

**UNIVERSIDAD COMPLUTENSE DE MADRID**

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



**EL MERCADO DE LA VIVIENDA EN ESPAÑA**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**  
**PRESENTADA POR**  
**Raquel Arévalo Tomé**

Bajo la dirección de los doctores:  
Javier Ruiz-Castillo Ucelay y Gregorio Serrano

**Madrid, 2001**

**ISBN: 84-669-2242-3**

T 25228

**UNIVERSIDAD COMPLUTENSE DE MADRID  
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES  
DEPARTAMENTO DE FUNDAMENTOS DEL ANÁLISIS ECONÓMICO II**

**TESIS DOCTORAL**



**UNIVERSIDAD COMPLUTENSE**



5314024525

**EL MERCADO DE LA VIVIENDA EN ESPAÑA**

**RAQUEL ARÉVALO TOMÉ**

**DIRECTOR: JAVIER RUIZ-CASTILLO UCELAY  
CODIRECTOR: GREGORIO SERRANO**

**ABRIL 2001**

**25228**



**BIBLIOTECA**

519022645  
i 26487279



*A mi familia*

# Índice General

Agradecimientos	i
Introducción	iii
1 Representatividad y calidad de la información sobre la vivienda en la Encuesta de Presupuestos Familiares	1
1.1 Introducción	2
1.2 Censo de viviendas <i>versus</i> Encuesta de Presupuestos Familiares	4
1.2.1 Ámbito poblacional	5
1.2.2 Ámbitos geográfico y temporal	6
1.2.3 Proceso de elaboración de datos	7
1.2.4 Definición y clasificación de las viviendas ocupadas	8
1.2.5 Características de las viviendas	9
1.3 Revisión de la información disponible en la EPF	19
1.3.1 Régimen de tenencia y calificación legal	20
1.3.2 Año de construcción y ocupación	22
1.3.3 Precios	27
1.4 Conclusiones	29
. Cuadros del capítulo I	35

<b>2 Construcción de indicadores del nivel de acondicionamiento de la vivienda</b>	<b>49</b>
2.1 Introducción . . . . .	50
2.2 Aspectos metodológicos del Análisis de Correspondencias . . .	52
2.2.1 Análisis de Correspondencias Simples . . . . .	53
2.2.2 Análisis de Correspondencias Múltiples . . . . .	71
2.3 Aplicación del ACM a la vivienda . . . . .	73
2.3.1 Notación y relación con el ACS . . . . .	73
2.3.2 Datos y variables . . . . .	78
2.3.3 Programa computacional . . . . .	82
2.4 Resultados obtenidos . . . . .	83
2.4.1 Porcentaje de inercia explicada . . . . .	83
2.4.2 Correlación lineal entre el índice y las características de la vivienda . . . . .	84
2.4.3 Pesos de las modalidades en el índice . . . . .	86
2.4.4 Limitaciones de la variable índice de calidad . . . . .	89
2.5 Aplicaciones del índice de calidad . . . . .	90
2.5.1 Asociación entre las características de la vivienda y el nivel de calidad . . . . .	91
2.5.2 Comparación de la calidad de la vivienda en 1980-81 y 1990-91 . . . . .	92
2.6 Conclusiones . . . . .	96
. Cuadros del capítulo II	101
. Apéndice I: Ejemplos del Análisis de Correspondencias	107
. Apéndice II: Deducción de las ecuaciones de transición	125

ÍNDICE GENERAL	3
. Apéndice III: Resultados de SAS	129
<b>3 Determinantes del alquiler de la vivienda</b>	<b>131</b>
3.1 Introducción . . . . .	132
3.2 Especificación del modelo . . . . .	137
3.2.1 Modelo empírico . . . . .	141
3.2.2 Determinación de la forma funcional . . . . .	143
3.2.3 Estimación del proceso de actualización del alquiler . . . . .	145
3.3 Modelo hedónico para 1980 . . . . .	147
3.3.1 Datos y variables disponibles . . . . .	147
3.3.2 Proceso de estimación . . . . .	151
3.3.3 Interpretación de los resultados . . . . .	154
3.4 Modelo con índice de calidad para 1980 . . . . .	157
3.4.1 Proceso de estimación . . . . .	159
3.4.2 Interpretación de los resultados . . . . .	161
3.5 Modelo con índice de calidad para 1990 . . . . .	166
3.5.1 Datos y variables disponibles . . . . .	166
3.5.2 Proceso de estimación . . . . .	169
3.5.3 Interpretación de los resultados . . . . .	172
3.6 Conclusiones . . . . .	176
. Cuadros del capítulo III	183
. Apéndice I: Determinación de la forma funcional del alquiler	189
. Apéndice II: Relación entre las variables explicativas disponibles y el alquiler	193
. Apéndice III: Estimación de los modelos de alquiler	201

. Apéndice IV: Datos oficiales sobre la inflación y estimación del <i>tenure discount</i>	219
<b>4 Valoración del flujo de servicios de la vivienda</b>	<b>223</b>
4.1 Introducción . . . . .	224
4.2 Especificación del modelo . . . . .	228
4.3 Valoración del parque de viviendas a precios de 1980 . . . . .	231
4.3.1 Datos y variables . . . . .	232
4.3.2 Resultados de la estimación . . . . .	235
4.3.3 Influencia de las observaciones sin CL . . . . .	236
4.3.4 Comparación del alquiler medio observado e imputado	240
4.4 Valoración del parque de viviendas a precios de 1990 . . . . .	246
4.4.1 Datos y variables . . . . .	248
4.4.2 Resultados de la estimación . . . . .	249
4.4.3 Influencia de las observaciones sin CL . . . . .	250
4.4.4 Comparación del alquiler medio observado e imputado	253
4.5 Conclusiones . . . . .	257
. Cuadros del capítulo IV	263
. Apéndice I: Relación entre el régimen de tenencia y las variables explicativas	277
. Apéndice II: Análisis sobre las variables en arrendamiento sin CL	283
Extensiones	305
Referencias bibliográficas	311

# Agradecimientos

En esta página quiero dejar constancia de mi agradecimiento a todas aquellas personas e instituciones que me han apoyado en la realización de este trabajo.

En primer lugar a Javier Ruiz-Castillo y Gregorio Serrano por su respectiva labor de director y codirector de tesis.

También quiero agradecer el continuo esfuerzo y dedicación de Javier Ruiz-Castillo en el seguimiento de esta tesis. Su motivación y espíritu crítico han logrado transmitirme el sentido del rigor y la responsabilidad en la labor investigadora. Trabajar con él supone un permanente estímulo de exigencia y superación del que espero seguir aprovechándome.

En segundo lugar, quiero agradecer al Departamento de Fundamentos del Análisis Económico II de la Universidad Complutense por el apoyo que siempre recibí para llevar mi trabajo a buen puerto. A la Universidad Carlos III de Madrid le debo mucho: fue mi fuente de financiación durante cuatro años, me permitió hacer uso de sus instalaciones y equipos informáticos durante algunos años más, y, lo más importante, me dio la posibilidad de conocer a personas tan distintas y enriquecedoras como: Edita Pereira, Eva Rodríguez, María José Caride, Xulia González, Raúl Moreno, Begoña Álvarez, Toño y Pilar Poncela, Antonio R.Sampaio, Gema Álvarez, Eva Presa, Raquel Carrasco, Ada Fiteni, Bernarda Zamora, Daniel Miles, Eduardo Giménez, Fidel Castro, Noemí Padrón, Coral del Río, Olga Alonso, Carmen Vargas, Fer-

nando del Río, Mikel Pérez, Fato, Fuco, Mar Rodríguez, Xosé M. González, Mercedes Sastre, Lola Robles, Julio, Pedro Delicado, Fernando Lorenzo, o Eva Senra. A la Universidad de Vigo le agradezco haberme acogido a mí, al noventa por ciento de los mencionados anteriormente, y a otros como José M. Chamorro, Pilar Abad, Ana Martínez, Marisol Otero, Olga Cantó, Pilar Piñeiro, Fernando Comesaña, María de los Angeles Quintás, Lorenzo, Montse y Javier, Toni y Elena, Adela, Gloria, Luci, Carlos Gradín, Manolo, Philippe, Calu, Jaime y Susi, y a nuestro Tuco.

Por último, y de forma muy especial, deseo agradecer las constantes muestras de paciencia, cariño y apoyo de José M. Chamorro, Eva Rodríguez, Edita Pereira, Xulia González y María José Caride. En ellos, y en mi familia, he encontrado siempre la comprensión y el ánimo que he necesitado en las distintas etapas del largo camino que ha requerido este trabajo.

# Introducción

El acceso a una vivienda digna y apropiada es una necesidad básica para toda unidad familiar. Constituye un derecho incuestionable en una sociedad desarrollada, que en nuestro país se refleja en el artículo 47 de la Constitución de 1978. De esta forma, el acceso a los servicios que proporciona una vivienda, bien en régimen de propiedad o arrendamiento, se reconoce como un derecho fundamental, lo que implica una protección de carácter prioritario por parte del sector público.

La importancia de la vivienda tiene su reflejo tanto en el ámbito familiar, puesto que supone el mayor gasto y/o inversión que realiza el hogar en su ciclo de vida<sup>1</sup>, como en el ámbito económico y social de cualquier país. En este sentido, se pueden enumerar distintos ejemplos que muestran la incidencia de este bien duradero en la actividad cotidiana de nuestra sociedad. Para empezar, el gasto en los servicios que proporciona la vivienda es un factor que contribuye a la elaboración de un indicador económico de referencia, tan obligada como habitual, en multitud de actuaciones como es el índice de Precios de Consumo (IPC).

Por otra parte, la vivienda ha sido objeto de intervención gubernamental en la mayoría de los países desarrollados como un elemento característico del

---

<sup>1</sup>En Taltavull (2000) se analizan los fenómenos y problemas relacionados con la vivienda en nuestro país desde la óptica de la perspectiva familiar.

llamado Estado del Bienestar. En España no existe una Política de Viviendas propiamente dicha, la intervención pública en el sector de la vivienda se ha llevado a cabo, principalmente, con dos tipos de actuaciones: 1) el control de alquileres que representa la regulación del precio nominal de los servicios de la vivienda, y 2) los sistemas de protección oficial (PO) establecidos tanto para viviendas el régimen de propiedad como en de arrendamiento<sup>2</sup>. Otro tipo de instrumentos al alcance del sector público son: la regulación de distintos aspectos de las instituciones del mercado hipotecario, por ejemplo, el tipo de interés; los programas de renovación urbana que pueden conllevar, por ejemplo, la adquisición de suelo urbanizable por parte de las comunidades, o la remodelación de barrios degradados, etc.; favorecer el incremento del flujo de créditos dirigidos a la vivienda por medio de la intervención en el mercado de capitales; tratamiento fiscal de la vivienda como, por ejemplo, vía impuestos sobre la propiedad inmobiliaria o reducción del impacto impositivo de los contribuyentes debido a la adquisición de una vivienda, etc.

Con independencia del efecto que pueda tener la política de viviendas en el sector de la construcción, los objetivos sociales que deben motivar este tipo de intervenciones son: i) facilitar el acceso a los servicios que proporciona una vivienda digna a cada familia, lo que implica priorizar a aquellas que disponen de menores posibilidades económicas, y ii) garantizar un peso razonable del coste en vivienda dentro del presupuesto familiar. En González-Páramo (1992) y López (1992a) se presenta una discusión en profundidad sobre los tres tipos de objetivos (eficiencia económica, redistribución y estabilidad) que deben justificar la actuación pública en el mercado de la vivienda. En este sentido, la importancia de la vivienda también radica en que puede ser

---

<sup>2</sup>Viviendas de titularidad pública alquilada por debajo del coste, así como las subvenciones a la vivienda dirigidas a economías domésticas de renta baja.

reflejo de la desigualdad existente entre los hogares de nuestro país.

En definitiva, los instrumentos fiscales, junto con otras medidas públicas, pueden ser una herramienta eficaz para utilizar criterios de equidad a lo hora de intentar garantizar una distribución más justa de dos factores complementarios. Por una parte, la atención se debe centrar en la distribución de la renta, como indicador por excelencia del grado de equidad de una sociedad, pero también, en unas necesidades básicas que deben estar cubiertas, entre ellas, el acceso a un nivel mínimo de los servicios de la vivienda.

Desde esta perspectiva de equidad social, la motivación de esta tesis es disponer de alguna medida objetiva que permita comparar, en términos reales y monetarios, la totalidad de los hogares de nuestro país en relación a los servicios que proporciona la vivienda que ocupan. El principal reto, por tanto, es adjudicar un valor flujo (ya sea real o monetario) a un bien duradero tan complejo como es la vivienda. Para llevarlo a cabo es preciso, en primer lugar, solventar los problemas derivados de enfrentarse a un bien que se puede caracterizar por multitud de aspectos. Efectivamente, la vivienda es un bien muy heterogéneo debido, entre otras cosas, a la diversidad de características físicas de que puede estar compuesto, el dispar grado de conservación observado, el efecto en los precios de múltiples de factores como: la localización, la disponibilidad de suelo o las distintas condiciones de acceso a la vivienda (mercado libre o sistema de protección oficial), que unido a la variedad de regímenes de tenencia (propiedad, arrendamiento, cesión) obliga a afrontar la difícil tarea de determinar alguna medida homogénea y válida para comparar el nivel, real y monetario, de los servicios de todo tipo de viviendas.

Con este planteamiento, los objetivos de esta tesis empírica se concretan en dos: i) sintetizar el conjunto de servicios o características físicas de la

vivienda ocupada con fines residenciales de nuestro país en un indicador de calidad, y ii) valorar, a precios corrientes, el flujo de servicios de estas viviendas así como el efecto de su localización.

La fuente de información que se utiliza para realizar este trabajo es la Encuesta de Presupuestos Familiares (EPF) elaborada por el Instituto Nacional de Estadística (INE) en 1980-81 y 1990-91<sup>3</sup>. Como es sabido, estas encuestas proporcionan información microeconómica de los hogares españoles con representatividad nacional, incluyendo a Ceuta y Melilla. Las EPF contienen datos detallados sobre el parque residencial (disponibilidad de diversos servicios de las viviendas, localización, nivel de equipamiento, régimen de tenencia y calificación legal, precio de adquisición o alquiler...); las características de los ocupantes (edad, sexo, educación, relación con el sustentador principal, actividad económica,...); la cuantía y tipo de ingresos percibidos en la unidad familiar, así como la descomposición del gasto del hogar en toda una gama de bienes de consumo corriente y de duraderos.

Para hacer operativa la utilización de la vasta información de las EPF, se procede a su división en una serie de ficheros, fácilmente manejables que, con este fin, se encuentran disponibles en internet<sup>4</sup> y documentados en Alonso-Colmenares *et al.* (1994) y en Arévalo *et al.* (1995). Se trata de un total de 28 ficheros generados a partir de las cintas directamente proporcionadas por el INE que, tras un proceso de tratamiento y clasificación de la información (en la medida de lo posible hemos aplicado criterios y notación homogénea

---

<sup>3</sup>La descripción y metodología de las Encuestas de Presupuestos Familiares de 1980-81 y de 1990-91 se encuentran en INE, 1983a y 1983b y en INE, 1993a, 1993b y 1993c, respectivamente.

<sup>4</sup>Los ficheros referidos a la EPF de 1980-81, y 1990-91, están disponibles en <<http://www.eco.uc3m.es/epf80-81.html>> y <<http://www.eco.uc3m.es/epf90-91.html>>, respectivamente.

en ambos períodos), representan el punto de partida de la base de datos utilizada en la tesis.

En relación al valor del flujo de los servicios de la vivienda, la EPF informa del alquiler realmente pagado (en 1980-81 y 1990-91 respectivamente) en caso de que la vivienda esté en régimen de arrendamiento. En el caso en que la vivienda sea de un régimen distinto al arrendamiento, la EPF proporciona una autoestimación del encuestado a la hora de valorar el alquiler que, según él, correspondería pagar por una vivienda similar a la que ocupa. Se trata, por tanto, de una valoración subjetiva de los precios de mercado del arrendamiento a precios de 1980-81 y 1990-91 respectivamente. Una de las propuestas de esta tesis es determinar una medida objetiva del valor flujo que reporta una vivienda a sus ocupantes según la información contenida en las EPF. Por lo tanto, se podrá disponer de dos medidas, objetiva y subjetiva, que surgen de la misma fuente de información y se refieren a las mismas viviendas.

La importancia que conlleva utilizar estas fuentes de información radica en: i) la capacidad de la EPF para representar el parque de viviendas de uso residencial de nuestro país en cada período, y ii) las posibilidades de relacionar distintos aspectos referidos de los miembros del hogar con las características de la vivienda que ocupan. Además, la relevancia económica de las EPF en nuestro país, da un valor añadido a las aportaciones de esta tesis puesto que abre nuevas líneas de investigación de gran importancia. Entre ellas se pueden indicar dos. La primera está relacionada con el papel de la EPF en la elaboración del IPC, y más concretamente con la estimación que realiza el INE del consumo en viviendas para el régimen distinto del arrendamiento. Así, ante la alternativa de utilizar la autoestimación subjetiva directamente proporcionada por el hogar de la EPF, la tesis proporciona una

estimación objetiva de la valoración del mismo flujo de servicios. Por tanto, resultará, cuando menos, interesante analizar las consecuencias de plantear ambas alternativas.

Una segunda línea de investigación está relacionada con la posibilidad de efectuar una valoración del efecto, en 1980-81 y en 1990-91, de la intervención pública con fines sociales en el sector de la vivienda en relación a: las características de los miembros de los hogares beneficiados, la localización de las viviendas afectadas, grado de adecuación entre las características de la vivienda y el tamaño o necesidades del hogar, etc.

La elección de las fuentes de información, EPF de 1980-81 y 1990-91, condiciona el ámbito temporal a la hora de concretar los objetivos planteados (comparación de viviendas ocupadas en cada período en términos reales y monetarios), pero también hacen posible extender dichos objetivos a una comparación interesante: calidad y valor del flujo de los servicios de la vivienda ocupada en 1990-91 respecto a la ocupada en 1980-81.

Dado el condicionante temporal que supone trabajar con información de 1980-81 y 1990-91, seguidamente se procede a dar una visión panorámica de la situación del mercado de la viviendas en el período de referencia de la tesis. Con ello se pretende enmarcar tanto su contenido, como la interpretación de los resultados y sus posibles aportaciones, dentro del contexto temporal que le corresponde.

En España, el problema del acceso a la vivienda, que aun perdura en la actualidad, adquiere una gran relevancia al final de los años ochenta, coincidiendo con una etapa de intensa recuperación de la actividad económica que, en gran medida, se ve apoyada por una evolución favorable que, desde 1985, experimenta el sector de la construcción<sup>5</sup>. En términos generales, los

---

<sup>5</sup>Tras una etapa de auge coyuntural de la industria de la construcción entre 1970 y

problemas para acceder a la vivienda a finales de los ochenta se deben a la intensificación de: i) la rigidez propia que caracteriza a la oferta y ii) los factores especulativos que repercuten en un problema de escasez de suelo, lo que genera una considerable elevación del precio. Este período fue calificado como el del 'boom inmobiliario' español y fue objeto de múltiples estudios que tratan de explicar tanto su origen como sus efectos en la sociedad. El boom se explica como un 'boom de demanda' en el que se intensifica espectacularmente el número de transacciones en el mercado inmobiliario<sup>6</sup>. Así, por ejemplo, en el trabajo de Lasheras *et al.* (1994) se afirma: *'Todo parece indicar que, a pesar de la mayor oferta (de viviendas residenciales) a finales de los ochenta, estamos ante un mercado caracterizado por el exceso de demanda (donde) la renta y, probablemente, factores no fiscales (inversiones del exterior y afloramiento de dinero negro) podrían ser más responsables que la política fiscal del boom inmobiliario'* (pág. 149).

En relación al efecto de la renta, resulta convincente el razonamiento expresado en López (1997) a la hora de considerar los motivos del aumento de la demanda de servicios/*stock* de viviendas en la segunda mitad de los ochenta: *'...es de esperar que el proceso (gradual) de expansión económica materializado en el aumento de la renta disponible, tanto corriente como esperada, comporte un aumento de la demanda de servicios y dé lugar a un incremento de la demanda derivada del stock de vivienda y de los precios de éste. Este fenómeno de capitalización del aumento de la demanda en los*

---

1974, se inicia un período de crisis (1975-1984) hasta que en 1985 da comienzo su fase recuperación. Esta recuperación se mantiene hasta 1991, fecha a partir de la que la evolución del valor añadido bruto de la construcción muestra una tendencia decreciente.

<sup>6</sup>Esta situación contrasta con la 'crisis de demanda' que caracterizó al sector en los últimos años setenta: una gran parte de las viviendas construidas no se vendió y fue absorbida muy lentamente en los años siguientes.

*precios de las viviendas existentes constituye la señal para el relanzamiento de la construcción nueva, y el aumento del stock de capital residencial... (a esta explicación) hay que añadir la existencia de un incentivo al ahorro en vivienda que se mantuvo durante el período de referencia (entre 1985 y 1990) a unos niveles nada despreciables' (pág. 311).*

Por otra parte, la evidencia empírica presentada por Bover (1993) en relación a la estimación de una ecuación de precios de la vivienda para el período 1976-1991, le lleva a concluir: *'El crecimiento de la renta real per cápita ha sido responsable del 70 por ciento del incremento del precio real de la vivienda entre 1985 y 1990. La tasa de rendimiento también tiene un efecto importante, contribuyendo un 20 por ciento al crecimiento observado en los precios de la vivienda'*(pág. 65).

En esta etapa, la principal manifestación de la crisis al acceso de la vivienda se manifiesta en las grandes ciudades y en las familias de rentas medias y bajas. Este mayoritario estrato social encuentra graves problemas para afrontar el desproporcionado esfuerzo presupuestario que requiere la adquisición de una vivienda en propiedad. Se debe indicar que la propiedad es el régimen de tenencia que, tradicionalmente, ha caracterizado la preferencia de los hogares de nuestro país. De hecho, España destaca en relación a los países de nuestro entorno por un menor uso del régimen de arrendamiento y una tendencia a su reducción. En 1990 tan solo el 15 por ciento de las viviendas principales son alquiladas (un 23 por ciento en 1980), mientras que en países de nuestro entorno, como Francia y Reino Unido, el arrendamiento suponía en 1989 un 40 por ciento del parque residencial y en Alemania un 60 por ciento (véase Instituto Sindical de Estudios, 1989).

A este esquema panorámico se debe añadir una progresiva falta de eficacia de los instrumentos de la política de vivienda en uso que, a lo largo de la

segunda mitad de los ochenta, se ha traducido en: i) un descenso progresivo de la proporción de viviendas en PO, ii) un proceso regresivo en las ayudas fiscales, y iii) un ligero descenso de las viviendas públicas construidas. En estas circunstancias, el denominado 'problema de la vivienda' constituía una de las cuestiones con mayor repercusión en la sociedad española. Surge así, la iniciativa del entonces Ministerio de Obras Públicas y Transportes de auspiciar la elaboración del 'Informe para una Nueva Política de Vivienda' (1992). Este informe fue elaborado por un Comité de Expertos, compuesto por 22 especialistas, que trabaja con un objetivo explícito: *'analizar los problemas que plantea la vivienda en la sociedad española y los medios más adecuados para su solución en el momento actual (para) asesorar al Gobierno de la Nación sobre el rumbo que, al parecer de los expertos convocados, debe tomar la política de la vivienda en este país durante los próximos años'* (pág. 12). El contenido del informe se divide en tres partes que abordan los distintos aspectos que se esperan de él. Así, en la primera parte se detallan los criterios y objetivos generales (sociales, económicos, y urbanísticos) que deberían articular esa política y se presenta un diagnóstico de la situación y las necesidades de viviendas. En la segunda parte se establecen las propuestas concretas para utilizar una serie de instrumentos de política económica (intervención sobre el suelo, fomento del régimen de arrendamiento, elementos para la financiación y la fiscalidad de la vivienda). En la tercera parte se abordan el papel de los agentes de la política de vivienda para solucionar los problemas de alojamiento del sector de la población que, por sus escasos recursos, no puede hacer frente a las condiciones del mercado. En este apartado se hace un especial énfasis en las cuestiones de la promoción pública y en la necesidad de una coordinación entre las administraciones implicadas. Para un mayor detalle sobre el contenido del informe se recomienda López

(1992b) donde se presenta una excelente síntesis y valoración del mismo.

Las medidas presupuestarias que complementan al resto de instrumentos de regulación pública en materia de vivienda, son de dos tipos: ayudas directas y ayudas indirectas. Las segundas se derivan de un tratamiento fiscal favorable a la vivienda habitual en el sistema impositivo español, en particular, en el impuesto sobre la renta. Las principales líneas de cambio de la política fiscal que se proponen en el Informe para una Nueva Política de Vivienda (1992) son: i) racionalizar y ordenar el conjunto de la tributación derivada de la propiedad del vivienda, ii) corregir la falta de neutralidad fiscal en relación a la forma de tenencia que opera a favor de la propiedad frente al arrendamiento, iii) reconsiderar el papel de las deducciones fiscales cristalizadas en tipos de efectivos menores para quienes disfrutaban de ellas; iv) reducir los elevados costes de transacción derivados de la transmisión de viviendas, y v) considerar más el fomento de la adquisición de vivienda en las ayudas directas, preferentemente en la subvención a los intereses de los préstamos hipotecarios contraídos.

Por su parte, la ayuda directa se articula mediante la legislación contenida en leyes que, generalmente, son de carácter plurianual y constituyen los denominados Planes de Vivienda. Estos planes contienen tanto ayudas para los promotores, como a la adquisición de suelo, a la rehabilitación, al alquiler, o a la adquisición de vivienda. En relación a estas últimas, hasta 1978 las actuaciones protegibles tenían multitud de regímenes de aplicación que, en esa fecha, se reducen a la vivienda de PO. Su promoción puede ser pública o privada, de tal forma que la primera de ellas se reserva para la construcción de viviendas destinadas a los estratos de la población con menor renta. A partir de 1989 la legislación contempla la posibilidad de ayudas directas a la adquisición de viviendas usadas, a partir de las actuaciones en viviendas

tasadas. Las condiciones para el acceso a este tipo de viviendas son menos restrictivas que las de adquisición de PO, aunque las ayudas son menores.

