 (.10)	-.011 (.10)
<i>STUDYLEVEL</i>	2.174 (10.32)	.363 (9.42)	1.271 (10.32)	1.949 (8.48)	.245 (7.32)	1.051 (8.48)
<i>BESTUDYING</i>	.482 (5.36)	.097 (3.66)	.067 (5.36)	.201 (1.46)	.245 (7.32)	.015 (1.46)
<b>Observaciones</b>	1202			1181		
<b>LR</b>	747.24 (p-valor=0.00)			808.83 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.3539			.4043		

(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

Regresores	La Rioja			Ceuta y Melilla		
	$\hat{\beta}$	$dy/dx^{(*)}$	$\varepsilon$	$\hat{\beta}$	$dy/dx^{(*)}$	$\varepsilon$
Constante	-2.870 (3.06)	---	---	-3.624 (3.07)	---	---
<i>ITIC</i>	.879 (1.82)	.108 (1.92)	1.439 (1.82)	1.296 (2.03)	.283 (2.23)	2.158 (2.03)
<i>HOUSINT</i>	.947 (2.84)	.159 (1.97)	.204 (2.84)	.591 (1.18)	.147 (1.00)	.182 (1.18)
<i>EQ</i>	-.097 (.39)	-.012 (.39)	-.023 (.39)	.320 (1.03)	.070 (1.02)	.110 (1.03)
<i>MALE</i>	.263 (1.95)	.033 (1.89)	.119 (1.95)	.041 (.21)	.009 (.21)	.019 (.21)
<i>AGE</i>	-.042 (6.94)	-.005 (6.73)	-2.124 (6.94)	-.033 (3.41)	-.007 (3.62)	-1.412 (3.41)
<i>HABITAT<sup>(+)</sup></i>	-.063 (1.85)	-.008 (1.85)	-.209 (1.85)	---	---	---
<i>STUDYLEVEL</i>	2.286 (7.02)	.282 (5.97)	1.298 (7.02)	1.601 (3.66)	.351 (3.39)	.940 (3.66)
<i>BESTUDYING</i>	.538 (3.70)	.082 (2.77)	.072 (3.70)	.580 (2.40)	.154 (1.83)	.106 (2.40)
<b>Observaciones</b>	549			203		
<b>LR</b>	398.89 (p-valor=0.00)			149.54 (p-valor=0.00)		
<b>Pseudo R<sup>2</sup></b>	.4109			.3681		

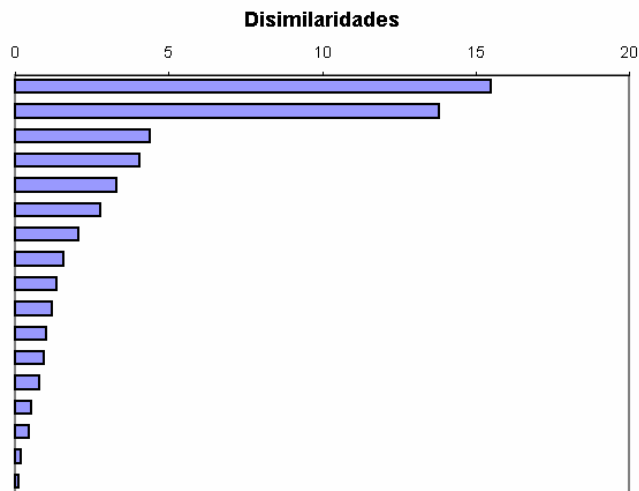
(\*) Para el caso de las variables ficticias, el efecto marginal representa el cambio discreto de 0 a 1.

(+) Esta variable no tiene variabilidad en Ceuta y Melilla. De considerarla en la regresión provocaría multicolinealidad exacta.

**APÉNDICE 3: Conglomerados Jerárquicos de CCAA para Elasticidades de Uso de Internet con Respecto al Gasto en Equipamiento, la Edad y el Nivel de Estudios**

Nodo	Ramificación a la izquierda	Ramificación a la derecha	Frecuencia	Peso	Nivel
35	33	34	18	18.000	15.498
34	32	30	11	11.000	13.793
33	27	31	7	7.000	4.377
32	29	24	5	5.000	4.060
31	25	26	4	4.000	3.292
30	28	23	6	6.000	2.792
29	2	19	3	3.000	2.058
28	13	21	4	4.000	1.568
27	6	22	3	3.000	1.340
26	12	17	2	2.000	1.215
25	7	11	2	2.000	1.006
24	8	14	2	2.000	0.934
23	9	18	2	2.000	0.793
22	1	15	2	2.000	0.516
21	4	20	3	3.000	0.450
20	5	16	2	2.000	0.186
19	3	10	2	2.000	0.123

**Histograma de los Niveles de los Nodos**



En este histograma es obvio el punto de corte en tres grupos debido a la clara diferencia de disimilaridades que se observa entre la segunda y la tercera barra.

### APÉNDICE 4: Resultados de las Estimaciones RETINA

Este apéndice se corresponde con los resultados expuestos en la Tabla 10 del apartado 4.4. La variable dependiente son las observaciones individuales en logaritmos del índice IDD.

Variable	Coefficiente
Constante	2,11 (10,45)
Interacciones	
$\ln(STUDYLEVEL) * \ln(EDAD)$	0,12 (10,31)
$1/\ln(EDAD)^2$	-10,79 (13,24)
$\ln IRPCPRO * \ln USOIN$	0,02 (14,45)
Pendientes específicas	
$CLUSTER1 * \ln(EQ) * \ln(STUDYLEVEL)$	-0,15 (4,13)
$CLUSTER1 * [1/\ln(EDAD)^2]$	-0,15 (1,83)
$CLUSTER2 * \ln(AGE)^2$	-0,04 (6,39)
$CLUSTER2 * \ln(EQ) * \ln(STUDYLEVEL)$	-0,16 (4,30)
$CLUSTER3 * \ln(EQ) * \ln(STUDYLEVEL)$	-0,17 (4,64)
$CLUSTER3 * \ln(STUDYLEVEL)^2$	-0,03 (2,21)
$CLUSTER3 * [\ln(STUDYLEVEL)/\ln(EDAD)]$	0,17 (1,98)
$CLUSTER3 * \ln(STUDYLEVEL) * \ln(AGE)$	-0,006 (1,01)
$HOUSINT * \ln(STUDYLEVEL)^2$	0,06 (2,44)
$WORK1 * \ln(EQ)$	-0,02 (1,70)
$WORK2 * \ln(EQ)$	-0,20 (8,71)
$WORK2 * \ln(AGE)$	0,25 (0,45)
$WORK2 * \ln(STUDYLEVEL) * \ln(EDAD)$	-0,01 (0,77)
$WORK2 * \ln(EDAD)^2$	-0,05 (0,52)
$WORK3 * \ln(EQ)$	0,06 (3,01)
$WORK3 * \ln(STUDYLEVEL) * \ln(EDAD)$	0,009 (0,94)
$WORK3 * \ln(STUDYLEVEL)^2$	0,02 (1,01)
$WORK3 * \ln(EDAD)^2$	-0,02 (1,02)
$WORK4 * \ln(EQ)$	1,37 (2,88)
$WORK4 * \ln(STUDYLEVEL) * \ln(EDAD)$	-0,02 (5,03)

$WORK4 * \ln(EDAD)^2$	-0,03 (7,81)
$WORK4 * [\ln(EQ)/\ln(EDAD)]$	-4,84 (2,43)
$SIZE6 * \ln(STUDYLEVEL) * \ln(AGE)$	-0,004 (1,46)
$SIZE6 * \ln(EDAD)^2$	-0,005 (7,38)
Observaciones	18673
R <sup>2</sup> ajustado	0,704
Error estándar de la estimación	0,257
RCMSPE	0257
AIC	-2,718

# CAPÍTULO IV

## Comercio Electrónico en España

## 1. INTRODUCCIÓN

Tal como se viene sosteniendo en los capítulos precedentes, durante los últimos 10 años se ha considerado a Internet como un nuevo canal de comunicación y entretenimiento. Además, Internet está adquiriendo importancia como canal de distribución. Su estudio contribuirá a entender los mercados virtuales y responder a cuestiones tales como que si puede la competencia en los mercados virtuales conducir a precios más homogéneos, o qué factores influyen en la formación de los precios en la Red.

Dentro de las principales características que diferencian a los mercados virtuales de los mercados convencionales destaca el hecho de que habrían menos fricciones. A su vez, una de las formas de crear poder de mercado en el caso del comercio virtual es reducir lo que se denomina *costes de búsqueda*, pues Internet ha incrementado sustancialmente la disponibilidad de la información de precios y productos, permitiendo a los consumidores identificar la mejor opción y mejorar su posición frente a los vendedores on-line. Es decir que los consumidores actuales han ido mejorando y fortaleciendo su posición a medida que el Comercio electrónico se ha desarrollado. En la literatura existente se puede destacar a Clemons, Hann y Hitt (1999), Combers y Patel (1997) y Brynjolfsson y Smith (1999) que analizan la conducta del consumidor a partir de la aparición del comercio electrónico.

Algunos de estos autores sostienen que la única característica destacable es la cercanía de este mercado a la competencia perfecta. En la versión extrema se podría decir que Internet es un mercado en donde la localización del vendedor es irrelevante, y donde los consumidores están muy informados de los precios y de los productos ofrecidos, y los vendedores están atomizados con beneficios iguales a cero. Aunque al

mismo tiempo, hay evidencias claras de que Internet no es un mercado completamente eficiente (Brynjolfsson y Smith, 1999).

Los mercados online parecen más eficientes en términos de precios y de costes. Algunos estudios han encontrado una dispersión de precios apreciable (Bakos, 2001). Esta dispersión puede ser explicada, en parte, por la heterogeneidad de factores específicos tales como la confianza o el conocimiento del sitio Web o de la marca.

El análisis del nivel de fricción de los mercados de Internet se hace desde dos puntos de vista: o bien comparando las características de los mercados electrónicos y los convencionales, o bien analizando la conducta dentro de los mercados electrónicos. El presente capítulo se centra en la segunda opción.

En los mercados electrónicos, la eficiencia se mide en cuatro dimensiones: el nivel de precios, su dispersión, las elasticidades de los bienes a los precios y los costes (Smith, Bailey y Brynjolfsson, 1999). En cuanto a la creación de los precios, habrá factores que no han influido demasiado en la formación de los precios de los mercados convencionales, como por ejemplo el efecto de los costes de búsqueda que presiona los precios a la baja e intensifica el intento de diferenciación para poder mantener los precios por encima del coste marginal (Peddibhotla, 2000). A su vez, el servicio post-venta para ciertos tipos de bienes o incluso el efecto del tamaño de la Red en el equilibrio podrán en principio ser tratados con elementos de la teoría microeconómica tradicional sobre todo utilizando los supuestos tradicionales (por ejemplo la información perfecta), pero yendo más allá de la creencia común de que los precios en Internet son bajos porque los consumidores tienen la facilidad de encontrarlos rápidamente. En la investigación reciente de Shapiro y Varian (1999) se analiza bajo qué condiciones ocurre todo esto.

En resumen y de acuerdo a lo que se viene comentando, se observa que la literatura actual está formada por estudios teóricos enfocados a la distribución de los bienes y los costes de búsqueda, otros en la formación de los precios, en el comportamiento del consumidor o en las estrategias de venta basadas en supuestos adicionales (por ejemplo la estructura del mercado y la transparencia).

En el presente capítulo se propone una aproximación empírica utilizando una muestra de datos de compradores por la Red en España, se mide el impacto sociodemográficos en el comercio electrónico y se contrastan sus implicaciones en la demanda de bienes y servicios en la Red. Se utiliza la misma base de datos que en los capítulos anteriores (ITIC-2003) recogida por el INE (Instituto Nacional de Estadística) de 18.948 observaciones sobre individuos para toda España, y con ello se determinan las características de la demanda de distintos bienes a través de comercio electrónico.

Una vez estimado el primer modelo y evaluadas sus implicaciones, se especifica y estima un segundo modelo para medir los efectos de los determinantes del gasto en bienes y servicios a través del comercio electrónico demostrando luego que la inclusión de pendientes específicas y no linealidades a través de recomendaciones del algoritmo RETINA (Pérez-Amaral, Gallo y White, 2003) mejora la especificación original en la predicción fuera de la muestra.

## **2. COMERCIO ELECTRÓNICO Y USO DE INTERNET**

Los efectos de Internet sobre el comercio son difíciles de analizar. Incluso los trabajos teóricos necesitan basarse en observaciones empíricas. Durante los últimos años, en España este sector ha ido evolucionando tanto en tamaño como en profesionalización. En algunos sondeos realizados para el sector en toda España, se

puede observar que en 2005 empleaba a más de 48.000 personas de forma permanente, además el volumen de negocio actual en sectores como el de telemarketing duplica ya la cifra alcanzada en el año 1999, situándose actualmente en casi 1.000 millones de euros anuales<sup>1</sup>.

En España la mayoría de los estudios existentes hacen referencia a cuestiones de regulación del mercado, pero pocos realizan estimaciones econométricas de funciones de gasto o funciones de demanda para analizar las influencias de los distintos factores, y realizar predicciones. Sin embargo, existen algunos estudios descriptivos que muestran y explican las cifras obtenidas en encuestas (Fundación AUNA, 2003, red.es, 2004, etc.).

Para entender el perfil de los demandantes de bienes a través de la Red, es importante considerar que los determinantes de la probabilidad de comprar a través de Internet incluyen información acerca de los determinantes de la demanda de uso de Internet (Cerno y Pérez-Amaral, 2005). Un regresor en un modelo de respuesta binaria explica el impacto en la probabilidad de que el agente elija una opción particular (Train, 2002). Al explicar aquí el comportamiento y los determinantes de la demanda de bienes a través de Internet, hay que considerar al acceso y uso de Internet como una variable explicativa.

En la primera parte de este capítulo se contribuye a la literatura existente al especificar un modelo para los determinantes de la demanda de bienes vía comercio electrónico para España utilizando un modelo Logit en donde la variable endógena, que tomará valores uno y cero, refleja si una persona ha comprado a través de Internet o no respectivamente. Dicha decisión está determinada por distintos atributos que se definen

---

<sup>1</sup> Citado en <http://www.n-economia.com/frindex.html>.

a continuación. Con este modelo luego se interpreta en términos probabilísticos la posibilidad de adquisición de bienes y servicios a través de la Red.

## 2.1. ANÁLISIS DESCRIPTIVO Y DEFINICIÓN DE VARIABLES

En una primera descripción del conjunto de datos se observa que sólo el 3,7% de la población española compra habitualmente por Internet (últimos tres meses). A continuación se presenta el Gráfico 1 donde se pueden ver los porcentajes del número de transacciones de bienes y servicios adquiridos a través de Internet:

**Gráfico1:** Productos comprados a través de Internet en los últimos 12 meses



Se observa que para la muestra estudiada los productos más demandados son las actividades de ocio tales como los viajes (41,1%) y las entradas a espectáculos (26,5%), mientras que en contrapartida están la adquisición de acciones y las apuestas que serán las menos demandadas, con un 4,9% y un 0,9% respectivamente.

### 2.1.1. CARACTERÍSTICAS DE LOS DATOS

Los datos se refieren a una parte de la muestra utilizada en los capítulos dos y tres recogida en 2003 para toda España. Se trata de los 5.273 individuos que se declaran usuarios de Internet. A continuación se presentan las Tablas 1 y 2 donde se definen las variables y se presentan estadísticos descriptivos.

**Tabla 1. Definición de Variables**

<i>Variable</i>	<i>Definición</i>
<i>COMPRA</i>	Ficticia =1 si el individuo compra por Internet
<i>IS_Q</i>	Quintil de Renta (variable secuencial 1-5) (=1 si se encuentra en el quintil más bajo... =5 si se halla en el quintil más alto)
<i>VIVINT</i>	Ficticia =1 si el individuo accede a Internet desde el hogar
<i>USO</i>	Cantidad de sitios donde usa Internet (de 1 a 4)
<i>USOIN</i>	Intensidad de uso de Internet (horas semanales)
<i>USOR</i>	Intensidad de uso del ordenador (horas semanales)
<i>SEXH</i>	Ficticia =1 si el individuo es hombre
<i>EDAD</i>	Edad del individuo
<i>POBLAC</i>	Tamaño poblacional relativo (nivel provincial) <sup>2</sup>
<i>TOTMIEM</i>	Cantidad de miembros en el hogar del individuo

**Tabla 2. Estadísticos Descriptivos**

<i>Variable</i>	<i>Desviación</i>			
	<i>Media</i>	<i>Típica</i>	<i>Mínimo</i>	<i>Máximo</i>
<i>COMPRA</i>	0,126	0,336	0	1
<i>IS_Q</i>	4,030	0,850	1	5
<i>VIVINT</i>	0,615	0,487	0	1
<i>USO</i>	1,490	0,653	1	4
<i>USOIN</i>	38,001	29,805	0	70
<i>USOR</i>	46,642	29,257	0	70
<i>SEXH</i>	0,519	0,499	0	1
<i>EDAD</i>	33,862	12,465	15	88
<i>POBLAC</i>	4,237	4,429	0,163	13,277
<i>TOTMIEM</i>	3,333	1,192	1	6

<sup>2</sup> Se refiere al tamaño poblacional de cada provincia a la que pertenece el individuo dividido por la población total de España.

Al comparar con los resultados de los análisis descriptivos de los capítulos dos y tres, se puede ver que la proporción de los usuarios de Internet que además compran productos a través de la Red es mayor que con la muestra completa, aunque sigue siendo bastante baja (12,6%). Además también se aprecia que pertenecen mayoritariamente a los dos quintiles de ingresos superiores, ya que la media de la variable  $IS\_Q$  es 4,030 teniendo en cuenta su rango de variación (de 1 a 5). Más de la mitad posee acceso a Internet en el hogar (61,5%), aunque en media utilizan Internet en más de un sitio (la media de la variable  $USO$  es 1,49). En cuanto a la intensidad, se puede ver que son usuarios medianamente intensos cuanto se observa la media de horas de uso semanal (38,001). El ordenador es más utilizado (46,642). Algo más de la mitad de la muestra de usuarios es de sexo masculino (51,9%) y tiene una edad promedio de 33,9 años.

A partir de este primer análisis a continuación se propone estimar un modelo para los determinantes de la compra de bienes a través de Internet.