Las medidas concretas en que se articulan las ayudas directas a la adquisición de vivienda son: i) subvenciones a fondo perdido, ii) subvenciones de tipo de interés a través de convenios entre las Administraciones Públicas y las entidades de crédito, y iii) préstamos cualificados. La obtención de estas medidas estaba condicionada al cumplimiento de determinados requisitos. En relación a estos requisitos, se debe indicar que hasta 1976, con el Real Decreto de 30 de julio de 1976 sobre política de vivienda social, no se establecen condiciones para acceder a estas ayudas en relación al nivel de renta de la población beneficiada, ni se establecen diferencias en los precios máximos en función de la localización de la vivienda, con el agravio comparativo que esto generaba. También en 1976 se introducen condiciones sobre el tamaño de las viviendas, de forma que este tipo de ayudas, aún en la actualidad, se conceden en función de: i) la relación entre la renta familiar y el salario mínimo internacional, ii) el tamaño máximo de la vivienda, y iii) el precio máximo de venta por áreas de ubicación. La determinación de estos requisitos se modifican en los distintos Planes de Vivienda de aplicación. En Gracia y Mas (2000) se presenta un resumen comparativo de las normas contenidas en los distintos Planes de Viviendas establecidos en España desde 1978 hasta el 2001.

Con posterioridad a la fecha de referencia de la tesis se establecen los Planes de Vivienda de 1992-95, de 1996-99 y de 1998-2001 respectivamente, que, en gran medida, han contribuido a fomentar la preferencia por la vivienda en propiedad (al dirigir buena parte de las ayudas a esta finalidad) y a fomentar el alquiler únicamente para aquellos niveles de renta más precarios. De forma que la legislación contenida en los Planes de Viviendas, al igual que la po-

lítica de viviendas en su conjunto, no ha propiciado la solución al problema del acceso a la vivienda que, aún en la actualidad, subsiste en nuestro país.

En el resto del apartado introductorio se presenta una breve descripción del contenido de los cuatro capítulos que componen esta tesis. Con ello se pretende guiar al lector sobre el proceso seguido en cada apartado para alcanzar los objetivos planteados sobre la comparación, en términos reales y monetarios, del flujo de servicios de la totalidad de las viviendas ocupada en España según las EPF disponibles. En síntesis, el esquema del contenido de la tesis es el siguiente. En el **primer capítulo** se analiza la calidad y capacidad de la información proporcionada en la EPF para caracterizar y representar al conjunto de viviendas españolas a nivel nacional. Se trata de un análisis descriptivo, que se considera necesario en cualquier trabajo empírico con objeto de chequear la información antes de interpretar los resultados generados con posterioridad. En el **segundo capítulo** se construyen indicadores ordinales del nivel del flujo de servicios del total de las viviendas contenido en las EPF. Estos indicadores constituyen una medida objetiva que permite llevar a cabo el primer objetivo de la tesis (comparar la disponibilidad de los servicios, en términos reales, del conjunto de viviendas ocupadas en España según la EPF de 1980-81, según la EPF de 1990-91 y el cambio experimentado en ambos períodos), así como realizar un análisis sobre la distribución del acondicionamiento físico de la vivienda según: el régimen de tenencia, el período de ocupación, la comunidad autónoma y el tamaño de municipio donde se ubica la vivienda. En el **tercer capítulo** se identifican aquellos factores proporcionados en la EPF que, junto con el nivel de servicios de la vivienda, explican el alquiler, o valor corriente, observado en el mercado libre de arrendamientos. En el **capítulo cuatro** se considera el efecto estimado de los determinantes del alquiler libre para llevar a cabo el segundo objetivo de

la tesis: adjudica un valor monetario, a precios corriente, al flujo de servicios de la totalidad del parque residencial representado en las EPF. En este capítulo también se realiza la comparación entre dicha imputación monetaria y el valor corriente directamente proporcionado en la EPF: alquiler realmente pagado en caso de arrendamiento o alquiler autoestimado por el hogar en el régimen de propiedad o cesión.

Dados los objetivo prácticos de la tesis y el hilo conductor de los cuatro capítulos se ha obviado presentar un apartado de conclusiones finales. No obstante, en se presenta un apartado de extensiones donde se proponen varias líneas futuras de investigación en relación a las aportaciones principales de esta tesis: disponer de dos medidas objetivas del flujo de servicios que proporciona el stock de vivienda residenciales en nuestro país. La primer representa el nivel de calidad de dichos servicios mientras que la otra representa su valor monetario, a precios corrientes de 1980 y 1990, una vez considerada el efecto de la localización.

Seguidamente se expone con mayor detalle el contenido y principales resultados de cada uno de los capítulos de la tesis.

## CAPITULO I

En este capítulo se evalúa, desde un punto de vista descriptivo, la capacidad de las EPF de 1980-81 y 1990-91 para representar el total de las viviendas ocupadas en España en dichos períodos. La referencia obligada para evaluar esta representatividad es el exhaustivo registro de viviendas proporcionado en el Censo de Viviendas (CV) realizado por el INE cada diez años. Los dos últimos censos disponibles se refieren a 1981 y 1991 respectivamente lo que hace posible una doble comparación EPF *versus* CV: i) número de viviendas ocupadas en cada período, y ii) el cambio experimentado en dicho período. El análisis se limita a las viviendas principales, la definición de las secunda-

rias en ambas fuentes no resulta compatible. Las variables susceptibles de comparación son: la antigüedad de la vivienda, el régimen de tenencia así como la Comunidad Autónoma y tamaño de municipio donde se ubica la vivienda.

Los principales resultados que se ponen de manifiesto en este capítulo en relación a la representatividad de las EPF *versus* CV son: 1) existen problemas para conocer la verdadera antigüedad del edificio en ambas fuentes de datos así como una mayor dificultad en la EPF para identificar las viviendas de más reciente construcción; 2) la EPF también subestima el número de viviendas ubicadas en los núcleos urbanos de menor tamaño; 3) existe una gran similitud en ambas fuentes en lo que respecta a la distribución del régimen de tenencia y cambio experimentado en la década de estudio; y 4) es necesario utilizar el factor de elevación en la EPF, no solo para mostrar la equivalencia población de la muestra, sino también para corregir una considerable subestimación muestral producida en las Comunidades de Madrid y Cataluña.

También en este capítulo se realiza un estricta revisión la calidad de la información sobre la vivienda disponible en este tipo de encuestas. Así se da cuenta, con todo detalle, de las variables y del número de hogares afectados por falta de información. También se detectan algunas incoherencias debido, por ejemplo, a que el hogar proporciona un año de construcción del edificio posterior al de la ocupación de forma que, en algunas ocasiones, se evidencia un error de transcripción susceptible de corrección automática. La totalidad de los casos afectados por algún tipo de problema se identifican a través de variables ficticias. Esta identificación permiten aplicar una serie de filtros para prescindir de los hogares implicado en los distintos análisis realizados en el desarrollo de la tesis. Con ello se evita una posible distorsión de los

resultados obtenidos, no obstante, en función de los objetivos de cada análisis, se diseñan distintos mecanismos de imputación que hacen posible: i) disponer de información completa para el conjunto de viviendas del parque residencial, y ii) controlar y evaluar el efecto de dichas imputaciones en la interpretación de los correspondientes resultados obtenidos.

## CAPITULO II

En este capítulo se lleva a cabo el primer objetivo de la tesis. Se construye un índice de calidad de la vivienda que sintetiza el conjunto de variables que caracteriza a los servicios que proporciona cada vivienda. Se trata de una variable ordinal que permite comparar, en términos de acondicionamiento, la totalidad de las viviendas representadas en la EPF.

La técnica multivariante que se utiliza para elaborar este índice es el Análisis de Correspondencias Múltiples<sup>7</sup> (ACM). Se trata de un instrumento estadístico, de carácter descriptivo, que proporciona un reducido número de indicadores (nuevas variables incorreladas entre sí) que son combinación lineal de un elevado número de variables originales de forma que la sustitución de variables por indicadores no supone pérdida de información alguna. Por construcción, cada indicador es óptimo desde el punto de vista de recoger la máxima variabilidad residual, (no recogido por el resto de indicadores) y es susceptible de ser interpretado en términos de las variables originales. Aunque en el desarrollo del capítulo se presenta una completa exposición metodológica del ACM, acompañada de diversos ejemplos que facilitan su comprensión, esta parte puede ser obviada por parte del lector que prefiera limitarse a los resultados de la aplicación a la vivienda.

Esta aplicación se realiza sobre tres conjuntos de viviendas: las representadas en la EPF de 1980-81, en la EPF de 1990-91 y en ambas EPF conside-

---

<sup>7</sup>Greenacre (1984, 1990 y 1991) y Lebart et al. (1984).

radas conjuntamente. Las variables originales en el primer período se refieren a las distintas modalidades disponibles sobre: la instalación de servicios higiénicos, agua caliente y/o fría, superficie construida en metros cuadrados, antigüedad del edificio en número de años, posesión de calefacción, garaje, teléfono y energía eléctrica. Además de estas variables, en la EPF de 1990-91 también se proporciona información sobre: posesión de piscina, zona deportiva, jardín, otros servicios comunitarios, ascensor, aire acondicionado, tipo de edificio en relación al número de viviendas que lo compone, tipo de energía, o combustible, utilizada para cocinar, para la calefacción y para calentar el agua, en su caso. El análisis conjunto de las dos EPF obliga a limitar el estudio a las ocho variables disponibles en ambas muestras.

En las tres aplicaciones se obtienen los siguientes resultados comunes: i) el primer indicador explica más del 73 por ciento de la variabilidad total observada en el conjunto de características de la vivienda y se puede definir como un índice de calidad de la vivienda; ii) el segundo indicador en importancia recoge más del 10 por ciento del total a explicar y se identifica con la variable dicotómica disponibilidad de energía eléctrica dentro de la vivienda.

En un análisis pormenorizado de cada variable, el resultado más interesante lo proporcionan las variables 'calefacción' y 'teléfono'. Efectivamente, al incluir ambos servicios en el ACM aplicado a la vivienda de la EPF de 1980-81, y de 1990-91 respectivamente, se revela un cambio conceptual en la calidad de la vivienda de ambos períodos. Así, la vivienda que dispone de calefacción y/o teléfono en 1980-81 destaca sobre la vivienda media con una calidad superior, sin embargo, el efecto de no disponer de estos servicios en 1990-91 es una notoria reducción de la calidad de la vivienda en relación a la media del correspondiente conjunto de viviendas contemporáneas.

Salvo la particularidad temporal que se acaba de comentar sobre la cale-

facción y el teléfono, las características que definen a la vivienda de calidad media en ambos períodos son: una antigüedad entre 20 y 30 años, una superficie construida entre 61 y 90 metros cuadrados, disponibilidad de luz, un baño (tal vez no completo), agua caliente, y sin garaje. Además, de vivienda media de 1990-91 también se puede caracterizar por: estar situada en edificios con más tres viviendas, sin ascensor, con servicios comunitarios, sin aire acondicionado, sin jardín, ni piscina o zona deportiva, con cocina y calentador de agua de butano, y donde se recurre a aparatos eléctricos y móviles para caldear la vivienda.

Las aportaciones de este capítulo se pueden concretar en que se construyen tres variables índices que permiten sintetizar y analizar la heterogeneidad sobre el nivel de servicios existente en tres conjuntos de viviendas: las representadas en la EPF de 1980-81, de la EPF de 1990-91 y de ambas EPF en su conjunto.

En el siguiente capítulo se comprueba la capacidad de los dos primeros índices para explicar, a través de un modelo de regresión, la variedad de alquileres observados en las viviendas en arrendamiento libre ocupadas en 1980-81 y 1990-91, respectivamente. También en el capítulo III se constatan las ventajas que reporta su utilización en el modelo del alquiler en relación a considerar el conjunto de las variables proporcionadas directamente por la EPF.

Por su parte, el índice de calidad construido con ambas EPF estriba en que hace posible evaluar el cambio (en nivel y distribución) que experimenta la vivienda ocupada en 1990-91 respecto a la ocupada en la década anterior. Un análisis descriptivo de este indicador pone de manifiesto dos cosas: 1) el grado de heterogeneidad de la distribución del acondicionamiento de la vivienda residencial en nuestro país, y 2) una mejora generalizada (en el segundo

período en relación al primero) tanto en media como en distribución, en todas y cada una de las particiones de viviendas consideradas: por comunidades autónomas, por el tamaño de municipio donde se ubica la vivienda, por régimen de tenencia y calificación legal de la vivienda.

La construcción del indicador de calidad común para ambas EPF también ha permitido comprobar, de una forma sencilla, que el tipo viviendas que experimentan un mayor incremento de la calidad en 1990-91 son aquellas que en 1980-81 presentaban una posición más desfavorable<sup>8</sup>. En este sentido se puede decir que en España se ha producido una reducción de la desigualdad en 1990-91 en lo que al acondicionamiento de la vivienda se refiere. Este resultado también se obtiene en Del Río y Ruiz-Castillo (1996) en relación a la desigualdad relativa analizada con datos sobre los gastos del hogar proporcionados por ambas EPF.

### CAPITULO III

El objetivo de este capítulo es identificar los factores que determinan el valor de mercado del flujo de servicios de la vivienda observado en la EPF de 1980-81 y de 1990-91 a precios corrientes.

Como se ha comentado, las EPF recogen el alquiler corriente realmente pagado por el inquilino de las viviendas en arrendamiento. Sin embargo, este precio no reflejará una valoración corriente de mercado en el caso en que la vivienda esté afectada por algún tipo de intervención pública: PO o control de alquileres. Por tanto, el valor del flujo de servicios que debemos explicar es el que corresponde a las viviendas en régimen de arrendamiento 'libre' en sentido estricto (AL a partir de ahora), es decir, cuyo primer contrato se celebró con posterioridad a la Ley de Arrendamientos Urbanos (LAU) de

---

<sup>8</sup>Deducir esta afirmación a través de un análisis de las variables originales en ambas EPF, en lugar del de una única variable síntesis, hubiera sido realmente complicado.

1964.

En este capítulo se identifican tres determinantes del alquiler libre observado en las EPF de 1980-81 y 1990-91: 1) la localización de la viviendas, 3) el año de ocupación y 4) las características de los servicios que proporciona la vivienda. El efecto de estos factores en el alquiler mensual es estimado de forma robusta a través de un modelo lineal (con variable dependiente el logaritmo del alquiler mensual) donde se aplica el método de Peña y Yohai (1995) para detectar observaciones atípicas influyentes.

En el proceso de especificación y estimación del modelo de alquiler en el capítulo se pueden observar las siguientes puntualizaciones. Las variables sobre localización representan un efecto cruzado entre la provincia y el tamaño del municipio donde se ubica la vivienda. Por su parte, el año de ocupación permite estimar el denominado '*tenure discount*'<sup>9</sup>. Se corresponde con aquel descuento que concede el propietario al inquilino que renueva su contrato de arrendamiento, en relación al precio que cobraría en el mercado a un nuevo arrendatario. La existencia del *tenure discount* muestra la preferencia, por parte del propietario, de mantener a los 'buenos' inquilinos una vez estimada la incertidumbre que supondría su sustitución en lo que respecta al pago y mantenimiento del inmueble.

En el modelo presentado en este capítulo se estima un efecto marginal negativo del número de años de ocupación (*tenure discount* interanual) en el alquiler observado en 1980-81 y 1990-91 respectivamente. La interpretación de este efecto es que la tasa de actualización aplicada en la renovación de los contratos es menor a la inflación experimentada en el mercado de nuevos arrendamientos. Además, al comparar la estimación de este descuento interanual con la evolución oficial de la inflación del alquiler (evolución del

---

<sup>9</sup>Börsch-Supan (1986), Miron (1990) y Hubert (1995).

IPC-Base 92 de la rúbrica 33: 'alquiler') proporcionada por el INE, se constata un mayor uso de las cláusulas de actualización en 1990-91 que en 1980-81, así como un cambio estructural producido a mediados de los setenta y a principios de los ochenta.

En relación al tercer factor determinante del alquiler (las características de los servicios de la vivienda) se analizan dos modelos alternativos: 1) un modelo tradicional de regresión hedónica<sup>10</sup>, donde, además de las variables independientes de localización y ocupación, se incluye directamente el conjunto de servicios de la vivienda disponibles en la EPF; y 2) un modelo con índice de calidad (IC), donde dichos servicios se sustituyen por el IC, construido en el capítulo anterior, que sintetiza el nivel de los servicios proporcionados por la totalidad de las viviendas representadas en la EPF de 1980-81 y 1990-91, respectivamente.

Aunque se limita la notación 'hedónico' al primer modelo, conceptualmente el modelo con IC también lo es puesto que la esencia de los modelos hedónicos es la expresión del precio de un bien heterogéneo en función de sus características. La diferencia está en que con el modelo tradicional se obtiene la estimación marginal de cada atributo mientras que con el modelo con IC se obtiene una estimación conjunta de todas ellas. Se pueden enumerar las siguientes ventajas derivadas de utilizar el IC: i) evita un problema de multicolinealidad entre las variables explicativas del alquiler, ii) permite estimar el efecto de variables de baja frecuencia, iii) facilita la estimación de un efecto cruzado entre la calidad de la vivienda y el años de ocupación, iv) supone una forma de depurar la información sobre las características de la vivienda que no ayuden a explicar el alquiler<sup>11</sup>.

---

<sup>10</sup>En Rosen (1974) se establecen los fundamentos microeconómicos de la regresión hedónica en condiciones de competencia perfecta.

<sup>11</sup>El IC, por construcción, está incorrelado con el resto de indicadores que contribuyen

Por su parte, los modelos hedónico y con IC estimados para 1980-81 muestran una equivalencia total en relación a: la bondad de ajuste (en ambos modelos se explica un 60 por ciento de la variabilidad total observada en el alquiler del AL), el efecto estimado de las variables de localización y del *tenure discount*, así como a la identificación de las mismas observaciones atípicas influyentes en el modelo. Estos resultados, además de mostrar una evidencia de la capacidad explicativa y simplificadora del IC, justifican que la especificación del modelo de alquiler que se presenta en este capítulo para 1990-91 se haya limitado únicamente al modelo con IC<sup>12</sup>.

Además de la interpretación de los efectos que explican el alquiler, las aportaciones del capítulo III son dos: 1) la identificación de los determinantes del alquiler observado en el sector de AL en 1980-81 y 1990-91 respectivamente, y 2) la identificación de las observaciones con un comportamiento atípico e influyente en la estimación del modelo.

#### CAPITULO IV

En este capítulo se adjudica, con criterios objetivos, un valor al flujo de los servicios de cada una de las viviendas que compone el parque residencial de nuestro país. Este valor imputado corresponde a una estimación del alquiler corriente que el hogar debería pagar mensualmente por una vivienda similar a la que ocupa en caso de acudir al mercado de nuevos arrendamientos en 1980 y 1990 respectivamente.

Para llevar a cabo esta imputación se estiman los efectos de los determinantes del alquiler libre (que fueron identificados en el capítulo anterior) a

---

a explicar la variabilidad total de las variables de la vivienda.

<sup>12</sup>El IC de 1990-91 sintetiza un total de dieciocho características de la viviendas, el de 1980-81 solo ocho. La estimación hedónica tradicional del modelo para 1990-91, así como la estimación del efecto cruzado entre calidad de la vivienda y ocupación sería realmente complicado.

través de un modelo que tiene en cuenta el posible sesgo de selección muestral en la información disponible. Efectivamente, nótese que el valor corriente del flujo de servicios sólo se observa en el régimen de AL, por tanto, en la medida en que las decisiones sobre las características de la vivienda y el régimen de tenencia se tomen simultáneamente, puede haber un problema de selección muestral. Así, en el modelo se considera la decisión de tenencia: AL u Otro régimen, como un factor endógeno en la decisión sobre las características de la vivienda. El modelo es estimado por máxima verosimilitud y como valores iniciales se consideran el resultado del modelo de Heckman (1979) en dos etapas con el objeto de obtener estimadores eficientes además de consistentes.

Las observaciones que se consideran para estimar el modelo son la totalidad de las viviendas representadas en la EPF, salvo las incluidas en alguno de los tres tipos siguientes: a) aquellas viviendas que no disponen de información sobre alguna de las variables explicativas del modelo<sup>13</sup>, b) aquellas viviendas que se identifican como observaciones atípicas en el modelo de alquiler del capítulo III, y b) aquellas viviendas en régimen de arrendamiento pero con calificación legal (CL) desconocida<sup>14</sup>, un 10 por ciento de total del arrendamiento en 1980-81 y un 31,6 por ciento en 1990-91. En relación a estas últimas (y debido al elevado número de viviendas involucradas, sobre todo en el segundo período) se realiza un análisis pormenorizado que permite comprobar que no hay evidencia<sup>15</sup> en contra de: i) el carácter aleatorio de la falta de información sobre la CL y ii) el efecto de estas variables en la

---

<sup>13</sup>El cómputo de las viviendas afectadas por falta de información se detalla en el capítulo I.

<sup>14</sup>Al no disponer de la calificación legal de este tipo de viviendas, no es posible la identificación entre AL y Otros régimen: arrendamiento de renta congelada (si la ocupación es posterior a la LAU de 1964) o arrendamiento de PO.

<sup>15</sup>Para un nivel de confianza del 95 por ciento.

estimación de los parámetros del modelo. Para llegar a este resultado, en primer lugar se estima un modelo de regresión logístico que permite estimar la CL de estas viviendas en: AL u Otro régimen. Posteriormente, a través de un procedimiento *bootstrap* se identifica el efecto de incluir estas observaciones, con CL estimada, en el modelo. La reestimación *bootstrap* que se aplica permite aproximar de forma independiente la ganancia de eficiencia, derivada de incluir un mayor número de observaciones, del efecto sesgo en la estimación de los parámetros.

Una vez concluida la fase de estimación de los efectos de los determinantes del alquiler (calidad de la vivienda, localización y ocupación), se dispone de un mecanismo objetivo para imputar un valor corriente a la totalidad del parque residencial representado en la EPF de 1980-81 y 1990-91. Para ello basta con computar los efectos estimados de la calidad y la localización a las especificaciones propias de cada vivienda<sup>16</sup>, así como considerar como año de ocupación 1980 y 1990, respectivamente.

El contenido del capítulo IV se completa con un análisis comparativo entre el alquiler medio imputado, vía modelo, y el alquiler medio observado en la EPF: alquiler realmente pagado (si la vivienda es de AL, de arrendamiento de renta antigua o de arrendamiento de PO) y el alquiler autoestimado por el ocupante de la vivienda (si la vivienda está en régimen de propiedad libre, propiedad de PO o en caso de cesión). A su vez, esta comparación sugiere un análisis pormenorizado de la influencia del nivel de calidad de la vivienda en los precios observados en los distintos regímenes de tenencia. De esta forma se se ponen de manifiesto los siguientes resultados: i) una asociación entre

---

<sup>16</sup>En el caso en que una vivienda no dispone de información sobre alguna de las variables relevantes para estimar su alquiler, se le asigna el valor medio que presentan las viviendas del mismo régimen de tenencia en la variable en cuestión, para posteriormente, imputar el precio correspondiente.

tenencia y calidad ya identificado en el capítulo II, ii) una gran coherencia entre la autoestimación por parte del hogar del valor de sus viviendas y el indicador de calidad construido en esta tesis con objeto de sintetizar los servicios de la vivienda; iii) el descuento medio que obtienen el arrendamiento de renta antigua, en relación al alquiler corriente que deberían pagar en el mercado, es superior al del arrendamiento de PO; iv) el hogar en propiedad o cesión infraestima el valor de mercado de su vivienda en 1980-81, mientras que en 1990-91 lo sobreestima. Este último resultado representa una evidencia empírica sobre la supuesta sobrestimación de la autoestimación de la EPF de 1990-91 que llevó al INE a tomar la decisión de eliminar la vivienda en propiedad y cesión del sistema del IPC-Base 1992, cuando en el IPC-Base 1976 y Base 1983 sí fueron incluidos. En Ruíz-Castillo *et al.* (1999) se discute este cambio de criterio seguido por el INE y se analizan sus efectos llegando a estimar el valor del sesgo en la medición de la inflación.

En este sentido, se define una de las posibles líneas de investigación futuras que se comentan en el apartado de extensiones de esta tesis: estimar el efecto que produciría en la inflación oficial proporcionada por el INE, el hecho de considerar el alquiler corriente imputado, con los criterios objetivos presentados en esta tesis, para las viviendas en régimen de propiedad y cesión.

# Capítulo 1

## Representatividad y calidad de la información sobre la vivienda en la Encuesta de Presupuestos Familiares

## 1.1 Introducción

El contenido de esta tesis se fundamenta en analizar la información proporcionada por la Encuesta de Presupuestos Familiares (EPF) de 1980-81 y 1990-91 en relación a la vivienda ocupada en ambos períodos. Al igual que el Censo de Vivienda (CV), las EPF de 1980-81 y 1990-91 son elaboradas por el Instituto Nacional de Estadística (INE) el cual facilita los factores de elevación poblacional que permiten hacer esta información representativa del exhaustivo cómputo de viviendas recogido en los CV de 1981 y 1991 respectivamente. Este capítulo persigue dos objetivos de carácter descriptivo: i) constatar la capacidad de las EPF para representar el parque de viviendas de nuestro país en ambos períodos, y ii) revisar la calidad de la información sobre la vivienda proporcionada en las EPF y la coherencia entre distintas variables de interés.

Para llevar a cabo el primer objetivo se realiza una comparación entre las dos EPF y los registros censales en los años 1981 y 1991. Se pretende constatar el grado de similitud de ambas fuentes al medir distintos aspectos de una misma realidad: el *stock* de viviendas en cada período y el cambio producido en la década que los delimita. El CV, a diferencia de la EPF, no dispone de información sobre las viviendas como el precio, la calificación legal o el año de ocupación, no obstante las características susceptibles de comparación -EPF *versus* CV- se refieren tanto a la antigüedad del inmueble, como a su localización, tamaño de municipio y Comunidad Autónoma a la que pertenece, y el régimen de tenencia en que es ocupada. Este ejercicio ha permitido detectar las reservas que se deben tener en cuenta al trabajar con datos proporcionados por una encuesta. Un buen ejemplo de ello es la dificultad detectada para conocer la 'antigüedad del edificio' en ambas fuentes. Aunque en gran medida el INE dispone y utiliza medios para corregir

ciertos errores detectados en el Censo de Vivienda a través del Censo de Edificios, la corrección en la EPF no se realiza por lo que se requiere la aplicación de algún criterio subjetivo por parte del investigador interesado en ello.