## 2.2. EL MODELO LOGIT

El modelo logit pertenece al grupo de modelos de probabilidad no lineales, en donde la especificación utilizada garantiza una predicción comprendida en el rango 0-1. Las funciones de distribución cumplen este requisito, siendo una de las más habituales la función de distribución logística, que da lugar a este modelo. La especificación es

$$Y_i = \frac{e^{\beta_k X_{ki}}}{1 + e^{\beta_k X_{ki}}} + u_i \quad (4.1)$$

Este modelo sirve para medir el efecto de los factores que influyen en la probabilidad de que ocurra el acontecimiento objeto de estudio ( $Y_i = 1$ ).

En cuanto a la interpretación de los parámetros estimados, el signo de los mismos indica la dirección en que se mueve la probabilidad cuando aumenta la variable explicativa, pero la cuantía del parámetro no coincide con la magnitud de la variación en la probabilidad, ya que cuando aumenta en una unidad la variable explicativa los incrementos en la probabilidad no son siempre iguales (Train, 2002).

Una interpretación sencilla es la que se obtiene a través de la linealización del modelo:

$$E(Y_i) = P_i = \frac{e^{\beta_k X_{ki}}}{1 + e^{\beta_k X_{ki}}}$$

$$\frac{P_i}{(1 - P_i)} = e^{\beta_k X_{ki}} \quad (4.2)$$

A la expresión (4.2) se la conoce como ratio de ventaja (odd-ratio en inglés). Su interpretación es el número de veces que es más probable que ocurra el fenómeno de estudio de que no ocurra. El campo de variación nunca es negativo y su interpretación se realiza en función de que el valor sea igual, menor o superior a la unidad. A su vez, el interés de esta medida adquiere sentido cuando se comparan las ventajas para distintos valores de la variable explicativa. Tomando logaritmos del ratio de ventaja, se linealiza la ecuación del modelo logit respetando el objetivo de que los valores de la variable dependiente caigan dentro del rango 0-1:

$$\ln\left(\frac{P_i}{(1-P_i)}\right) = \beta'_k X_{ki} \quad (4.3)$$

La nueva expresión (4.3) representa en una escala logarítmica la diferencia entre las probabilidades de que ocurra el suceso de estudio y su contraria. La estimación de este tipo de modelos, debido a su no linealidad, se puede realizar por máxima verosimilitud.

### 2.3. COMPRA POR INTERNET: ESPECIFICACIÓN Y RESULTADOS

A continuación se propone un modelo para los determinantes de la compra por Internet, considerando determinantes sociodemográficos tales como la renta, el sexo, la edad y el hábitat poblacional. También se consideran las características del individuo respecto al servicio de Internet tales como el acceso desde el hogar, el uso desde otros sitios además del hogar y la intensidad de utilización del servicio medida en horas semanales. La especificación, considerando esos atributos, quedará entonces:

$$\begin{aligned} \ln\left(\frac{P_i}{1-P_i}\right) = & \beta_0 + \beta_1 IS\_Q_i + \beta_2 VIVINT_i + \beta_3 USO_i + \beta_4 USOINT_i + \beta_5 USOR_i + \\ & + \beta_6 SEXO\_H_i + \beta_8 EDAD_i + \beta_9 EDAD_i^2 + \beta_{10} POBLAC_i + \beta_{11} TOTMIEM_i + u_i \end{aligned} \quad (4.4)$$

Donde  $P_i = \Pr(COMPRA_i)$  es la probabilidad de que el individuo  $i$ -ésimo compre a través de Internet. A continuación, en la siguiente tabla se presentan los resultados de las estimaciones de los parámetros y de los ratios de ventaja:

**Tabla 3. Resultados del Modelo Logit para la Compra por Internet**

Variable Dependiente: Compra por Internet	Coefficiente	Ratio de Ventaja	z
Constante	-6,19	---	13,67
<i>IS_Q</i>	0,23	1,26	3,75
<i>VIVINT</i>	0,48	1,62	4,39
<i>USO</i>	0,39	1,48	5,84
<i>USOIN</i>	0,01	1,01	6,31
<i>USOR</i>	0,01	1,01	2,61
<i>SEXH</i>	0,38	1,46	4,24
<i>EDAD</i>	0,09	1,09	4,01
<i>EDAD</i> <sup>2</sup>	-0,001	0,99	4,22
<i>POBLAC</i>	0,04	1,04	4,77
<i>TOTMIEM</i>	-0,08	0,92	2,16

Observaciones: 5.223  
 Log-Verosimilitud: -1773,45  
 $\chi^2_{(10)}$  : 397,09 (Prob. = 0,000)  
 Pseudo R<sup>2</sup>: 0,1007

La significatividad conjunta del modelo es alta, la razón de verosimilitudes que se obtiene es de 397,09, valor elevado para una chi-cuadrado con 10 grados de libertad. La bondad de ajuste del modelo se la considera también como buena luego de ver el grado de acierto de las predicciones dentro de la muestra:

**Tabla 4. Predicciones del Modelo Estimado**

		Valor Real		
		<i>COMPRA = 0</i>	<i>COMPRA = 1</i>	<i>Total</i>
Valor Predicho	<i>COMPRA = 0</i>	3.063	208	3.271
	<i>COMPRA = 1</i>	1.521	451	1.972
	<i>Total</i>	4.584	659	5.243

El modelo predice acertadamente 3.514 de las 5.243 observaciones (67,02%). Debido a la gran cantidad de ceros dentro de la muestra, se utiliza como umbral la

proporción de unos en la variable endógena (0,126). Con esto, la especificidad (porcentaje de aciertos iguales a 0) es de 66,81% y la sensibilidad (porcentajes de aciertos iguales a 1) de 68,43%.

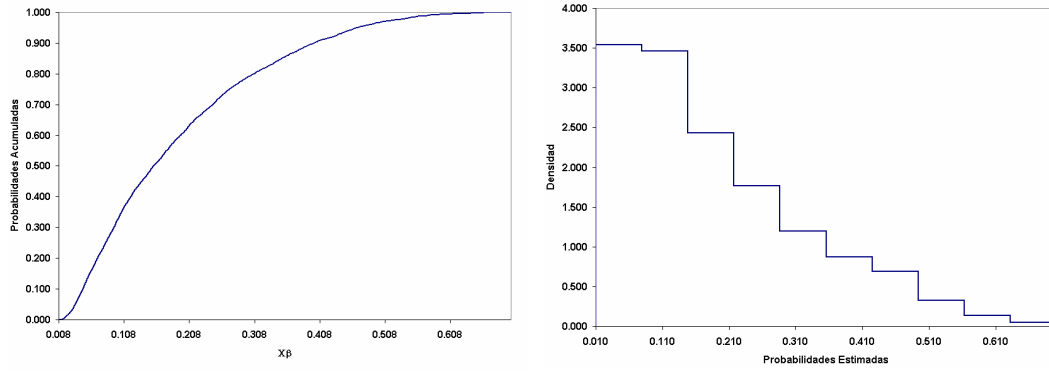
Hay que mantener siempre en mente que se está utilizando una muestra condicionada a los usuarios de Internet, de manera que los resultados obtenidos y las interpretaciones realizadas serán sólo aplicables a esta parte de la muestra. Analizando los valores de los ratios de ventaja, se puede decir:

- Exceptuando la cantidad de miembros en el hogar, todas las demás variables tendrán efectos positivos en la probabilidad de comprar por Internet.
- El acceso a Internet en la vivienda (*VIVINT*) y la cantidad de lugares de uso de Internet (*USO*), ambas variables referidas a los hábitos de Internet del individuo, tienen los coeficientes de mayor valor. El acceso a Internet en la vivienda tendrá el valor mayor en el ratio (1,62). Le sigue la cantidad de sitios de uso de Internet con un ratio de 1,48. Es decir que el acceso a Internet en el hogar contribuye en la probabilidad de adquirir productos a través de la Red en un 62%, y el utilizar el servicio en varios sitios además del hogar contribuirá a incrementar la probabilidad de comprar en 48%. Estos resultados tienen lógica al considerar que antes de comprar por Internet, primero hay que conectarse a la Red, o bien desde el hogar, o bien desde otros sitios.
- Se observa que pertenecer al sexo masculino tiene influencia en la probabilidad de comprar por Internet. Su ratio de ventaja es de 1,46. Esto sugiere mayor afinidad del género masculino con las compras a través de la Red.
- La influencia de los niveles de renta también es positiva. La variable que contiene sus efectos tiene un ratio de ventaja de 1,26. Tendrá también su lógica debido a que al haber una mayor renta habrá un mayor gasto. Es decir que

también tendrá una influencia positiva en la probabilidad de comprar por Internet.

- El hábitat poblacional (*POBLAC*), las intensidades de uso de Internet (*USOIN*) y del ordenador (*USOR*) también impactarán de forma positiva en la probabilidad de compra por Internet, pero de manera menor que las detalladas en los puntos anteriores. Los ratios de ventaja serán más cercanos a 1 (1,04 para la variable *POBLAC*, y 1,01 tanto para la variable *USOIN* y *USOR*).
- El caso de la variable *EDAD* es específico debido a que está incluida dentro del modelo tanto en niveles como al cuadrado. Su derivada es una función lineal en *EDAD*:  $\partial \hat{COMPR} / \partial EDAD = 0,09 - (2) \cdot (0,001) EDAD$ . Es decir que a medida que aumenta la edad disminuyen los aumentos de la probabilidad de comprar por Internet. El punto de inflexión donde el efecto total de la edad empieza a hacer decrecer la probabilidad sería sobre los 45 años.
- La única variable de las consideradas que disminuye la probabilidad de comprar bienes y servicios en Internet será la variable referida al total de miembros en el hogar (*TOTMIEM*). Su ratio de ventaja está por debajo de 1 (0,92). Es decir que a mayor cantidad de miembros en el hogar, menor probabilidad de adquirir bienes en la Red.

A continuación en el Gráfico 2 se observan las probabilidades acumuladas e individuales calculadas a partir del modelo estimado:



**Gráfico2:** Probabilidades Estimadas Acumuladas e Individuales de Comprar en Internet

Siempre teniendo en cuenta que para cada individuo habrá un efecto distinto sobre la probabilidad de comprar productos en Internet, a primera vista se observa que los individuos con mayor probabilidad de adquirir productos en Internet se encuentran en la cola de la distribución, debido a que las probabilidades mayores tienen una densidad menor (gráfico de la derecha).

En cuanto a los efectos marginales, más arriba se ha visto que las variables con mayor ratio de ventaja en la probabilidad serán el acceso a Internet en la vivienda, la renta, el uso de Internet y el pertenecer al sexo masculino. Estas cuatro variables eran las que más influyen positivamente en la probabilidad de comprar en Internet, que es lo que observa al cuantificar los efectos de las mismas y al estudiar la evolución de sus efectos marginales (ver Apéndice 1). Esta cuantificación no es directa, además que se obtendrá un valor probabilístico para cada uno de los individuos.

En el Apéndice 1 se observan los efectos marginales desde dos puntos de vista: a

través de la función obtenida con la derivada  $\frac{\partial P_i}{\partial x_{ik}} = P_i(1 - P_i)\hat{\beta}_k$  (gráficos de la

izquierda) y en la diferencia de las dos funciones de probabilidad predichas para los 5.243 individuos ordenados estos por nivel ascendente de renta y que se interpreta como el efecto de cada variable explicativa sobre las probabilidades (gráficos de la derecha).

En todos los casos considerados se observa que la probabilidad de que un individuo de la muestra utilice Internet para adquirir productos y servicios irá aumentando con los valores de las cuatro variables consideradas.

Por último se evalúa la variación en las probabilidades pronosticadas observando el cambio que se produce en estas cuando el individuo adquiere una conexión a Internet desde el hogar, pasar de utilizar Internet cada vez desde más lugares o pasar de un quintil de ingresos inferior a uno superior. A continuación se presentan las siguientes tablas:

**Tabla 5. Probabilidades de Comprar en Internet en los Distintos Niveles de Renta**

Quintil de Renta	1°	2°	3°	4°	5°
Probabilidad	0,0815	0,1004	0,1232	0,1503	0,1821
Diferencia	0,0189	0,0228	0,0271	0,0318	

**Tabla 6. Probabilidades de Comprar en Internet para la Cantidad de Sitios de Uso**

Sitios de Uso de Internet	1	2	3	4
Probabilidad	0,1283	0,1786	0,2431	0,3217
Diferencia	0,0503	0,0645	0,0786	

**Tabla 7. Probabilidades de Comprar en Internet para el Acceso desde el Hogar**

Internet en el Hogar	No	Sí
Probabilidad	0,1171	0,1765
Diferencia	0,0479	

De estas tablas se puede decir:

- Las diferencias muestran que el mayor incremento en la probabilidad de comprar por Internet se debe a que los individuos utilicen el servicio en muchos

sitios, siendo este incremento de un 7,86% cuando se pasa de utilizar Internet en tres sitios a utilizarlo en cuatro.

- Pasar a un quintil de ingresos superior no siempre provoca el mismo incremento en la probabilidad de comprar en Internet, aunque vemos que esta diferencia de probabilidades va incrementándose a medida que se va subiendo de categoría de renta
- Como es lógico, el contar con el servicio de Internet desde el hogar incrementa la probabilidad de comprar productos en la Red, con un incremento superior a los experimentados al pasar a quintiles superiores de ingresos (4,79% contra el mayor incremento al pasar del quintil 4º al 5º con una diferencia del 3,18%).

Como conclusión de esta primera parte se puede decir que, dentro de los determinantes considerados para la compra de bienes y servicios en la Red, los más importantes serán las características de los individuos como tener acceso al servicio desde la vivienda, y la cantidad de lugares de uso del mismo además del hogar. Curiosamente la intensidad de uso de Internet (horas semanales conectado a la Red) impactará positivamente, pero no con la magnitud que lo harán estas dos variables mencionadas (acceso desde la vivienda y cantidad de lugares de uso). Las otras dos variables que provocarán impactos positivos en la probabilidad serán la pertenencia al género masculino y el nivel de renta, con impactos importantes tanto una como otra variable, pero no con el mismo grado que las dos mencionadas anteriormente. Esto tiene sentido debido a que antes de comprar bienes y servicios en la Red hay que tener acceso a la misma, y además hacer uso del servicio. Una vez cubierta esta parte, los factores como el nivel de renta o el género también influirán positivamente en la probabilidad, pero en menor medida.

Una vez estudiados el perfil de quién compra en la Red dentro de los que tienen acceso a la misma, a continuación se elabora un modelo de gasto individual en comercio electrónico para los usuarios de Internet.

### 3. MODELOS DE DEMANDA DE COMERCIO ELECTRÓNICO

La razón más citada de por qué se puede esperar que los mercados electrónicos sean más eficientes que los mercados físicos es principalmente por la reducción de información asimétrica y la de los costes de búsqueda. El enfoque original lo debemos a Hotelling (1929) cuando plantea que para los grandes consumidores los costes de búsqueda hacen que los precios se sitúen por encima de los costes marginales en equilibrio. En la misma línea Bakos (1997) sostiene que si los mercados electrónicos logran disminuir dichos costes, esto conducirá a una disminución de precios tanto para bienes homogéneos como diferenciados.

Se parte de la idea de un conjunto de bienes demandados en Internet considerados dentro de la muestra (ver Gráfico 1) como el conjunto cerrado y convexo de consumo del individuo, y suponiendo que las preferencias se ordenan satisfaciendo determinadas propiedades habituales<sup>3</sup> de manera tal de que exista una función de utilidad  $u : X \rightarrow R$  tal que  $x \succ y$  si y solo si  $u(x) > u(y)$  (Mas-Colell *et al.*, 1995). Suponiendo entonces la existencia de esta función de utilidad podemos analizar la conducta del consumidor y derivar su función de demanda partiendo de la hipótesis básica de que el individuo siempre elige un conjunto de opciones asequibles a la cesta por la que muestra mayor preferencia.

---

<sup>3</sup> Ver Mas-Colell, Whinston y Green (1995). Comparar con Varian (1992).

Si  $M$  es la renta de la que dispone el individuo y  $P = (p_1, p_2, \dots, p_k)$  es el vector de precios de los bienes 1, 2, ...,  $k$ , el conjunto presupuestario del consumidor a través de la Red vendrá dado por  $B = \left\{ x_i \in X : \sum_{i=1}^k p_i x_i \leq M \right\}$  y el problema a resolver se formula como:

$$v(P, M) = \max u(x) \text{ tal que } \sum_{i=1}^k p_i x_i = M$$

en donde  $v(P, M)$  es la función indirecta de utilidad y que representa a la utilidad máxima alcanzada para una cesta de bienes sujeta a los precios y a la renta.

La función homogénea de grado cero en  $P$  y en  $M$  que relaciona el vector de precios y la renta con la cesta demandada se denomina función de demanda del consumidor. Si aplicamos la teoría de la dualidad del consumo, y dado que  $v(P, M)$  es creciente en  $M$ , se puede invertir la función y despejar la renta  $M$  en función del nivel de utilidad, y así obtener otra función  $e(P, u)$  que se denomina función del gasto y que indica el coste mínimo de alcanzar el nivel fijo de utilidad. Aplicando el Lema de Shephard obtenemos la función de demanda hicksiana:

$$\frac{\partial e(P, u)}{\partial p_i} = h(P, u)$$

La función de demanda hicksiana nos indica la cesta minimizadora del gasto necesaria para alcanzar el nivel de utilidad  $u$  a los precios  $P$ . Pero el problema con que nos encontramos al intentar estimar esta función de demanda es que la misma es no

observable, debido a que depende de la utilidad, que no lo es. En cambio, las funciones de demanda expresadas en función de los precios y de la renta sí son observables. A este tipo de función de demanda se le llama función de demanda marshalliana,  $x_i(P, M)$ , y es posible demostrar las siguientes relaciones (Varian, 1992):

$$x_i(P, M) = h_i[P, v(P, M)]$$

$$h_i(P, u) = x_i[P, e(P, u)]$$

Es decir que la demanda marshalliana correspondiente al nivel de renta  $M$  es idéntica a la demanda hicksiana correspondiente al nivel de utilidad  $v(P, M)$ , o la demanda hicksiana correspondiente al nivel de utilidad  $u$  es idéntica a la demanda marshalliana correspondiente al nivel de gasto  $e(P, u)$ .

Estas identidades demuestran que cualquier cesta demandada de bienes puede expresarse como una solución del problema de maximización de la utilidad o bien como un problema de minimización del gasto. Esto es lo que la teoría denomina *dualidad*.

A partir de las ideas plasmadas en este apartado, a continuación se especificará y estimará una función de demanda de bienes y servicios y una función de gasto a través de Internet que sea lineal, contando con la información de los usuarios utilizada en los apartados anteriores y con inclusión de nuevas variables que a continuación se describen.