En este sentido, se enmarca el segundo de los objetivos del capítulo: revisar la calidad de la información disponible en las EPF. Esta revisión entraña la aplicación de una serie de filtros para prescindir de información que pudiera distorsionar las conclusiones de un análisis posterior. En el proceso de depuración de la base de datos se ha creado una serie de variables ficticias que permiten identificar cada una de las correcciones y filtros aplicados con el fin de controlar y evaluar las consecuencias de su aplicación en los posteriores análisis.

La estructura del capítulo que nos ocupa se divide en las siguientes partes. En un segundo apartado se realiza la comparación entre los CV de 1981 y 1991, y las EPF de 1980-81 y 1990-91. La exposición del análisis se hace una leve mención a los ámbitos poblacional, geográfico y temporal que abarca cada fuente estadística, el proceso de elaboración de los datos así como la comparación 'CV versus EPF' sobre aquellas variables que lo hacen posible. En el segundo apartado se presenta la revisión de la información disponible en las EPF y los filtros aplicados para crear la base de datos utilizada en la tesis. En el cuarto apartado se resumen las conclusiones más destacadas sobre la representatividad de las EPF y la calidad de la información disponible.

## 1.2 Censo de viviendas *versus* Encuesta de Presupuestos Familiares

Tanto en los Censos de Viviendas como en las Encuestas de Presupuestos Familiares se puede encontrar un gran detalle sobre datos estadísticos relativos al número y características de la vivienda española. La información proporcionada por ambas fuentes es recopilada por el INE con objetivos y períodos de referencia diferentes. El CV tiene como objetivo fundamental la determinación del stock, su distribución geográfica y las características de la vivienda en la fecha censal. Se realizan cada diez años desde 1950, por lo que se dispone en la actualidad de cinco estadísticas censales<sup>1</sup>. El CV representa un registro exhaustivo del número de viviendas de nuestro país. Constituye una base para la formación de un registro de viviendas y sirve de marco de muestreo para la realización de otras investigaciones estadísticas en aspectos relativos tanto a la vivienda como a la población. De hecho, el CV se utiliza como base para determinar la muestra que compone la EPF, de tal forma que cada encuesta informa de la totalidad del stock de viviendas una vez se aplica el correspondiente factor de elevación proporcionado por el INE.

Por su parte, el objetivo prioritario de la EPF es la actualización de los bienes y servicios que integran la cesta de la compra de los hogares españoles, así como sus ponderaciones a los efectos de la elaboración del Índice de Precios de Consumo. A partir de la encuesta referida al año 1980-81, el INE integra en la EPF la información contenida en la Encuesta de Equipamiento y Nivel Cultural de las Familias con el propósito de mejorar las estimaciones de los gastos en bienes de consumo duradero que hasta entonces se habían infravalorado. Además, se enriquece notablemente la información ofrecida

---

<sup>1</sup>El próximo CV estará disponible en el 2001.

sobre equipamiento y condiciones de la vivienda con una ventaja adicional: la posibilidad de realizar estudios que relacionen la posesión de los bienes duraderos con el amplio conjunto de variables sobre el hogar disponible en la EPF.

### 1.2.1 **Ámbito poblacional**

En el CV la unidad básica fundamental es la vivienda en un sentido amplio, esto es, todo recinto destinado a habitación humana y aquellos otros que, sin tener esa finalidad, están efectivamente habitados en la fecha censal. Por el contrario, no se incluyen las viviendas en construcción si no están habitadas, a no ser que su terminación esté pendiente de ligeros detalles que no impedirían ser ocupadas en el momento censal. Tampoco se incluyen las viviendas que están en proceso de demolición o están vacías por haber sido consideradas en estado ruinoso. En el CV de 1981 se clasifican las viviendas familiares<sup>2</sup> atendiendo a su disponibilidad y a la utilización de las mismas a lo largo del año, dando lugar a la distinción entre vivienda principal, secundaria y desocupada (disponible, o no, para alquiler o venta). En el CV de 1991 en la clasificación 'vivienda desocupada' se engloba tan solo las viviendas no disponibles por trámites de venta o alquiler, mientras que las disponibles eventuales<sup>3</sup> se consideran en 'otro tipo'. A efectos censales, tanto la vivienda principal como la secundaria se consideran ocupadas dado que están disponibles para ello aún cuando en el momento de su identificación no se encuentren habitadas.

---

<sup>2</sup>Frente a las viviendas *familiares*, cuyo destino es ser ocupadas por familias, el censo distingue entre *alojamientos* (pajares, molinos, garajes,...) y viviendas *colectivas* (hoteles, pensiones, campamentos militares,...) no concebidos en un principio con fines residenciales habituales o definitivos.

<sup>3</sup>Generalmente las ocupadas sólo en determinados períodos del año.

La información proporcionada por la EPF se refiere únicamente a las viviendas ocupadas. Se considera vivienda a todo recinto ocupado por uno, o más, hogares<sup>4</sup> y que en el momento de la encuesta no se utiliza totalmente para fines diferentes a los residenciales. Las viviendas se clasifican también en principales y secundarias según el grado de utilización temporal de la misma a lo largo del año, como se define más adelante.

### **1.2.2 Ámbitos geográfico y temporal**

Tanto el Censo como la EPF extienden su investigación a todo el territorio español, esto es, las cincuenta provincias peninsulares e insulares más Ceuta y Melilla.

La fecha de referencia de los Censos de interés en este estudio se centra en el 1 de marzo de 1981 y 1991 respectivamente. La recogida de información tiene lugar entre el 1 de marzo y el 31 de mayo de dichos años. El marco básico de referencia para la localización de las viviendas es el Censo de Edificios y Locales cuya información es explorada en su totalidad, sin aplicar ningún tipo de técnicas de muestreo.

La EPF 1980-81 (1990-91) se realiza durante el año comprendido entre el 1 de abril de 1980 (1990) y el 31 de marzo de 1981 (1991). El número de encuestas realizadas se distribuye homogéneamente en cuatro trimestres en todo el ámbito espacial. Con ello se capta todo tipo de consumo de carácter estacional, pero el desfase temporal entre el primer y el último hogar entrevistado impide contabilizar las variaciones en el stock producidas entre trimestres. Además, parte de la información del CV se recoge dos meses después que la EPF, lo que explica que el parque de viviendas ocupado

---

<sup>4</sup>Se entiende por hogar al conjunto de personas que comparten gastos con cargo a un mismo presupuesto y que, por lo general, mantienen una relación de carácter familiar.

según la EPF sea inferior al que aparece en el censo. Como se verá más adelante, esta circunstancia se observa de forma clara para las viviendas de más reciente construcción.

### 1.2.3 Proceso de elaboración de datos

Los Censos de Viviendas recogen de forma exhaustiva el número de viviendas familiares, alojamientos, y viviendas colectivas, así como la clasificación de las familiares en principales, secundarias y desocupadas. Las restantes características de las viviendas censadas se obtienen por muestreo, al igual que la totalidad de la información recogida en la EPF.

Los resultados censales sobre las características de la vivienda, se obtienen mediante una muestra<sup>5</sup> aproximada del 25 por ciento del total de viviendas que se realiza en función del tamaño de municipio atendiendo a su población de derecho. A cada vivienda se le hace corresponder un código municipal, junto al provincial, lo que constituye la identificación de las viviendas y permite la adjudicación de un factor de elevación que hacen a la muestra representativa.

Las viviendas investigadas en la EPF suponen un 0,23 por ciento del total en 1981 y un 0,19 por ciento en 1991. Con la base de la información censal, la selección de la muestra<sup>6</sup> responde a criterios geográficos, demográficos y

---

<sup>5</sup>La técnica aplicada es un muestreo estratificado con selección sistemática de las viviendas con arranque aleatorio (véase INE, 1982 e INE, 1991).

<sup>6</sup>Se realiza un muestreo bietápico con estratificación de las unidades de la primera etapa, diseñándose una muestra independiente para cada provincia. Las unidades de la segunda etapa son todas las viviendas familiares existentes en las secciones censales seleccionadas en la etapa anterior. Los criterios de estratificación de la primera etapa son de carácter geográfico de acuerdo a la importancia demográfica de los municipios, así como de carácter socioeconómico dentro de cada estrato. Las secciones muestrales se seleccionan dentro de

socioeconómicos de tal forma que el factor de elevación que permite la estimación de los datos poblacionales, se elabora al nivel del tamaño de municipio en cada provincia.

#### 1.2.4 Definición y clasificación de las viviendas ocupadas

Se entiende por 'vivienda' todo recinto estructuralmente separado e independiente que por la forma en que fue construido, transformado o adaptado, está concebido para ser habitado por personas. La clasificación de 'viviendas ocupadas' que presenta el CV entre principales y secundarias es compatible con la de la EPF y su definición es la siguiente:

- Principal: cuando es utilizada toda o la mayor parte del año como residencia habitual por sus ocupantes.
- Secundaria: cuando es utilizada solamente parte del año de forma estacional, periódica o esporádica y no constituye la residencia habitual de una o varias familias.

El número de viviendas de cada tipo se muestra en el Cuadro 1.1 donde aparece el stock de viviendas ocupadas de cada tipo según la EPF y el CV así como la diferencia entre ambas fuentes. El *stock* que recogen las dos EPF es menor que el de los respectivos CV, lo que se puede deber a que, como se comentó en el apartado anterior, parte de la información de los CV se recoge dos meses después que la de las EPF. Sin embargo, destaca el dispar comportamiento de las viviendas principales y secundarias: el número de viviendas cada estrato con probabilidad proporcional a su tamaño medido en número de viviendas. La vivienda se selecciona en cada sección muestral con igual probabilidad mediante un muestreo sistemático con arranque aleatorio (véase INE, 1983a e INE, 1993c).

secundarias supone un 56 por ciento (65 por ciento) menos en la EPF que en el Censo en 1991 (1981), mientras que el porcentaje correspondiente a las viviendas principales no supera el 4 por ciento en ambos años. Ahora bien, como se advierte en las publicaciones del INE, la distinción censal entre viviendas secundarias y desocupadas presenta dificultades prácticas. En general, la carga política y social del término vivienda desocupada reduce su consideración en favor de las viviendas secundarias. Este factor debe influir en que el número de viviendas secundarias del CV supere en más de un millón doscientas mil a las que aparecen en la EPF, mientras que la diferencia en las principales está en torno a las cuatrocientas mil en ambos años, cifra que supone un 5,7 por ciento del incremento interanual del stock de viviendas principales censadas.

El Cuadro 1.2 permite analizar el cambio producido en el stock de viviendas en esta década tanto en el Censo como en la EPF. Según la fuente censal hay algo más de dos millones trescientas mil viviendas ocupadas en 1991 que en 1981. La EPF cuantifica este incremento en aproximadamente cuatrocientas mil viviendas menos. Esta divergencia se debe, casi en su totalidad, a las viviendas secundarias, cuyos problemas de interpretación ya se han comentado en el párrafo anterior. Por el contrario, el incremento de las viviendas principales según el Censo y la EPF es muy similar, en torno al millón trescientas mil viviendas. En términos relativos, la EPF capta el mismo fenómeno que el Censo: el parque de viviendas principales se incrementa en algo más de un 12,5 por ciento.

### **1.2.5 Características de las viviendas**

Ciertas características que aparecen en la EPF requieren una adaptación a los criterios de clasificación que aparecen publicados en el CV para su

comparación. En el CV de 1991, las viviendas secundarias, desocupadas y otras se consideran conjuntamente lo que imposibilita la comparación de sus características con las de las viviendas secundarias del CV de 1981 y la EPF del año correspondiente<sup>7</sup>. Esto obliga a limitar el estudio comparativo a las viviendas principales. Las características analizadas se refieren a la ubicación de la vivienda, el régimen de tenencia y el año de construcción del edificio.

### Ubicación de la vivienda

Las características publicadas en el CV referentes a la localización geográfica de la vivienda son la provincia o Comunidad Autónoma y el tamaño del municipio al que pertenece.

### Comunidad Autónoma

La distribución porcentual de la vivienda principal por Comunidades Autónomas aparece en el Cuadro 1.3. Para la EPF se presenta la distribución muestral y la poblacional, obtenida tras la aplicación de los factores de elevación que facilita el INE. Tanto en la EPF de 1980-81 como en la de 1990-91, puede apreciarse claramente el nivel de corrección que proporcionan estos factores, destacando Madrid y Cataluña como las CCAA donde las muestras subestimaban en mayor medida el número de viviendas principales. En Castilla-León, Castilla la Mancha y Aragón ocurre lo contrario: la muestra sobrestima este tipo de stock.

Las CCAA aparecen ordenadas de mayor a menor porcentaje de viviendas

---

<sup>7</sup>Además, la falta de información adecuada sobre las características de la vivienda secundaria en el CV de 1991, ha llevado al INE a tomar la decisión de que en el próximo censo, correspondiente al año 2001, y salvo la pregunta sobre *clase de vivienda* todas las demás variables se refieran únicamente a la vivienda principal.

en 1981. Se constata que ésta ordenación coincide para las cuatro columnas referentes a datos poblacionales, salvo una permutación entre Andalucía y Cataluña en 1991. Según el Censo y la EPF, Andalucía pasa a ser la CCAA que presenta el mayor número de viviendas principales en 1991, seguida de Cataluña cuando en 1981 la situación era la contraria. Más del 54 por ciento de las viviendas principales ocupadas en 1981 y 1991 se localizan en las comunidades de Cataluña, Andalucía, Madrid y Valencia.

### **Tamaño de municipio**

Se distinguen los siguientes modalidades en relación a la población de derecho:

- Zona Rural: municipio con menos de 2.000 habitantes.
- Zona Intermedia: municipio entre 2.001 y 10.000 habitantes.
- Zona Urbana: municipio con más de 10.000 habitantes.

La vivienda localizada en los municipios de menor tamaño están mejor representadas en el CV que en la EPF: en el Cuadro 1.4 se muestra claramente que la EPF subestima la zona rural y, por el contrario, sobrestima la zona intermedia y, de forma más notoria, la urbana. Aplicar el factor de elevación no corrige este sesgo: pasar de los datos muestrales a la población supone obtener un ligero menor peso (inferior al 2,3 por ciento) en la zona rural a favor de las zonas mayores.

Como se observa en el Cuadro 1.5, la distorsión Censo-EPF se produce con mayor intensidad en 1991 que en 1981, en especial en el medio rural donde la diferencia se aproxima a un 1 millón de viviendas ocupadas en 1981 y sobrepasa 1.3 millones en 1991.

Para analizar el cambio producido en la década de estudio se recurre al Cuadro 1.6. Aunque en media el número de viviendas principales según el

Censo y la EPF se incrementa entre 1981 y 1991 en cifras similares (alrededor de 1,3 millones) ambas fuentes difieren en lo que se refiere a su localización. Así, el incremento en el número de viviendas rurales se aproxima al 2 por ciento según el Censo mientras la EPF registra una reducción del 27 por ciento. Por el contrario, la EPF de 1990-91 registra un mayor aumento que el Censo en las zonas más densamente pobladas, sobre todo en la zona urbana donde el incremento es de 1,3 millones, cuando para el Censo es de 1 millón.

Según el Censo, la distribución porcentual de las viviendas principales se mantiene en la década de los ochenta salvo el incremento de un 2 por ciento producido en la zona urbana al tiempo que una reducción similar en la rural (véase Cuadro 1.3). Este fenómeno se aprecia igualmente en la EPF poblacional con una magnitud próxima al 4 por ciento. La proporción de viviendas en la zona intermedia permanece estable los dos años de estudio, en torno al 17 por ciento del total de viviendas según el Censo y al 19 por ciento según las EPF.

### **Régimen de tenencia**

Los conceptos por los cuales el hogar disfruta de la vivienda que ocupa, son los siguientes:

- Propiedad: donde a su vez se distingue entre,
  - a) Totalmente pagada, incluyendo donaciones y herencias.
  - b) Pendientes de pago.
- Arrendamiento: cuando alguno de los ocupantes satisface por el uso de la vivienda una cantidad, anual o mensual, en metálico o en especie, e independientemente de que la persona que percibe esa cantidad sea el

propietario o un arrendatario y de que medie, o no, contrato entre los agentes.

- Cedidas por la empresa: viviendas facilitadas (gratuita o semigratuitamente) por el patrono, empresa u organismo para el que trabaja alguno de los ocupantes de la vivienda durante el tiempo en que preste sus servicios.
- Otra forma: viviendas facilitadas por otros hogares, instituciones públicas o privadas, u otro tipo que no se ajuste a los casos anteriores.

La subestimación aproximada del 4 por ciento en las viviendas principales de las dos EPF respecto a los Censos, aparece desglosada por el régimen de tenencia en el Cuadro 1.7. Según este criterio, la distribución del *stock* en ambas estadísticas, es muy similar, sobre todo en 1991. Ambas fuentes reflejan que el 78 por ciento de las viviendas son propiedad de sus ocupantes en 1991, el 15 por ciento arrendadas y el 7 por ciento restante corresponde a algún tipo de cesión.

La subdivisión hecha dentro de la vivienda en propiedad permite observar que el CV registra mayor proporción de 'pendientes de pago' que la EPF. No debe sorprender este hecho si se tienen en cuenta los períodos de recogida de información de ambos tipos de datos y la concentración de la subestimación cometida por la EPF en las viviendas más nuevas, como se muestra en el Cuadro 1.10.

Aunque las cuantías difieren entre ambas fuentes, de acuerdo con el Cuadro 1.8 se observa que los CV y las EPF permiten obtener las mismas conclusiones en cuanto al cambio en el stock de viviendas según el régimen de tenencia:

- El número de viviendas ocupadas en propiedad se incrementa en más de 1,5 millones en 1991 respecto a diez años antes. Este resultado es debido al aumento en las viviendas pagadas, donadas o heredadas ya que el número de viviendas pendientes de pago se reduce.
- El cambio de las viviendas en arrendamiento y cedidas por la empresa, tiene un signo negativo. Las primeras presentan una mayor magnitud en la reducción para la década de estudio, suponiendo en 1991 un 15 por ciento del total de viviendas, como se veía en el cuadro anterior. Por su parte, la vivienda cedida por la empresa representa poco más del 1 por ciento.
- Las viviendas cedidas consideradas dentro del epígrafe 'otras forma', experimentan en 1991 un aumento respecto al stock que suponían en 1981.

### **Año de construcción del edificio**

Tanto en el Censo como en la EPF el hogar es la unidad informante del año de construcción del edificio. En general, este dato no es conocido con exactitud, sobre todo cuando la vivienda data de fechas muy alejadas en el tiempo. Muestra de ello son las discrepancias encontradas en el Censo entre las distintas viviendas de un mismo edificio, y con respecto a los datos proporcionados por el Censo de Edificios. Para los casos donde no se proporciona el año de construcción, el Censo tiene medios para adjudicar una edad, generalmente a partir del Censo de Edificios<sup>8</sup>. En cambio, en la EPF

---

<sup>8</sup> Este es uno de los motivos por los que en el próximo Censo del 2001 se introducirá una novedad: la búsqueda de una óptima coordinación entre la preguntas de viviendas y las de edificios lo que supondrá, entre otras cosas, la imposibilidad práctica de incoherencias entre ambas informaciones.

se informa explícitamente de la carencia del dato, lo que ocurre en el 0,3 por ciento del total de viviendas principales en 1980-81 y en un 1,4 por ciento en 1990-91.

En el Cuadro 1.9<sup>9</sup> se compara el Censo de Edificios y el Censo de Viviendas. En él se proporciona la distribución respecto al año de construcción del total de 17,20 millones de viviendas familiares registradas en 1991. La persona encuestada en el CV tiende a retrasar la antigüedad de su vivienda cuando ésta se construyó con posterioridad a 1941. Por ello, el porcentaje de las construidas con anterioridad es superior al que proporciona el Censo de Edificios. En cualquier caso, es clara la similitud de ambas distribuciones.

Nada hace pensar que este problema no ocurra igualmente en la EPF. En concreto, en el Cuadro 1.10 se observa que para 1981 el Censo registra cerca de 0,5 millones más de viviendas que la EPF para la construcción anterior a 1941. Así, las viviendas construidas antes de 1941 suponen el 26 por ciento del total de viviendas censadas, mientras que la EPF 1980-81 sobrevalora este dato hasta el 32 por ciento, por tanto 6 puntos porcentuales de diferencia.

Adicionalmente, en la EPF de 1990-91 la variable que nos ocupa se define como el año de terminación de la construcción del edificio o de la última transformación fundamental del mismo. La verdadera edad del edificio se adelanta en el tiempo en la medida que se realizan obras de renovación o rehabilitación del mismo, lo cual ocurrirá con mayor probabilidad en las viviendas de más edad. Se añade un matiz al contenido de la variable que puede justificar el mayor número de viviendas antiguas en el Censo de 1991 que el registrado en la EPF 1990-91, que es cercana a 2 millones de viviendas

---

<sup>9</sup>No es posible realizar una comparación directa entre la EPF y el Censo de Edificios debido a que este último no cuenta con información desagregada de las viviendas familiares. Considera conjuntamente las viviendas principales, secundarias, desocupadas disponibles y desocupadas no disponibles.

para una antigüedad superior a 30 años. La infravaloración más notoria de la EPF 1990-91 *versus* el CV 1991, se produce en este primer tramo y es del orden de 32 por ciento cuando en 1980-81 se daba una sobrevaloración próxima al 17 por ciento.

Como consecuencia de este dispar comportamiento de la EPF en ambos años de estudio, se explica el dato más contradictorio de este trabajo comparativo mostrado en el Cuadro 1.11. En efecto, se observa que la reducción de viviendas ocupadas en 1991 respecto a 1981, para la construcción anterior a 1951, es superior a 1,8 millones según la EPF mientras que los datos censales reflejan tan solo una reducción de 0,6 millones. Únicamente en el tramo de las viviendas más nuevas, el sentido de las discrepancias el censo y la EPF coinciden para 1981 y 1991: el CV contabiliza más viviendas principales ocupadas que la EPF. A excepción de las viviendas de 'otra forma' y para 1981, este resultado es generalizado en todos los demás regímenes de tenencia (véase Cuadros 1.12). Sin duda, ésto es debido a los períodos de recogida de información de las encuestas, como se daba cuenta en el apartado 'Ambito Temporal' de este trabajo. La antigüedad más frecuente de la vivienda ocupada en 1991 y 1981 es entre 10 y 20 años. Resultan ser, en media, un 28 por ciento del total de viviendas según el CV y la EPF. Analizando según el tipo de tenencia (Cuadro 1.13), el problema de infravaloración de las viviendas más nuevas de la EPF 1980-81 comentado anteriormente, se localiza en el 69 por ciento de las viviendas totales que componen la propiedad. El stock de viviendas 'cedidas por razón de trabajo' se infravalora en la totalidad de tramos estudiados mientras que, si la cesión se debe a 'otros motivos', la EPF de 1980-81 sobrestima el número de viviendas para cualquier antigüedad de la misma, aunque destacadamente en las de antigüedad superior a 50 años. Por el contrario, el régimen de 'arrendamiento' se encuentra sobrestimado

por la EPF respecto al CV de 1981, salvo para las viviendas de menos de 11 años en construcción.

El comportamiento del año de construcción en la EPF 1990-91 *versus* CV 1991, resulta homogéneo ante el desglose por régimen de tenencia, en los siguientes puntos: a) para la vivienda más antigua el CV contabiliza mayor número de viviendas que la EPF y b) La vivienda más nueva está también infravalorada por la EPF. Las razones pueden ser distintas, en el primer caso se debería a que el año facilitado en la EPF por el encuestado se refiere en mayor medida a la última reparación fundamental del edificio que a su construcción propiamente dicha. El segundo caso se ha justificado anteriormente por motivos de periodicidad de la realización de las encuestas.

En general, tanto en 1981 como en 1991 y respecto a CV y EPF, se puede deducir que más del 65 por ciento de las viviendas ocupadas en propiedad cuentan con menos de 30 años; esta cifra se aproxima al 53 por ciento para las cedidas, por razón de trabajo u otra forma, y al 40 por ciento para alquiler. En concreto, las viviendas con 10 o menos años de antigüedad representa un mayor porcentaje dentro de la propiedad (un 16,4 por ciento de las mismas, dato censal de 1991), el segundo lugar lo ocupa otra forma de cesión distinta a razón de trabajo (16,1 por ciento), siguiendo las viviendas cedidas por las empresas (11,7 por ciento). Así pues son las viviendas arrendadas las que tienen un menor porcentaje de viviendas de más reciente construcción (8,6 por ciento). Los equivalentes datos de la EPF 1990-91 son 16,1 por ciento para propiedad, 12,6 para otra cesión, 9,7 para cesión por trabajo y 4,7 para arrendamiento. Estas cifras muestran una menor representación de las viviendas más nuevas en la EPF que en el CV en todos los regímenes de tenencia, circunstancia que se deduce igualmente con los datos disponibles del año 1981, como se muestra en los Cuadros 1.14 y 1.15. La vivienda

construida en la década de los cuarenta destaca por ser la de menor nivel de ocupación en los años de referencia y para todos los regímenes de tenencia. La década de expansión económica de los 60 tiene un gran peso porcentual en ocupación de viviendas para todos los tipos de tenencia. Comparativamente con otros regímenes, las viviendas arrendadas son las que registran un mayor porcentaje de viviendas construidas antes de 1941<sup>10</sup>, del orden del 37,2 por ciento en 1981 según el CV (36,6 por ciento según la EPF) y del 28,7 por ciento en 1991 (el 24,4 por ciento en la EPF).

A pesar de la dificultad de esta variable, en el Cuadro 1.11 se obtiene una conclusión lógica en cuanto al cambio de la distribución entre 1981 y 1991. Según la EPF y el CV, las viviendas ocupadas de mayor antigüedad tienden a ser reemplazadas por las de reciente construcción. En 1991 las primeras ven reducido su número en términos porcentuales a favor de las segundas. En los Cuadros 1.14 y 1.15, donde se desglosa esta información por el régimen de tenencia según el Censo y la EPF respectivamente, se obtiene este mismo resultado para cada uno de los regímenes de tenencia de la vivienda principal. Los dos tipos de viviendas principales donde se ve reducida su ocupación en 1991 respecto a 1981 son la vivienda cedida por razón de trabajo y la arrendada. Esta reducción es sistemática para todos los tramos de antigüedad de las viviendas en las primeras, con la única excepción de la década de los 70 para el arrendamiento. En propiedad y otra cesión, aunque se eleva el número de viviendas ocupadas en 1991, se observa una mayor discrepancia cuando se analiza la antigüedad del edificio con una tendencia clara a abandonar las viviendas más antiguas a favor de ocupar las

---

<sup>10</sup>Salvo en la EPF 1980-81, donde son las viviendas cedidas por entidades distintas a la empresa, las que poseen un mayor porcentaje (41,12 por ciento) de viviendas construidas antes de 1941.

de construcción más reciente.

### **1.3 Revisión de la información disponible en la EPF**

En este apartado se pretende revisar la calidad de la información proporcionada por los hogares que componen las Encuestas de Presupuestos Familiares de 1980-81 y 1990-91 en referencia a sus viviendas: principal y secundaria(s). El interés se centra en una serie de variables clave: 'año de construcción del edificio' (AC)<sup>11</sup>; 'año de ocupación' (AO) que informa de la fecha de ocupación de la vivienda arrendada y del año de compra, herencia, donación o cesión según el caso; 'régimen de tenencia y calificación legal' (RTC); 'precio de alquiler' (PA) que corresponde al valor pagado mensualmente si el régimen es el arrendamiento y un valor imputado equivalente en otro régimen; y 'precio de compra'(PC) en caso de vivienda en propiedad y supone el valor corriente en el momento de la adquisición o a su estimación en caso de herencia.

La forma de evaluar la calidad de la información disponible consiste en contabilizar los errores, o incoherencias muestrales encontradas y, en su caso, explicar las correcciones efectuadas haciendo constar la subjetividad de las mismas. Esta revisión sugiere la aplicación de una serie de filtros con el fin de prescindir de aquellas viviendas que no reporten información o que puedan distorsionar los distintos puntos a analizar. Estos filtros se pueden definir en

---

<sup>11</sup>En el apartado anterior se comentan los problemas a la hora de recordar el 'año de construcción del edificio' por parte de los hogares que componen una encuesta. También allí se advierte del matiz que se añade al contenido de esta variable en la EPF 1990-91, ya que puede expresar el año de la última renovación sustancial realizada en la vivienda.

relación al conjunto de viviendas seleccionadas en los mismos:

- 1º Filtro Viviendas con información sobre calificación legal: régimen libre (L) o protección oficial<sup>12</sup>.
- 2º Filtro Viviendas con información del año de construcción anterior, o igual, al año de ocupación ( $AC \leq AO$ ).
- 3º Filtro Viviendas con información sobre el valor de compra o su estimación ( $PC \neq 0$ ) y/o el alquiler mensual real o estimado ( $PA \neq 0$ ).

Seguidamente se procede a comentar y cuantificar la importancia de cada uno de ellos en relación a las variables involucradas.

### **1.3.1 Régimen de tenencia y calificación legal**

El régimen de tenencia hace referencia a las alternativas en que se disfruta la vivienda por parte de sus ocupantes (propiedad, arrendamiento o cesión) mientras que la calificación legal (CL) muestra la información que proporciona la EPF sobre las siguientes opciones:

- Renta libre (L).
- Protección oficial (PO). Se refiere a viviendas, en propiedad o arrendamiento, cuyo valor no responde al mercado libre y que la EPF de 1980-81 recoge bajo los siguientes conceptos: subvenciones, renta limitadas (grupo 1º o 2º), viviendas sociales, viviendas de protección oficial (Real Decreto de 10 noviembre de 1978), o viviendas con renta limitada pero se desconoce el tipo. Por su

---

<sup>12</sup>Todas la viviendas disponen de información sobre el régimen de tenencia: propiedad, arrendamiento o algún tipo de cesión.

parte, en la EPF de 1990-91 se distinguen tres grupos: protección oficial privada, protección oficial pública y otras formas de vivienda protegida.

- No sabe.

La información conjunta sobre el régimen de tenencia y la calificación legal se considera en una variable denominada RTC que se presenta en los Cuadros 1.16 y 1.17.

La totalidad de las viviendas contenidas en ambas muestras presentan información sobre el régimen de tenencia, aunque no ocurre lo mismo con la variable CL. El número de hogares al que se eleva la no disponibilidad muestral de información sobre CL en 1980-81 es de 2.509 y de 3.451 en 1990-91, que representan 1.072.997 y 1.941.931 viviendas respectivamente a efectos poblacionales. Para evaluar la importancia relativa de esta falta de información, en el Cuadro 1.16 se cuantifica en términos porcentuales en relación a la totalidad de viviendas contenidas en la EPF.

Más del 10 por ciento de los hogares que componen la EPF de 1980-81 no proporcionan la calificación legal de su viviendas y este porcentaje supera el 15 por ciento en la EPF de 1990-91. En el Capítulo IV de esta tesis se propone un método estadístico que permite adjudicar una probabilidad estimada de pertenecer a renta libre a toda vivienda en arrendamiento sin CL. En relación a esta probabilidad se definen una regla de clasificación discriminante: libre frente PO, que permite recuperar información con un determinado error asociado debido a su condición de clasificación estimada.

### Primer filtro

Se prescinde de las viviendas 'sin CL', y la nueva distribución de la variable 'régimen de tenencia y calificación legal' en cada período aparece en el Cuadro 1.17.

El análisis desagregado según vivienda principal y secundaria permite apreciar el comportamiento diferente de la variable RTC en cada caso. Más del 90 por ciento de las viviendas secundarias están en régimen de propiedad en 1980-81 y más del 96 por ciento en 1990-91. Además, la proporción de éstas que corresponde a PO es muy reducido en comparación con la vivienda principal en propiedad. Esta observación es igualmente apreciable para las viviendas en arrendamiento.

### **1.3.2 Año de construcción y ocupación**

Si bien la información sobre el año de construcción de la vivienda (AC), y salvo los problemas analizados en el Apartado I, es una pregunta homogénea para todo tipo de viviendas no ocurre lo mismo con el año de ocupación (AO). Así, la EPF recoge dos tipos de información sobre la ocupación dependiendo del régimen de tenencia: i) para las viviendas en arrendamiento se pregunta por la fecha de ocupación, que con toda probabilidad debería coincidir con la de la realización del contrato, mientras que, ii) para la vivienda que se disfruta en concepto de propiedad se pregunta por el año de la compra, herencia o donación correspondiente.

Respecto a la calidad de esta información, en principio, se podría presuponer un menor error en la variable AO que en AC debido a la aproximación más reciente en el tiempo y la implicación directa del entrevistado en tal acontecimiento. Interrelacionando la información disponible sobre ambas variables se detectan dos tipos de problemas: i) el entrevistado ignora cuando fue construida y/o ocupada la vivienda, y ii) incoherencia en el sentido de encontrar que el 'año de ocupación' es anterior al 'año de construcción' ( $AC > AO$ ). En el Cuadro 1.18 se muestran las viviendas afectadas, así como su desglose por tipo de vivienda (principal y secundaria) para cada período

muestral.

El 81 por ciento de los datos disponibles en 1990-91 son correctos, cuando en 1980-81 suponían un 90 por ciento. Resulta relevante el incremento porcentual de los errores por incoherencia producido en 1990-91 para las viviendas principales. Además, en el Cuadro 1.19 (donde se muestra muestra esta información según los regímenes de tenencia: Propiedad, Arrendamiento y Otros<sup>13</sup>) se comprueba una clara asociación entre este tipo de error y el régimen de propiedad: en 1980-81 aproximadamente el 91 por ciento de este tipo de errores es cometido por los propietarios de la vivienda y una década después este porcentaje se eleva al 98 por ciento.

Caben dos circunstancias que podrían justificar la mayor frecuencia de 'error por incoherencia' y su relación con la 'propiedad':

- (1) Para el régimen de 'propiedad por compra' la variable AO no es más que una aproximación al año de ocupación puesto que corresponde al año en que se adquiere la vivienda. En este caso es habitual comenzar a pagar una vivienda con fecha previa a la finalización de su construcción, lo que explicaría la diferencia con respecto a la vivienda en arrendamiento, u otras, para las que AO se define como 'año de ocupación'.
- (2) En 1990-91 el 'año de construcción' puede reflejar la última renovación sustancial realizada en la vivienda, lo que explicaría la diferencia cuantitativa de este tipo de error en ambas encuestas, a la vez que una mayor frecuencia en el régimen de propiedad.

---

<sup>13</sup>El epígrafe 'Otros' engloba los regímenes distintos de propiedad y arrendamiento, esto es, vivienda cedida gratuita o semigratuitamente por instituciones públicas o privadas, por razones de trabajo, etc.

Por otra parte la falta de dato también está asociada al régimen 'Otros' con cerca del 88 y 96 por ciento, respectivamente, en cada período. Tan solo para un 12 por ciento de las viviendas con el régimen 'Otros' se dispone de la fecha de ocupación en 1980-81 mientras que para 1990-91 se desconoce esta información en todos los tipos de cesión.

### Segundo filtro

Se pretende considerar el mayor número de viviendas posible y a su vez limitarlo a aquellas que dispongan de una 'correcta' información sobre ambas variables, por ello se aplica el siguiente criterio subjetivo:

- Se prescinde de la totalidad de hogares sin dato en AC y/o AO.
- Se modifican directamente las variables AC y AO en 23 datos puntuales de 1980-81 y 1 dato de 1990-91<sup>14</sup>. Se debe a casos donde claramente se ha cometido un error de transcripción al anotar el año de ocupación y/o el año de construcción. Dos ejemplos representativos de este tipo de errores son: 1) se observa año 997 y se modifica por 1.997, y 2) se observa 9.181 y se interpreta 1.981.
- Se recuperan 685 datos muestrales por incoherencia en la información en 1980-81 y 2.113 en 1990-91<sup>15</sup>. La intervención consiste en asignar un año de construcción igual al de ocupación de acuerdo con los siguientes criterios en relación al número de años en que AC supera a AO ( $d =$

---

<sup>14</sup> Como información útil para los usuarios de las EPF se proporcionan los Números de Registro que identifican estos hogares en las encuestas. Para la EPF 1980-81 son: 1.196, 3.892, 4.407, 4.875, 5.062, 5.348, 10.264, 11.931, 12.811, 13.023, 13.517, 13.530, 14.393, 14.643, 15.062, 16.121, 16.403, 17.465, 18.398, 19.472, 19.475, 19.476 y 23.791. En términos poblacionales, estos hogares son representativos de 11.006 viviendas. Para la EPF 1990-91 el único hogar afectado corresponde al registro 6.877 y representa a 200 viviendas.

<sup>15</sup> Son representativos de 296.994 viviendas en 1980-81 y de 1.063.132 en 1990-91.

$AC - AO$ ), esto es:

- a.- Cuando  $0 < d \leq 10$  se atrasa el año de construcción haciéndolo coincidir con el de ocupación ( $AC^* = AO$ ). Afecta a 674 viviendas en 1980-81 y a 1.995 en 1990-91. Para justificar esta decisión se remite a los comentarios sobre la mayor fiabilidad de la variable de ocupación que de la construcción, así como a la consideración de ser  $d$  una diferencia en años lo suficientemente pequeño para plantearse recuperar la información disponible.
- b.- Cuando  $d$  es un múltiplo de 10 igualmente de hace  $AC^* = AO$ . Se consideran errores de transcripción en la cifra de posición centésima del año de construcción. Afecta a 11 viviendas en 1980-81 y a 118 en 1990-91.

Este criterio se aplica con carácter general para las viviendas principales y secundarias. Se construye una variable ficticia que permite identificar las viviendas modificadas y filtradas con el fin de controlar el posible efecto de estas decisiones en investigaciones posteriores.

Una vez aplicado el filtro, el número de viviendas con año de construcción superior o igual al de ocupación se reduce a 9.925.956 en 1980-81 y a 11.259.864 en 1990-91. En el Cuadro 1.20 aparecen las distribuciones del año de construcción y ocupación para distintos períodos<sup>16</sup>. En primer lugar, se observa que el parque de viviendas ocupado con fines residenciales en 1990-91 es considerablemente de menor antigüedad que el de 1980-81: cerca de 1,3 millones de viviendas construidas antes de 1920 son abandonadas en la década de los ochenta. Además, el 42 por ciento de las viviendas ocupadas en

---

<sup>16</sup>Para la EPF 1980-81 el tramo comprendido 'entre 1973 y 1980' incluye la información recogida hasta marzo del año 1981.

1980-81 tenían una antigüedad superior a 25 años, porcentaje que se reduce a un 36 por ciento en 1990-91.

En ambas EPF se constata un mismo hecho que puede evidenciar un lento ajuste entre la oferta y la demanda de este bien duradero: más del 33 por ciento del total de las viviendas son ocupadas en los últimos 10 años, cuando la construcción en las respectivas décadas corresponde a menos de la mitad. En el Cuadro 1.21 se muestra una información más detallada sobre el proceso conjunta de la ocupación y la antigüedad asociada donde, además, se aprovecha la circunstancia de trabajar con dos EPF que permiten observar un mismo fenómeno con una década de diferencia, aunque bajo la restricción de ser viviendas ocupadas en 1980-81 y 1990-91 respectivamente. Así, en el Cuadro 1.21 la información se limita a las viviendas que, habiendo sido ocupadas antes de 1981, se mantienen ocupadas en 1990-91. Representa el stock de viviendas anterior a la década de los ochenta que la EPF permite analizar desde dos momentos distintos en el tiempo: principios de los ochenta y principios de los noventa. Se pretende responder a la pregunta ¿qué variación porcentual ( $\Delta 90$ ) supone la ocupación en 1990-91 respecto a la ocupación en 1980-81? Dada la consideración agregada de los datos<sup>17</sup>, una variación positiva indica existencia de vacantes en 1980, mientras que un signo negativo indica un abandono neto del número de viviendas habitadas en la década de los ochenta.

La EPF de 1990-91 permite contabilizar que el número de viviendas ocupadas, y por tanto construidas, antes de 1981 es de 7.481.483 (que representen un 66,46 por ciento del total muestral, véase Cuadro 1.20) lo que supone casi un 25 por ciento menos que las viviendas ocupadas en 1980 que ascienden

---

<sup>17</sup>No es posible la identificación de los hogares que, estando ocupados en 1980-81, se mantienen ocupados en 1990-91.

a 9.925.956; Según el desglose realizado, para todos los períodos se produce una reducción, más o menos considerable, a excepción del último período. Efectivamente, para el período '1973-80' se muestra como cerca de seiscientas mil (587.120) viviendas construidas y no ocupadas en 1980-81 se ocupan en esta década. Lo que puede confirmar la tendencia a no ocupar de forma inmediata la viviendas de más reciente construcción, o la dificultad de la EPF para incluir en la muestra a este tipo de viviendas.

No resulta sorprendente las relaciones entre antigüedad y ocupación que se observan en dos sentidos: i) la tasa de abandono de las viviendas aumenta con su antigüedad, y ii) comparando viviendas construidas en el mismo período de clasificación, el porcentaje de abandono es de mayor magnitud a medida que la vivienda fue ocupada con anterioridad en el tiempo. Sin duda esto es debido a las circunstancias personales de sus ocupantes, entre las que supuestamente destaca el envejecimiento.

### **1.3.3 Precios**

La información sobre precios que proporciona la EPF hace referencia tanto al precio total de la vivienda en propiedad en el momento de la adquisición<sup>18</sup>, como al alquiler mensual correspondiente al régimen de arrendamiento. Adicionalmente, para la vivienda en propiedad, acceso a la propiedad o algún tipo de cesión, se solicita al entrevistado la estimación del alquiler equivalente que debería pagar por su vivienda en el mercado libre. Se trata de un alquiler bruto imputado por parte del hogar, donde se debe considerar la vivienda sin muebles e incluir los gastos de comunidad.

---

<sup>18</sup>En el caso de la propiedad por herencia se anota el precio al que se valoró la vivienda en la herencia; en caso de no existir dicha valoración, se pide una estimación de la misma en el momento correspondiente.

En el Cuadro 1.22 se muestra la disponibilidad del precio de adquisición para la vivienda principal y la secundaria. Se cuenta con esta información en prácticamente la totalidad de viviendas para ambos períodos de estudio.

Por su parte, el alquiler mensual pagado por el hogar en régimen de arrendamiento es proporcionado en más de un 98 y 97 por ciento en 1980-81 y 1990-91 respectivamente, como se muestra en el Cuadro 1.23<sup>19</sup>. El alquiler imputado del resto de viviendas está disponible en una proporción similar se cuenta con una similar proporción de datos disponibles.

### Tercer filtro

Este filtro supone prescindir de las viviendas sin información sobre precios.

En el Cuadro 1.24 se muestra los precios medio y mediana del metro cuadrado según el régimen de tenencia y calificación legal (RTC) del parque de viviendas, principales y secundarias, una vez se aplican los tres filtros expuestos a las dos EPF. Como se puede observar, todos los precios se ven incrementados en 1990-91 respecto a 1980-81 así, por ejemplo el precio del metro cuadrado adquirido en el régimen de propiedad libre eleva un 203 por ciento (un 166 por ciento en la PO), mientras que en el arrendamiento libre el alquiler del metro cuadrado se eleva un 241 por ciento (un 172 por ciento en la PO). En contra de lo que cabría esperar, destaca la similitud entre los precios observados del metro cuadrado en la vivienda con calificación de PO y libres, e incluso, la superioridad de los primeros respecto a los segundo en los siguientes casos: i) precio medio de adquisición en 1980-81 y alquiler medio autoestimado por el hogar en ambos períodos y ii) precios mediana, tanto de alquiler como de adquisición en ambas EPF, salvo en el alquiler libre

---

<sup>19</sup>La desagregación en viviendas principales y secundarias apenas muestra diferencias para este tipo de información por lo que se muestran conjuntamente.

en 1990-91. Por otra parte se observa que en el régimen de arrendamiento los alquileres medios autoestimados por los ocupantes del hogar llegan a representar cerca del doble del valor observado en el mercado de arrendamiento. Ante estos resultados cabría preguntarse sobre el papel de la calidad de la vivienda en la diversidad de precios observados, e incluso, la repercusión de su distribución en términos de valorar la política de viviendas (sistema de protección oficial en este caso) puesto que el objetivo último de este tipo de intervenciones es facilitar el acceso a una vivienda 'digna' (calidad mínima) a determinados hogares a través de un precio inferior al que se obtendría en el mercado libre para viviendas similares. Estas cuestiones sobrepasan los objetivos descriptivos de este capítulo, sin embargo en el Capítulo II se describe la forma de sintetizar el conjunto de atributos de las viviendas de las EPF en indicadores de calidad. Posteriormente, en los Capítulos III y IV se estimará el efecto de dicha calidad sobre el valor del flujo de los servicios de cada vivienda, lo que, como veremos, será relevante para dar sentido a cualquier tipo de comparación, en términos de calidad y/o precio, entre distintos grupos de viviendas, e incluso entre viviendas de distintas EPF.

## **1.4 Conclusiones**

La fuente de información utilizada en esta tesis es la Encuesta de Presupuestos Familiares (EPF) realizada por el Instituto Nacional de Estadística (INE) en 1980-81 y 1990-91. El Capítulo I tiene un carácter descriptivo centrado en la información que proporciona esta base de datos sobre la vivienda. Su contenido se divide en dos partes que responden a dos objetivos distintos: i) medir la representatividad de las viviendas de la EPF a nivel nacional, y ii) analizar la calidad de la información disponible.

El objetivo principal de este capítulo es calibrar la capacidad de las EPF para recoger información sobre las principales características del parque de vivienda que representan. Para ello se realiza un análisis comparativo respecto al exhaustivo registro de viviendas de los dos últimos Censo de Vivienda (CV) elaborado por el INE en 1981 y 1991. El resultado más relevante de esta comparación es haber constatado la necesidad de utilizar los factores de elevación poblacional, proporcionados por el INE, cuando se trabaja con la EPF en el campo de la economía aplicada. Se ha comprobado que la representatividad de ambas EPF está condicionada a dichos factores porque corrigen una subestimación muestral en las comunidades de Madrid y Cataluña así como una sobrestimación en comunidades como Castilla y León o Castilla la Mancha.

Las variables de la vivienda susceptibles de comparación entre el CV y la EPF son: la localización de la vivienda por comunidad autónoma y tamaño de municipio, el régimen de tenencia: propiedad, arrendamiento o cesión, y la antigüedad del inmueble según el año de construcción del edificio. El estudio comparativo se limita a las viviendas principales ocupadas en 1981 y 1991 en todo el territorio nacional, se abandona el análisis de las viviendas secundarias debido a la incompatibilidad práctica que supone su definición en ambas fuentes estadísticas. Este hecho se refleja en los datos cuanto se obtiene una subestimación de este tipo de viviendas en la encuesta respecto al CV del 65 por ciento en 1981 y del 56 por ciento en 1991. Las conclusiones más destacadas de la comparación CV versus EPF son las siguientes:

Las EPF de 1980-81 y 1990-91 subestiman el número de viviendas principales ocupadas respecto al registro de este tipo de viviendas que proporciona el CV. En ambos años la subestimación es similar y no supera el 4 por ciento, pudiéndose justificar en parte porque el período de recogida de información

del CV se termina dos meses más tarde que el de la EPF.

De forma general, las variables que muestran un comportamiento equivalente en ambas fuentes son el régimen de tenencia y la localización de la vivienda por Comunidades Autónomas. La variable tamaño de municipio presenta un sesgo sistemático en ambos períodos y las mayores diferencias se deben a la antigüedad del inmueble debido a las dificultades que presenta la recogida de información en ambas fuentes.

En particular, los puntos coincidentes en cuanto al *stock* de viviendas ocupadas y el cambio producido en la década de estudio según ambas fuentes son:

- El número de viviendas principales se ha incrementado en más de un 12 por ciento entre 1981 y 1991.
- La distribución porcentual por Comunidades Autónomas presenta una gran similitud en el CV y la EPF: más del 54 por ciento de las viviendas principales ocupadas en 1981 y 1991 se localizan en las comunidades de Cataluña, Andalucía, Madrid y Valencia.
- En 1991 el 78 por ciento de las viviendas principales están en régimen de propiedad, el 15 por ciento en arrendamiento y el 7 por ciento restante corresponde a algún tipo de cesión.
- Entre los cambios observados, el mayor incremento en el *stock* de viviendas ocupadas se debe a la propiedad (más de 1,5 millones). El cambio de arrendamiento y vivienda cedida por razón de trabajo, experimenta una considerable reducción en 1991 ya que de acuerdo con ambas fuentes, en 1981 ambos regímenes de tenencia representaban el 24 y el 25 por ciento del total respectivamente. Esta reducción se produce en todos los tramos de antigüedad de

las viviendas cedidas por la empresa y, con excepción de la década de los 70, también en el arrendamiento.

- Las viviendas principales más antiguas tienden a ser sustituidas por las más nuevas. La antigüedad más frecuente de la vivienda ocupada en 1991 y 1981 es entre 10 y 20 años suponiendo una media de un 28 por ciento del total de viviendas. Con independencia del régimen de tenencia, la vivienda construida en la década de los 40 es la de menor nivel de ocupación en ambos períodos de estudio. Más del 65 por ciento de las viviendas en propiedad tienen una antigüedad menor a 30 años, en torno al 53 por ciento entre las cedidas y un 40 por ciento si es arrendada. Comparativamente con otros regímenes de tenencia, el arrendamiento presenta una mayor proporción de viviendas construidas antes de 1941.

La información más discordante entre ambas fuentes estadísticas hace referencia a las siguientes puntualizaciones:

- El Censo de Edificios permite constatar que el CV tiende a sobrestimar la antigüedad de las viviendas familiares. Por su parte, la EPF infravalora el número de viviendas principales respecto al CV cuando la antigüedad es menor a 11 años. Para las viviendas más antiguas se observa un comportamiento dispar entre la EPF de 1980-81 y 1990-91. La primera sobrevalora el dato censal en un 16,6 por ciento, mientras que la segunda lo infravalora en un 32,4 por ciento. Esto origina el dato más discordante de este análisis: según la Encuesta de 1991 se abandonan 1,7 millones de viviendas construidas antes de 1941 y ocupadas en 1981, cuando el Censo registra sólo un abandono de 0,6 millones.

- La EPF de 1980-81 contabiliza cerca de un millón de viviendas principales menos que el Censo en la zona rural. Por el contrario, sobrevalora (en más de medio millón) el número de viviendas localizadas en municipios de más de 2.000 habitantes, tanto más cuanto mayor es el tamaño del municipio. Ambos fenómenos se repiten en 1990-91 de forma aún más acusada.
- El Censo muestra un menor cambio porcentual que la Encuesta en el éxodo 'rural-urbano' respecto al total de viviendas principales. Según la primera fuente se reduce en un 2 por ciento la representación rural con un incremento en igual cuantía de la zona urbana; en la Encuesta este cambio supone cerca de un 4 por ciento. La zona intermedia (de 2.001 a 10.000 habitantes) mantiene la misma proporción en 1981 y 1991 según ambas estadísticas, un 17 por ciento en el CV y un 19 por ciento en la EPF.

En lo que respecta al segundo objetivo del capítulo: analizar la calidad de la información que proporcionan las EPF en relación a la disponibilidad de datos y su coherencia interna, las principales conclusiones son:

- La totalidad de viviendas dispone de información sobre el régimen de tenencia mientras que los hogares que no proporcionan la calificación legal de las viviendas principales en 1980-81 y 1990-91 suponen, respectivamente, un 10 y un 15 por ciento del total.
- La falta de dato en el año de construcción y/o de ocupación en ambas EPF supone algo menos de un 7 por ciento del total de viviendas (principal más secundaria) y está asociada al régimen de cesión. Por su parte, la incoherencia debida a declarar el año de construcción posterior

al año de ocupación está asociada al régimen de propiedad representando un 12 por ciento en 1990-91, cuando en 1980-81 tan solo era un 2,7 por ciento. Esto se puede deber a que en la EPF 1990-91 la variable 'año de construcción' puede hacer referencia a la fecha de la última renovación sustancial de la vivienda.

- El número de hogares en propiedad que no proporciona información sobre el precio de adquisición de su vivienda supone el 1,32 por ciento del total en 1980-81 y el 0,08 por ciento en 1990-91. En relación a los precios de alquiler, la información en relación al alquiler está disponible en el 99 por ciento de las viviendas arrendadas en 1980-81 y en el 97,59 por ciento en 1990-91, mientras que las rentas imputadas a las viviendas en propiedad son proporcionadas prácticamente por la totalidad de los propietarios o benefactores de algún tipo de cesión en ambos períodos de análisis.