### 3.1. DESCRIPCIÓN DE VARIABLES

Modelizar funciones de demanda o de gasto en las telecomunicaciones generalmente no suele ser una tarea fácil, debido a que con frecuencia la mayoría de los datos disponibles son binarios y a veces no se tiene información de los precios pagados por los consumidores. Además los datos pueden presentar una importante heterogeneidad. Asimismo, no se tendrá gran confianza acerca de una forma funcional ni sobre la homogeneidad de los datos.

Antes de estimar las funciones de demanda y de gasto a través de Internet, en este apartado se va a describir en detalle las variables que se utilizarán para tal fin. Luego se recogerá la heterogeneidad en los datos utilizando una metodología exploratoria habitual: el algoritmo k-Medias (Mac Queen, 1967; Hartigan y Wong, 1979) para identificar los distintos grupos con características similares en los datos.

La información de cantidades pagadas, bienes demandados por los consumidores o distintos costes asumidos por los vendedores se extrae a partir de variables *proxy* que contengan estos efectos. Las variables adicionales a las ya expuestas (comprar en Internet –*COMPRA*–, quintil de renta –*IS\_Q*–, acceso a Internet desde el hogar –*VIVINT*–, cantidad de sitios en donde usar Internet –*USO*–, intensidad de uso del ordenador –*USOR*–, Intensidad de uso de Internet –*USOIN*–, género –*SEXH*–, edad –*EDAD*–, tamaño poblacional –*POBLAC*– y cantidad de miembros en la familia –*TOTMIEM*–) se detallan a continuación en el Cuadro 3, se puede observar que en su mayoría son variables ficticias. Otra cosa que se destaca es que ahora el índice de renta individual tiene sus valores originales, es decir ya no está categorizado:

**Tabla 8. Definición de Variables Adicionales**

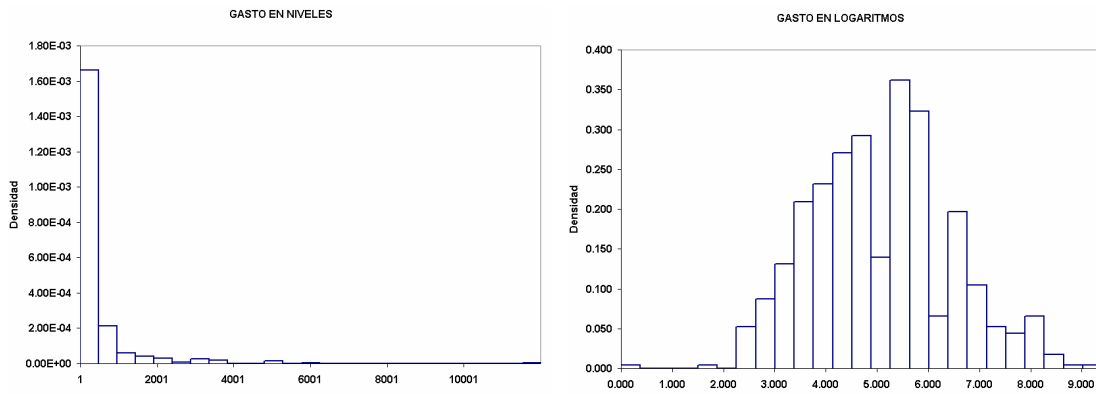
<i>Variable</i>	<i>Definición</i>
$G_i$	Gasto realizado en productos por la Red (en euros).
$IS_i$	Índice de renta individual <sup>4</sup> .
$P_1$	Ficticia =1 si el individuo adquirió productos del hogar.
$P_2$	Ficticia =1 si adquirió música.
$P_3$	Ficticia =1 si adquirió libros.
$P_4$	Ficticia =1 si adquirió ropa.
$P_5$	Ficticia =1 si adquirió software.
$P_6$	Ficticia =1 si adquirió material informático.
$P_7$	Ficticia =1 si adquirió artículos electrónicos.
$P_8$	Ficticia =1 si adquirió acciones y títulos financieros.
$P_9$	Ficticia =1 si adquirió paquetes de viajes.
$P_{10}$	Ficticia =1 si adquirió entradas a espectáculos.
$P_{11}$	Ficticia =1 si compró lotería o realizó apuestas online.
$P_{12}$	Ficticia =1 si adquirió otros productos o servicios.
$CANT_i$	Número de transacciones para adquisiciones en Internet.
$FP1_i$	Ficticia =1 si la forma de pago es con tarjeta de crédito
$FP2_i$	Ficticia =1 si la forma de pago es con transferencia bancaria
$FP3_i$	Ficticia =1 si la forma de pago es contra reembolso
$FP4_i$	Ficticia =1 si la forma de pago es por suscripción

A continuación se hace un análisis gráfico-descriptivo de las variables comparándolas con el gasto en la Red. La variable referida al gasto en euros en la Red ( $G_i$ ) se refiere al gasto realizado durante los últimos tres meses. La respuesta de los encuestados era de tres tipos: en pesetas, en euros y en categorías predefinidas de gastos (también en pesetas y en euros). A partir de esta información se elaboró la variable  $G_i$ , llevando todas las observaciones a valor euro (1 euro = 166,36 pesetas). En las observaciones referidas a intervalos de gasto, se utilizó el punto medio del mismo.

Observando el histograma de la variable gasto  $G_i$  en niveles y en logaritmos, se evidencia la existencia de valores extremos:

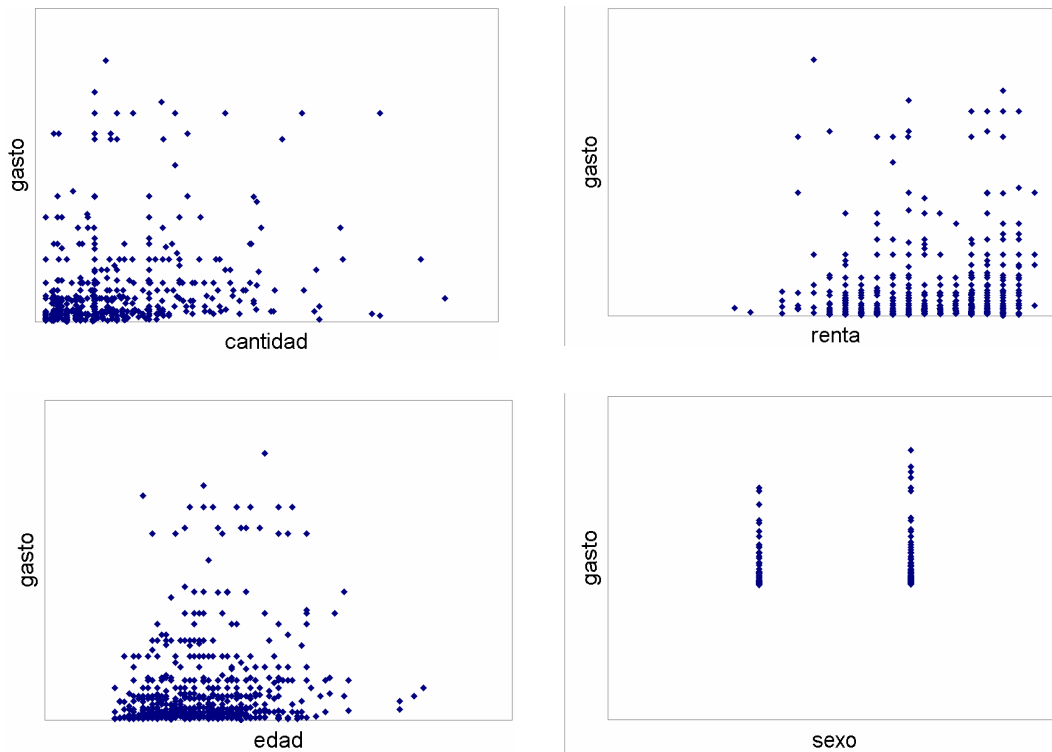
<sup>4</sup> Explicada en detalle en el capítulo 2, esta variable está construida a partir de agregar ponderaciones de capital humano y no humano.

**Gráfico 3.** Histograma de Gasto en Internet (en niveles y en logaritmos)



Haciendo uso del algoritmo de detección y exclusión de valores extremos de Peña y Yohai (1999) incorporado en RETINA y aplicándolo al logaritmo del gasto, se detectan y excluyen momentáneamente de la muestra 22 valores extremos. Luego se realizan gráficos bivariantes de la variable Gasto con algunas variables explicativas:

**Gráfico 4.** Gasto vs. Variables Explicativas Seleccionadas sin Valores Extremos



El método de recogida de la información realizada por el INE se basó en una combinación de procedimientos CATI<sup>5</sup> (entrevista telefónica) y PAPI<sup>6</sup> (entrevista personal). Para cada sección de la muestra se dispuso de dos listas de reservas para sustituir las incidencias en las viviendas originalmente seleccionadas. Luego de la recogida de datos se utilizaron distintos tipos de algoritmos de filtro para depurar la muestra. Es decir que es difícil pensar que estos valores extremos equivalgan a valores mal registrados. Además, al ser una muestra de datos de corte transversal, la heterogeneidad puede estar presente. Será necesario recogerla en los posteriores modelos que se vayan a ajustar.

A continuación se procede a detectar grupos de individuos con perfiles homogéneos. Para ello se utiliza la probabilidad individual de comprar en Internet estimada en el apartado anterior, y el k-medias. El algoritmo de k-Medias de Mac Queen (1967) es un clásico método estadístico para detectar subgrupos homogéneos dentro de una muestra. El criterio de homogeneidad que se utiliza es minimizar la suma de cuadrados dentro de los subgrupos para todas las variables dada por:

$$\min SCDG = \sum_{g=1}^G \sum_{j=1}^P \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2 = \min \sum_{g=1}^G \sum_{j=1}^P n_g s_{jg}^2$$

donde  $x_{ijg}$  es el valor para el individuo  $i$  de la variable  $j$  clasificado dentro del grupo  $g$ , y  $\bar{x}_{jg}$  es la media de esta variable dentro del grupo. Se puede ver que esto es equivalente a minimizar la suma ponderada de las varianzas de las variables en los grupos. Estas varianzas grupales se las puede considerar una buena medida de

<sup>5</sup> Acrónimo de “Computer Assited Telephonic Interview”.

<sup>6</sup> Acrónimo de “Paper and Pencil interview”.

heterogeneidad, porque al minimizarlas se obtendrían grupos más homogéneos. El algoritmo k-Medias funciona con las siguientes etapas (Peña, 2002):

- a) Seleccionar los  $k$  puntos más alejados entre sí como centros de los grupos iniciales.
- b) Calcular las distancias euclídeas de cada elemento a los centros de los  $k$  grupos y asignando a cada elemento el grupo de cuyo centro esté más próximo.
- c) Definir el criterio de homogeneidad (en nuestro caso será minimizar la  $SCDG$ ) e ir reasignando.
- d) Si luego de la reasignación el criterio no mejora, terminar el proceso.

En este caso se ha agrupado la muestra en 5 grupos a partir de intervalos de probabilidad calculada a partir de los resultados de la estimación del modelo (4.4). Aunque esta cantidad podría reducirse o aumentar dependiendo del grado de agregación y de homogeneidad pretendido dentro de cada grupo, siendo el máximo de grupos igual a la cantidad de observaciones. Se tienen entonces los siguientes grupos:

**Tabla 9. Usuarios de Internet Agrupados por Probabilidad de Compra en la Red**

GRUPO	Probabilidad de Compra en Internet	Tamaño
Menos Propensos (Grupo 1)	0,010 – 1,101	1.757
Poco Propensos (Grupo 2)	0,101 – 0,191	1.341
Propensión Media ( Grupo 3)	0,191 – 0,298	1.037
Propensión Alta (Grupo 4)	0,299 – 0,429	684
Propensión muy Alta (Grupo 5)	0,430 – 0,701	411

Incluyendo estos grupos en forma de constantes específicas dentro del modelo de demanda de bienes en la Red, se recoge la heterogeneidad de los datos observada más arriba.

### 3.2. MODELO DE DEMANDA DE BIENES ADQUIRIDOS EN LA RED

En este apartado se estimarán los parámetros de una función de demanda de bienes y servicios en Internet. Para especificar dicho modelo se utiliza como variable endógena la variable  $CANT_i$  referida a la cantidad de bienes que el individuo ha adquirido en la Red. Esta variable es el número de productos que los individuos han declarado comprar en Internet. Al ser esta una variable cuyos datos son números enteros no negativos, se utilizará una especificación acorde. Aunque se podría utilizar una regresión mínimo cuadrática, sin embargo por el tipo de datos (gran cantidad de ceros, datos enteros y ninguno negativo) y por la forma que tiene la distribución de la variable endógena, los modelos que más se utilizan son el de Poisson y el binomial negativo, este último utilizado sobre todo cuando los datos tienen sobredispersión. Aquí se especifica una función de demanda de bienes en Internet con un modelo de Poisson para la variable endógena  $CANT_i$  en función de la renta y de características tales como cuatro formas de pago consideradas (dentro de las características económicas), el sexo, la edad, el nivel de estudios y la población a la que pertenece (características sociodemográficas). Para recoger la heterogeneidad incluiremos además constantes específicas referidas a cuatro de los cinco grupos detallados en la Tabla 9.

El modelo entonces queda especificado de la siguiente manera:

$$\ln \lambda_i = CANT_i = \alpha_0 + \alpha_1 IS_i + \alpha_2 NIVELEST_i + \alpha_3 SEXO\_H_i + \alpha_4 EDAD_i + \\ + \alpha_5 POBLAC_i + \sum_{k=1}^4 \phi_{ik} FP_k + \sum_{m=1}^5 \xi_{im} GR_m + u_i$$

en donde  $i = 1, \dots, 5.218$ . Estos son sólo los que sí compran en Internet, las variables son las que están descritas más arriba en las Tablas 1 y 8 (índice de renta individual, nivel de estudios alcanzado, pertenencia al género masculino, edad y población),  $FP_k$  se refiere a las cuatro formas de pago de las compras consideradas, y  $GR_m$  se refiere a los grupos por categoría de probabilidad de comprar a través de Internet, y donde  $\lambda_i$  es el número de veces que el individuo  $i$  ha comprado por Internet en un período dado, y equivale a la esperanza condicional de una variable aleatoria con distribución de Poisson:

$$\Pr(CANT|\lambda) = \frac{e^{-\lambda} \lambda^{CANT}}{CANT!} \quad (4.5)$$

Este modelo de regresión se estima por máxima verosimilitud a partir de la siguiente función de verosimilitud:

$$L = \prod_{i=1}^{5243} \Pr(CANT_i | \lambda_i) \\ \ln L = \sum_{i=1}^{5243} (CANT_i \ln(\lambda_i) - \lambda_i) \quad (4.6)$$

### 3.2.1. RESULTADOS Y DISCUSIÓN

La siguiente tabla muestra los resultados de la estimación:

**Tabla 10. Resultados del Modelo de Demanda**

Variable Dependiente	Coefficiente Estimado	Ratio de Incidencia	<i>z</i>
<i>CANT</i>			
Constante	-3,63	---	17,08
<i>IS</i>	0,42	1,51	4,67
<i>NIVELEST</i>	-2,20	0,11	4,74
<i>SEX_H</i>	0,12	1,13	1,93
<i>EDAD</i>	0,002	1,00	0,77
<i>POBLAC</i>	0,10	1,01	1,69
<i>FP1</i>	2,42	11,23	36,00
<i>FP2</i>	0,82	2,27	10,93
<i>FP3</i>	1,32	3,75	19,79
<i>FP4</i>	0,36	1,42	2,72
<i>GR1</i>	-0,79	0,45	7,69
<i>GR2</i>	-0,22	0,81	2,84
<i>GR3</i>	-1,14	0,32	5,14
<i>GR4</i>	0,11	1,01	0,13
<i>GR5</i>	---	---	---

Observaciones: 5.218  
 Log-Verosimilitud: -1.914,34  
 $\chi^2_{(13)} : 3635,19$  (Prob. = 0,000)  
 Pseudo R<sup>2</sup>: 0,4870

La regresión de Poisson es susceptible de ser interpretada tanto a través de los coeficientes como en términos del ratio de incidencia. Recordando siempre que la variable endógena se refiere a “cantidades”, a los coeficientes estimados se los puede interpretar como la diferencia entre los logaritmos de las cantidades esperadas, es decir  $\hat{\beta} = \log E(y|x+1) - \log E(y|x)$ . Así, también se puede interpretar el parámetro estimado como el logaritmo del “ratio” o cociente de las cantidades esperadas, es decir

$\log\left[E(y|x+1)/E(y|x)\right]$ . El valor del ratio de incidencia se lo obtiene haciendo el cálculo de  $e^{\hat{\beta}}$ .

A partir de los resultados de la Tabla 10 podemos decir que:

- El número esperado de artículos adquiridos en Internet se incrementa notablemente con la posibilidad de pagar con tarjeta de crédito. Su ratio de incidencia es el más grande de todos con diferencia (11,23). Esto quiere decir que si el individuo posee una tarjeta de crédito (*FP1*), incrementa su número esperado de artículos adquiridos en Internet en mucho más del doble.
- La incidencia del resto de las formas de pago, aunque bastante menor que la que produce el pago con tarjeta de crédito, también es importante. Los pagos con transferencia bancaria (*FP2*) y contra reembolso (*FP3*) son los que inciden en 2,27 y 3,75 respectivamente. Observamos que el impacto del pago contra reembolso es mayor que el pago con transferencia bancaria. El pago por suscripción (*FP4*) es también importante, pero su incidencia es la menor de todas (1,42).
- El nivel de renta individual (representado por la variable *IS*) también tiene una incidencia positiva en la cantidad de bienes y servicios adquiridos por comercio electrónico. Su incidencia es de 1,51. Es decir que el incremento unitario el valor del índice *IS* equivale a un aumento en la cantidad esperada de transacciones para adquirir bienes y servicios en la Red del 51%.
- En cuanto a las constantes específicas de los cinco grupos de individuos, hemos eliminado del modelo al grupo del último intervalo de probabilidad (el de probabilidades más altas) para evitar problemas de multicolinealidad exacta en la función de verosimilitud. Vemos que a medida que aumenta la probabilidad de comprar en Internet, como es lógico, los ratios de incidencia tienden a

aumentar. Esto quiere decir que las personas más propensas a comprar en Internet adquieren mayor cantidad de productos mientras que con los menos propensos ocurrirá lo contrario.

- Atributos sociodemográficos como la edad, el nivel de estudio o el tamaño de la población tendrán una escasa influencia en la cantidad esperada de comercio en la Red.