. Cuadros del capítulo I

Cuadro 1.1: Stock de viviendas ocupadas según el tipo

	EPF 1980-81	%	CENSO 1981	%	Diferencia	(%)*
<b>Total</b>	<b>10.689.787</b>	<b>100</b>	<b>12.330.654</b>	<b>100</b>	<b>-1.640.867</b>	<b>-13,31</b>
Principales	10.024.166	93,77	10.430.895	84,59	-406.729	-3,90
Secundarias	665.621	6,23	1.899.759	15,41	-1.234.138	-64,96
	EPF 1990-91	%	CENSO 1991	%	Diferencia	(%)*
<b>Total</b>	<b>12.582.955</b>	<b>100</b>	<b>14.659.991</b>	<b>100</b>	<b>-2.077.036</b>	<b>-14,17</b>
Principales	11.298.509	89,79	11.736.376	80,06	-437.867	-3,73
Secundarias	1.284.446	10,21	2.923.615	19,94	-1.639.169	-56,07

\* Diferencia porcentual de la EPF respecto al Censo en cada tipo de vivienda

Cuadro 1.2: Cambio del stock de viviendas según el tipo

	CENSO 1991	CENSO 1981	Cambio	(%)*
<b>Total</b>	<b>14.659.991</b>	<b>12.330.654</b>	<b>2.329.337</b>	<b>18,89</b>
Principales	11.736.376	10.430.895	1.305.481	12,52
Secundarias	2.923.615	1.899.759	1.023.856	53,89
	EPF 1990-91	EPF 1980-81	Cambio	(%)*
<b>Total</b>	<b>12.582.955</b>	<b>10.689.787</b>	<b>1.893.168</b>	<b>17,71</b>
Principales	11.298.509	10.024.166	1.274.343	12,71
Secundarias	1.284.446	665.621	618.825	92,97

\* Cambio porcentual que supone 1991 respecto a 1981 en cada tipo de vivienda

Cuadro 1.3: Distribución porcentual de las CCAA respecto al total de viviendas principales

	1980-81			1990-91		
	CENSO	EPF*	EPF**	CENSO	EPF*	EPF**
<b>Total</b>	<b>10.430.895</b>	<b>10.024.166</b>	<b>23.971</b>	<b>11.736.376</b>	<b>11,298,509</b>	<b>21,155</b>
Cataluña	16,84	16,22	9,88	16,46	15,99	7,77
Andalucía	15,85	15,99	18,41	16,62	16,61	17,37
Madrid	12,82	12,19	5,29	12,81	12,56	3,61
Valencia	10,03	10,05	7,38	10,35	10,10	8,06
Castilla-León	7,11	7,28	13,93	6,83	7,11	14,95
Galicia	7,01	7,25	6,59	6,72	6,91	8,22
País Vasco	5,51	5,48	5,02	5,37	5,41	6,43
Castilla-Mancha	4,46	4,54	7,53	4,34	4,54	8,01
Aragón	3,43	3,50	5,43	3,31	3,40	5,22
Canarias	3,14	3,20	3,61	3,40	3,48	3,65
Asturias	3,12	3,25	2,88	2,97	3,01	2,09
Extremadura	2,76	2,84	3,88	2,71	2,91	3,92
Murcia	2,41	2,48	1,90	2,53	2,51	2,49
Baleares	1,92	2,04	1,99	1,97	1,88	2,03
Cantabria	1,32	1,36	2,20	1,32	1,34	1,71
Navarra	1,30	1,29	1,52	1,31	1,29	1,73
La Rioja	0,72	0,72	1,44	0,72	0,67	1,69
Ceuta y Melilla	0,27	0,32	1,10	0,26	0,31	1,04
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

\* Datos poblacionales. \*\* Datos muestrales.

Cuadro 1.4: Distribución porcentual del tamaño de municipio respecto al total

	1980-81			1990-91		
	CENSO	EPF*	EPF**	CENSO	EPF*	EPF**
<b>Total</b>	<b>10.430.895</b>	<b>10.024.166</b>	<b>23.971</b>	<b>11.736.376</b>	<b>11.298.509</b>	<b>21.155</b>
Rural	20,09	11,24	12,1	18,18	7,26	9,5
Intermedio	16,62	18,98	16,6	16,64	19,40	18,9
Urbano	63,29	69,79	71,3	65,18	73,34	71,5
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

\* Datos poblacionales. \*\* Datos muestrales.

Cuadro 1.5: Stock de las viviendas principales según el tamaño de municipio.

	<b>EPF 1980-81</b>	<b>CENSO 1981</b>	<b>Diferencia</b>	<b>(%)*</b>
<b>Total</b>	<b>10.024.166</b>	<b>10.430.895</b>	<b>-406.729</b>	<b>-3,90</b>
Rural	1.126.317	2.096.069	-969.752	-46,27
Intermedio	1.902.150	1.732.680	169.470	9,78
Urbano	6.995.699	6.602.146	393.553	5,96
	<b>EPF 1990-91</b>	<b>CENSO 1991</b>	<b>Diferencia</b>	<b>(%)*</b>
<b>Total</b>	<b>11.298.509</b>	<b>11.736.376</b>	<b>-437.867</b>	<b>-3,73</b>
Rural	820.818	2.133.745	-1.312.927	-61,53
Intermedio	2.191.406	1.952.583	238.823	12,23
Urbano	8.286.285	7.650.048	636.237	8,32

\* Diferencia porcentual que supone la EPF respecto al Censo según el tamaño de municipio

Cuadro 1.6: Cambio en el stock de viv. principales según el tamaño de municipio

	<b>CENSO 1991</b>	<b>CENSO 1981</b>	<b>Cambio</b>	<b>(%)*</b>
<b>Total</b>	<b>11.736.376</b>	<b>10.430.895</b>	<b>1.305.481</b>	<b>12,52</b>
Rural	2,133.745	2.096.069	37.676	1,80
Intermedio	1.952.583	1.732.680	219.903	12,69
Urbano	7.650.048	6.602.146	1.047.902	15,87
	<b>EPF 1990-91</b>	<b>EPF 1980-81</b>	<b>Cambio</b>	<b>(%)*</b>
<b>Total</b>	<b>11.298.509</b>	<b>10.024.166</b>	<b>1.274.343</b>	<b>12,71</b>
Rural	820.818	1.126.317	-305.499	-27,12
Intermedio	2.191.406	1.902.150	289.256	15,21
Urbano	8.286.285	6.995.699	1.290.586	18,45

\* Cambio porcentual que supone 1991 respecto a 1981 según el tamaño de municipio

Cuadro 1.7: Stock de viviendas principales según el régimen de tenencia

	EPF 1980-81	%	CENSO 1981	%	Diferencia	(%)*
<b>Total</b>	<b>10.024.166</b>	<b>100</b>	<b>10.430.895</b>	<b>100</b>	<b>-406.729</b>	<b>-3,90</b>
Propiedad	6.928.150	69,11	7.629.659	73,14	-701.509	-9,19
<i>Pagada</i>	<i>5.380.343</i>	<i>53,67</i>	<i>5.763.828</i>	<i>55,26</i>	<i>-383.485</i>	<i>-6,65</i>
<i>Pendiente de pago</i>	<i>1.547.806</i>	<i>15,44</i>	<i>1.865.831</i>	<i>17,89</i>	<i>-318.025</i>	<i>-17,04</i>
Arrendamiento	2.297.105	22,92	2.168.661	20,79	128.444	5,92
Cedida por empresa	219.163	2,19	321.621	3,08	-102.458	-31,86
Otras formas	579.748	5,78	310.954	2,98	268.794	86,44
	EPF 1990-91	%	CENSO 1991	%	Diferencia	(%)*
<b>Total</b>	<b>11.298.509</b>	<b>100</b>	<b>11.736.376</b>	<b>100</b>	<b>-437.867</b>	<b>-3,73</b>
Propiedad	8.789.287	77,79	9.194.491	78,34	-405.204	-4,41
<i>Pagada</i>	<i>7.504.317</i>	<i>66,42</i>	<i>7.361.235</i>	<i>62,72</i>	<i>143.082</i>	<i>1,94</i>
<i>Pendiente de pago</i>	<i>1.284.971</i>	<i>11,37</i>	<i>1.833.256</i>	<i>15,62</i>	<i>-548.285</i>	<i>-29,91</i>
Arrendamiento	1.694.184	14,99	1.780.695	15,17	-86.511	-4,86
Cedida por empresa	134.081	1,19	189.529	1,61	-55.448	-29,26
Otras formas	680.957	6,03	571.661	4,87	109.296	19,12

\* Diferencia porcentual que supone la EPF respecto al Censo según el régimen de tenencia

Cuadro 1.8: Cambio en el stock de viviendas principales según régimen de tenencia

	CENSO 1991	CENSO 1981	Cambio	(%)*
<b>Total</b>	<b>11.736.376</b>	<b>10.430.895</b>	<b>1.305.481</b>	<b>12,52</b>
Propiedad	9.194.491	7.629.659	1.564.832	20,51
<i>Pagada</i>	<i>7.361.235</i>	<i>5.763.828</i>	<i>1.597.407</i>	<i>27,71</i>
<i>Pendiente de pago</i>	<i>1.833.256</i>	<i>1.865.831</i>	<i>-32.575</i>	<i>-1,75</i>
Arrendamiento	1.780.695	2.168.661	-387.966	-17,89
Cedida por la empresa	189.529	321.621	-132.092	-41,07
Otras formas	571.661	310.954	260.707	83,84
	EPF 1990-91	EPF 1980-81	Cambio	(%)*
<b>Total</b>	<b>11.298.509</b>	<b>10.024.166</b>	<b>1.274.343</b>	<b>12,71</b>
Propiedad	8.789.287	6.928.150	1.861.137	26,86
<i>Pagada</i>	<i>7.504.317</i>	<i>5.380.343</i>	<i>2.123.974</i>	<i>39,48</i>
<i>Pendiente de pago</i>	<i>1.284.971</i>	<i>1.547.806</i>	<i>-262.835</i>	<i>-16,98</i>
Arrendamiento	1.694.184	2.297.105	-602.921	-26,25
Cedida por la empresa	134.081	219.163	-85.082	-38,82
Otras formas	680.957	579.748	101.209	17,46

\* Cambio porcentual que supone 1991 respecto a 1981 según el régimen de tenencia

Cuadro 1.9: Porcentaje\* de viviendas familiares de 1991 según año de construcción

	CENSO DE VIVIENDAS	CENSO DE EDIFICIOS	DIFERENCIA POCENTUAL
<b>Total</b>	<b>100</b>	<b>100</b>	<b>0</b>
Antes de 1921	13,55	11,03	2,52
Entre 1921 y 1940	6,07	5,12	0,95
Entre 1941 y 1950	4,56	5,07	-0,51
Entre 1951 y 1960	9,83	10,37	-0,54
Entre 1961 y 1970	21,66	22,22	-0,56
Entre 1971 y 1980	26,00	27,16	-1,16
Después de 1980	18,33	19,03	-0,70

\* Respecto a 17.200.000 viviendas.

Cuadro 1.10: Stock de viviendas principales según año de construcción del edificio

	EPF 1980-81	%	CENSO 1981	%	Diferencia	(%)*
<b>Total</b>	<b>10.024.166</b>	<b>100</b>	<b>10,430,895</b>	<b>100</b>	<b>-406.729</b>	<b>-3,90</b>
Menos de 1941	3.203.040	31,95	2.747.802	26,34	455.238	16,57
Entre 1941 y 1950	622.772	6,21	553.155	5,30	69.617	12,59
Entre 1951 y 1960	1.353.263	13,50	1.491.890	14,30	-138.627	-9,29
Entre 1961 y 1970	2.815.161	28,08	2.866.783	27,48	-51.622	-1,80
Entre 1971 y 1980	1.997.817	19,93	2.771.265	26,57	-773.448	-27,91
Sin información	32.113	0,32	-	-	32.113	-
	<b>EPF 1990-91</b>	<b>%</b>	<b>CENSO 1991</b>	<b>%</b>	<b>Diferencia</b>	<b>(%)*</b>
<b>Total</b>	<b>11.298.509</b>	<b>100</b>	<b>11.736.376</b>	<b>100</b>	<b>-437.867</b>	<b>-3,73</b>
Menos de 1941	1.439.973	12,74	2.131.051	18,16	-691.078	-32,43
Entre 1941 y 1950	502.377	4,45	538.750	4,59	-36.373	-6,75
Entre 1951 y 1960	1.357.841	12,02	1.296.297	11,05	61.544	4,75
Entre 1961 y 1970	2.925.259	25,89	2.801.198	23,87	124.061	4,43
Entre 1971 y 1980	3.319.531	29,38	3.195.626	27,23	123.905	3,88
Menos de 1991	1.594.302	14,11	1.773.454	15,11	-179.152	-10,10
Sin información	159.226	1,41	-	-	159.226	-

\* Diferencia porcentual que supone la EPF respecto al Censo según el año de construcción del edificio

Cuadro 1.11: Cambio del stock de viviendas principales según año de construcción

	<b>CENSO 1991</b>	<b>CENSO 1981</b>	<b>Cambio</b>	<b>(%)*</b>
<b>Total</b>	<b>11.736.376</b>	<b>10.430.895</b>	<b>1.305.481</b>	<b>12,52</b>
Menos de 1941	2.131.051	2.747.802	-616.751	-22,45
Entre 1941 y 1950	538.750	553.155	-14.405	-2,60
Entre 1951 y 1960	1.296.297	1.491.890	-195.593	-13,11
Entre 1961 y 1970	2.801.198	2.866.783	-65.585	-2,29
Entre 1971 y 1980	3.195.626	2.771.265	424.361	15,31
Más de 1981	1.773.454	-	1.773.454	-
	<b>EPF 1990-91</b>	<b>EPF 1980-81</b>	<b>Cambio</b>	<b>(%)*</b>
<b>Total</b>	<b>11.298.509</b>	<b>10.024.166</b>	<b>1.274.343</b>	<b>12,04</b>
Menos de 1941	1.439.973	3.203.040	-1.763.067	-55,04
Entre 1941 y 1950	502.377	622.772	-120.395	-19,33
Entre 1951 y 1960	1.357.841	1.353.263	4.578	0,34
Entre 1961 y 1970	2.925.259	2.815.161	110.098	3,91
Entre 1971 y 1980	3.319.531	1.997.817	1.321.714	66,16
Más de 1981	1.594.302	-	1.594.302	-
Sin información	159.226	32.113	127.113	395,83

\* Cambio porcentual que supone 1991 respecto a 1981 según el año de construcción del edificio

Cuadro 1.12. Stock de viviendas principales por año de construcción según régimen de tenencia.

	EPF 1980-81		CENSO 1981		DIFERENCIA	
	n° viviendas	%	n° viviendas	%	n° viviendas	%
TOTAL	10,024,166	-	10,430,895	-	-406,729	-3.39
PROPIEDAD	6,928,150	100	7,629,659	100	-701,509	-9.19
Menos de 1941	2,046,658	29.54	1,788,504	23.44	258,154	14.43
Entre 1941 y 1950	365,360	5.27	353,599	4.63	11,761	3.32
Entre 1951 y 1960	820,201	11.84	983,446	12.89	-163,245	-16.60
Entre 1961 y 1970	2,023,612	29.21	2,127,846	27.89	-104,234	-4.90
Más de 1970	1,663,007	24.00	2,376,264	31.15	-713,257	-30.02
Sin información	9,312	0.13	-	-	9,312	-
ARRENDADAS	2,297,105	100	2,168,661	100	128,444	5.92
Menos de 1941	840,257	36.58	806,889	37.21	33,368	4.14
Entre 1941 y 1950	198,501	8.64	153,125	7.06	45,376	29.63
Entre 1951 y 1960	425,279	18.51	364,586	16.81	60,693	16.65
Entre 1961 y 1970	602,289	26.22	565,877	26.09	36,412	6.43
Más de 1970	209,042	9.10	278,184	12.83	-69,142	-24.85
Sin información	21,737	0.95	-	-	21,737	-
CEDIDAS	219,163	100	321,621	100	-102,458	-31.86
Menos de 1941	77,741	35.47	81,061	25.20	-3,320	-4.10
Entre 1941 y 1950	24,203	11.04	29,762	9.25	-5,559	-18.68
Entre 1951 y 1960	33,445	15.26	79,624	24.76	-46,179	-58.00
Entre 1961 y 1970	54,371	24.81	81,142	25.23	-26,771	-32.99
Más de 1970	29,094	13.28	50,032	15.56	-20,938	-41.85
Sin información	309	0.14	-	-	309	-
OTRA FORMA	579,748	100	310,954	100	268,794	86.44
Menos de 1941	238,383	41.12	71,350	22.95	167,033	234.10
Entre 1941 y 1950	34,708	5.99	16,669	5.36	18,039	108.22
Entre 1951 y 1960	74,339	12.82	64,233	20.66	10,106	15.73
Entre 1961 y 1970	134,889	23.27	91,917	29.56	42,972	46.75
Más de 1970	96,674	16.68	66,785	21.48	29,889	44.75
Sin información	755	0.13	-	-	755	-

\*Diferencia porcentual que supone la EPF respecto al Censo en 1981

Cuadro 1.13. Stock de viviendas principales por año de construcción según régimen de tenencia.

	EPF 1990-91		CENSO 1991		DIFERENCIA	
	n° viviendas	%	n° viviendas	%	n° viviendas	(%)*
TOTAL	11,298,509		11,736,376		-437,867	-3.73
PROPIEDAD	8,789,287	100	9,194,491	100	-405,204	-4.41
Menos de 1941	934,539	10.63	1,500,442	16.32	-565,903	-37.72
Entre 1941 y 1950	310,941	3.54	372,764	4.05	-61,823	-16.59
Entre 1951 y 1960	1,032,627	11.75	936,353	10.18	96,274	10.28
Entre 1961 y 1970	2,198,981	25.02	2,172,662	23.63	26,319	1.21
Entre 1971 y 1980	2,895,017	32.94	2,705,470	29.42	189,547	7.01
Más de 1980	1,415,424	16.10	1,506,800	16.39	-91,376	-6.06
Sin información	1,758	-	-	-	1,758	-
ARRENDADAS	1,694,184	100	1,780,695	100	-86,511	-4.86
Menos de 1941	413,227	24.39	511,696	28.74	-98,469	-19.24
Entre 1941 y 1950	148,482	8.76	122,502	6.88	25,980	21.21
Entre 1951 y 1960	221,649	13.08	238,822	13.41	-17,173	-7.19
Entre 1961 y 1970	547,005	32.29	441,536	24.80	105,469	23.89
Entre 1971 y 1980	250,369	14.78	313,695	17.62	-63,326	-20.19
Más de 1980	80,371	4.74	152,444	8.56	-72,073	-47.28
Sin información	33,081	1.95	-	-	33,081	-
CEDIDAS	134,081	100	189,529	100	-55,448	-29.26
Menos de 1941	13,774	10.27	35,225	18.59	-21,451	-60.90
Entre 1941 y 1950	12,968	9.67	15,365	8.11	-2,397	-15.60
Entre 1951 y 1960	21,436	15.99	35,841	18.91	-14,405	-40.19
Entre 1961 y 1970	28,392	21.18	44,937	23.71	-16,545	-36.82
Entre 1971 y 1980	20,284	15.13	35,956	18.97	-15,672	-43.59
Más de 1980	12,987	9.69	22,205	11.72	-9,218	-41.51
Sin información	24,240	18.08	-	-	24,240	-
OTRA FORMA	680,957	100	571,661	100	109,296	19.12
Menos de 1941	78,435	11.52	83,688	14.64	-5,253	-6.28
Entre 1941 y 1950	29,986	4.40	28,119	4.92	1,867	6.64
Entre 1951 y 1960	82,129	12.06	85,281	14.92	-3,152	-3.70
Entre 1961 y 1970	150,880	22.16	142,063	24.85	8,817	6.21
Entre 1971 y 1980	153,862	22.59	140,505	24.58	13,357	9.51
Más de 1980	85,520	12.56	92,005	16.09	-6,485	-7.05
Sin información	100,145	14.71	-	-	100,145	-

\*Diferencia porcentual que supone la EPF respecto al Censo en 1991

Cuadro 1.14. Cambio en viviendas principales por año de construcción según régimen de tenencia.

	CENSO 1991	CENSO 1981	CAMBIO	
	n° viviendas	n° viviendas	n° viviendas	(%)*
TOTAL	11,736,376	10,430,895	1,305,481	12.52
PROPIEDAD	9,194,491	7,629,659	1,564,832	20.51
Menos de 1941	1,500,442	1,788,504	-288,062	-16.11
Entre 1941 y 1950	372,764	353,599	19,165	5.42
Entre 1951 y 1960	936,353	983,446	-47,093	-4.79
Entre 1961 y 1970	2,172,662	2,127,846	44,816	2.11
Entre 1971 y 1980	2,705,470	2,376,264	329,206	13.85
Más de 1980	1,506,800	-	1,506,800	
ARRENDADAS	1,780,695	2,168,661	-387,966	-17.89
Menos de 1941	511,696	806,889	-295,193	-36.58
Entre 1941 y 1950	122,502	153,125	-30,623	-20.00
Entre 1951 y 1960	238,822	364,586	-125,764	-34.50
Entre 1961 y 1970	441,536	565,877	-124,341	-21.97
Entre 1971 y 1980	313,695	278,184	35,511	12.77
Más de 1980	152,444	-	152,444	-
CEDIDAS	189,529	321,621	-132,092	-41.07
Menos de 1941	35,225	81,061	-45,836	-56.55
Entre 1941 y 1950	15,365	29,762	-14,397	-48.37
Entre 1951 y 1960	35,841	79,624	-43,783	-54.99
Entre 1961 y 1970	44,937	81,142	-36,205	-44.62
Entre 1971 y 1980	35,956	50,032	-14,076	-28.13
Más de 1980	22,205	-	22,205	
OTRA FORMA	571,661	310,954	260,707	83.84
Menos de 1941	83,688	71,350	12,338	17.29
Entre 1941 y 1950	28,119	16,669	11,450	68.69
Entre 1951 y 1960	85,281	64,233	21,048	32.77
Entre 1961 y 1970	142,063	91,917	50,146	54.56
Entre 1971 y 1980	140,505	66,785	73,720	110.38
Más de 1980	92,005	-	92,005	

\*Cambio porcentual que supone el Censo de Viviendas de 1991 respecto al de 1981

Cuadro 1.15. Cambio en viviendas principales por año de construcción según régimen de tenencia.

	EPF 1990-91	EPF 1980-81	CAMBIO	
	n° viviendas	n° viviendas	n° viviendas	(%)*
TOTAL	11,298,509	10,024,166	1,274,343	12.71
PROPIEDAD	8,789,287	6,928,150	1,861,137	26.86
Menos de 1941	934,539	2,046,658	-1,112,119	-54.34
Entre 1941 y 1950	310,941	365,360	-54,419	-14.89
Entre 1951 y 1960	1,032,627	820,201	212,426	25.90
Entre 1961 y 1970	2,198,981	2,023,612	175,369	8.67
Entre 1971 y 1980	2,895,017	1,663,007	1,232,010	74.08
Más de 1980	1,415,424	-	1,415,424	-
Sin información	1,758	9,312	-7,554	-
ARRENDADAS	1,694,184	2,297,105	-602,921	-26.25
Menos de 1941	413,227	840,257	-427,030	-50.82
Entre 1941 y 1950	148,482	198,501	-50,019	-25.20
Entre 1951 y 1960	221,649	425,279	-203,630	-47.88
Entre 1971 y 1980	250,369	209,042	41,327	19.77
Más de 1980	80,371	-	80,371	-
Sin información	33,081	21,737	11,344	52.19
CEDIDAS	134,081	219,163	-85,082	-38.82
Menos de 1941	13,774	77,741	-63,967	-82.28
Entre 1941 y 1950	12,968	24,203	-11,235	-46.42
Entre 1951 y 1960	21,436	33,445	-12,009	-35.91
Entre 1961 y 1970	28,392	54,371	-25,979	-47.78
Entre 1971 y 1980	20,284	29,094	-8,810	-30.28
Más de 1980	12,987	-	12,987	-
Sin información	24,240	309	23,931	-
OTRA FORMA	680,957	579,748	101,209	17.46
Menos de 1941	78,435	238,383	-159,948	-67.10
Entre 1941 y 1950	29,986	34,708	-4,722	-13.60
Entre 1951 y 1960	82,129	74,339	7,790	10.48
Entre 1961 y 1970	150,880	134,889	15,991	11.85
Entre 1971 y 1980	153,862	96,674	57,188	59.16
Más de 1980	85,520	-	85,520	-
Sin información	100,145	755	99,390	13164.24

\*Cambio porcentual que supone la EPF 1990-91 respecto a la EPF 1980-81

Cuadro 1.16: Porcentaje\* de viviendas según el RTC

RTC	EPF 1980-81			EPF 1990-91		
	Principal	Secundaria	Total	Principal	Secundaria	Total
Prop. con CL	57,56	5,03	62,59	60,08	8,47	68,55
Prop. sin CL	7,25	0,63	7,88	9,77	1,33	11,10
Arrend. con CL	19,40	0,32	19,72	9,21	0,18	9,39
Arrend. sin CL	2,09	0,06	2,15	4,25	0,08	4,33
Otro	7,47	0,19	7,66	6,48	0,15	6,63
Total	93,77	6,23	100	89,79	10,21	100

\* Respecto al total de viviendas: 10.689.787 en 1980-81 y 12.582.955 en 1990-91.

Cuadro 1.17: Distribución porcentual\* de viviendas según información disponible

Nº total	EPF 80-81			EPF 90-91		
	Principal	Secundaria	Total	Principal	Secundaria	Total
	9.025.241	591.549	9.616.790	9.533.608	1.107.416	10.641.024
Prop. L	45,48	75,23	47,31	51,22	83,71	54,61
Prop. PO	22,69	15,58	22,25	28,07	12,55	26,45
Arrend. L	14,41	5,63	13,87	7,68	1,55	7,04
Arrend. PO	8,57	0,16	8,05	4,48	0,53	4,07
Otro	8,85	3,40	8,52	8,55	1,66	7,83
TOTAL	100	100	100	100	100	100

\* Respecto al total de viviendas con información sobre CL.

Cuadro 1.18: Información disponible\* sobre el AC y AO

Tipo de dato	EPF 1980-81			EPF 1990-91		
	Principal	Secundaria	Total	Principal	Secundaria	Total
Correcto	84,34	5,73	90,07	72,66	8,38	81,04
Sin datos	6,69	0,22	6,91	6,75	0,17	6,92
Incoherencia	2,74	0,28	3,02	10,38	1,66	12,04
Total	93,77	6,23	100	89,79	10,21	100

\* En términos porcentuales respecto al total de viviendas: 10.689.787 en 1980-81 y 12.582.955 en 1990-91.

Cuadro 1.19: Disponibilidad de AC y AO por régimen de tenencia\*

Tipo de dato	EPF 1980-81				EPF 1990-91			
	Prop.	Arrend.	Otros	Total	Prop.	Arrend.	Otros	Total
Correcto	74,60	23,69	1,71	100	83,75	16,25	-	100
Sin datos	7,74	3,61	87,65	100	0,52	3,81	95,67	100
Incoherencia	90,88	7,01	2,11	100	97,59	2,41	-	100
Total	70,47	21,87	7,66	100	78,65	14,72	6,63	100

\* En términos porcentuales respecto a cada tipo de dato

Cuadro 1.20: Número y porcentaje de viviendas según información filtrada

Período	AC				AO			
	EPF 1980-81		EPF 1990-91		EPF 1980-81		EPF 1990-91	
	Número	%	Número	%	Número	%	Número	%
≤20	2.186.008	22,02	889.766	7,90	127.717	1,29	17.888	0,16
21-40	900.907	9,08	637.643	5,66	613.773	6,18	154.638	1,37
41-56	1.076.556	10,85	966.866	8,59	1.121.041	11,29	554.920	4,93
57-64	1.494.062	15,05	1.576.127	14,00	1.589.632	16,02	1.057.289	9,39
65-72	2.744.913	27,65	2.840.426	25,23	3.101.365	31,24	2.316.654	20,58
73-80	1.523.510	15,35	2.867.345	25,45	3.372.428	33,98	3.380.094	30,03
81-85	-	-	1.039.853	9,24	-	-	2.043.449	18,13
≥86	-	-	441.838	3,93	-	-	1.734.932	15,41
<b>TOTAL</b>	<b>9.925.956</b>	<b>100</b>	<b>11.259.864</b>	<b>100</b>	<b>9.925.956</b>	<b>100</b>	<b>11.259.864</b>	<b>100</b>

Cuadro 1.21: Distribución de las viviendas ocupadas con anterioridad a 1980-81

AO	Año	AC						Total
		-20	21-40	41-56	57-64	65-72	73-80	
≤20	80	127.717	401.612	487.343	373.206	434.410	361.720	2.186.008
	Δ90	-85,99	-83,00	-72,86	-68,17	-68,55	-48,72	-69,84
21-40	80	-	212.161	168.532	143.911	193.141	183.162	900.907
	Δ90	-	-59,29	-56,58	-55,69	-40,77	-29,79	-48,24
41-56	80	-	-	465.166	199.267	206.938	205.185	1.076.556
	Δ90	-	-	-24,87	-43,62	-34,25	-26,66	-30,48
57-64	80	-	-	-	873.248	332.808	288.006	1.494.062
	Δ90	-	-	-	-12,69	-21,12	-17,07	-15,41
65-72	80	-	-	-	-	1.934.068	810.845	2.744.913
	Δ90	-	-	-	-	-13,81	-30,19	-18,65
73-80	80	-	-	-	-	-	1.523.510	1.523.510
	Δ90	-	-	-	-	-	38,54	38,54
Total	80	127.717	613.773	1.121.041	1.589.632	3.101.365	3.372.428	9.925.956
	Δ90	-85,99	-74,81	-50,50	-33,49	-25,30	0,23	-24,63

Δ90: variación porcentual en el número de viviendas registradas en la EPF 1990-91 respecto a la EPF 1980-81.

Cuadro 1.22: Porcentaje\* de viviendas según la información sobre el precio de compra

Tipo de dato	EPF 1980-81			EPF 1990-91		
	Principal	Secundaria	Total	Principal	Secundaria	Total
Disponible	90,85	7,83	98,68	87,69	12,23	99,92
No disponible	1,12	0,20	1,32	0	0,08	0,08
Total	91,97	8,03	100	87,69	12,31	100

\* Respecto al total de viviendas en propiedad: 7.533.066 en 1980-81 y 10.022.816 en 1990-91.

Cuadro 1.23: Porcentaje de viviendas según información sobre precio de alquiler

Tipo de dato	EPF 1980-81				EPF 1990-91			
	Imputado*	%	Real**	%	Imputado*	%	Real**	%
Disponible	8.343.595	99,90	2.313.789	98,98	10.853.092	99,97	1.685.127	97,59
No disponib.	8.505	0,10	23.898	1,02	3.177	0,03	41.559	2,41
Total	8.352.100	100	2.337.687	100	10.856.269	100	1.726.686	100

\*Para las viviendas en propiedad y cedidas. \*\*Para el total de viviendas en arrendamiento.

Cuadro 1.24: Precios medios del metro cuadrado según régimen de tenencia

	EPF 1980-81					
	Renta real		Renta imputada		Precio de adquisición	
	AL	APO	PL	PPO	PL	PPO
Nº viviendas*	1.286.389	763.597	4.497.422	2.132.658	4.466.416	2.126.525
Importe medio	54	50	88	107	6.187	6.976
Importe mediana	31	40	63	97	2.500	4.357
	EPF 1990-91					
	Renta real		Renta imputada		Precio de adquisición	
	AL	APO	PL	PPO	PL	PPO
Nº viviendas*	711.779	411.465	5.474.749	2.786.792	5.469.783	2.787.899
Importe medio	184	136	311	361	18.768	18.584
Importe mediana	113	91	221	300	8.125	11.111

\* Viviendas principales y secundarias tras los filtros 1, 2 y 3.

## Capítulo 2

# Construcción de indicadores del nivel de acondicionamiento de la vivienda

## 2.1 Introducción

La Encuesta de Presupuestos Familiares (EPF) realizadas por el Instituto Nacional de Estadística (INE) en 1980-81 y 1990-91 respectivamente, proporcionan datos detallados sobre el acondicionamiento físico de las viviendas españolas ocupadas con fines residenciales. El objetivo de este capítulo es utilizar esta información multivariante para obtener una visión global de todas las variables que caracterizan la vivienda ocupada en España en cada período, así como disponer de algún criterio de comparación entre ellas.

Para ello se construyen medidas que sintetizan el conjunto de características de las viviendas disponibles en las EPF bajo el criterio optimizador de recoger la máxima variabilidad observada. La técnica multivariante que se utiliza es el Análisis de Correspondencias Múltiples<sup>1</sup> (ACM) que permite la interpretación de un elevado número de variables, posiblemente correlacionadas, a través de un reducido número de indicadores independientes entre sí. Se trata de un instrumento estadístico de carácter descriptivo que no incorpora ningún criterio subjetivo que pudiera condicionar los resultados obtenidos.

En este capítulo se aplica el ACM para resumir la información sobre el acondicionamiento de tres conjuntos de viviendas: las recogidas en la EPF de 1980-81, en la EPF de 1990-91 y en ambas encuestas consideradas conjuntamente. Los indicadores sintéticos que se obtienen son, por construcción, una combinación lineal de las características físicas que componen cada vivienda. Cada indicador lleva asociado una medida de la pérdida de información que traería consigo limitar el estudio a dicho indicador, de tal forma que la consideración conjunta de todos ellos anularía dicha pérdida.

La aplicación del ACM a la vivienda de la EPF revela que un único

---

<sup>1</sup>Veáse Greenacre (1984) y Tenenhaus y Young, (1985).

indicador llega a explicar un elevado porcentaje de la variabilidad observada en los respectivos tipos de viviendas: un 86 por ciento de las recogidas en la EPF de 1980-81, un 73 por ciento en la EPF de 1990-91 y cerca de un 88 por ciento en el análisis de las dos encuestas conjuntas. Además, esta nueva variable sintética se identifica con un indicador del nivel de calidad de cada vivienda en relación a la calidad media correspondiente a cada período.

Posteriormente, el contenido del capítulo se centra en la explotación descriptiva de la información contenida en estos índices de calidad con la pretensión de:

i) Caracterizar, en relación a las variables físicas asociadas, tanto la vivienda media de los distintos período como la vivienda tipo que corresponde a cada partición cuartílica de la distribución de la variable Índice de Calidad.

ii) Analizar a las viviendas ocupadas en 1980-81 y 1990-91 bajo sus respectivos concepto de calidad para responder a preguntas como: ¿las características de la vivienda tienen el mismo peso relativo sobre la calidad en ambos períodos?, y ¿cuál es la ordenación o posición relativa de cada viviendas respecto a la vivienda de calidad media en el período correspondiente?

iii) Analizar, en términos de un mismo concepto de calidad, el grado de heterogeneidad de la distribución de la calidad de la vivienda residencial en nuestro país, así como el cambio producido en 1990-91 respecto a 1980-81. En otras palabras, se pretende responder a preguntas como: ¿existe evidencia de que la vivienda en 1990-91 es de mejor calidad que la ocupada diez años antes?; ¿es homogéneo y/o sistemático el cambio experimentado en las distintas Comunidades Autónomas?, ¿se puede decir lo mismo en relación a los distintos tamaños de municipio donde se ubica la vivienda?, ¿y en relación al tipo de régimen de tenencia (propiedad, arrendamiento y cesión) o calificación legal (mercado libre frente a algún sistema de Protección Oficial)?

El resto del capítulo se divide en cinco apartados. En el siguiente apartado se expone la técnica multivariante del Análisis de Correspondencias, se comienza con el Análisis Simple para presentar posteriormente el Análisis Múltiple como su generalización cuando se trabaja con más de dos variables. Con objeto de complementar del contenido analítico de esta parte, en el Apéndice 2.1 se muestran algunos ejemplos que permiten ilustrar la interpretación de los resultados obtenidos en este tipo de análisis. En el tercer apartado se describe la aplicación del Análisis de Correspondencias Múltiples a los datos referentes a la vivienda principal ocupada en España en 1980-81 y 1990-91. En el cuarto apartado se presentan los resultados obtenidos de dicha aplicación, a partir de los cuales se define una variable indicador de la calidad de las viviendas. En el apartado quinto se utiliza esta variable para comparar las viviendas con distinto año de ocupación -1980-81 y 1990-91- lo que permite analizar el cambio producido en la calidad de las viviendas durante esta década. Dicha comparación también se presenta desagregada en las particiones siguientes: según la Comunidad Autónoma, el tamaño de municipio y el régimen de tenencia y la calificación legal de las viviendas. El sexto y último apartado se reserva para comentar los resultados más relevantes.

## 2.2 Aspectos metodológicos del Análisis de Correspondencias

El Análisis de Correspondencias (AC) es una técnica multivariante aplicable a variables de tipo cualitativo, cuyos valores discretos se denominan *clases* o *modalidades*. Su objetivo es descriptivo y consiste en analizar las relaciones entre variables en términos de las asociaciones o similitudes de sus modali-

dades. Cada modalidad se puede representar como un punto en un espacio multidimensional donde es posible interpretar la estructura del conjunto de ellos. Las coordenadas de los puntos hacen referencia a las frecuencias relativas de las modalidades y la asociación entre dos modalidades se refleja a través de su distancia chi-cuadrado. Cuando la dimensión de este espacio es mayor que dos, su interpretación se dificulta considerablemente y se hace necesaria una técnica que sintetice las relaciones existentes entre las variables. El AC permite una aproximación óptima del espacio original bajo el criterio de recoger la máxima variabilidad de la nube de puntos, previa fijación de la dimensión del subespacio sobre el que se desea realizar la aproximación. El análisis proporciona una medida de la pérdida de información que esta simplificación trae consigo, de forma que si la dimensión del subespacio coincide con la del espacio original, no se produce pérdida alguna; simplemente se habría aplicado un cambio de base en el espacio de las modalidades representadas originalmente.

### 2.2.1 Análisis de Correspondencias Simples

Cuando el AC se utiliza para analizar únicamente las relaciones entre dos variables cualitativas se habla del Análisis de Correspondencia Simple (ACS). Se aplica a una tabla de frecuencias cuya dimensión depende del número de modalidades de ambas variables, llamadas variable fila y variable columna respectivamente.

El objetivo del ACS es interpretar la posible asociación entre ambas variables comparando las distribuciones multinomiales de las distintas modalidades fila respecto a las modalidades columna, o viceversa. La comparación entre dos modalidades de una misma variable se realiza en términos de similitud. Precisa del cómputo de perfiles, esto es, frecuencias relativas al total

de cada modalidad, que se pueden expresar en términos porcentuales o en proporciones. Dos modalidades fila (columna) serán similares si presentan el mismo perfil y, en la medida en que se observen diferencias significativas se habla de asociación entre modalidades fila y columna, y por tanto, asociación entre variables.

Para facilitar el seguimiento del ACS, a continuación se introducen una serie de nociones que permiten considerar a los vectores perfiles como puntos en un espacio multidimensional con una interpretación geométrica. En el Apéndice I se muestran una serie de ejemplos sencillos que ayudan a su comprensión.

### Notación

Normalmente, para la aplicación del ACS se parte de unos datos dispuestos en una *Tabla de Contingencia* que contiene las frecuencias conjuntas ( $n_{ij}$ ) de cada par de modalidades de dos variables, donde 'i' hace referencia a la  $i$ -ésima modalidad de la variable fila y 'j' a la  $j$ -ésima modalidad de la variable columna, con  $i = 1, 2, \dots, I$  y  $j = 1, 2, \dots, J$ , de forma que  $N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$  es el total de observaciones.

En términos relativos, se define la *Matriz de Frecuencias*  $F$  cuyo elemento genérico es  $f_{ij} = n_{ij}/N$ , de forma que las *Frecuencias Marginales* fila y columna son, respectivamente  $f_{i.} = \sum_{j=1}^J f_{ij}$  y  $f_{.j} = \sum_{i=1}^I f_{ij}$ .

La matriz  $F$  se transforma en la *Matriz de Perfiles Fila*  $P_I$  (o *Perfiles Columna*  $P_J$ ) cuando dichas frecuencias se expresan en relación al total de los elementos de cada fila  $f_{i.}$  (o al total de cada columna  $f_{.j}$ ). Así, el perfil de la modalidad  $i$ -ésima de la variable fila es el vector  $p_i \equiv [f_{i1}/f_{i.} \quad f_{i2}/f_{i.} \quad \dots \quad f_{iJ}/f_{i.}]$ , mientras que  $p_j \equiv [f_{1j}/f_{.j} \quad f_{2j}/f_{.j} \quad \dots \quad f_{Ij}/f_{.j}]^T$  es el perfil de la columna  $j$ .

Cada perfil fila (columna) tiene asociado un *Peso*  $f_i(f_j)$ , que no es más que su frecuencia marginal. Los respectivos *Perfiles Medios* se hallan a través de la suma de perfiles ponderada por los pesos y se denota por  $c$  para las filas y por  $r$  para las columnas, siendo  $c = \sum_{i=1}^I f_i p_i \equiv [f_{.1} f_{.2} \dots f_{.J}]$  y  $r = \sum_{j=1}^J f_j p_j \equiv [f_{1.} f_{2.} \dots f_{I.}]^T$ . Las matrices diagonales  $D_r$  y  $D_c$  contienen, respectivamente, los pesos de las filas (perfil columna medio) y las columnas (perfil fila medio). Las relaciones entre las matrices definidas son:

	Fila	Columna
Peso	$D_r \equiv \text{diag}(r)$	$D_c \equiv \text{diag}(c)$
Matriz de perfiles	$P_I = D_r^{-1} F$	$P_J = D_c^{-1} F^T$
Vector perfil medio	$c = F^T 1_J$	$r = F 1_I$

siendo  $1$  el vector  $1 \equiv [1 \dots 1]^T$ . En el Apéndice I se muestra un ejemplo donde se calculan estas matrices a partir de una tabla de contingencias.

### Interpretación gráfica<sup>2</sup>

Sea  $R^J$  el espacio de modalidades columna determinado por los vectores de una base canónica  $\{e_1, e_2, \dots, e_j, \dots, e_J\}$ <sup>3</sup>. Cada  $j$ -ésima modalidad columna representa un *punto vértice* de coordenadas  $(0_1, 0_2, \dots, 1_j, \dots, 0_J)$ . Los puntos vértice sirven como referencia para conocer las coordenadas de los  $I$  puntos perfiles en el espacio  $R^J$  ya que  $p_i = f_{i1}/f_i e_1 + f_{i2}/f_i e_2 + \dots + f_{iJ}/f_i e_J$ , para todo  $i = 1, 2, \dots, I$ . A este conjunto de puntos se le denomina *nube de puntos perfiles fila*  $N(I)$  y se localiza en un subespacio de dimensión  $(J - 1)$ , ya que la suma de coordenadas de  $p_i$  es la unidad  $\left(\sum_{j=1}^J f_{ij}/f_i = 1\right)$ .

Además de las coordenadas de los perfiles, la caracterización de los puntos  $N(I)$  en un espacio requiere de una ponderación de los mismos y una noción

<sup>2</sup>En el Ejemplo 1 del Apéndice I se ilustra el contenido de este apartado.

<sup>3</sup>Estos vectores definen las direcciones de los  $J$  ejes ortonormales que determinan el espacio de las modalidades columna.

de distancia que permita su interpretación en términos de similitudes de perfiles. Así, la nube de puntos  $N(I)$  en  $R^J$  se caracteriza por:

a) Las *coordenadas* del conjunto de puntos. La coordenada del punto  $i$  sobre el eje  $e_j$  es igual a  $(f_{ij}/f_{i.})$  mientras que la del perfil fila medio ( $c$ ) es  $(f_{.j})$ .

b) El *peso* de cada modalidad  $i$  :  $(f_{i.})$ .

c) La *métrica* inducida por la distancia chi-cuadrado, que asigna a cada eje  $e_j$  un peso  $(1/f_{.j})$ .

Equivalentemente, en el espacio  $R^I$  donde la dimensión está determinada por el número de modalidades fila ( $I$  puntos vértice), los perfiles de  $N(J)$  y el perfil columna medio ( $r$ ) tienen respectivamente las coordenadas  $(f_{ij}/f_{.j})$  y  $(f_{i.})$  respecto al eje  $e_i$ ; el peso de cada perfil columna es  $(f_{.j})$  y la métrica está dada por  $(1/f_{i.})$ .

La caracterización de los respectivos perfiles viene dada, en forma matricial, por:

	$N(I)$ en $R^J$	$N(J)$ en $R^I$
Coordenadas	$P_I$	$P_J$
Métrica	$D_c^{-1}$	$D_r^{-1}$
Pesos	$D_r$	$D_c$

### Distancia Chi-cuadrado entre perfiles

En este contexto, donde se dispone de  $I$  puntos que representan los perfiles fila en el espacio  $R^J$  de las columnas, o  $J$  puntos en  $R^I$ , la similitud entre dos perfiles se mide a través de su *distancia chi-cuadrado* ( $d$ ), siendo ésta una distancia euclídea ponderada<sup>4</sup> con la métrica definida por  $D_c^{-1}$  para los per-

<sup>4</sup>En un espacio Euclídeo Ponderado de dimensión  $E$ , el producto escalar entre dos puntos  $x$  e  $y$  se define como:

$$x^T D_q y = \sum_{e=1}^E q_e x_e y_e$$

files fila, y por  $D_r^{-1}$  para los perfiles columna. En la práctica, la ponderación de dimensiones mencionada produce el efecto de que las modalidades menos frecuentes tengan una contribución mucho mayor a la distancia interperfiles que las de frecuencias altas. Desde un punto de vista gráfico, el resultado de la ponderación es incrementar el valor de las coordenadas correspondientes a las categorías relativamente menos frecuentes frente al resto. De esta forma se refleja su menor similitud con el perfil medio, localizado de forma más próxima a las modalidades más frecuentes.

La expresión de la distancia chi-cuadrado entre dos perfiles en función de sus coordenadas es,  $d^2(i_1, i_2) = \sum_{j=1}^J \frac{1}{f_{.j}} (f_{i_1j}/f_{i_1.} - f_{i_2j}/f_{i_2.})^2$  para las filas en  $R^J$ , y

$$d^2(j_1, j_2) = \sum_{i=1}^I \frac{1}{f_{i.}} (f_{ij_1}/f_{.j_1} - f_{ij_2}/f_{.j_2})^2 \text{ para las columnas en } R^I.$$

De igual forma, las respectivas distancias de un punto perfil fila y perfil columna a su correspondiente perfil medio son,  $d^2(i, c) = \sum_{j=1}^J \frac{1}{f_{.j}} (f_{ij}/f_{i.} - f_{.j})^2$  y  $d^2(j, r) = \sum_{i=1}^I \frac{1}{f_{i.}} (f_{ij}/f_{.j} - f_{i.})^2$ , lo que en notación matricial equivale a  $(P_I - 1c^T)D_c^{-1}(P_I - 1c^T)^T$  y  $(P_J - 1r^T)D_r^{-1}(P_J - 1r^T)^T$ , respectivamente.

Las razones que justifican utilizar la distancia chi-cuadrado en lugar de la euclídea son las siguientes:

1.- La métrica considerada corrige dispersiones inherentes en las varianzas de los conjuntos de frecuencias expresadas como perfiles. Equivale a una estandarización que equipara el papel de las modalidades a la hora de medir

donde  $q_1, q_2, \dots, q_E$  son números reales positivos que representan los pesos relativos asignados a las respectivas dimensiones del espacio, por lo que la métrica queda definida por la matriz diagonal  $D_q$ . En la métrica  $D_q$ , el cuadrado de la *distancia chi-cuadrado* ( $d$ ) entre los puntos  $x$  e  $y$  es:

$$d^2(x, y) \equiv (x - y)^T D_q (x - y) = \sum_{e=1}^E q_e (x_e - y_e)^2$$

distancias entre perfiles (y por tanto asociaciones).

2.- La distancia chi-cuadrado posee la propiedad denominada 'equivalencia distribucional' (Benzécri et al., 1973), según la cual modalidades de una variable con perfiles similares y, por tanto, próximos en el espacio de perfiles, pueden ser agregados sin afectar a las distancias entre los demás pares de puntos en el espacio multidimensional. Es decir, no hay pérdida (o ganancia) de información cuando se agrupan (o dividen) categorías homogéneas.

### Estadístico Chi-cuadrado

El *estadístico chi-cuadrado* ( $\chi^2$ ) se computa para medir la significatividad estadística de la asociación entre variables bajo la hipótesis de independencia. Su expresión es:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{N(f_{ij} - f_{i.f.j})^2}{f_{i.f.j}}$$

Para juzgar la representatividad de su valor se utilizan la tablas de la distribución  $\chi^2$  de Pearson con  $(I - 1)(J - 1)$  grados de libertad, siendo  $I$  y  $J$  el número de modalidades de cada variable.

Su relación con la distancia chi-cuadrado entre un perfil  $p_k$  (bien sea fila o columna) y su perfil medio  $\bar{p}$  es:

$$\chi^2 = \sum_k^K \chi_k^2 = \sum_k^K n_k d^2(p_k, \bar{p}) = \sum_k^K n_k (p_k - \bar{p}) D_{\bar{p}}^{-1} (p_k - \bar{p})$$

donde  $n_k$  denota la frecuencia observada de la  $k$ -ésima modalidad perfil de una variable respecto a otra, y  $N = \sum_k^K n_k$ . El estadístico  $\chi_k^2$  permite contrastar la significatividad individual de que el perfil  $k$  sea diferente del perfil esperado, o medio, mientras que  $\chi^2$  considera el conjunto de ellos y, por tanto, la significatividad de la asociación entre las dos variables.

### Inercia de la nube de perfiles<sup>5</sup>

Un concepto muy utilizado en AC es la *Inercia Total* de la nube de perfiles que equivale a la noción estadística de varianza. Se define como la media ponderada del cuadrado de las distancias chi-cuadrado entre los perfiles de las modalidades de una variable y su perfil medio. La inercia de la nube de perfiles fila y columna coincide y sus respectivas expresiones son:

$$\begin{aligned} \text{Inercia}(N(I)) &= \sum_{i=1}^I f_i d^2(i, c) = \sum_{i=1}^I f_i \sum_{j=1}^J \frac{(f_{ij}/f_i - f_j)^2}{f_j} = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j} \\ \text{Inercia}(N(J)) &= \sum_{j=1}^J f_j d^2(j, r) = \sum_{j=1}^J f_j \sum_{i=1}^I \frac{(f_{ij}/f_j - f_i)^2}{f_i} = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j} \end{aligned}$$

En términos matriciales, la inercia total (*IT*, a partir de ahora) se puede expresar en función de la *Matriz de Errores Estandarizados* definida como

$$E = D_r^{-\frac{1}{2}}(F - rc^T)D_c^{-\frac{1}{2}}, \quad (2.1)$$

de forma que

$$IT = tr [E^T E] = tr [EE^T] \quad (2.2)$$

La *IT* tiene una interpretación geométrica ya que, mide el grado de dispersión de los puntos perfiles respecto a sus medias en un espacio apropiadamente ponderado. La relación de la inercia total con el estadístico chi-cuadrado es  $\chi^2 = N * IT$ , lo que pone de manifiesto que la *IT*, al igual que  $\chi^2$ , muestra la asociación entre la variable fila y columna. Cuando todos los perfiles fila (columna) son iguales, se concentran en el perfil medio y la inercia toma valor cero mostrando una independencia total entre ambas variables. La máxima asociación se obtiene cuando todos los perfiles de una variable se localizan exactamente en los vértices del espacio definido por los vectores de las modalidades de la otra variable.

<sup>5</sup>Véase el apartado B del Ejemplo 1 en el Apéndice I.

### Análisis respecto al perfil medio<sup>6</sup>

El análisis anterior es equivalente al que se puede realizar para los perfiles centrados  $N(\tilde{I})$  o  $N(\tilde{J})$ , es decir, considerando cada perfil con referencia al perfil medio. Este cambio en el sistema de referencia supone que el origen de coordenadas se traslada al punto perfil medio, lo que implica una translación de las coordenadas de los perfiles cuyo valor refleja ahora directamente la divergencia de cada perfil con el perfil medio, sin que por ello se vea afectada ni la posición relativa de los puntos ni sus pesos. Respecto al nuevo sistema, el perfil medio tiene coordenada cero y, si denotamos por  $\tilde{i}$  y  $\tilde{j}$  los perfiles fila y columna centrados, sus nuevas coordenadas (en un eje genérico  $j$  e  $i$ , de  $R^J$  y  $R^I$ ) son:  $(f_{ij}/f_i - f_{\cdot j})$  y  $(f_{ij}/f_j - f_i)$  respectivamente. Así, la caracterización de la nube centrada es

	$N(\tilde{I})$ en $R^J$	$N(\tilde{J})$ en $R^I$
Coordenadas	$P_I - 1c^T$	$P_J - 1r^T$
Métrica	$D_c^{-1}$	$D_r^{-1}$
Pesos	$D_r$	$D_c$

La inercia total se puede expresar en términos de los perfiles centrados  $N(\tilde{I})$  ya que la matriz  $E$  de la ecuación (2.1) es igual a,

$$E = D_r^{\frac{1}{2}}(P_I - 1c^T)D_c^{-\frac{1}{2}} \quad (2.3)$$

de forma que el elemento  $e_i e_i^T$  de la diagonal principal de la matriz  $EE^T$  es la distancia chi-cuadrado entre el perfil fila  $i$  y el perfil medio ( $c$ ), ponderado por su frecuencia relativa ( $r_i$ ) y su suma es la inercia total, que como indica la ecuación (2.2), en este caso será  $Inercia(N(\tilde{I})) = tr [D_r(P_I - 1c^T)D_c^{-1}(P_I - 1c^T)^T]$ . De forma análoga se puede interpretar el elemento  $e_j^T e_j$  de  $E^T E$  para las columnas, cuando se expresa en función de los perfiles centrados  $N(\tilde{J})$

$$E^T = D_c^{\frac{1}{2}}(P_J - 1r^T)D_r^{-\frac{1}{2}} \quad (2.4)$$

<sup>6</sup>Véase el apartado C del Ejemplo 1 en el Apéndice I.

y la  $Inercia(N(\tilde{J})) = tr [D_c(P_J - 1r^T)D_r^{-1}(P_J - 1r^T)^T]$ .