Resumiendo entonces, se ha considerado un modelo para los determinantes de la probabilidad de comprar en Internet. Para ello se ha utilizado la muestra de los que se declaran usuarios de Internet. Una vez obtenida, se ha agrupado a los individuos por su propensión a la compra en Internet y con esto se recogió la heterogeneidad en un modelo de demanda de bienes y servicios a través de Internet utilizando como variable endógena el número de bienes adquiridos a través de la Red. A continuación se especifica y se estima una función de gasto en comercio electrónico lineal en los parámetros.

### **3.3. FUNCIÓN LINEAL DE GASTO EN COMERCIO ELECTRÓNICO**

La predicción del gasto es importante desde muchos puntos de vista. Permite evaluar la conducta o descubrir perfiles del consumidor en Internet, sirve para determinar nuevas formas de financiación a partir de estos perfiles, etc. En este apartado se buscará obtener un modelo de predicción para el gasto. Para ello, a continuación se incluirán los grupos de la Tabla 9 en forma de constantes específicas dentro de un modelo MCO utilizando como variable endógena el logaritmo del gasto. La especificación es la siguiente:

$$\log(G_i) = \gamma_0 + \gamma_1 \log(IS_i) + \gamma_3 \log(EDAD_i) + \gamma_4 SEXO\_H + \sum_{g=1}^4 \delta_{ig} G_i + \sum_{p=1}^{12} \xi_{ip} P_i + u_i \quad (4.7)$$

Como se observa, se han incluido las constantes específicas  $\delta_{ig}$  y  $\xi_{ip}$  correspondientes a los cinco grupos de individuos del apartado anterior y a los doce productos considerados, respectivamente.

El fin ahora es conseguir un modelo para previsión que además mejorará la especificación. Para ello se utilizará RETINA, ya explicado en el capítulo anterior y con el cual se obtendrá un modelo con buena capacidad predictiva por fuera de la muestra. En la Tabla 11 se observan las variables que incluyó RETINA (los resultados de las estimaciones se encuentran en el Apéndice):

**Tabla 11:** Detalle de las Variables Utilizadas

<b>Variable Endógena</b>	$\ln(G)$
<i>Variables Continuas Originales</i>	$\ln(IS)$ , $\ln(EDAD)$
<i>Constantes Específicas</i>	$SEXO\_H$ , $G_g$ donde $g = 1, \dots, 5$
<i>Pendientes Específicas</i>	$SEXO\_H$ , $G_g$ , $P_p$ donde $p = 1, \dots, 12$

Recordar que la variable  $G_i$  hace referencia a los cinco grupos de individuos detectados en el apartado 3.1, y  $P_p$  hace referencia a los doce grupos de bienes y servicios considerados que se adquieren en Internet.

Hay que recordar lo dicho en el Capítulo III, que RETINA también es útil cuando no se tiene una hipótesis de especificación clara en cuanto a qué variables hay que considerar en el modelo. En este caso, la inclusión de constantes y pendientes específicas como regresores ha mejorado la explicación de las variaciones del gasto en

Internet, y eso se nota a partir del mejoramiento de la previsión fuera de la muestra, el AIC y la bondad de ajuste dentro de la muestra ( $\bar{R}^2$ ). En la siguiente Tabla 12 se presentan los principales resultados comparativos, en donde se destaca la mejora en la previsión fuera de la muestra con una importante reducción del Error Cuadrático Medio de Predicción Robusto (RCMSPE) de 1,195 a 0,689.

**Tabla 12:** Comparación entre el Modelo MCO (BLM) y el Modelo RETINA (URM) para la Función de Gasto

	MCO (BLM)	RETINA (URM)
Parámetros	8	33
AIC	0,356	-0,813
$\bar{R}^2$	0,260	0,779
RCMSPE	1,195	0,689

Se puede observar como la inclusión de constantes específicas, pendientes específicas y productos cruzados de los regresores mejoran los resultados globales. Además del RCMSPE, también se observa que el  $\bar{R}^2$  pasa de un valor de 0,260 a 0,779, que se puede considerar bueno, sobre todo teniendo en cuenta que esta es una medida que penaliza la inclusión de nuevos parámetros, observando que el número de parámetros aumenta de 8 a 33 (se ve en la Tabla 11 que las 12 variables ficticias referidas a los productos adquiridos sólo se incluirán como pendientes específicas). También el AIC mejora mucho pasando del valor 0,356 a -0,813.

De los resultados de las regresiones podemos decir:

- Las variables continuas originales consideradas en los BLM referidas al logaritmo de la edad y al logaritmo de la renta ( $\ln(EDAD)$  y  $\ln(IS)$ ) serán ambas significativas en el modelo MCO sin constantes específicas. Sin embargo, al incluir

las constantes específicas de los grupos vemos que sólo será significativa el  $\ln(EDAD)$ .

- Se observa que en el modelo RETINA los regresores originales sólo entrarán en forma de interacciones (productos cruzados entre variables explicativas continuas) o pendientes específicas (productos cruzados en donde una de las dos variables es binaria). Esto quiere decir que las correlaciones individuales más altas entre la variable endógena y los regresores se darán cuando estos últimos se refieran a interacciones o a pendientes específicas, y no como se presentan originalmente en el modelo MCO.
- Con respecto a los cuatro grupos de consumidores considerados, se observa que si bien en el modelo MCO son todos significativos (el  $G_V$  no lo incluimos evitando así problemas de multicolinealidad, pero su efecto está en el término constante), en el modelo RETINA el único grupo que aparece es el  $G_{II}$  en forma de pendiente específica.
- Se observa además que incluir variables binarias referidas a los productos es importante para recoger la heterogeneidad. Se puede observar que aparecerán 11 de las 12 en el modelo MCO en forma de pendientes específicas. Esto hace que el  $\bar{R}^2$  aumente (a pesar de ser una medida que penaliza la inclusión de regresores) hasta valores aceptables que indican un ajuste (0,779). A su vez, el RCMSPE disminuirá casi hasta la mitad pasando de 1,195 a 0,689.
- Como conclusión, se puede agregar que recoger la heterogeneidad e incluir en el modelo no linealidades en los regresores, puede mejorar sustancialmente tanto el ajuste como la habilidad predictiva, cuando se usan modelos sugeridos por RETINA. A su vez observamos que los productos adquiridos a través de Internet y considerados en la muestra serán buenos predictores de la demanda de comercio

electrónico y que los regresores continuos originales explicarán mejor dicha demanda a partir de ser no lineales.

#### **4. CONCLUSIONES**

En este capítulo se ha abordado la modelización de la demanda de comercio electrónico en España usando tres modelos: el primero referido al perfil del demandante de la Red teniendo en cuenta principalmente el nexo existente entre el comercio electrónico como tal y el acceso y uso de Internet. El segundo está referido a la cantidad de transacciones (demanda de bienes) en la Red. El tercero corresponde a una función lineal de gasto en Internet.

Los resultados de la primera especificación confirman la existencia de este vínculo y de que los determinantes del demandante serán su nivel de renta, la edad y el nivel de estudios alcanzado del individuo. El cálculo de las probabilidades de la compra a través de Internet revela que esta será más sensible a los hábitos de acceso y uso de Internet. En cambio el uso del ordenador, el género sexual y el hábitat, aunque también influyen, lo harán en menor medida.

El análisis gráfico bivariante del gasto en euros a través de la Red muestra la existencia de datos anómalos y heterogeneidad en los datos que deberán ser detectados y tratados para alcanzar una representación aceptable. Para ello se utiliza el algoritmo k-medias en donde se detectan 5 grupos de consumidores.

Posteriormente se usa un segundo modelo para los determinantes de la demanda de comercio electrónico. Para ello se especifica y estima un modelo de Poisson para la cantidad de bienes demandados a través de la Red. Se concluye que las distintas formas de pago de los productos y servicios de la Red (tarjeta de crédito, transferencia bancaria,

contra reembolso) influirán notoriamente en la cantidad de bienes adquiridos, seguidos por el nivel de renta y los distintos atributos sociodemográficos considerados.

El último modelo planteado para tratar la demanda de comercio electrónico será la especificación y estimación de una función lineal de gasto. Para tal tarea se utiliza el algoritmo RETINA para obtener un modelo flexible con buena predicción fuera de la muestra. RETINA generará e incluirá en el modelo nuevos regresores a partir de los originales, que mejorará sustancialmente el ajuste y disminuirá el RCMSPE casi hasta la mitad.

En cuanto a lo que se aprende a partir de estos modelos es que el perfil del individuo demandante de bienes y servicios en Internet en España es el de una persona joven, con estudios avanzados terciarios o universitarios, perteneciente a un ámbito poblacional mediano y con un nivel de renta mediano que le da acceso del crédito y le permite realizar pagos con tarjetas de crédito u otras maneras alternativas.

A su vez, con estos resultados se podría decir que los modelos planteados para los determinantes del comercio electrónico para este tipo de datos serían adecuados para realizar un estudio pormenorizado de los determinantes del comercio electrónico en España. Se considera que la metodología que se plantea para el tratamiento de datos de corte transversal con heterogeneidad en la muestra se podría también utilizar en otros contextos.

En futuras investigaciones sería interesante analizar paneles de datos e incluir dinámica dentro de los modelos especificados.

## **5. REFERENCIAS**

Angrist, J. (2004) "Treatment Effect Heterogeneity in Theory and Practice" *Economic Journal*, Royal Economic Society, vol. 114(494), pp. C52-C83

Bakos, Y. (1997) "Reducing Buyer Search Cost: Implications for Electronic Marketplaces" *Management Science*, vol.43, Issue 12

Bakos, Y. (2001) "The Emerging Landscape for Retail E-Commerce" *Journal of Economic Perspectives*, vol. 15, nº 1, pp. 69-80

Brynjolfsson, E. y Smith, M (1999) "Frictionless Commerce? A Comparison of Internet and Conventional Retailers" Working Paper, MIT Sloan School of Management

Cerno, L. y Pérez-Amaral, T. (2005) "Demand of Internet Access and Use in Spain" Documento de Trabajo 0506. Instituto Complutense de Análisis Económico (ICAE)

Cerno, L. y Pérez-Amaral, T. (2006) "Medición y Determinantes de la Brecha Tecnológica en España", mimeo

Fundación AUNA (2003) "Informe España 2003", disponible en [http://www.fundacionauna.com/areas/25\\_publicaciones/](http://www.fundacionauna.com/areas/25_publicaciones/)

Greene, W. (2000) "Econometric Analysis" 4<sup>th</sup>. Edition, Prentice Hall

Hartigan, J. y Wong, M. (1979) "A k-means Clustering Algorithm" *Applied Statistics*, 28, 100-108

Heckman, J. (1978) “Dummy Endogenous Variables in a Simultaneous Equation System” *Econometrica* 46: 931-959

Hotelling, H. (1929) “Stability in Competition” *The Economic Journal*, pp. 41-57

INE, “Encuesta Sobre Equipamiento y Uso de Tecnologías de Información y Comunicación de los Hogares del Segundo Trimestre de 2005” disponible en [www.ine.es/prensa/np388.pdf](http://www.ine.es/prensa/np388.pdf)

Mac Queen, J. (1967) “Some Methods for Classification and Analysis of Multivariate Observations” *Proc. Symposium of Mathematics, Statistics and Probability*, 5<sup>th</sup>, pp. 281-297, Berkeley

Maddala, G. (1983) “Limited Dependent and Qualitative Variables in Econometrics” Cambridge University Press

Marinucci, M. y Pérez-Amaral, T. (2005) “Econometric Modeling of Business Telecommunications Demand Using RETINA and Finite Mixtures” Documento de Trabajo 0501. Instituto Complutense de Análisis Económico (ICAE).

Mas-Colell, A., Whinston, M. y Green, J. (1995) “Microeconomic Theory” Oxford University Press

McKinsey, J. (2000) "Internet Pricing. A Creator of Value Not a Destroyer. McKinsey Marketing Practice" Documento disponible en [http://marketing.mckinsey.com/solutions/McK-Internet\\_Pricing.pdf](http://marketing.mckinsey.com/solutions/McK-Internet_Pricing.pdf)

Peddibhotla, N. (2000) "Are Economic Theories Good Enough to Understand E-Commerce?" en Wiseman, A. "The Internet Economy: Access, Taxes, and Market Structure", Brookings Institution Press, Washington D.C.

Peña, D. (2002) "Análisis de Datos Multivariantes", Mc Graw Hill

Peña, D. y Yohai, V. (1999) "A Fast Procedure for Outlier Diagnostic in Large Regression Problems" Journal of the American Economic Association, 94: 434-445

Pérez-Amaral, T. Gallo, G. y White, H. (2003) "A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA)" Oxford Bulletin of Economics and Statistics, 65:821-838

Red.es (2004) – Ministerio de Industria, Turismo y Comercio "Las TIC en los Hogares Españoles" disponible en <http://observatorio.red.es/estudios/consumo/index.html>

Shapiro, C. y Varian, H. (1998) "Information Rules: A Strategic Guide to the Network Economy", Harvard Business School Press

Smith, M., Bailey, J. y Brynjolfsson, E., (1999) “Understanding the Digital Markets: Review and Assessment” en Brynjolfsson, E. y Kahin, B. “Understanding the Digital Economy”, MIT Press

Train, K. (2002) “Discrete Choice Methods with Simulation” Cambridge University Press

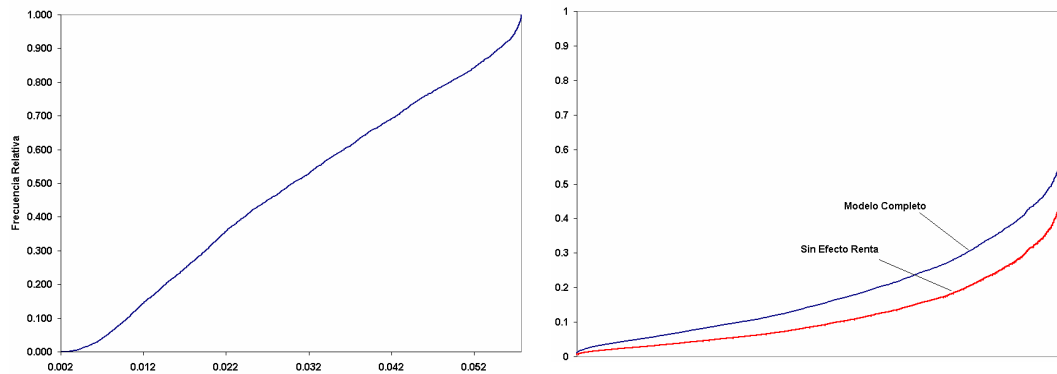
Varian, H. (1992) “Análisis Microeconómico” 3ra. Edición, Antoni Bosch ed.

Wilde, J. (2000) “Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors”. *Journal of Econometrics* 30: 415-443

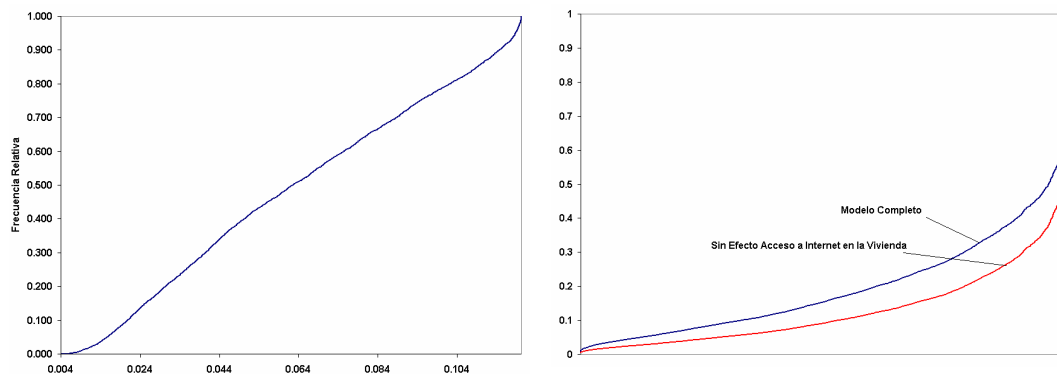
Zellner, A. (1962) “An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias” *Journal of the American Statistical Association*, 57, pp. 500-509

## APÉNDICE 1: GRÁFICOS DE EFECTOS MARGINALES (MODELO LOGIT)

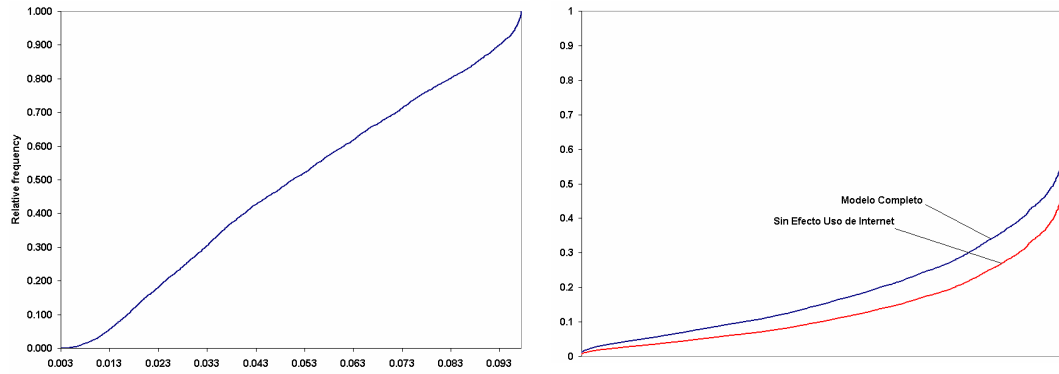
Los gráficos que se exponen a continuación fueron elaborados como sigue. Los de la derecha fueron creados a partir de los resultados del modelo (4.4) del apartado 2.3. Se estiman las probabilidades con todas las variables y luego sin la variable de estudio que corresponda. La diferencia existente entre las dos líneas equivale al efecto marginal, graficado a la izquierda.