Resulta inmediato demostrar que:

$$Inercia(N(\tilde{I})) = Inercia(N(I)) = Inercia(N(\tilde{J})) = Inercia(N(J))$$

### Búsqueda de subespacios óptimos<sup>7</sup>

Hasta aquí se ha definido el conjunto de puntos perfiles, centrados o no, en un espacio multidimensional donde la asociación entre perfiles se ha calculado a través de distancias chi-cuadrado, ponderando las dimensiones y teniendo en cuenta la frecuencia de cada punto a través de la asignación de un peso. Sin embargo, cuando el número de modalidades es grande, la interpretación gráfica se dificulta considerablemente.

El objetivo del ACS es identificar los subespacios de menor dimensionalidad donde mejor se reflejen las asociaciones entre los perfiles de la nube original. El criterio optimizador pretende conservar en la medida de lo posible las distancias entre perfiles existentes en el espacio multidimensional de partida. La aproximación de la nube centrada original a un subespacio óptimo  $S$ -dimensional, precisa de la descomposición en valores y vectores singulares de la Matriz de Errores Estandarizados ( $E$ ) definida en la ecuación (2.1). Es decir, se trata de realizar la transformación

$$E = UD_\alpha V^T = \sum_{k=1}^K \alpha_k u_k v_k \quad (2.5)$$

$$\text{con } U^T U = V^T V = Id$$

donde los *valores singulares* son los  $K$  elementos positivos de la matriz diagonal  $D_\alpha$ , el valor  $K$  indica el rango de la matriz  $E$ , por lo que  $K \leq \min \{I - 1, J - 1\}$ , y los *vectores singulares* por la izquierda y la derecha de  $E$  corresponden a las  $K$  columnas de las matrices  $U$  y  $V$ , respectivamente.

<sup>7</sup>En el Ejemplo 2 del Apéndice I se ilustra el contenido de este apartado.

El objeto de esta descomposición es identificar un espacio base de referencia  $R^K$ , cuya dimensión es la mínima necesaria para contener la estructura de los puntos perfiles de la matriz  $E$  en  $R^I$  y  $R^J$  respectivamente. Así, i) los vectores  $\{u_1, u_2, \dots, u_K\}$  definen una base canónica para las columnas de  $E$  en el espacio  $R^I$ , mientras que  $\{v_1, v_2, \dots, v_K\}$  forman una base canónica para las filas de  $E$  en  $R^J$ ; ii) los valores singulares  $\alpha_k$ , con  $k = 1, 2, \dots, K$ , muestran la magnitud de la proyección de  $E$  en cada una de las direcciones de los vectores base de  $R^K$ .

A través de las ecuaciones (2.3) y (2.4) y la descomposición de la matriz  $E$  en la ecuación (2.5), se puede deducir que la proyección ortogonal de los puntos de la nube  $N(\tilde{I})$  o  $N(\tilde{J})$  sobre los vectores base, determina las coordenadas de los perfiles centrados en  $R^K$ . Dichas coordenadas permiten expresar los perfiles centrados como combinaciones lineales, ya que:

a) La matriz de perfiles fila centrados es  $(P_I - 1c^T) = D_r^{-\frac{1}{2}} U D_\alpha V^T D_c^{\frac{1}{2}}$ .

Si denotamos:

$$\begin{aligned} N &= D_r^{-\frac{1}{2}} U \\ N_\alpha &= N D_\alpha \\ V_c &= D_c^{\frac{1}{2}} V, \end{aligned}$$

entonces

$$(P_I - 1c^T) = N D_\alpha V_c^T = N_\alpha V_c^T.$$

Por lo que  $(P_I - 1c^T) = \sum_{k=1}^K (\alpha_k n_k) v_{c_k}^T = \sum_{k=1}^K (n_{\alpha_k}) v_{c_k}^T$ , donde los vectores  $\{v_{c_1}, v_{c_2}, \dots, v_{c_K}\}$  forman una base ortonormal<sup>8</sup> para los puntos perfiles centrados  $N(\tilde{I})$  en el espacio  $R^J$ .

<sup>8</sup>El sentido de la ortonormalidad de  $V_c$  es con respecto a la métrica del espacio de  $N(\tilde{I})$  ya que  $V_c^T D_c^{-1} V_c = Id$ .

La proyección de la nube  $N(\tilde{I})$  sobre los vectores base del espacio  $R^K$  proporciona los *Factores de Proyección para las Filas* contenidos en las filas de la matriz:

$$(P_I - 1c^T)D_c^{-\frac{1}{2}}V = D_r^{-\frac{1}{2}}UD_\alpha$$

que permiten hallar las coordenadas de cada fila en  $R^K$ . Dichas coordenadas se pueden expresar como *Coordenadas Principales* ( $N_\alpha$ ), o como *Coordenadas Estandarizadas* ( $N$ ), siendo la diferencia entre ellas un factor de escala determinado por  $D_\alpha$ . Los valores singulares ( $\alpha_k$ ) muestran la magnitud de la proyección de  $N(\tilde{I})$  en la dirección del  $k$ -ésimo elemento de la base, de forma que  $IT = tr(D_\alpha^2) = \sum_{k=1}^K \alpha_k^2$ . La matriz  $N$  contiene las coordenadas estandarizadas de los perfiles fila al corresponder a puntos en desviaciones al perfil medio y estar normalizadas por la inercia proyectada sobre cada factor de proyección. Su relación con las coordenadas principales es:

$$N_\alpha = D_r^{-\frac{1}{2}}UD_\alpha = ND_\alpha . \quad (2.6)$$

b) Equivalentemente, la matriz de perfiles columnas centrados es

$$(P_J - 1r^T) = D_c^{-\frac{1}{2}}VD_\alpha U^T D_r^{\frac{1}{2}} .$$

Si denotamos:

$$\begin{aligned} M &= D_c^{-\frac{1}{2}}V \\ M_\alpha &= MD_\alpha \\ U_r &= D_r^{\frac{1}{2}}U , \end{aligned}$$

entonces

$$(P_J - 1r^T) = MD_\alpha U_r^T = M_\alpha U_r^T ,$$

por lo que  $(P_J - 1r^T) = \sum_{k=1}^K (\alpha_k m_k) u_{r_k}^T = \sum_{k=1}^K (m_{\alpha_k}) u_{r_k}^T$ , donde los vectores

$\{u_{r_1}, u_{r_2}, \dots, u_{r_K}\}$  forman una base ortonormal<sup>9</sup> para los puntos  $N(\tilde{J})$  en  $R^I$ .

La matriz de *Factores de Proyección para las Columnas* es:

$$(P_J - 1r^T)D_r^{-\frac{1}{2}}U = D_c^{-\frac{1}{2}}VD_\alpha \quad (2.7)$$

y las coordenadas principales ( $M_\alpha$ ) y estandarizadas ( $M$ ) de las filas en  $R^K$  se relacionan a través de:

$$M_\alpha = D_c^{-\frac{1}{2}}VD_\alpha = MD_\alpha \quad (2.8)$$

### Ecuaciones de transición

Una consecuencia inmediata de las ecuaciones (2.5), por una parte, y de las ecuaciones (2.6) y (2.8), por otra, es que los dos conjuntos de coordenadas (perfiles filas y columnas) están relacionadas. En el Apéndice II se deduce dicha relación que se resume en las siguientes *Ecuaciones de Transición*:

$$N = P_I M D_\alpha^{-1} \quad \text{ó, } N_\alpha = P_I M \quad (2.9)$$

$$M = P_J N D_\alpha^{-1} \quad \text{ó, } M_\alpha = P_J N \quad (2.10)$$

Así, por ejemplo, se pueden hallar las coordenadas estandarizadas  $n_k(i)$  de una modalidad fila  $i$  sobre el vector base  $k$  de  $R^K$ . Se calculan como las medias ponderadas de las coordenadas estandarizadas ( $m_1, m_2, \dots, m_K$ ) de las modalidades columna correspondientes a dicha fila, seguido de un cambio de escala que depende de la inercia proyectada sobre el eje ( $\lambda_k = \alpha_k^2$ ),

$$n_k(i) = \frac{1}{\alpha_k} \sum_{j=1}^J \frac{f_{ij}}{f_i} m_k(j) .$$

---

<sup>9</sup>El sentido de la ortonormalidad de  $U_r$  es con respecto a la métrica del espacio de  $N(\tilde{J})$  ya que  $U_r^T D_r^{-1} U_r = Id$ .

Equivalentemente, para las modalidades columna

$$m_k(j) = \frac{1}{\alpha_k} \sum_{i=1}^I \frac{f_{ij}}{f_{.j}} n_k(i) .$$

Cuando se dispone de una gran desagregación de modalidades (o un gran número de variables en el Análisis de Correspondencias Múltiples) las Ecuaciones de Transición resultan de gran utilidad, ya que posibilitan la localización de modalidades fila y columna en  $R^K$  a través de la aplicación del AC a un solo conjunto de perfiles (fila o columna). No obstante, se debe advertir del error que supondría interpretar, en términos de semejanza de perfiles, las distancias entre puntos que pertenecen a distintas variables, o en términos de asociación modalidades de la misma variable.

Efectivamente, cuando  $I \gg J$ , el análisis se suele limitar al conjunto de perfiles  $N(\tilde{J})$  para hallar los vectores base y factores de proyección de  $R^K$ . Una vez definido cada factor, de la forma que después comentaremos, tiene sentido aplicar la ecuación (2.9) para proyectar los perfiles  $N(\tilde{I})$  sobre cada factor e interpretar las distancias  $d(\tilde{i}, \tilde{i}')$ , para todo  $\tilde{i}$  distinto de  $\tilde{i}'$ , a partir de dicha definición. Uno de los elementos que posibilita la interpretación de cada uno de estos factores es la posición relativa de las modalidades columna, es decir, las coordenadas. Si se trabaja con el mismo tipo de coordenadas (estandarizadas o principales) en ambos conjuntos de perfiles no es posible dotar de significado a las distancias  $d(\tilde{i}, \tilde{j})$  para cualquier  $\tilde{i}$  y  $\tilde{j}$  de  $N(\tilde{I})$  y  $N(\tilde{J})$ , respectivamente. No obstante, cuando el interés es analizar las similitudes entre perfiles de una misma variable, lo relevante es la posición relativa de dichos perfiles y no el valor concreto de sus coordenadas.

En la práctica, para conocer las coordenadas que permiten interpretar factores, se realiza la descomposición espectral, o en valores propios (DVP),

de la matriz cuadrada

$$E^T E = D_c(P_J - 1r^T)D_r^{-1}(P_J - 1r^T)^T, \quad (2.11)$$

de forma que los vectores propios por la derecha e izquierda coinciden (véase ecuación 2.5):

$$E^T E = V D_\lambda V^T.$$

Los valores propios  $\lambda_1, \lambda_2, \dots, \lambda_K$ , que se consideran ordenados tal que  $\lambda_1 > \lambda_2 > \dots > \lambda_K$ , son los elementos de la diagonal principal de la matriz diagonal  $D_\lambda = D_\alpha^2$ . Indican la inercia proyectada sobre cada uno de los vectores propios asociados de manera que  $IT = \sum_{k=1}^K \lambda_k$ .

Se puede decir que la dirección del *primer vector propio* (el vector  $v_1$  asociado al mayor de los valores propios  $\lambda_1$ ) es óptima desde el punto de vista de representar la orientación lineal que recoge la máxima disparidad entre las modalidades columna en lo que a la variable fila se refiere, esto es, la dirección de proyección ortogonal para la que la inercia de la nube proyectada es máxima (véase L. Lebart, A. Morineau, K.M. Warwick, 1984). O de forma equivalente, la dirección que minimiza la distancia entre los puntos del espacio original y el factor de proyección (véase M. Greenacre, 1984).

El *segundo vector propio* ( $v_2$  asociado a  $\lambda_2$ ), es ortogonal al primero y representa la orientación lineal que captura la máxima disparidad residual, es decir, la no recogida en el primer eje de proyección, y así sucesivamente hasta que la inercia total está contenida, y descompuesta, en inercias proyectadas sobre un total de  $K$  ejes ortogonales. La consideración conjunta de todos ellos identifica el espacio  $R^K$  de forma que la nube de perfiles  $N(\tilde{J})$  se puede expresar en función de  $K$  factores de proyección y no sólo de  $I$  modalidades fila, con  $K \prec I$ . Se puede reducir el número de 'modalidades vértice'

respecto a las cuales son interpretables las distancias que permiten hablar de similitudes de 'modalidades perfiles' y en consecuencia de asociación entre dos variables, su posición relativa no ha variado<sup>10</sup>.

Para analizar la asociación entre las variables en el nuevo espacio de referencia es necesario considerar medidas adicionales a las coordenadas que permiten interpretar correctamente los factores obtenidos.

### Ayudas a la interpretación de resultados

La proyección sobre  $R^K$  de la nube  $N(\tilde{J})$  da lugar a los  $K$  factores de proyección  $(P_J - 1r^T)D_r^{-\frac{1}{2}}u_k$  (ecuación 2.7), con  $k = 1, 2, \dots, K$ , donde  $u_k$  es el vector propio asociado al valor propio  $\lambda_k$  obtenido de la DVP. Cada factor contiene las coordenadas principales de cada uno de los perfiles columna en la nueva base:  $m_{\alpha_k}(\tilde{j}) = \sqrt{\lambda_k/f_{.j}}v_k$  (ecuación 2.8) que hacen posible su representación gráfica. El análisis de estas coordenadas permite dotar de una interpretación al factor, pero además se deben considerar las siguientes medidas complementarias:

1. La *Tasa de Inercia* asociada a cada factor  $k$ . Es la proporción de inercia proyectada en el eje  $k$  ( $\lambda_k / \sum_{k=1}^K \lambda_k$ ) y mide la calidad de representación de la nube  $N(\tilde{J})$  en cada factor, es decir, su importancia relativa.
2. La *Contribución Relativa* ( $CR_k(\tilde{j})$ ) de un punto  $\tilde{j}$  en el factor  $k$ , mide la calidad de representación del punto en el eje, es decir, la participación del factor en la explicación del punto. La posición de cada punto está explicado por la totalidad de los factores, así  $\sum_{k=1}^K CR_k(\tilde{j}) = 1$ . Analíticamente se expresa en función de sus coordenadas y como proporción

---

<sup>10</sup>En el Ejemplo 2 del Apéndice I se desarrolla el cómputo de estos valores a través de la DVP de la matriz de perfiles centrados.

del total de inercia del punto,

$$CR_k(\tilde{j}) = \frac{m_{\alpha_k}^2(\tilde{j})}{\sum_{k=1}^K m_{\alpha_k}^2(\tilde{j})},$$

de forma que sus valores oscilan entre  $0 \leq CR_k(\tilde{j}) \leq 1$ , con las siguientes implicaciones:

- $CR_k(\tilde{j}) = 1$  si el punto  $\tilde{j}$  de la nube  $N(\tilde{J})$  se encuentra situado sobre el eje  $k$ , y por tanto, las proyecciones sobre los demás ejes son nulas.
- $CR_k(\tilde{j}) = 0$  si la proyección de  $\tilde{j}$  sobre  $k$  es nula, es decir, se encuentra situado en el subespacio ortogonal al eje  $k$ .

3. La *Contribución Absoluta* ( $CA_k(\tilde{j})$ ) de un punto  $\tilde{j}$  a un determinado factor  $k$ , mide la participación del perfil  $\tilde{j}$  en la inercia proyectada sobre el eje  $k$ , es decir, en la determinación de su orientación en el espacio. Analíticamente equivale al producto del peso del punto por su coordenada al cuadrado y se expresa en términos relativos a la inercia recogida por el factor:

$$CA_k(\tilde{j}) = \frac{1}{\lambda_k} f_j m_{\alpha_k}^2(\tilde{j}) \quad (2.12)$$

con  $j = 1, 2, \dots, J$  y  $k = 1, 2, \dots, K$ , de forma que  $\sum_{j=1}^J CA_k(\tilde{j}) = 1$ , ya que  $\lambda_k = \sum_{j=1}^J f_j m_{\alpha_k}^2(\tilde{j})$ .

Considerando estos conceptos, cada factor de proyección se interpreta, o define, teniendo en cuenta lo que representa, esto es, una visión parcial de las distancias entre los puntos perfiles. Que dos puntos estén próximos en la nube implica que deben estar próximos en la proyección, pero dos puntos

próximos en la proyección estarán próximos en la nube sólo si la calidad de la representación es buena. El proceso de interpretación de los factores debe seguir los siguientes pasos:

1°. Identificar el orden de importancia de los factores a través de la tasa de inercia proyectada sobre cada uno de ellos, ya que esta tasa informa de la parte de la relación entre variables que representa dicho factor.

2°. La interpretación individualizada de cada factor debe preceder a la de los planos formados por su consideración conjunta.

3°. Las modalidades que hayan intervenido en mayor medida en la orientación del factor de proyección (mayor contribución absoluta) son las que permiten interpretarlo, y así cada factor queda definido por la oposición de dichas modalidades proyectadas más extremas: coordenadas negativas frente a coordenadas positivas.

4°. En la medida en que las coordenadas extremas de los perfiles presenten una alta calidad de representación (contribución relativa próxima a uno), se estará indicando que la mayor parte de la divergencia del perfil con el perfil medio está recogido por ese factor. Posteriormente se analizará la oposición de coordenadas entre las que presentan calidades de representación media, etc.

5°. A la hora de interpretar los planos factoriales, también se tendrá en cuenta que los puntos más alejados del origen, son los más diferentes del perfil medio. Si se trabaja con coordenadas principales y se realiza el análisis de los perfiles fila  $N(\tilde{I})$  sobre  $R^K$  (equivalentemente para las columnas  $N(\tilde{J})$ ), las distancias relativas entre los perfiles filas (columnas) se interpretan en términos de similitudes respecto a las modalidades columna (filas). La posición de un perfil columna (fila) con respecto a las filas (columnas) debe analizarse en relación a la totalidad de estas últimas y en términos medios, como indica

la ecuación de transición.

6.º. En el análisis de correspondencias, es frecuente encontrar el llamado *Efecto Guttman* que se caracteriza por una disposición parabólica de las modalidades proyectadas sobre el plano formado por los dos primeros factores. Este efecto aparece cuando existe una estructura de orden en las modalidades fila (columna) asociada a una estructura de orden manifestada igualmente en las modalidades columnas (fila). Indica que la relación entre las modalidades se puede resumir con un solo factor, el cual debe ser interpretado según dicha ordenación. La interpretación del otro factor corresponde a la oposición entre los puntos extremos y medios de la ordenación.

**Aproximación a la nube de perfiles proyectada** Hasta ahora se ha visto que en el espacio  $R^I$  (sería equivalente en  $R^J$ ) se localizan los puntos de  $N(\tilde{J})$  ( $N(\tilde{I})$ ) reflejando las similitudes entre perfiles columna (fila) y respecto a su perfil medio. La distancia chi-cuadrado permite esta interpretación así como conocer el grado de dispersión o concentración de los puntos, es decir, la *IT* de la nube. Esta inercia está igualmente recogida en la totalidad de los ejes factoriales que componen la nueva base ortogonal de  $R^K$  que proporciona la DVP realizada, ya que  $K$  es el rango de la matriz de perfiles original. Una vez hallada la descomposición de la inercia total en  $K$  inercias de proyección (con  $\lambda_1 > \lambda_2 > \dots > \lambda_K$ ), en la medida en que los  $S$  primeros ejes ( $S < K$ ) conserven más fielmente las distancias existentes entre los puntos perfiles en el espacio original  $R^I$  ( $R^J$ ), el porcentaje de la inercia total proyectada sobre el subespacio  $R^S$  será mayor. Esto supondrá que la aproximación de la  $N(\tilde{J})$  ( $N(\tilde{I})$ ) a  $R^S$  lleva consigo una menor pérdida de información (asociación entre las modalidades perfiles y vértices) una vez se fija la dimensión  $S$ .

Se puede medir la importancia relativa de cada eje  $s$  en términos de su

capacidad para resumir la relación existente entre las dos variables consideradas y posteriormente se interpretan en base a los puntos vistos en el apartado anterior. Que la tasa de inercia de  $\lambda_1$  sea próxima a la unidad, refleja que la asociación entre modalidades filas y modalidades columna puede ser recogida casi en su totalidad de forma lineal y, por tanto, se justifica el interés en interpretar el primer factor. Además, puesto que la inercia proyectada sobre el primer eje ( $\lambda_1 = \alpha_1^2$ ) proporciona la máxima correlación lineal que puede encontrarse entre dos variables tras la asignación de valores numéricos a las modalidades (coordenadas), el coeficiente de correlación lineal de Pearson será muy próximo a la unidad ya que su valor es  $\alpha_1$  (véase Greenacre, 1991).

### 2.2.2 Análisis de Correspondencias Múltiples

Existe una gran variedad de técnicas basadas en distintos métodos que, habiendo sido propuestas en distintos países y bajo distintos nombres, son matemáticamente equivalentes: permiten establecer asociaciones entre un determinado número de observaciones definidas por un conjunto de variables cualitativas. En la medida en que existan correlaciones entre estas variables, se puede definir un número reducido de nuevas variables a través de ponderaciones óptimas de las originales que permiten caracterizar a las observaciones objeto de estudio.

Así, en América se dispone de los métodos 'Optimal Scaling', 'Optimal Scoring' y 'Appropriate Scoring'; en Canadá 'Dual Scaling'; en Holanda 'Homogeneity Analysis'; en Israel 'Scalogram Analysis'; y en Japón 'Quantification Method'. El término de ACM es de procedencia francesa, y se puede considerar como la generalización más simple del ACS cuando se consideran observaciones caracterizadas por más de dos variables de tipo discreto (véase Lebart, et al. (1984), Greenacre (1984, 1990 y 1991)). En Tenenhaus y

Young (1985) se revisan distintos métodos para analizar datos multivariantes de carácter cualitativo y se muestra cómo al aplicarse sobre los mismos datos se obtienen los mismos resultados que en el ACM, y, como veremos, resultan de apoyo a la hora de interpretar dichos resultados. Adicionalmente introduce la noción de 'diagrama dual' para sintetizar los aparentemente distintos métodos en uno solo.

Efectivamente, el ACM es una generalización del ACS cuando se dispone de más de dos variables y equivale a analizar las asociaciones entre cada par de modalidades de un determinado conjunto de variables. Cuando se considera el cruce de frecuencias de las modalidades de una misma variable, su asociación es perfecta ya que a cada sujeto le corresponde una única clase de cada variable, lo que distorsiona el valor de la inercia total a explicar. De igual forma, se pierde la intuición a la hora de interpretar los resultados obtenidos, ya que los conceptos geométricos asociados con el ACS se muestran inadecuados para el ACM.

A pesar de esto, se justifica su utilización por la fácil implementación computacional y porque el ACM asigna, al igual que el ACS, un valor escalar a cada modalidad de forma óptima: las coordenadas sobre el primer eje de proyección proporcionan la mayor discriminación entre los sujetos en lo que al conjunto de variables se refiere (véase Greenacre, 1991). La ventaja adicional del ACM sobre los demás métodos es que si se utiliza la Matriz Indicador  $Z$  que se define en el siguiente apartado, es posible asignar a cada observación un valor indicador correspondiente a la variable síntesis del conjunto de variables consideradas. En consecuencia, una vez interpretado cada indicador, la comparación entre las distintas observaciones es inmediata.

## 2.3 Aplicación del ACM a la vivienda

En este capítulo se realiza el ACM en relación a la información sobre las viviendas investigadas en las Encuestas de Presupuestos Familiares (EPF) de 1980-81 y 1990-91, que, como se comentó en el Capítulo I, son elaboradas por el Instituto Nacional de Estadística con ámbito nacional (incluidas Ceuta y Melilla). La utilización del ACM se concreta en tres aplicaciones en función del tipo de viviendas principales consideradas, estas son: las recogidas en la EPF de 1980-81, en la EPF de 1990-91, y en ambas EPF conjuntamente consideradas. Como se verá más adelante, este último análisis permitirá realizar comparaciones entre la vivienda ocupada en 1980-81 y 1990-91.

El objetivo de las aplicaciones del ACM es caracterizar a cada grupo de viviendas respecto al conjunto de servicios que posee de forma que se preserve la máxima dispersión entre ellas con un reducido número de variables síntesis o indicadores. Cada nueva variable indicador puede suponer una medida adecuada que permite la comparación en términos ordinales entre distintas viviendas o agrupaciones de ellas, por ello se debe interpretar y analizar su capacidad de síntesis en relación a los resultados obtenidos.

### 2.3.1 Notación y relación con el ACS

Para cada aplicación, se dispone de información sobre  $Q$  variables de tipo cualitativo referentes a un total de  $N$  viviendas. Sea  $Z$  la Matriz Indicador de dimensiones  $(N \times J)$ , donde  $J$  hace referencia al número total de clases o modalidades de las  $Q$  variables, siendo  $N \gg J$ . Se denota por  $J_q$  al número de modalidades de cada variable  $q$ ,  $q = 1, 2, \dots, Q$ , de manera que  $J = \sum_{q=1}^Q J_q$ .