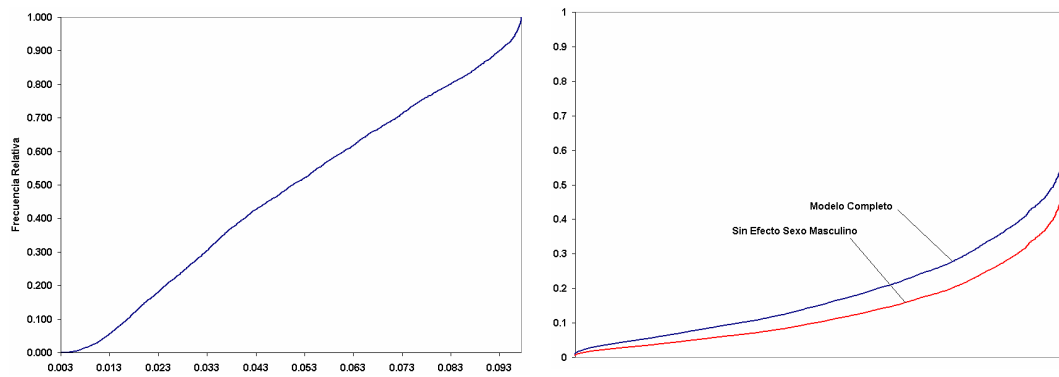
**Efecto Marginal de la Renta**



**Efecto Marginal del Acceso a Internet en la Vivienda**



Efecto Marginal del Uso de Internet



Efecto Marginal del Sexo Masculino

**APÉNDICE 2: FUNCIÓN LINEAL DEL GASTO CON Y SIN CONSTANTES ESPECÍFICAS (BLM)**

Variable Dependiente	log( <i>G</i> )	
	Observaciones	614
R <sup>2</sup> corregido	0,053	0,260
Error estándar de la estimación	1,341	1,184
CMSPE robusto	1,348	1,195
AIC	0,594	0,356
Constante	2,84 (4,26)	4,09 (6,21)
log( <i>EDAD</i> )	0,77 (4,37)	0,51 (3,25)
log( <i>IS</i> )	0,83 (3,06)	0,48 (1,19)
<i>SEXO_H</i>	0,17 (1,51)	0,02 (0,17)
<i>G<sub>I</sub></i>	---	-0,76 (3,19)
<i>G<sub>II</sub></i>	---	2,35 (8,01)
<i>G<sub>III</sub></i>	---	-0,75 (3,99)
<i>G<sub>IV</sub></i>	---	---

**APÉNDICE 3: FUNCIÓN LINEAL DE GASTO RETINA (URM)**

Observaciones	614	
R <sup>2</sup> ajustado	0,779	
Error estándar de la estimación	0,647	
RCMSPE	0,689	
AIC	-0,813	
<b>Variable</b>	<b>Coefficiente</b>	<b>Estadístico t</b>
Constante	7,87	5,65
Interacciones		
$1/(\log(EDAD))^2$	-29,28	3,70
$(\log(EDAD))^2$	-0,19	3,24
$1/\log(IS)$	-0,12	1,59
Pendientes Específicas		
$G_{II}/\log(EDAD)$	-3,73	5,53
$P_1/\log(EDAD)$	0,72	2,71
$P_1/[\log(EDAD)]^2$	-0,12	1,58
$P_2/[\log(EDAD) * \log(IS)]$	4,40	2,77
$P_2/[\log(EDAD)]^2$	-0,20	2,37
$P_2 * [\log(EDAD)/\log(IS)]$	-0,43	2,46
$P_2 * \log(IS)$	-9,54	4,35
$P_2 * [\log(IS)]^2$	-6,94	4,33
$P_3/\log(EDAD)$	3,04	13,63
$P_3 * \log(EDAD) * \log(IS)$	-0,45	2,01
$P_4 * [\log(EDAD)/\log(IS)]$	-0,11	1,46
$P_4/[\log(EDAD)]^2$	-0,07	0,90
$P_5 * \log(EDAD) * \log(IS)$	-0,67	2,91
$P_5/[\log(EDAD)]^2$	7,85	2,66
$P_5 * [\log(IS)/\log(EDAD)]$	6,04	1,87
$P_6 * \log(EDAD) * \log(IS)$	-0,21	3,36
$P_6/[\log(EDAD) * \log(IS)]$	-0,87	3,98
$P_7/\log(EDAD)$	2,59	2,57
$P_7 * \log(EDAD) * \log(IS)$	-0,14	1,84
$P_7/[\log(EDAD) * \log(IS)]$	1,47	1,23
$P_8 * [\log(EDAD)]^2$	0,25	1,97

$P_8 / [\log(EDAD)]^2$	100,78	1,92
$P_8 / \log(EDAD)$	-37,06	1,79
$P_9 * [\log(EDAD) / \log(IS)]$	-0,07	5,46
$P_9 * \log(IS)$	-0,95	6,21
$P_{10} / [\log(EDAD) * \log(IS)]$	-3,15	2,75
$P_{10} / \log(IS)$	0,90	2,49
$P_{10} * [\log(EDAD)]^2$	0,08	4,75
$P_{11} * [\log(EDAD)]^2$	0,05	2,63

# CAPÍTULO V

## Conclusiones

## **1. INTRODUCCIÓN**

La presente tesis analiza algunos aspectos de la economía digital en España, tomando como referencia el fenómeno Internet. En cuanto al área que compete a esta investigación, la economía digital, hace relativamente poco tiempo que se inició su estudio teórico sobre todo desde la perspectiva de la organización industrial y la regulación. Los primeros trabajos aplicados se citan en el Capítulo 1 en donde se observa que tienen una antigüedad de no más de 10 años. Concretamente en España se han empezado a realizar encuestas y recogidas de bases de datos estadísticos hace poco más de cuatro años, pero no se conocen estudios econométricos que los utilicen. Esta tesis pretende contribuir al estudio de la Nueva Economía en los ámbitos económico y sociocultural.

La contribución de esta tesis ha consistido en la investigación de tres fenómenos derivados directamente de lo anterior:

- Los determinantes del acceso y uso de Internet.
- La medición y análisis de una brecha tecnológica.
- La demanda de bienes y servicios por Internet.

Para su realización se utilizaron datos españoles a nivel individual y por Comunidades Autónomas. La metodología utilizada es susceptible de ser adaptada a otros datos y a otros ámbitos.

Las principales conclusiones se resumen en los próximos apartados que a continuación.

## 2. CONCLUSIONES DEL ACCESO Y USO DE INTERNET

Se considera a Internet no solo un fenómeno tecnológico sino también como un fenómeno social. Se encontró que en 2003 el 25,2% de los que declaran que acceden a Internet desde el hogar en España, sólo el 35,5% lo hace a través de una conexión por banda ancha. Sin embargo, estos porcentajes varían considerablemente por Comunidades Autónomas. Además, en un primer análisis gráfico se observan relaciones entre características sociodemográficas y el uso de Internet tal como ocurre en otros países.

### 2.1. Modelo para el Acceso a Internet desde el Hogar

- Se evidencia la importancia directa del nivel de renta y los atributos tecnológicos tales como el uso del PC o el poseer y usar el teléfono móvil. Los impactos negativos los provocan algunos determinantes sociodemográficos tales como el tamaño del hábitat y la edad.
- El modelo presenta buenos resultados de ajuste global tanto en significatividad como de bondad de ajuste. Los resultados del contraste de razón de verosimilitudes y el Pseudo  $R^2$  llevan a afirmarlo.
- El modelo tiene un 85,48% de aciertos totales, con un 59,50% de sensibilidad (aciertos iguales a uno), que es un porcentaje razonablemente alto dado el ajuste del modelo y el escaso porcentaje de poseedores de banda ancha en el hogar en la muestra total. Esto se observa en el cuadro de validación cruzada. Significará que los regresores propuestos tienen una

relación apreciable con la probabilidad de adquirir una conexión a Internet por banda ancha para el hogar.

## 2.2. Modelo para el Uso de Internet

- En este modelo también se utilizan los mismos atributos (económicos, tecnológicos y sociodemográficos) como influencia en el uso de Internet, medido este como la cantidad de sitios en los que se declara usar el servicio. Se observa que la mayoría de las variables tendrán efectos positivos en el uso de Internet.
- La renta tiene un efecto positivo importante debido a que al ser mayor provoca un aumento en la probabilidad de adquisición de bienes y servicios. Es decir que se podría asimilar a Internet a un bien normal o un bien de lujo, pero no a un bien de primera necesidad.
- Otros efectos positivos en el uso de Internet serán:
  - i. Los provocados por atributos tecnológicos tales como la posesión de un acceso a Internet en el hogar, debido a que tener el servicio disponible en el hogar crea un hábito de uso que lleva a utilizar cada vez más el servicio.
  - ii. Lo mismo ocurre con la frecuencia de uso del ordenador y con atributos sociodemográficos tales como encontrarse estudiando, o la pertenencia al género masculino, todos estos también con efectos positivos.

### 3. CONCLUSIONES DE LA MEDICIÓN Y ANÁLISIS DE LA BRECHA TECNOLÓGICA

Aquí primeramente se observa que la importancia aparente de la brecha digital entre países contrasta con la relativa escasez de estudios disponibles. Las principales preguntas que se intentan responder son:

¿Qué es y cómo se mide la brecha digital?

¿A qué se debe que exista?

¿Cómo se puede disminuir?

Se considera que, así como existen diferencias en el acceso a las Nuevas Tecnologías a nivel intercontinental, o entre distintos países, dicha diferencia también se da a nivel de Comunidad Autónoma en España e incluso a nivel individual. Generalmente el uso de Internet se considera un buen indicador del uso de la tecnología digital, y primeramente al observar los primeros gráficos se puede ver que algunas Comunidades Autónomas tienen mayor porcentaje de usuarios que otras. Es un primer indicio de la existencia de diferencias en el grado de digitalización de las mismas.

Primeramente, para medir estas diferencias, se propone un índice sintético que dependerá de los pesos que se utilicen, que medirá el consumo en tecnología y el nivel de digitalización, según la ponderación que se utilice, donde se destaca que Madrid es la Comunidad Autónoma con mayor consumo de tecnología y mayor nivel de digitalización. Le siguen Cataluña, el País Vasco y Canarias. Las Comunidades que menos consumen tecnología y que en términos relativos están menos digitalizadas son Extremadura, Galicia y Castilla-La Mancha.

La media del índice de consumo en equipamiento tiene un valor mayor que la del nivel de digitalización. Esta discrepancia, en primer lugar se da por las diferencias existentes en las ponderaciones de los dos índices. En el caso de la medición del nivel de digitalización por ejemplo, al dársele un mayor peso a Internet de banda ancha y otros componentes más avanzados, los valores del indicador tienden a disminuir. Esto evidencia la diferencia existente entre el consumo de cualquier tipo de tecnología, y la “inclinación” hacia aquellas más avanzadas.

Luego, considerando una medida de brecha tecnológica como la dispersión entre los valores de los índices, si la misma se mide en cuanto al consumo destinado a equipamiento (ITIC) será menor que la existente en cuanto a nivel de digitalización y afinidad con las nuevas tecnologías (IDD). Esto también se lo puede observar en la suma total de los porcentajes absolutos de variación con respecto a las medias (55,31% y 107,02%). Esto se debe a que un incremento en el nivel de renta per capita como el que experimentaron las regiones de España después de su ingreso en la UE se tradujo, como es lógico, en un mayor consumo en equipamiento. Sin embargo el nivel de digitalización tendrá otros determinantes además de la renta.

Para especificar adecuadamente un modelo para medir el efecto de los determinantes que provocan las diferencias en las Comunidades Autónomas, se recogerá la heterogeneidad a partir del agrupamiento de las Comunidades Autónomas de acuerdo a la elasticidad del uso de Internet respecto al gasto en tecnología, la edad y el nivel de estudios. A partir de este planteamiento se acabarán obteniendo tres grupos de Comunidades con elasticidades similares.

Una vez recogida la heterogeneidad, se modeliza econométricamente el nivel de digitalización individual en función de lo que se consideran sus determinantes. La hipótesis que se quiere comprobar es que la brecha tecnológica es en realidad el

resultado de otras brechas existentes tales como la brecha educativa, la brecha de edad o la brecha de la renta, entre otras. Es decir que las desigualdades estructurales existentes otros ámbitos como el económico o el sociodemográfico contribuirán en gran manera a incrementar la brecha tecnológica en España. La comprobación de este supuesto luego se abordará desde tres líneas de análisis descritas a continuación.

### **3.1. Modelo de regresión múltiple de los determinantes del nivel de digitalización individual en España**

- La elasticidad de la renta per capita es significativa y tiene el signo positivo esperado. El caso de la variable referida a la edad del individuo la elasticidad tiene signo negativo y es igual a 0,07. En la literatura la edad suele ser una variable con efecto inverso en el acceso y uso de nuevas tecnologías. Sin embargo en muestras de individuos más familiarizados con las Nuevas Tecnologías, la edad suele tener un efecto positivo.
- En cuanto al nivel de estudios alcanzados, la variable será significativa con un efecto positivo.
- El caso de las variables referidas a la frecuencia de uso del ordenador y conexión a Internet, el efecto será que la frecuencia de conexión a Internet tendrá un efecto positivo de 0,09.
- Para el caso de la pertenencia al género masculino, el efecto será significativo y positivo de 0,01.

- Para el caso de que el individuo se halle estudiando en ese momento, sólo será significativa en la muestra condicionada a los usuarios con un valor positivo de 0,03.

### **3.2. Modelo de regresión Cuantílica de los Determinantes del Nivel de Digitalización Individual en España**

La justificación adicional del uso de esta metodología se centra en responder a la pregunta de si deben las políticas de reducción de brecha digital centrarse en las mismas causas para toda la población y si los determinantes observables del nivel de digitalización individual son estadísticamente significativos y tienen todos los mismos impactos entre distintas porciones de la distribución del nivel de digitalización. Las conclusiones son las siguientes:

- Se observa que la situación es distinta del resto para los cuantiles 0,10 y 0,25 referidos a los valores más bajos de digitalización individual. Por ejemplo, la capacidad productiva provincial sólo será significativa en el cuantil 0,10, mientras que la renta per capita lo será además solo en el cuantil 0,25. Es decir que, en las provincias o regiones en donde el aporte al PIB nacional se incremente el 1% hará disminuir el nivel de digitalización un 0,17% *sólo* en los individuos con niveles más bajos del mismo. En el resto de la distribución esta variable no influye.
- El caso de la renta per capita tendrá elasticidades de 0,40 en el cuantil 0,10 y de 0,12 en el cuantil 0,25. En el resto de los cuantiles tampoco influirá. La curiosidad aquí es que a medida que se va aumentando el

nivel de digitalización individual (es decir, pasándose de cuantiles inferiores a superiores) estas variables pierden significatividad.

- Al observar las elasticidades con respecto al nivel de estudios alcanzado, se observa que ocurre lo opuesto, ya que los valores de elasticidad serán mayores en los cuantiles superiores de la distribución (0,21% tanto para el cuantil 0,75 y 0,90).
- La variable edad no será significativa en el cuantil más alto de la distribución. Lo contrario ocurre con el cuantil más bajo, en donde el impacto es el mayor de todos con un valor negativo de 0,11%.
- Luego, las variables referidas a la frecuencia de uso del ordenador y de conexión a Internet tendrán a lo largo de toda la distribución impactos positivos, con más o menos intensidad según qué cuantil se trate. Es el caso de la frecuencia del uso del PC en donde la mayor elasticidad se encuentra en la mediana (cuantil 0,50) con un valor de 0,09%. Para el caso de la frecuencia de conexión a Internet el mayor impacto será en el cuantil 0,25 con un valor de 0,11%.

### **3.3. Modelo de Previsión del Nivel de Digitalización Individual en España**

En el modelo anterior se evidencia la posibilidad de que la brecha digital sea en realidad la consecuencia de brechas existentes de edad, de educación o de renta. Esto es importante para la toma de decisiones de política y para ello se plantea un modelo de predicción que permita evidenciar la reducción de la brecha digital obtenida al reducir las otras brechas consideradas.

Se usa RETINA para especificar un modelo con mejor capacidad predictiva permitiendo la inclusión de no linealidades e interacciones. Entre las conclusiones de este último modelo se puede decir lo siguiente:

- En cuanto a la sensibilidad del grado de digitalización respecto a la renta, cabe notar que si esta fuera la misma para todos los usuarios de Internet (renta promedio), el total de las diferencias de los valores del índice de cada CA respecto de la media total disminuiría desde 107,02% a 72,60%. Es decir que utilizando la renta promedio se disminuiría la brecha tecnológica en un 34,42%. O sea que igualar a su media todas las observaciones del nivel de renta provoca un descenso en la brecha digital y demuestra que es una variable muy relacionada con el fenómeno TIC.
- También se observa que las mayores diferencias positivas son provocadas en las comunidades de mayor nivel de digitalización tales como Madrid (12,8%) y Cataluña (6,3%), y las negativas en las comunidades de menor digitalización tales como Extremadura (-8,4%) y Castilla-La Mancha (-8,0%). Esto podría deberse a que regiones con menor tendencia a lo digital sean a su vez regiones con bajo nivel de renta per capita y que obviamente serán menos propensos al uso de las TIC que los usuarios de regiones de mayor nivel de renta.
- Si bien algo parecido ocurrirá con el resto de determinantes considerados tales como la Capacidad Productiva, el Nivel de Estudios, la Edad y la Población, que al ser estas variables reemplazadas por sus valores medios provocan reducciones en la brecha digital de 12,82%, 16,12%, 14,82% y 12,02% respectivamente, en todos los casos los

mayores impactos por encima de la media serán Madrid y Cataluña, mientras que los más bajos serán Extremadura y Castilla-La Mancha. Sin embargo los efectos serán distintos y variarán de acuerdo a la Comunidad considerada.

#### **4. CONCLUSIONES DEL ANÁLISIS DEL COMERCIO ELECTRÓNICO**

En primer lugar se destacan las diferencias existentes entre los mercados convencionales y los mercados virtuales. Entre otras, se tiene que habrá un menor coste de transacción, de búsqueda y selección adversa, y por lo tanto menos fricciones<sup>1</sup>. Luego de un análisis descriptivo, para analizar estos temas se ha propuesto tres modelos cuyas conclusiones se detallan a continuación.

##### **4.1. Modelo para los Impactos Sociodemográficos en el Comercio Electrónico**

- Se especifica un modelo logit binario cuya variable dependiente es una variable ficticia que será igual a 1 cuando el individuo responde a que ha comprado en Internet durante los últimos tres meses. Exceptuando la cantidad de miembros en el hogar, todas las demás variables tendrán efectos positivos en la probabilidad de comprar por Internet. El acceso a Internet en el hogar será la variable más influyente. Esto último tiene lógica al considerar que antes de comprar por Internet, primero hay que conectarse a la Red, o bien desde el hogar, o bien desde otros sitios.

---

<sup>1</sup> Una *fricción de mercado* serían todas las circunstancias que propician que el precio no esté en equilibrio.

- También se observa que el pertenecer al sexo masculino tiene una gran influencia en la probabilidad de comprar por Internet. Esto viene a confirmar lo observado en los informes recientes donde se expone la mayor afinidad del género masculino por las compras a través de la Red. En cuanto a la renta, su influencia también es positiva. Tendrá también su lógica debido a que al haber una mayor renta habrá una mayor capacidad de compra. Sin embargo no es la variable que más influye aunque no deja de ser positiva su influencia en la probabilidad de comprar por Internet.
- Luego también el hábitat poblacional y las intensidades de uso de Internet y el ordenador también impactarán positivamente en la probabilidad de compra por Internet, aunque de manera bastante menor que las detalladas en los puntos anteriores.
- También para el caso de la variable referida a la edad la interpretación es particular debido a que está incluida dentro del modelo tanto linealmente como al cuadrado. A medida que aumenta la edad aumenta la probabilidad de comprar por Internet, pero cada vez menos. La edad en donde empieza a decrecer la probabilidad de acuerdo a la muestra y a estos resultados es sobre los 45 años.
- La única variable que disminuye la probabilidad de comprar bienes y servicios en Internet será el total de miembros en el hogar. Es decir que a mayor cantidad de miembros en el hogar, menor probabilidad de adquirir bienes en la Red.

## 4.2. Modelos para las Transacciones a través de Internet y Previsión del Gasto

También se propusieron dos modelos adicionales para analizar otros aspectos de la demanda:

- **Un primer modelo para la cantidad de transacciones en Internet:**

Este caso correspondería a los determinantes de una función de demanda, ya que se refiere a transacciones de compra de bienes y contratación de servicios realizadas a través de la Red. Por las características de la variable endógena (cantidad de bienes adquiridos, de 0 a 12) se especifica y estima una regresión de Poisson en cuyos resultados se ve que el número esperado de e-transacciones se incrementa notablemente si el pago se realiza con tarjeta de crédito. También se observa que la incidencia del resto de las formas de pago, aunque bastante menor que la que produce el pago con tarjeta de crédito, es importante. En cuanto al nivel de renta también tiene una incidencia positiva. Además se ve que los atributos sociodemográficos como la edad, el nivel de estudio o la población tendrán una escasa influencia en la cantidad esperada de comercio en la Red.

- **Un segundo modelo para la previsión del gasto:** El fin ahora es conseguir un modelo para previsión que entre otras cosas además mejorará la especificación. Para ello se utilizará RETINA, con el cual se obtendrá un modelo con buena capacidad predictiva por fuera de la muestra. La inclusión de constantes y pendientes específicas como regresores ha mejorado el modelo del gasto del consumidor español en Internet, que se nota a partir de la mejor previsión fuera de la muestra,

medida por el AIC y la bondad de ajuste medida por el  $\bar{R}^2$ . Curiosamente se observa que los regresores originales sólo entrarán en el modelo de previsión dentro de interacciones o pendientes específicas.

## 5. FUTURAS LÍNEAS DE INVESTIGACIÓN

A partir de la cada vez mayor información estadística, las futuras líneas de investigación podrían tender a responder para España las mismas preguntas planteadas en la literatura para otros países:

- a. Obligación del Servicio Universal: ¿Por qué algunos hogares y pequeñas empresas no se suscriben al servicio de Internet de banda ancha estando este al alcance? ¿Tiene sentido, y en su caso, cómo incluir Internet de Banda Ancha dentro del Servicio Universal?
- b. Conocimiento y disponibilidad del servicio de Internet: ¿Cómo evalúa el consumidor la calidad del servicio de Internet de banda ancha dentro de los alcanzables (ADSL, Cable módem, wi-fi, etc)?
- c. Precio del servicio de Internet: ¿Es realmente el precio el atributo más importante?, ¿es el coste del servicio de Internet lo que más valoran los consumidores a la hora de suscribirse, o también lo serán otros atributos?

- d. Dinámica de la brecha digital: ¿La brecha digital va en aumento?, ¿las políticas aplicadas para su reducción son efectivas?, ¿existen cambios estructurales que hacen que no siempre dependa de los mismos determinantes?
- e. Dinámica del comercio electrónico en España: ¿Se evoluciona igual que en otros países hacia el comercio virtual? ¿Son los mismos determinantes de otros países los que provocan que en España se compre o venda a través de Internet? ¿Están los mercados virtuales debidamente regulados de manera tal que lleven al consumidor a hacer uso del comercio virtual? ¿Qué se puede decir sobre las externalidades y fricciones provocadas por los mercados virtuales?
- f. Teletrabajo: ¿Se tiende al teletrabajo en España tal como está ocurriendo en otros países avanzados? ¿Son más o menos productivas las horas trabajadas? ¿Proveen las empresas españolas un buen soporte a sus teletrabajadores como ocurre en otros países? ¿es la vinculación con la tecnología avanzada el principal atributo del teletrabajo? ¿Existe relación entre brecha digital y teletrabajo?

Por último, en el apartado siguiente se expone un cuadro resumen de todos los modelos especificados y estimados en esta tesis.

## CUADRO RESUMEN DE LAS ECUACIONES ESTIMADAS

ECUACIÓN – MODELO	OBJETIVOS	RESPUESTA	CARACTERÍSTICAS GENERALES
<u>Acceso a Internet desde el Hogar</u> (Capítulo II, sección 4.1)	<ul style="list-style-type: none"> <li>• Determinantes en España</li> <li>• Descubrimiento de Impactos Económicos, Tecnológicos y Sociodemográficos</li> </ul>	Se evidencia la importancia directa del nivel de renta y los atributos tecnológicos.	Modelo Probit Binario con sesgo de selectividad Muestra: 18.948 individuos, 4.470 con conexión a Internet en el hogar. LR = -9229,138 (P-valor = 0,00) Pseudo R <sup>2</sup> = 0,4591
<u>Uso de Internet general</u> (Capítulo II, sección 4.2)	<ul style="list-style-type: none"> <li>• Determinantes en España</li> <li>• Descubrimiento de Impactos Económicos, Tecnológicos y Sociodemográficos</li> </ul>	La mayoría de los coeficientes indicarán efectos positivos en la cantidad de sitios de uso de Internet.	Modelo de Poisson inflado de ceros Muestra: 18.816 individuos LR = 1.140,72 (P-valor = 0,00) Pseudo R <sup>2</sup> = 0,3912 Vuong = 39,69 (P-valor = 0,00)
<u>Índice de Digitalización</u> <u>Índice de Consumo en Tecnología</u> (Capítulo III, sección 2.1)	<ul style="list-style-type: none"> <li>• Propuesta de medición de la brecha digital entre las CCAA de España</li> <li>• Aportación a la búsqueda de políticas acordes de reducción de diferencias entre CCAA.</li> </ul>	Se encuentra clara evidencia de diferencias existentes entre las CCAA en donde algunas (Madrid, Cataluña, etc) son bastante superiores en tecnología y digitalización que otras (Extremadura, Galicia, etc)	Índice sintético creado a partir de “cestas tecnológicas” proporcionando un peso a cada una. $I_{mc_i} = \left\{ \sum_{j=1}^n \left[ q_{ij} D_{ij} \prod_{j=1}^k (1 + q_{ijk} D_{ijk}) \right] \right\}$ <small>i=1, ..., 18.948</small>
<u>Uso de Internet por CCAA</u> (Capítulo III, secciones 3.1 y 3.2)	<ul style="list-style-type: none"> <li>• Recogimiento de la heterogeneidad a partir del agrupamiento de CCAA por elasticidades del uso de Internet respecto al consumo de tecnología, la edad y el nivel de estudios.</li> </ul>	Se obtiene un conglomerado de tres grupos de CCAA con elasticidades similares.	Modelo de Poisson para cada CCAA (en total 18 con Ceuta y Melilla juntas). En todas las regresiones hay buenas medidas de bondad de ajuste y significatividad individual y global.
<u>Nivel de Digitalización Individual I</u> (a nivel global y a nivel de usuario de Internet) (Capítulo III, sección 4.1)	<ul style="list-style-type: none"> <li>• Estudio de los determinantes del nivel de digitalización individual</li> </ul>	Derivará en conclusiones similares o diferentes según se trate de una muestra completa o restringida sólo a los usuarios de Internet	Modelo MCO con estimador consistente de la varianza (White) Muestra: 18.796 individuos R <sup>2</sup> = 0,424 AIC = -2,058 RCMSPE = 0,358 F (P-valor) = 0,000
	<ul style="list-style-type: none"> <li>• Comprobar si las políticas de</li> </ul>	Se observa que la situación es distinta del	Regresión cuantílica para los cuantiles

<p><u>Nivel de Digitalización Individual II</u> (en distintas partes de la distribución) (Capítulo III, sección 4.2)</p>	<p>reducción de brecha digital deben ser la mismas para toda la población</p> <ul style="list-style-type: none"> <li>• Evaluar los impactos de los determinantes del nivel de digitalización en distintas partes de la población.</li> </ul>	<p>resto para los cuantiles 0,10 y 0,25 referidos a los valores más bajos de digitalización individual</p>	<p>0,10, 0,25, 0,50, 0,75 y 0,90 Muestra: 18.796 individuos Pseudo R<sup>2</sup> (0,10) = 0,1438 Pseudo R<sup>2</sup> (0,25) = 0,2213 Pseudo R<sup>2</sup> (0,50) = 0,3229 Pseudo R<sup>2</sup> (0,75) = 0,3061 Pseudo R<sup>2</sup> (0,90) = 0,1750</p>
<p><u>Nivel de Digitalización Individual III</u> (modelo de previsión) (Capítulo III, sección 4.3)</p>	<ul style="list-style-type: none"> <li>• Comprobar la hipótesis de que la brecha digital es el resultado de la existencia de otras <i>brechas</i></li> </ul>	<p>Tener en cuenta la heterogeneidad y la inclusión de no linealidades en las especificaciones mejora tanto el ajuste como la capacidad predictiva del modelo. Se observa que gran parte de la brecha digital es explicada por la existencia de otras brechas</p>	<p>Estimación RETINA Parámetros: 17 Transformaciones (W): 10 Pendientes Específicas: 6 <math>\bar{R}^2 = 0,444</math> RCMSPE(1000)= 0,352 AIC = -2,089</p>
<p><u>Compra de Bienes y Servicios en Internet</u> (Capítulo IV, sección 2.3)</p>	<ul style="list-style-type: none"> <li>• Determinantes de la compra a través de la Red.</li> <li>• Búsqueda de un perfil del comprador por Internet</li> </ul>	<p>Exceptuando el término constante y la cantidad de miembros en el hogar, todas las demás variables consideradas en el modelo tendrán efectos positivos en la probabilidad de comprar por Internet</p>	<p>Modelo Logit Binario Observaciones: 5.223 Log-Verosimilitud: -1773,45 <math>\chi^2_{(10)} : 397,09</math> (Prob. = 0,000) Pseudo R<sup>2</sup>: 0,1007</p>
<p><u>Demanda de Transacciones en Internet</u> (Capítulo IV, sección 3.2)</p>	<ul style="list-style-type: none"> <li>• Estimación de una función de demanda de bienes y servicios a través de la Red</li> <li>• Medición de los determinantes</li> </ul>	<p>La demanda de transacciones aumentará al pagar con tarjeta de crédito u otras formas de pago alternativas. El nivel de renta también tiene una incidencia positiva.</p>	<p>Modelo de Poisson Muestra: 5.218 Log-Verosimilitud: -1.914,34 <math>\chi^2_{(13)} : 3635,19</math> (P-valor = 0,000) Pseudo R<sup>2</sup>: 0,4870</p>
<p><u>Función Lineal de Gasto en Internet</u> (Capítulo IV, sección 3.3)</p>	<ul style="list-style-type: none"> <li>• Estimación de una función de gasto a través de la Red</li> <li>• Medición de los determinantes</li> </ul>	<p>La inclusión de constantes y pendientes específicas como regresores ha mejorado la explicación de las variaciones del gasto en Internet, y eso se nota a partir del mejoramiento de la previsión fuera de la muestra, el AIC y la bondad de ajuste.</p>	<p>Estimación RETINA Parámetros: 33 Transformaciones (W): 3 Pendientes Específicas: 29 <math>\bar{R}^2 = 0,779</math> RCMSPE (1000) = 0,689 AIC = -0,813</p>

# ANEXO

Base de Datos

Diseño de Encuesta

# Diseño de Registro Encuesta TIC-H 2003

Versión: 16-Enero-2004

## Tabla CUESTIONARIO

NOMBRE	TAMAÑO	Posición		PREGUNTA	DESCRIPCION	VALOR ES	VALORES VÁLIDOS
		Inicial	Posición Final				
CIPRO	2	3	3	P1.1	LADILLOS	52	1=SI / 6=NO
TELEV	1	4	4	P9	Identificar por de la vivienda	2	1=SI / 6=NO
RECEP1	1	4	4	P9	Recibe canales a través de antena convencional	3	1=SI / 6=NO/ blanco
RECEP2	1	5	5	P10	Recibe canales a través de antena parabólica	3	1=SI / 6=NO/ blanco
RECEP3	1	6	6	P11	Recibe canales a través de cable	3	1=SI / 6=NO/ blanco
RECEP4	1	7	7	P12	Recibe canales de televisión digital terrestre	3	1=SI / 6=NO/ blanco
ORDE1	1	8	8	P13	Ordenador de sobremesa en la vivienda	2	1=SI / 6=NO
ORDE2	1	9	9	P14	Algún residente de la vivienda tiene ordenador portátil	2	1=SI / 6=NO
ORDE3	1	10	10	P15	Algún residente de la vivienda tiene otro tipo de ordenador	2	1=SI / 6=NO
TELEF1	1	11	11	P16	Teléfono fijo ( se incluyen inalámbricos ) en la vivienda	2	1=SI / 6=NO
TELEF2	1	12	12	P17	Algún residente de la vivienda tiene teléfono móvil	2	1=SI / 6=NO
MUSIC1	1	13	13	P18	Cadena musical, equipo de alta fidelidad o laserdisk	2	1=SI / 6=NO
MUSIC2	1	14	14	P19	Radio en la vivienda	2	1=SI / 6=NO
VIDEO	1	15	15	P20	Video en la vivienda	2	1=SI / 6=NO
DVD	1	16	16	P21	DVD en la vivienda	2	1=SI / 6=NO
FAX	1	17	17	P22	Fax en la vivienda	2	1=SI / 6=NO
AUT_NAV	1	18	18	P23	Algún residente de la vivienda tiene automóvil con sistema de navega	2	1=SI / 6=NO
VIV_INTER	1	19	19	P24	Dispone de acceso a Internet en la vivienda	2	1=SI / 6=NO
INTENO1	1	20	20	P25.1	Motivo1 de NO disponer acceso a Internet en la vivienda	4	1=SI / 6=NO / 9=NS/NR/blanco
INTENO2	1	21	21	P25.2	Motivo2 de NO disponer acceso a Internet en la vivienda	4	1=SI / 6=NO / 9=NS/NR/blanco
INTENO3	1	22	22	P25.3	Motivo3 de NO disponer acceso a Internet en la vivienda	4	1=SI / 6=NO / 9=NS/NR/blanco
INTENO4	1	23	23	P25.4	Motivo4 de NO disponer acceso a Internet en la vivienda	4	1=SI / 6=NO / 9=NS/NR/blanco
INTENO5	1	24	24	P25.5	Motivo5 de NO disponer acceso a Internet en la vivienda	4	1=SI / 6=NO / 9=NS/NR/blanco
L_INTENO6	120	25	144	P25.6.1	Especificación de otros motivos de No acceso a Internet en vivienda	-	Texto/ blanco
INTENO6	1	145	145	P25.6.2	Motivo6 de NO disponer acceso a Internet en la vivienda (otros motiv	4	1=SI / 6=NO / 9=NS/NR/blanco
INTEFOR1	1	146	146	P26.1	Forma de conexión a Internet en la vivienda (1)	4	1=SI / 6=NO / NS/NR=9 / blanco
INTEFOR2	1	147	147	P26.2	Forma de conexión a Internet en la vivienda (2)	4	1=SI / 6=NO / NS/NR=9 / blanco
INTEFOR3	1	148	148	P26.3	Forma de conexión a Internet en la vivienda (3)	4	1=SI / 6=NO / NS/NR=9 / blanco
INTEFOR4	1	149	149	P26.4	Forma de conexión a Internet en la vivienda (4)	4	1=SI / 6=NO / NS/NR=9 / blanco
L_INTEFOR5	120	150	269	P26.5.1	Especificación de otras formas de conexión a Internet	-	Texto/ blanco
INTEFOR5	1	270	270	P26.5.2	Forma de conexión a Internet en la vivienda (5, otras formas)	4	1=SI / 6=NO / NS/NR=9 / blanco
EQINTE1	1	271	271	P27.1	Equipo utilizado para conexión a Internet en la vivienda (1)	4	1=SI / 6=NO / NS/NR=9 / blanco
EQINTE2	1	272	272	P27.2	Equipo utilizado para conexión a Internet en la vivienda (2)	4	1=SI / 6=NO / NS/NR=9 / blanco
EQINTE3	1	273	273	P27.3	Equipo utilizado para conexión a Internet en la vivienda (3)	4	1=SI / 6=NO / NS/NR=9 / blanco
EQINTE4	1	274	274	P27.4	Equipo utilizado para conexión a Internet en la vivienda (4)	4	1=SI / 6=NO / NS/NR=9 / blanco
L_EQINTE5	120	275	394	P27.5.1	Especificación de equipo utilizado para conexión a Internet en la vivie	-	Texto/ blanco
EQINTE5	1	395	395	P27.5.2	Equipo utilizado para conexión a Internet en la vivienda (5, otras form	4	1=SI / 6=NO / NS/NR=9 / blanco
PROSEG1	1	396	396	P28.1	Programas de seguridad instalados (1)	4	1=SI / 6=NO / NS/NR=9 / blanco
PROSEG2	1	397	397	P28.2	Programas de seguridad instalados (2)	4	1=SI / 6=NO / NS/NR=9 / blanco
PROSEG3	1	398	398	P28.3	Programas de seguridad instalados (3)	4	1=SI / 6=NO / NS/NR=9 / blanco
L_PROSEG4	120	399	519	P28.4.1	Especificación de otros programas de seguridad	-	Texto/ blanco
PROSEG4	1	519	519	P28.4.2	Programas de seguridad instalados (4, otros programas)	4	1=SI / 6=NO / NS/NR=9 / blanco
UTIL_OR	1	520	520	P29	Utilización de ordenador (últimos 3 meses)	3	1=SI / 6=NO/ NS/NR=9
CUR_OR	1	521	521	P30	Realización de curso relativo uso ordenadores (al menos 4 horas)	4	1=SI / 6=NO / NS/NR=9 / blanco
OTR_CUR	1	522	522	P31	¿Cuándo fue el último curso de estas características? 1, 6, 9 / blan	4	1, 6, 9 / blanco
FREC_OR	1	523	523	P32	Frecuencia de uso de ordenador (últimos 3 meses)	5	1 - 4 / blanco
LUGAR1	1	524	524	P33.1	Lugar de utilización de ordenador (últimos 3 meses) (1- vivienda)	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGAR2	1	525	525	P33.2	Lugar de utilización de ordenador (últimos 3 m) (2- centro de trabajo)	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGAR3	1	526	526	P33.3	Lugar de utilización de ordenador (últimos 3 meses) (3- centro de estu	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGAR4	1	527	527	P33.4	Lugar de utilización de ordenador (últimos 3 meses) (4- otro lugar)	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG1	1	528	528	P34.1	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(1- texto	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG2	1	529	529	P34.2	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(2- calcul	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG3	1	530	530	P34.3	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(3- base de	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG4	1	531	531	P34.4	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(4- pres	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG5	1	532	532	P34.5	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(5- visor	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG6	1	533	533	P34.6	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(6- juegc	4	1=SI / 6=NO / NS/NR=9 / blanco
PROG7	1	534	534	P34.7	Prog.o aplicaciones usados en el ordenador(últimos 3 meses)(7- espe	4	1=SI / 6=NO / NS/NR=9 / blanco
L_PROG8	120	535	654	P34.8.1	Especificación de otra aplicación usada con el ordenador	-	Texto/ blanco
PROG8	1	655	655	P34.8.2	Prog.o aplicaciones usadas en el ordenador(últimos 3 meses)(8-otra	4	1=SI / 6=NO / NS/NR=9 / blanco
USO_INTER	1	656	656	P35	Ha usado Internet (últimos 3 meses)	2	1=SI / 6=NO
NOUSO1	1	657	657	P36.1	Motivos de NO usar Internet (últimos 3 meses) (1)	3	1=SI / 6=NO/ blanco
NOUSO2	1	658	658	P36.2	Motivos de NO usar Internet (últimos 3 meses) (2)	3	1=SI / 6=NO/ blanco
NOUSO3	1	659	659	P36.3	Motivos de NO usar Internet (últimos 3 meses) (3)	3	1=SI / 6=NO/ blanco
NOUSO4	1	660	660	P36.4	Motivos de NO usar Internet (últimos 3 meses) (4)	3	1=SI / 6=NO/ blanco
NOUSO5	1	661	661	P36.5	Motivos de NO usar Internet (últimos 3 meses) (5)	3	1=SI / 6=NO/ blanco
L_NOUSO6	120	662	781	P36.6.1	Especificación de motivos de NO usar Internet (últimos 3 meses)	-	Texto/ blanco
NOUSO6	1	782	782	P36.6.2	Motivos de NO usar Internet (últimos 3 meses) (6- otros motivos)	3	1=SI / 6=NO/ blanco
LUGARIN1	1	783	783	P37.1	Lugar de uso de Internet (últimos 3 meses) (1- vivienda)	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGARIN2	1	784	784	P37.2	Lugar de uso de Internet (últimos 3 meses) (2- centro de trabajo)	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGARIN3	1	785	785	P37.3	Lugar de uso de Internet (últimos 3 meses) (3- centro de estudios)	4	1=SI / 6=NO / NS/NR=9 / blanco
LUGARIN4	1	786	786	P37.4	Lugar de uso de Internet (últimos 3 meses) (4- otro lugar)	4	1=SI / 6=NO / NS/NR=9 / blanco
FREC_INTER	1	787	787	P38	Frecuencia de uso de Internet (últimos 3 meses)	6	1-4, 9 / blanco
TIM_INTER1	1	788	788	P39	Tiempo aproximado de uso de Internet (última semana)	8	1-6, 9 / blanco
TIM_INTER2	1	789	789	P40	Tiempo aproximado de uso de Internet (últimos 3 meses)	8	1-6, 9 / blanco
SERV1	1	790	790	P41.1	Servicio, correo electrónico	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV2	1	791	791	P41.2	Servicio, Chats	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV3	1	792	792	P41.3	Servicio, teléfono	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV4	1	793	793	P41.4	Servicio, mensajes	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV5	1	794	794	P41.5	Servicio, bases de datos	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV6	1	795	795	P41.6	Servicio, ventas	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV7	1	796	796	P41.7	Servicio, compras	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV8	1	797	797	P41.8	Servicio, buscar información	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV9	1	798	798	P41.9	Servicio, reservas de turismo	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV10	1	799	799	P41.10	Servicio, medios de comunicación	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV11	1	800	800	P41.11	Servicio, ocio	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV12	1	801	801	P41.12	Servicio, páginas web	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV13	1	802	802	P41.13	Servicio, descargar formularios	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV14	1	803	803	P41.14	Servicio, enviar formularios	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV15	1	804	804	P41.15	Servicio, cursos de educación	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV16	1	805	805	P41.16	Servicio, cursos dirigidos a búsqueda de empleo	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV17	1	806	806	P41.17	Servicio, otros cursos	4	1=SI / 6=NO / NS/NR=9 / blanco
SERV18	1	807	807	P41.18	Servicio, cursos de idiomas	4	1=SI / 6=NO / NS/NR=9 / blanco
L_SERV19	120	808	927	P41.19.1	Servicio, literal de otros	-	Texto/ blanco
SERV19	1	928	928	P41.19.2	Servicio, otros	4	1=SI / 6=NO / NS/NR=9 / blanco
TRA_INT1	1	929	929	P42.1	Motivo de trabajo, envío de solicitudes	4	1=SI / 6=NO / NS/NR=9 / blanco
TRA_INT2	1	930	930	P42.2	Motivo de trabajo, búsqueda de información	4	1=SI / 6=NO / NS/NR=9 / blanco
TRA_INT3	1	931	931	P42.3	Motivo de trabajo, envío y recepción	4	1=SI / 6=NO / NS/NR=9 / blanco
TRA_INT4	1	932	932	P42.4	Motivo de trabajo, comunicación	4	1=SI / 6=NO / NS/NR=9 / blanco
TRA_INT5	1	933	933	P42.5	Motivo de trabajo, acceso a servidor	4	1=SI / 6=NO / NS/NR=9 / blanco
L_TRA_INT6	120	934	1053	P42.6.1	Motivo de trabajo, literal otros	-	Texto/ blanco
TRA_INT6	1	1054	1054	P42.6.2	Motivo de trabajo, otros	4	1=SI / 6=NO / NS/NR=9 / blanco
SEGU1	1	1055	1055	P43.1	Problemas de seguridad, correos electrónicos	4	1=SI / 6=NO / NS/NR=9 / blanco
SEGU2	1	1056	1056	P43.2	Problemas de seguridad, virus	4	1=SI / 6=NO / NS/NR=9 / blanco
SEGU3	1	1057	1057	P43.3	Problemas de seguridad, fraude tarjetas de crédito	4	1=SI / 6=NO / NS/NR=9 / blanco
SEGU4	1	1058	1058	P43.4	Problemas de seguridad, ficheros o correos legales	4	1=SI / 6=NO / NS/NR=9 / blanco
L_SEGU5	120	1059	1178	P43.5.1	Problemas de seguridad, literal otros	-	Texto/ blanco

SEGU5	1	1179	-	1179	P43.5.2	Problemas de seguridad, otros	4	1=SI / 6=NO/ NS/NR=9 / blanco
COMPRAS1	1	1180	-	1180	P44	Compras a través de Internet (últimos 3 meses)	3	1=SI / 6=NO/ NS/NR=9
COMPRAS2	1	1181	-	1181	P45	Compras a través de Internet (alguna vez)	4	1=SI / 6=NO/ NS/NR=9 / blanco
CAUSANO1	1	1182	-	1182	P46.1	Causas de no comprar por Internet, prefiere en una tienda	4	1=SI / 6=NO/ NS/NR=9 / blanco
CAUSANO2	1	1183	-	1183	P46.2	Causas de no comprar por Internet, preocupa la seguridad	4	1=SI / 6=NO/ NS/NR=9 / blanco
CAUSANO3	1	1184	-	1184	P46.3	Causas de no comprar por Internet, preocupa privacidad	4	1=SI / 6=NO/ NS/NR=9 / blanco
CAUSANO4	1	1185	-	1185	P46.4	Causas de no comprar por Internet, no necesidad	4	1=SI / 6=NO/ NS/NR=9 / blanco
CAUSANO5	1	1186	-	1186	P46.5	Causas de no comprar por Internet, no buen funcionamiento	4	1=SI / 6=NO/ NS/NR=9 / blanco
L_CAUSANO6	120	1187	-	1306	P46.6.1	Causas de no comprar por Internet, literal otros	4	Texto / blanco
CAUSANO6	1	1307	-	1307	P46.6.2	Causas de no comprar por Internet, otros	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD1	1	1308	-	1308	P47.1	Productos/servicios comprados en Internet (3 últimos meses), hogar	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD2	1	1309	-	1309	P47.2	Productos/servicios comprados en Internet (3 últimos meses), música	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD3	1	1310	-	1310	P47.3	Productos/servicios comprados en Internet (3 últimos meses), libros	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD4	1	1311	-	1311	P47.4	Productos/servicios comprados en Internet (3 últimos meses), ropa	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD5	1	1312	-	1312	P47.5	Productos/servicios comprados en Internet (3 últimos meses), software	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD6	1	1313	-	1313	P47.6	Productos/servicios comprados en Internet (3 últimos meses), materia	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD7	1	1314	-	1314	P47.7	Productos/servicios comprados en Internet (3 últimos meses), electrón	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD8	1	1315	-	1315	P47.8	Productos/servicios comprados en Internet (3 últimos meses), accion	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD9	1	1316	-	1316	P47.9	Productos/servicios comprados en Internet (3 últimos meses), viajes	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD10	1	1317	-	1317	P47.10	Productos/servicios comprados en Internet (3 últimos meses), entrada	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD11	1	1318	-	1318	P47.11	Productos/servicios comprados en Internet (3 últimos meses), loterías	4	1=SI / 6=NO/ NS/NR=9 / blanco
L_PROD12	120	1319	-	1438	P47.12.1	Productos/servicios comprados en Internet (3 últimos meses), loterías	4	1=SI / 6=NO/ NS/NR=9 / blanco
PROD12	1	1439	-	1439	P47.12.2	Productos/servicios comprados en Internet (3 últimos meses), lit. otros	4	1=SI / 6=NO/ NS/NR=9 / blanco
GASTO1	5	1440	-	1444	P48.1	Gasto en Internet (3 últimos meses), en euros	10000	1-99999/ blanco
GASTO2	7	1445	-	1451	P48.2	Gasto en Internet (3 últimos meses), en pesetas	1E+07	1-9999999/ blanco
GASTO3	1	1452	-	1452	P48.3	Gasto en Internet (3 últimos meses), intervalos	10	1-9/ blanco
FOR_PA1	1	1453	-	1453	P49.1	Formas de pago a través de Internet, tarjetas	4	1=SI / 6=NO/ NS/NR=9 / blanco
FOR_PA2	1	1454	-	1454	P49.2	Formas de pago a través de Internet, transferencia	4	1=SI / 6=NO/ NS/NR=9 / blanco
FOR_PA3	1	1455	-	1455	P49.3	Formas de pago a través de Internet, contra reembolso	4	1=SI / 6=NO/ NS/NR=9 / blanco
FOR_PA4	1	1456	-	1456	P49.4	Formas de pago a través de Internet, suscripción	4	1=SI / 6=NO/ NS/NR=9 / blanco
L_FOR_PA5	120	1457	-	1576	P49.5.1	Formas de pago a través de Internet, literal otros	4	Texto / blanco
FOR_PA5	1	1577	-	1577	P49.5.2	Formas de pago a través de Internet, otros	4	1=SI / 6=NO/ NS/NR=9 / blanco
ECURSA	1	1578	-	1578	P50.1	Estudios en curso	2	1=SI / 6=NO
NIVELST	1	1579	-	1579	P50.2	Estudios terminados, código	6	1-6. 9
TRABAJA	1	1580	-	1580	P51	Trabajo en la semana pasada	2	1=SI / 6=NO
SIT_PROF	1	1581	-	1581	P52	Situación profesional en el trabajo principal	3	1 / 6 / blanco
SIT_LAB	1	1582	-	1582	P53	Situación laboral durante la semana pasada	6	1 / 5 / blanco
TOT_MIEM	2	1583	-	1584	LADILLOS	Total de registros de esta vivienda en la tabla MIEMBROS	30	1-30
TOT_MEN	1	1585	-	1585	Var. Aux.Ladillos	Total de registros de esta vivienda en la tabla MEMORES	10	0-9
SERIAL	5	1586	-	1590		Número de cuestionario	-	00001-99999
TIPO_ENTREV	1	1591	-	1591		Tipo de entrevista de la muestra	2	1=CATI, 3=PAPI
HABITAT (ESTRATO)	1	1592	-	1592	LADILLOS	Código de hábitat	7	0-6 (*)
SUBSTRATO	2	1593	-	1594	LADILLOS	Código de subhábitat	2	Afanumérico
SEXO	1	1595	-	1595	LADILLOS	Sexo de la persona informante	2	1=H, 2=M
EDAD	2	1596	-	1597	LADILLOS	Edad de la persona informante	85	15 - 99
FACTOR_HOG	3	1598	-	1606		Factor de elevación del hogar	-	4 enteros y 5 decimales
FACTOR_PERS	3	1607	-	1615		Factor de elevación del informante	-	5 enteros y 4 decimales

(\*) VARIABLE HABITAT:

- Estrato 0 : Capitales de provincia con más 500.000 habitantes
- Estrato 1 : Resto de capitales de provincia
- Estrato 2 : Municipios ( no capitales de provincia) con más de 100000 habitantes
- Estrato 3 : Municipios ( no capitales de provincia) con más de 50000 y menos de 100000 habitantes
- Estrato 4 : Municipios con más de 20000 y menos de 50000 habitantes.
- Estrato 5 : Municipios con más de 10000 y menos de 20000 habitantes.
- Estrato 6 : Municipios con menos de 10000 habitantes.

## Cuestionario

### I. IDENTIFICACIÓN Y SELECCIÓN DE LA PERSONA A ENCUESTAR

#### 1.- Identificación de la vivienda

1. Provincia _____	<input type="text"/>	<input type="text"/>
2. Nº orden de la sección _____	<input type="text"/>	<input type="text"/>
3. Identificador de la vivienda _____	<input type="text"/>	<input type="text"/>

#### 2.- Identificación del cuestionario

1. Total de cuestionarios _____	<input type="text"/>
2. Número de cuestionario _____	<input type="text"/>

#### Instrucciones para el Entrevistador:

**En primer lugar:** Lea lo que sigue al informante (y pase luego a cumplimentar la Tabla de Miembros de la Vivienda, con las personas que el informante le vaya enumerando): **‘En primer lugar dígame el nombre y apellidos de las personas que viven habitualmente aquí. Por favor, para que no se le olvide ninguna, dígamelas por orden de edad, primero las personas mayores y luego los menores’.**

**A continuación:** Pregúntele: **‘¿Hay alguna otra persona que no esté en este momento pero que viva habitualmente aquí?’** Si hay alguna, debe añadirse a continuación.

**Finalmente:** Pregúntele: **‘¿Hay alguna persona que no tenga ninguna residencia habitual y que viva actualmente en esta vivienda?’** Si hay alguna, debe añadirse a continuación.

#### Naturaleza, características y finalidad

La Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares (TIC-H.2003) es una investigación estadística que realiza el Instituto Nacional de Estadística (INE) siguiendo las recomendaciones de la Oficina Estadística de la Unión Europea (EUROSTAT).

Esta investigación se dirige a los hogares con el fin de obtener datos comparativos sobre la disponibilidad en los mismos de dicho equipamiento para comunicarse y obtener o enviar información, así como sobre el uso que los miembros del hogar hacen de dichas tecnologías.

#### Legislación

##### Secreto Estadístico

Serán objeto de protección y quedarán amparados por el **secreto estadístico**, los datos personales que obtengan los servicios estadísticos, tanto directamente de los informantes como a través de fuentes administrativas (art. 13.1 de la Ley de la Función Estadística Pública de 9 de mayo de 1989, (LFEP). Todo el personal estadístico tendrá la obligación de preservar el secreto estadístico (art. 17.1 de la LFEP).

Los servicios estadísticos podrán solicitar datos de todas las personas físicas y jurídicas nacionales y extranjeras, residentes en España (artículo 10.1 de la LFEP).

Todas las personas físicas y jurídicas que suministren datos, tanto si su colaboración es obligatoria como voluntaria, **deben contestar de forma veraz, exacta, completa y dentro del plazo** a las preguntas ordenadas en la debida forma por parte de los servicios estadísticos (art. 10.2 de la LFEP).

Para velar por el cumplimiento de estas normas la LFEP (art. 48) otorga al INE capacidad sancionadora.

### 3. TABLA DE MIEMBROS DE LA VIVIENDA

1. NOMBRE Y APELLIDOS	2. ¿Hay alguna persona de las mencionadas, que sea trabajador interno? (Haga para cada persona las marcas que correspondan)		3. ¿Alguna persona (que no sea trabajador interno) está desplazada de su residencia habitual por estudios, trabajo, enfermedad o viaje? (Haga para cada persona las marcas que correspondan)	
	a. TRABAJADOR INTERNO	b. ¿Trabaja interno en esta vivienda?	a. DESPLAZADO TEMPORAL	b. ¿Es ésta su vivienda habitual de la que está desplazada?
Entrevistador: lea antes las instrucciones en el anverso. Anote con letras mayúsculas, claras y respetando los peines	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>
	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 3</i>	1 <input type="checkbox"/> SI → <i>Tache pregunta 3</i> 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>pasa a 4</i>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>

**Entrevistador:** no olvide, antes de seguir adelante, preguntar por otras personas ausentes y por residentes actuales.

<b>4. PERSONA N°:</b> <i>Entrevistador: anote un número correlativo en las casillas de las filas que no están tachadas</i>	<b>5. SEXO</b>	<b>6. Dígame el mes y el año de nacimiento</b>	<b>7. ES ENCUESTABLE</b> <i>Entrevistador: compruebe si la persona tiene 15 años o más en el momento de la entrevista</i>	<b>8. ENCUESTABLE N°:</b> <i>Entrevistador: anote un número correlativo en las casillas de las filas que no están tachadas</i>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>	1 <input type="checkbox"/> Varón 6 <input type="checkbox"/> Mujer	<div style="text-align: center;"> <div style="border: 1px solid black; width: 40px; height: 15px; margin: 0 auto;"></div>           Mes  <div style="border: 1px solid black; width: 80px; height: 15px; margin: 0 auto;"></div>           Año         </div>	1 <input type="checkbox"/> SI → 6 <input type="checkbox"/> NO → <i>FIN (Tache el resto de la fila)</i>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>

## SELECCIÓN DE LA PERSONA A ENCUESTAR

Entrevistador, léale al informante 'A continuación voy hacer la selección de la persona a encuestar'.

1º) Anote aquí el 'Identificador de la Vivienda' que figura en la portada del cuestionario:

### 4. IDENTIFICADOR DE LA VIVIENDA \_\_\_\_\_

Haga una marca en la fila correspondiente a este n° en la *Tabla de Selección*:

#### TABLA DE SELECCIÓN

IDENTIFICADOR DE LA VIVIENDA	TOTAL DE PERSONAS ENCUESTABLES									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	1	3	1	8	9	1
2	1	1	2	3	5	2	7	7	8	10
3	1	2	1	2	4	1	6	6	7	9
4	1	1	3	1	3	6	5	5	6	8
5	1	2	2	4	2	5	4	4	5	7
6	1	1	1	3	1	4	3	3	4	6
7	1	2	3	2	5	3	2	2	3	5
8	1	1	2	1	4	2	1	1	2	4

2º) Consulte en la *Tabla de Miembros de la Vivienda* de este cuestionario el último valor de la columna '8. ENCUESTABLE N°:', y anótelos aquí:

### 5. TOTAL DE PERSONAS ENCUESTABLES \_\_\_\_\_

Haga una marca en la columna correspondiente a este n° en la *Tabla de Selección*.

3º) Busque en la *Tabla de Selección* el n° que corresponde para la fila y columna marcadas, haga un círculo alrededor de él y anote el número a continuación:

### 6. N° DE PERSONA ENCUESTABLE SELECCIONADA \_\_\_\_\_

4º) Consulte de nuevo la *Tabla de Miembros de la Vivienda* y busque la persona que tenga ese n° en la columna '8. ENCUESTABLE N°:'. Haga una marca en la fila de la persona seleccionada y anote aquí el dato de 'NOMBRE y APELLIDOS' y el de 'PERSONA N°':

### 7. NOMBRE ..... PERSONA N° \_\_\_\_\_

5º) Según sea:

A) Si la persona seleccionada coincide con la que está entrevistando indíquesele y siga adelante.

B) Si no es la persona que está entrevistando, infórmele de ello. Pregúntele quién de los dos conoce mejor los bloques de Equipamiento e Internet en la vivienda y de Uso por niños del Ordenador e Internet. Si es el informante actual complímelo con él esos bloques. Si es la persona seleccionada, pregunte por el momento adecuado, y antes de dar por terminada la entrevista, no olvide pedir un teléfono de contacto y hacer las anotaciones correspondientes en la última página (preguntas 64 y 65). Recuerde que si esa persona va a estar ausente debe intentar entrevistarla por teléfono.

## II. EQUIPAMIENTO DE LA VIVIENDA PRINCIPAL EN PRODUCTOS DE TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

*Entrevistador, léale al informante: 'A continuación le voy a preguntar si la vivienda tiene algunos productos como radio, televisión, ordenador, etc. Tenga en cuenta que solo nos referimos a la vivienda principal y que debe incluir aparatos en reparación y comprados o encargados, aunque no estén instalados todavía. También se incluyen si son alquilados o puestos a disposición de la vivienda de cualquier otra forma'.*

### 8. ¿Tiene televisión?

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6 → pasar a pregunta 13

**A continuación le voy a preguntar por la forma de recepción para la televisión:**

### 9. ¿Recibe canales a través de antena convencional?

*(A través de este sistema se pueden ver canales como TV1, TV2, Antena3, Tele5, Canal+, etc.)*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### 10. ¿Recibe canales a través de antena parabólica?

*(A través de este sistema se pueden ver Canal Satélite Digital, Vía Digital, emisoras extranjeras y emisoras nacionales recibidas por satélite)*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### 11. ¿Recibe canales a través de cable?

*(Televisión digital codificada cuya señal se recibe por cable óptico y necesita un aparato descodificador. A través de este sistema se pueden ver canales como Madritel, Euskaltel, Ono, Menta, Retecal, R y Supercable)*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### 12. ¿Recibe canales de televisión digital terrestre, como son Onda6, La Otra, etc?

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### 13. ¿Tiene ordenador de sobremesa? Pero no incluya ordenadores portátiles

*(También llamado PC)*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### 14. ¿Algún residente de la vivienda tiene ordenador portátil?

*(Ordenador de tamaño pequeño-medio, que se puede transportar como un maletín y apoyar en el regazo)*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**15. ¿Algún residente de la vivienda tiene otro tipo de ordenador, por ejemplo PDA, palmtop, pocket PC?**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**16. ¿Tiene teléfono fijo? Se incluyen inalámbricos**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**17. ¿Algún residente de la vivienda tiene teléfono móvil? No se incluyen inalámbricos**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**18. ¿Tiene cadena musical, equipo de alta fidelidad o laserdisk?**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**19. ¿Tiene radio? Se incluyen radiocassette y transistor que se usen en la vivienda**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**20. ¿Tiene vídeo?**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**21. ¿Tiene DVD?**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**22. ¿Tiene Fax?**

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

**23. ¿Algún residente de la vivienda tiene automóvil con sistema de navegación de tráfico?**

*(Sistema informático de localización GPS y gestión de tráfico en tiempo real )*

SI \_\_\_\_\_  1

NO \_\_\_\_\_  6

### III.- ACCESO A INTERNET DE LA VIVIENDA

#### 24. ¿Dispone su vivienda de acceso a Internet?

(Puede tratarse de cualquier tipo de acceso desde cualquier medio: ordenador, teléfono móvil, televisión con Internet, teléfono fijo con pantalla, etc.)

SI \_\_\_\_\_  1 → pasar a pregunta 26

NO \_\_\_\_\_  6

#### 25. ¿Por qué motivos no dispone su vivienda de acceso a Internet? Le leo posibles causas:

- |  | SI                         | NO                         | NS/NR                      |                         |
|--|----------------------------|----------------------------|----------------------------|-------------------------|
| 1. ¿Porque acceden desde otro lugar? _____                           | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 | } pasar al<br>bloque IV |
| 2. ¿Porque no quieren, no lo necesitan o lo desconocen? _____        | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                         |
| 3. ¿Porque la conexión cuesta demasiado? _____                       | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                         |
| 4. ¿Porque no tienen conocimientos para utilizarlo? _____            | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                         |
| 5. ¿Porque no tienen ordenador o no tienen el equipo adecuado? _____ | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                         |
| 6. ¿Por otros motivos? (Si dice que SI, debe especificar)            |                            |                            |                            |                         |
| _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                         |

#### 26. ¿Cuál de las siguientes formas de conexión a Internet tiene su vivienda?

- |  | SI                         | NO                         | NS/NR                      |
|--|----------------------------|----------------------------|----------------------------|
| 1. ¿Línea telefónica convencional? _____                         | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 2. ¿Línea ADSL? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 3. ¿Línea RDSI? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 4. ¿Red de cable (hilo óptico)? _____                            | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 5. ¿Otras formas de conexión? (Si dice que SI, debe especificar) |                            |                            |                            |
| _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |

#### 27. ¿Cuál de los siguientes equipos utiliza para conectarse a Internet en su vivienda?

- |  | SI                         | NO                         | NS/NR                      |
|--|----------------------------|----------------------------|----------------------------|
| 1. ¿Ordenador de sobremesa? _____                                | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 2. ¿Ordenador portátil? _____                                    | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 3. ¿Otro tipo de ordenador, PDA, palmtop, pocket PC? _____       | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 4. ¿Teléfono móvil (WAP, GPRS)? _____                            | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 5. ¿Otras formas de conexión? (Si dice que SI, debe especificar) |                            |                            |                            |
| _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |





**37. Dígame si en los últimos tres meses ha usado Internet desde cada uno de los lugares que le enumero a continuación:**

- |  | SI                         | NO                         | NS/NR                      |
|--|----------------------------|----------------------------|----------------------------|
| 1. ¿Desde la vivienda? _____           | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 2. ¿Desde el centro de trabajo? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 3. ¿Desde el centro de estudios? _____ | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 4. ¿Desde otros lugares? _____         | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |

**38. ¿Con qué frecuencia ha usado Internet en los últimos tres meses? Tenga en cuenta que nos referimos a lo habitual, y no debe considerar vacaciones ni otras situaciones excepcionales.**

- |  |                            |                       |
|--|----------------------------|-----------------------|
| Diariamente, al menos 5 días por semana _____            | <input type="checkbox"/> 1 | } pasar a pregunta 39 |
| Todas las semanas, pero no diariamente _____             | <input type="checkbox"/> 2 |                       |
| Al menos una vez al mes, pero no todas las semanas _____ | <input type="checkbox"/> 3 | } pasar a pregunta 40 |
| No todos los meses _____                                 | <input type="checkbox"/> 4 |                       |
| NO SABE / NO RECUERDA _____                              | <input type="checkbox"/> 9 |                       |

**39. ¿Cuánto tiempo aproximadamente ha usado Internet en la última semana? Le leo unos intervalos de tiempo y Vd. me dice el que corresponde:**

- |                                    |                            |                       |
|------------------------------------|----------------------------|-----------------------|
| ¿Una hora o menos? _____           | <input type="checkbox"/> 1 | } pasar a pregunta 41 |
| ¿Más de 1 hora y hasta 5? _____    | <input type="checkbox"/> 2 |                       |
| ¿Más de 5 horas y hasta 10? _____  | <input type="checkbox"/> 3 |                       |
| ¿Más de 10 horas y hasta 20? _____ | <input type="checkbox"/> 4 |                       |
| ¿Más de 20 horas y hasta 50? _____ | <input type="checkbox"/> 5 |                       |
| ¿Más de 50 horas? _____            | <input type="checkbox"/> 6 |                       |
| NO SABE / NO RECUERDA _____        | <input type="checkbox"/> 9 |                       |

**40. ¿Cuánto tiempo aproximadamente ha usado Internet en los últimos tres meses? Le leo unos intervalos de tiempo y Vd. me dice el que corresponde:**

- |                                    |                            |
|------------------------------------|----------------------------|
| ¿Una hora o menos? _____           | <input type="checkbox"/> 1 |
| ¿Más de 1 hora y hasta 5? _____    | <input type="checkbox"/> 2 |
| ¿Más de 5 horas y hasta 10? _____  | <input type="checkbox"/> 3 |
| ¿Más de 10 horas y hasta 20? _____ | <input type="checkbox"/> 4 |
| ¿Más de 20 horas y hasta 50? _____ | <input type="checkbox"/> 5 |
| ¿Más de 50 horas? _____            | <input type="checkbox"/> 6 |
| NO SABE / NO RECUERDA _____        | <input type="checkbox"/> 9 |

**41. Para cada uno de los servicios de Internet que le enumero a continuación, dígame si lo ha usado en los últimos tres meses desde cualquier lugar y usando cualquier medio:**

- A. COMUNICACIONES** SI NO NS/NR
1. ¿Para correo electrónico? \_\_\_\_\_  1  6  9
2. ¿Para Chats, Conversaciones o Foros? \_\_\_\_\_  1  6  9
3. ¿Para telefonar a través de Internet o videoconferencias? \_\_\_\_\_  1  6  9
4. ¿Para mensajes a móviles? (SMS) \_\_\_\_\_  1  6  9

- B. VENTAS / INVERSIONES**
5. ¿Para banca electrónica y actividades financieras? Bolsa, etc. \_\_\_\_\_  1  6  9
6. ¿Para ventas de productos y servicios? Nuevos, de stock, subastas, etc. \_\_\_\_\_  1  6  9
7. ¿Para compras de productos y servicios? \_\_\_\_\_  1  6  9

- C. BÚSQUEDA DE INFORMACIÓN Y SERVICIOS EN LA RED**
8. ¿Para buscar información sobre productos y servicios? \_\_\_\_\_  1  6  9
9. ¿Para hacer reservas de turismo? Viajes, alojamiento, etc. \_\_\_\_\_  1  6  9
10. ¿Para medios de comunicación, como escuchar la radio, tv, leer periódicos, revistas, etc.? \_\_\_\_\_  1  6  9
11. ¿Para servicios de ocio, como descargar o usar juegos, música, etc.? \_\_\_\_\_  1  6  9

- D. RELACIÓN CON ORGANISMOS DE LA ADMINISTRACIÓN**
12. ¿Para obtener información de páginas web de la Administración? \_\_\_\_\_  1  6  9
13. ¿Para descargar formularios oficiales? \_\_\_\_\_  1  6  9
14. ¿Para enviar formularios cumplimentados? \_\_\_\_\_  1  6  9

- E. EDUCACIÓN Y FORMACIÓN**
15. ¿Para cursos de educación reglada, del colegio, la universidad, etc.? \_\_\_\_\_  1  6  9
16. ¿Para cursos dirigidos a la búsqueda de oportunidades de empleo? \_\_\_\_\_  1  6  9
17. ¿Para otro tipo de cursos? \_\_\_\_\_  1  6  9

- F. SERVICIOS DE SALUD**
18. ¿Para realizar trámites o buscar información sobre problemas de salud? \_\_\_\_\_  1  6  9

- G. OTROS SERVICIOS**
19. ¿Para otros servicios o consultas? (Si dice que SI, debe especificar)
- \_\_\_\_\_  1  6  9



**46. ¿Alguno de los motivos siguientes ha sido causa de no realizar compras a través de Internet en los últimos tres meses?**

- |  | SI                         | NO                         | NS/NR                      |                          |
|--|----------------------------|----------------------------|----------------------------|--------------------------|
| 1. ¿Porque prefiere comprar personalmente en una tienda? _____   | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 | } pasar al<br>bloque VII |
| 2. ¿Porque le preocupa la seguridad?<br>(Al dar detalles de la tarjeta de crédito a través de Internet)_____ | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                          |
| 3. ¿Porque le preocupa la privacidad?<br>(Al dar detalles personales a través de Internet) _____             | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                          |
| 4. ¿Porque no ha tenido necesidad? _____   | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                          |
| 5. ¿Porque cree que no funciona bien la recepción y<br>devolución de las compras hechas por Internet? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                          |
| 6. ¿Por otros motivos? (Si dice que SI, debe especificar)  |                            |                            |                            |                          |
| _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |                          |

**47. Dígame si compró o encargó a través de Internet en los últimos tres meses alguno de los productos o servicios que le enumero a continuación (le recuerdo que no lo incluya si lo realizó por motivo de trabajo):**

- |   | SI                         | NO                         | NS/NR                      |
|---|----------------------------|----------------------------|----------------------------|
| 1. ¿Productos para el hogar? Comida, menaje, etc._____                            | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 2. ¿Videos o música? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 3. ¿Libros, revistas o material de aprendizaje electrónico? _____                 | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 4. ¿Ropa o material deportivo? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 5. ¿Software de ordenador? Incluido videojuegos _____                             | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 6. ¿Material informático? _____   | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 7. ¿Equipamiento electrónico? Incluidas cámaras _____                             | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 8. ¿Compra de acciones o valores, seguros u otros servicios<br>financieros? _____ | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 9. ¿Viajes o alojamiento de vacaciones? _____                                     | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 10. ¿Entradas de espectáculos? _____  | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 11. ¿Loterías o apuestas? _____   | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |
| 12. ¿Otros productos o servicios? (Si dice que SI, debe especificar)              |                            |                            |                            |
| _____   | <input type="checkbox"/> 1 | <input type="checkbox"/> 6 | <input type="checkbox"/> 9 |

**Entrevistador:** Compruebe las respuestas dadas en esta pregunta (47) y si sólo ha contestado SI a la subpregunta 8, pase al bloque VII.

**48. ¿Cuánto dinero aproximadamente ha pagado Ud. por compras a través de Internet en los últimos tres meses? No incluya compras por motivo de trabajo ni de servicios financieros: acciones, seguros, fondos de inversión, etc.**

**Entrevistador:** Si duda el informante, dígame que en euros o en pesetas, como prefiera. Anótelos en el peine adecuado. En caso de duda o si le dice que no puede saberlo ni siquiera aproximadamente, pregúntele si por favor puede contestar en cuál de los intervalos, que le va a leer a continuación, se situaría. Si finalmente no puede seleccionar un intervalo, marque la casilla 'NO SABE/NO RECUERDA'.



## VIII. TABLA DE USO DE TIC POR LOS NIÑOS (10 A 14 AÑOS)

**Entrevistador:** consulte la Tabla de Miembros de la Vivienda y compruebe si hay alguna persona que tenga entre 10 y 14 años en el momento de la entrevista. Si no hay ninguna, pase a cumplimentar el bloque IX. Si hay alguna, transcriba para todas ellas, el 'NOMBRE y APELLIDOS' y 'PERSONA N°' y léale al informante: 'A continuación le voy a hacer algunas preguntas sobre el uso de ordenadores e Internet por los niños de 10 a 14 años'. Pregunte para cada uno de ellos y cumplimente la columna correspondiente.

<b>54. Nombre</b> <b>Persona N°:</b>	..... _____ □□	..... _____ □□	..... _____ □□
<b>55. ¿Ha utilizado el ordenador en su vivienda en los últimos tres meses?</b>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 57</i>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 57</i>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 57</i>
<b>56. ¿Para qué lo ha utilizado?</b>			
1. ¿Para ocio, música juegos, etc.?	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
2. ¿Para trabajos escolares?	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
3. ¿Otros usos? (Especificar)	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
<b>57. ¿Ha utilizado Internet en los últimos tres meses?</b>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 60</i>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 60</i>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9 } <i>pasa a 60</i>
<b>58. ¿Para qué ha utilizado Internet?</b>			
1. ¿Para ocio, música juegos, etc.?	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
2. ¿Para trabajos escolares?	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
3. ¿Otros usos? (Especificar)	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
<b>59. ¿Desde qué lugar ha usado Internet?</b>			
1. Vivienda	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
2. Centro de estudios	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
3. Otros lugares (Especificar)	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	..... <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9
<b>60. ¿Dispone de teléfono móvil propio?</b>	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9	SI <input type="checkbox"/> 1 NO <input type="checkbox"/> 6 NS/NR <input type="checkbox"/> 9