A cada vivienda le corresponde una, y sólo una, de las modalidades de cada variable, de forma que el término general  $z_{ij}$  de  $Z$  toma valor uno, si

la vivienda  $i$  tiene el atributo  $j$ , y cero en caso contrario, con  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, J$ . Además, por construcción el  $\text{rango}(Z) \leq J - Q$  y se cumplen las relaciones siguientes:  $\sum_{j=1}^J z_{ij} = Q$ ;  $\sum_{i=1}^N z_{ij} = N_j$ , siendo  $N_j$  la frecuencia de la modalidad  $j$ ; para cada  $q$ -variable  $\sum_{j_q=1}^{J_q} N_{j_q} = N$ ; y  $NQ = \sum_{i=1}^N \sum_{j=1}^J z_{ij}$ .

Para caracterizar las  $N$  viviendas de la matriz  $Z$  respecto a las modalidades de las variables columna, de acuerdo con el objetivo mencionado, se utiliza la técnica del ACM. El ACM se presenta en el apartado siguiente como una generalización del ACS, por lo que se hace una continua mención a los términos descritos en la sección anterior.

### ACM como una generalización del ACS

En este trabajo, donde se dispone de múltiples variables con  $N \gg J$ , el ACM consiste en aplicar el ACS a la matriz  $Z$  a través de la proyección de  $N(\tilde{J})$  sobre un espacio base  $R^k$ . Así, la matriz de frecuencias relativas es  $F = Z/NQ$ . A partir de ella se define el perfil columna de cada modalidad  $P_j$  como la  $j$ -ésima columna de  $F$  expresada en relación al total de sus elementos. El elemento  $i$ -ésimo del perfil  $P_j$  es igual a  $z_{ij}/N_j$ ; el peso de cada  $P_j$  es igual a  $N_j/NQ$ , elemento de la diagonal principal de la matriz  $D_c$ ; el perfil columna medio es  $r = F1_N$  con elemento igual a  $1/N$  que permite definir la métrica del espacio  $D_r^{-1}$ .

Se utiliza la ecuación (2.4) para definir la matriz de errores estandarizados  $E$  en función de los perfiles  $N(\tilde{J})$ . En el ACM la inercia total depende del número de variables y modalidades. La DVP de la matriz  $E^T E$  definida en la ecuación (2.11) permite hallar su valor y su descomposición en  $K$  inercias:  $\lambda_1, \lambda_2, \dots, \lambda_K$ <sup>11</sup>. Como se ha expuesto anteriormente, estas particiones de la

<sup>11</sup>La DVP de la matriz cuadrada de la ecuación (2.11) es,  $E^T E = VD_\lambda V^T$  con  $V^T V = Id$ . Se puede demostrar que el valor de la inercia total es  $IT = \text{tr}(E^T E) = \text{tr}(D_\lambda) =$

$IT$  se consideran ordenadas tal que  $\lambda_1 > \lambda_2 > \dots > \lambda_K$  y corresponden a las inercias proyectadas sobre los ejes ortogonales que identifican la base del espacio  $R^K$ . Los factores de proyección de  $N(\tilde{J})$  sobre  $R^K$  (ecuación(2.7)) proporcionan las coordenadas estandarizadas de las modalidades que están contenidas en las filas de la matriz  $M = D_c^{-\frac{1}{2}}V$  tal como indica la ecuación (2.8). Para conocer las coordenadas principales de cada vivienda  $i$  en  $R^K$ , se aplica la ecuación de transición (2.9) que, en este caso, se reduce a hallar la matriz  $N_\alpha = ZM$ , de forma que se pueden definir  $K$  nuevas *variables síntesis* como combinaciones lineales de las características de la vivienda que denotamos por  $w_k$  a partir de ahora. Su valor es  $w_k(i) = n_{\alpha_k}(i)$  para toda vivienda  $i$ , con  $i = 1, 2, \dots, N$ . Por tanto:

$$w_k(i) = \sum_{j=1}^J z_{ij} m_k(j) \quad \forall k = 1, 2, \dots, K \quad (2.13)$$

Estas variables ( $w_k$ ) son independientes entre sí, debido a la ortogonalidad de los factores, o ejes de proyección, y su correlación con las ponderaciones de las modalidades (coordenadas  $m_k(j)$ ) es independiente de las restricciones de normalización, las cuales se imponen a posteriori.

Como se expuso en la sección del ACS, cada una de las variables síntesis es susceptible de interpretación en relación a: i) la proporción de varianza total que viene explicada por cada variable síntesis (tasa de inercia recogida en cada eje); ii) la correlación entre la nueva variable  $w_k$  y las variables originales (contribución absoluta); y iii) las ponderaciones relativas del conjunto de características de la vivienda en cada  $w_k$ , esto es, los valores  $m_k(j)$ .

Antes de presentar los resultados obtenidos y su interpretación, se deben tener en cuenta las implicaciones de aplicar el ACS a la matriz  $Z$  que se consideran en el siguiente apartado.

---


$$(J - Q)/Q.$$

### Interpretación del ACM it versus ACS

La disposición de la información sobre las  $Q$  variables cualitativas en la matriz de datos  $Z$ , aumenta artificialmente la dimensión del espacio original, pero permite una generalización del ACS cuando  $Q > 2$  sin afectar a los resultados óptimos buscados: máxima discriminación entre las viviendas respecto a las variables que las definen (véase página 131-132 de Greenacre, 1984). Desde un punto vista práctico, a la hora de interpretar los resultados obtenidos en el ACM se deben tener en cuenta las siguientes consideraciones respecto a su discrepancia con el ACS, (para mayor detalle véase Greenacre, 1990 y 1991):

- 1º- La dimensión del espacio base  $R^K$ , donde se puede reproducir la estructura de los perfiles fila y columna de la matriz  $Z$ , es  $K \leq J - Q$ .
- 2º- En el ACM la inercia total está determinada desde el momento que se definen las variables cualitativas contenidas en la matriz  $Z$ , ya que  $IT = (J - Q)/Q$ . Por tanto, aquí no representa la dispersión de la nube de puntos original como en ACS, aunque coincide con este análisis en ser una cantidad a explicar.
- 3º- En el caso múltiple, las tasas de inercia proyectadas sobre cada eje son siempre muy pequeñas, lo que muestra una idea pesimista de la verdadera capacidad de resumir información de los factores de proyección, y por tanto de las variables síntesis  $w_k$ . Para conocer su verdadera representatividad Benzécri (1979) propone considerar sólo los  $P$  ejes relevantes, definiendo a éstos como los asociados a los valores propios  $\lambda_p > \frac{1}{Q}$ , con  $p = 1, 2, \dots, P$  y  $P \leq K$ . Igualmente propone corregir los valores propios con la transformación

$$\lambda_p^c = \left( \frac{Q}{Q-1} \right)^2 \left( \lambda_p - \frac{1}{Q} \right)^2 \quad (2.14)$$

y hallar las proporciones de inercia explicadas en relación a su total  $\sum_{p=1}^P \lambda_p^c$ . Se corrige de esta forma la parte trivial de la inercia ( $1/Q$ ) que toda matriz  $Z$  posee aún en caso de total independencia entre viviendas y características.

4º- Los resultados obtenido de aplicar el ACM en referencia a las inercias principales y su descomposición en contribuciones, relativas y absolutas, no tienen un significado geométrico como en el ACS. Las primeras sólo permiten identificar el orden de importancia de las distintas modalidades según su representación en cada eje, mientras que las contribuciones absolutas deben ser interpretadas como correlaciones cuadradas. La comparación entre las técnicas Homogeneity Analysis y ACM que realiza Greenacre (Greenacre, 1991, apartado 8, página 205) muestra que las inercias  $\lambda_k$  obtenidas de este último análisis representa la media de las correlaciones lineales al cuadrado ( $\rho_k^2$ ) entre las ponderaciones  $m_k(j)$  del conjunto de  $J$  modalidades (ecuación 2.13) y las variables  $w_k(i)$ , para todo  $k = 1, 2, \dots, K$ , es decir  $\lambda_k = \rho_k^2$ . En la misma referencia bibliográfica se demuestra que las contribuciones absolutas obtenidas de la ecuación (2.12) permiten hallar las correlaciones  $\rho_{kq}$  entre cada factor, o variable síntesis  $k$ , y cada variable  $q$  a través de la relación:

$$\rho_{kq}^2 = Q \sum_{j \in J_q} \lambda_k C A_k(\tilde{j}) = Q \sum_{j \in J_q} f_{.j} * m_{\alpha_k}^2(\tilde{j}) \quad (2.15)$$

de forma que  $\rho_k^2 = \frac{1}{Q} \sum_{q=1}^Q \rho_{kq}^2$ .

5º- En el ACM, la interpretación de distancias entre puntos proyectados es similar a la del ACS en cuanto que los puntos más alejados del origen son los más importantes, porque son los perfiles más diferentes del perfil medio. Por ello debemos tener en cuenta no solo el peso de la

modalidad en la variable síntesis, sino también la correlación de cada variable original con la nueva variable. En cuanto a la interpretación de la posición relativa de las modalidades debe diferenciarse si las características pertenecen o no a la misma variable. En el primer caso su distancia no es interpretable en términos de semejanza, en esto difiere del ACS. La posición relativa de dos modalidades de distintas variables se interpreta en términos de asociación respecto a las viviendas que representan. Su distancia será menor cuanto mayor sea el número de viviendas que comparten ambos servicios conjuntamente.

### 2.3.2 Datos y variables

Como ya se ha comentado, en este capítulo se realizan tres aplicaciones del ACM sobre la vivienda. Las observaciones que definen la población de estudio en cada aplicación son, respectivamente, las viviendas principales ocupadas en España contenidas en la EPF de 1980-81, en la EPF de 1990-91 y en ambas encuestas analizadas conjuntamente<sup>12</sup>.

La selección de variables a incluir en el análisis se realiza en función de su relevancia *a priori* para determinar la calidad de una vivienda. Así, las  $Q = 8$  variables proporcionadas en la EPF de 1980-81 son: antigüedad del edificio<sup>13</sup>, disponibilidad de servicios higiénicos, metros cuadrados construidos, agua corriente, teléfono, calefacción, garaje y luz eléctrica. En la EPF de 1990-91, además de las variables anteriores, también se proporciona información sobre las diez variables siguientes: tipo de edificio, disponibilidad de aire acondicionado, ascensor, jardín, piscina, zona deportiva, otros servi-

---

<sup>12</sup>El tamaño muestral de la EPF de 1980-81 (1990-91) es de 23.971 (21.155) viviendas principales representativas de 10.024.166 (11.298.509).

<sup>13</sup>Variable calculada a partir del 'año de construcción del edificio'.

cios comunitarios, tipo de energía para calentar el agua, para la calefacción y para cocinar. Por su parte, en el análisis que considera ambas EPF conjuntamente se debe prescindir del segundo tipo de viviendas referido a la vivienda ocupada en 1990-91 por no estar disponibles en la EPF 1980-81.

En gran medida, la falta de información puede condicionar el tamaño muestral disponible. En este sentido, las únicas variables afectadas por ausencia de datos son 'antigüedad del edificio'<sup>14</sup> y 'metros cuadrados construidos'<sup>15</sup>. En ambos períodos se crea una modalidad independiente para cada variable con el fin de identificar la influencia de la 'no respuesta' en el análisis. Aunque se obtienen unos resultados muy similares antes y después de añadir estas modalidades, se decide prescindir de las viviendas sin información, con ello se evita una posible distorsión de los resultados del análisis. Por consiguiente, el tamaño muestral es de 23.898 en 1980-81 y de 20.799 en 1990-91 que representan un total de 9.992.051 y 11.105.215 viviendas a nivel poblacional respectivamente.

Posteriormente, se imputará un valor 'antigüedad' y 'metros cuadrados' a cada una de las viviendas que no dispone de estas variables con objeto de disponer de las correspondientes variables indicador, resultado del ACM, para el total de viviendas de la EPF. El criterio que se sigue para ello es adjudicar la antigüedad media (metros cuadrados medios) de las viviendas de igual régimen de tenencia<sup>16</sup>.

---

<sup>14</sup>En la EPF de 1980-81 hay 73 hogares que no proporcionan el 'año de construcción del edificio'. El número de casos se eleva a 289 en la EPF de 1990-91.

<sup>15</sup>Con 67 casos en las viviendas principales de la EPF de 1990-91 y ningún caso en la EPF de 1980-81.

<sup>16</sup>En concreto, estas imputaciones son: a) en 1980-81 se asignan los siguientes años de construcción: 1932 a propiedad libre (13 casos); 1966 a propiedad de Protección Oficial (3 casos); 1924 a propiedad sin calificación legal (3 casos); 1910 a arrendamiento libre con ocupación previa a la 1965 (7 casos); 1937 a arrendamiento libre con ocupación posterior

Como se ha comentado anteriormente, el ACM requiere que las variables sean de carácter cualitativo. La definición de las  $Q$  variables utilizadas en cada aplicación del ACM y su división en  $J$  modalidades se presenta en el Cuadro 2.1 donde se proporciona la frecuencia relativa de las mismas en cada período. Seguidamente se exponen algunas de las puntualizaciones en relación al análisis descriptivo que ha permitido decidir la codificación final de las variables, cuya notación aparece en la primera columna del Cuadro 2.1.

El análisis descriptivo se inicia con el máximo nivel de desagregación de las variables sobre la vivienda de la EPF de 1980-81 y de la EPF de 1990-91<sup>17</sup>. Posteriormente, con un criterio general, a la hora de determinar el número de modalidades de cada variable, se ha tenido en cuenta que: i) un número reducido de modalidades conlleva una pérdida de información, de forma que los factores del ACM sólo pondrían de manifiesto los aspectos más generales del análisis, y ii) un incremento excesivo en el número de clases puede conducir a una sobredimensión del problema que se traducirá en un

---

a 1964 (21 casos); 1928 a arrendamiento de Protección Oficial (13 casos); 1927 a arrendamiento sin calificación legal (8 casos) y 1931 a algún tipo de cesión (5 casos). b) en 1990-91 se asigna: i) para el 'año de construcción': 1963 a propiedad libre (3 casos); 1951 a arrendamiento ocupado después de 1964 (7 casos); 1925 al arrendamiento de renta antigua, es decir, ocupado antes de 1965 (16 casos); 1964 a arrendamiento de PO (11 casos); 1946 a arrendamiento sin CL (22 casos); 1960 a la vivienda en cesión (230 casos), y ii) para 'metros cuadrados construidos' se asigna: 115 a propiedad libre (7 casos); 93 a propiedad de PO (1 caso); 103 a la propiedad sin CL (2 casos); 90 al arrendamiento de renta antigua (1 caso); 87 al arrendamiento sin CL (2 casos); 92 al arrendamiento sin CL (2 casos), y 97 a algún tipo de cesión (52 casos).

<sup>17</sup>Las variables se encuentran documentadas en Alonso y Colmenar (1994) y en Arévalo et al. (1995) respectivamente, y disponibles en la página web <<http://www.eco.uc3m.es/epf80-81.html> (epf90-91.html)/>.

elevado número de factores, y por tanto de variables indicador  $w(i)$ , con pérdida del aspecto sintético buscado.

De forma particular, las variables continuas han tenido el siguiente tratamiento en el análisis. Como se ha comentado, este tipo de variables debe ser discretizado para su consideración en el ACM. En este trabajo se dispone de dos variables continuas de interés, 'antigüedad del edificio' y 'el número de metros cuadrados construidos'. Debido a que en ninguna de ellas existe un umbral natural que justifique una determinada división en clases, se procedió al análisis de sus distribuciones. Se observó una irregularidad de las mismas que sugirió aplicar, en la medida de lo posible, los siguientes criterios:

i) Definir las modalidades en función de los intervalos determinados por los valores menos frecuentes.

ii) Se ha buscado más un corte de la distribución en clases de igual efectivo, que en intervalos de igual amplitud. Este 'equilibrio' buscado en las frecuencias de las modalidades tiene por objeto no distorsionar de forma artificial los resultados del análisis, ya que, las frecuencias de los perfiles tienen un papel condicionante en los resultados del análisis, en concreto en las coordenadas y contribución absoluta tal y como se muestra en las ecuaciones (2.8) y (2.12).

Efectivamente, si una modalidad es relativamente poco frecuente con un significado claramente discrepante con la(s) modalidad(es) que caracterizan a la mayoría de las vivienda<sup>18</sup>, en el análisis se refleja como una cualidad muy alejada de las modalidades asociadas al perfil medio y, posiblemente con gran influencia en la orientación del factor de proyección correspondien-

---

<sup>18</sup>Un ejemplo podría ser para la variable binaria 'disponibilidad de luz eléctrica', ya que, las frecuencias de la modalidad 'no luz' en 1980-81 y 1990-91 son respectivamente 0,9 y 0,15 por ciento.

te. Sin embargo, se debe evitar provocar este resultado con una discretización incorrecta (de las variables discretas en general y de las continuas en particular) que suponga obtener poca frecuencia en modalidades sin un significado sustancialmente discrepante del resto.

### **2.3.3 Programa computacional**

El software empleado para la ejecución del ACM que posibilita la construcción de indicadores de vivienda es el programa SAS en su versión 6, contexto UNIX. El procedimiento CORRESP permite aplicar análisis de correspondencias tanto simple como múltiple. Con la opción MCA se ejecuta un Análisis de Correspondencias Múltiple, cuyos resultados corresponden a la consideración de una matriz disyuntiva binaria (matriz  $Z$ ) donde las filas representan los hogares, y las columnas las variables incluidas en el análisis. Esta matriz de datos de partida corresponde a un formato SAS previamente elaborado con el mismo programa, de forma que a las modalidades de cada variable cualitativa se le adjudicó previamente un 'label', o etiqueta, con el fin de ser identificadas después, tanto en los resultados numéricos como gráficos. SAS dispone de multitud de opciones para normalizar las coordenadas fila y columna, lo cual muestra la gran flexibilidad de este programa a la vez que la necesidad de un cuidadoso conocimiento del programa y de la técnica que garantice la correcta interpretación de los resultados proporcionados. La opción elegida ha sido la representación de las columnas en coordenadas estandarizadas, aunque, dados los intereses de este trabajo de investigación, posteriormente se aplicará una particular normalización que se detalla más adelante.

## 2.4 Resultados obtenidos

En este apartado se presentan e interpretan los resultados obtenidos en la aplicación del ACM a las viviendas que componen los tres conjuntos muestrales: la EPF 1980-81, la EPF 1990-91 y ambas encuestas conjuntamente que a partir de ahora se denotará como EPF 80-90. En el Apéndice III se muestran los resultados directamente proporcionados por el programa SAS<sup>19</sup> para cada uno de los análisis realizados. Seguidamente se procede a la interpretación de los mismos en los términos comentados en la parte metodológica del AC, es decir, se analizará: i) la importancia relativa de los distintos indicadores que sintetizan el conjunto de características de la vivienda, ii) la correlación de cada indicador con el las variables originales, y iii) la ponderación de cada variables en los indicadores obtenidos.

### 2.4.1 Porcentaje de inercia explicada

La descomposición de la inercia total en los distintos ejes de proyección que se muestra en el Apéndice III precisa de la transformación de Benzécri (ecuación 2.14) para deducir la verdadera representatividad de los factores de proyección. Los resultados de esta corrección se muestran en el Cuadro 2.2. Como se puede apreciar, en ninguno de los análisis la pérdida de información supera el 17 por ciento cuando la reducción de dimensión se limita al plano formado por los dos primeros ejes, ni el 27 por ciento si se reduce a la recta. Por tanto, la simplificación que se puede conseguir al pasar de las  $Q$  variables (con  $J$  modalidades) a tan sólo una variable indicador es realmente relevante. En concreto, el primer indicador obtenido para las viviendas de la EPF de 1980-81 y EPF 80-90 explica más del 86 y el 87 por ciento de la variabilidad

---

<sup>19</sup>Versión 6 del Programa SAS en contexto UNIX.

observa en los respectivos conjuntos de viviendas, donde el número de características analizadas es  $Q = 8$  y  $J = 23$ . Este porcentaje supone algo más de un 73 por ciento en las viviendas de la EPF de 1990-91, donde  $Q = 18$  y  $J = 52$ .

Respecto a las modalidades proyectadas sobre el plano, en los tres análisis se observa el denominado *Efecto Guttman* comentado en el apartado 'Ayudas a la Interpretación de Resultados'. Se caracteriza por la estructura parabólica que presentan las variables con más de dos modalidades. En este caso están, por ejemplo, las variables: servicios higiénicos, metros cuadrados construidos, antigüedad del edificio o agua. El *Efecto Guttman* refleja que mientras que el primer eje sintetiza la estructura del orden de sus proyecciones, el segundo eje representa una oposición entre las modalidades extremas (poca frecuencia) y modalidades medias (de mayor frecuencia) de una misma variable. Dada esta interpretación del segundo factor, de aquí en adelante el análisis se limita al primero y, por tanto, a la primera variable indicador de la vivienda, ya que contiene la información esencial de la variabilidad entre viviendas en lo que a sus características físicas se refiere. Para ello se cambiará la terminología de factor de proyección por variable síntesis o índice; inercia por varianza; y coordenadas por ponderación o peso de las modalidades en las nuevas variables índice (ecuación 2.13) que se denotarán por:  $I_{80}$ ,  $I_{90}$ , y  $I_{8090}$  respectivamente, reflejando con ello la muestra que genera su obtención.

#### **2.4.2 Correlación lineal entre el índice y las características de la vivienda**

Como se ha mencionado, la aplicación del ACM proporciona la descomposición de la varianza total de las viviendas en tantas partes como variables síntesis (índices), y cada una de esas varianzas muestra la media de las corre-

laciones al cuadrado entre los índices de los hogares y los respectivos pesos asignados a las características de las viviendas. En lo referente a la primera variable índice, esta correlación media ( $\rho_1$ ) se halla directamente de los resultados mostrados en el Apéndice III como  $\rho_1 = \sqrt{\lambda_1}$  de forma que se obtienen los siguientes valores: 0,57 en el análisis para la EPF de 1980-81; 0,67 para la EPF de 1990-91; y 0,58 para la EPF 80-90. Por su parte, la correlación del primer indicador y cada una de las  $q$ -variables que caracterizan las viviendas ( $\rho_{1q}$ ) se obtiene a través de la ecuación (2.15) y sus valores se muestran en el Cuadro 2.3.

Se puede observar que las  $Q = 8$  variables comunes en los tres períodos de análisis tienen un comportamiento similar respecto al primer factor: las variables 'servicios higiénicos' y 'disponibilidad de agua' son las más influyentes a la hora de determinar el índice; en una posición intermedia se encuentran las variables 'teléfono', 'antigüedad del edificio' y 'calefacción' que muestran una mayor relación con el primer indicador que las variables 'garaje' y 'superficie construida', mientras que la variable 'luz eléctrica' es la que muestra una destacada menor correlación con dicho indicador. Esta última variable es la que presenta mayor divergencia entre las frecuencias de sus modalidades, en 1990-91, por ejemplo, sólo el 0,15 de las viviendas no tienen luz eléctrica. A pesar de las pocas viviendas que no disponen de este atributo, se ha decidido su inclusión en el ACM considerando la relevancia de esta información a la hora de discriminar viviendas respecto a los servicios instalados. Todas las variables, salvo 'servicios higiénicos' y 'agua', ganan importancia en la determinación del índice en 1990-91, respecto a la década anterior, entre ellas destaca la variable 'garaje' que en 1990-91 presenta una correlación lineal con el primer indicador de 0,58 mientras que en 1980-81 era de 0,41. Esto quiere decir que la posesión de garaje tiene más importancia en

el segundo período que en el primero en lo que respecta a la determinación del la primera variable indicador que, como se verá más adelante, se le puede interpretar como un 'índice de calidad' de la vivienda.

### 2.4.3 Pesos de las modalidades en el índice

Con objeto de interpretar el primer indicador, corresponde ahora realizar el análisis de los pesos relativos de las modalidades<sup>20</sup>, sin olvidar la contribución de cada variable a la construcción de dicho indicador que viene dada por su correlación ( $\rho_{1q}$ ). Respecto a la comparación entre los tres ACM realizados, se debe tener en cuenta que los pesos de las modalidades están en una escala de medida distinta en cada período de referencia. Se trata de valores estandarizados y la escala depende directamente de la inercia total explicada por el índice. Así, dicha comparación sólo tiene sentido en términos relativos dentro de cada período. En el Cuadro 2.4 se presenta, para cada muestra, la ordenación de modalidades de menor a mayor peso tras haber aplicado una normalización que conserva tanto el signo del peso de la modalidad correspondiente, como el valor cero referido al índice medio en cada conjunto de viviendas. La normalización consiste en asignar el valor arbitrario -10 a la modalidad de menor peso y el valor 10 al equivalente valor extremo de signo positivo, conservando los pesos relativos de los servicios que componen la vivienda<sup>21</sup>. Como se puede observar, el valor máximo no aparece en nin-

<sup>20</sup>Las coordenadas de las modalidades de cada variables que se presentan en el Apéndice III.

<sup>21</sup>Si se denota por  $m_j$  al peso de la modalidad  $j$  que procede del análisis, y por  $m'_j$  al que corresponde tras la normalización, con  $j = 1, 2, \dots, J$ , su relación viene dada por la ecuación:

$$m'_j = m_j \frac{10}{\bar{m}}$$

siendo  $\bar{m} = \max\{|m_1|, |m_2|, \dots, |m_J|\}$

guno de los tres períodos, lo cual indica que no hay ninguna modalidad que compense positivamente el valor extremo negativo tomado por la modalidad de peso -10.

Teniendo en cuenta cada variable individualmente, la ordenación de sus modalidades presenta un peso negativo para el caso de su peor característica, signo positivo en su mejor clase, y una gradación acorde con su significado en el caso de existir más de dos modalidades. Este resultado es sistemático en los tres análisis realizados y para todas las variables involucradas. La no posesión de servicios higiénicos, agua y luz, resultan ser las modalidades extremas con influencia negativa. Además, las dos primeras son las más correladas con el índice, lo que indica que son las variables más importantes a la hora de su interpretación. Por otro lado, y en lo que se refiere a las variables comunes en los tres análisis, 'la disponibilidad de garaje' y 'servicios higiénicos con más de un baño completo' son las características que están asociados a los valores positivos extremos. En 1990-91, donde se considera un mayor número de características, las variables con mayor peso sobre el índice son: 'zona deportiva', 'gas ciudad o gas natural para la cocina', 'ascensor', 'piscina', y 'cocina eléctrica'.

Los aspectos más relevantes que se observan al comparar ambos períodos son:

a) En 1980-81, la existencia de calefacción aporta un peso al índice de 2,79 , mientras que la no calefacción se separa en menor medida del hogar medio (-1,73). La correlación de esta variable con el índice es de 0,59 en este año, mientras que en 1990-91 se eleva a 0,65. Además, en este año su comportamiento se invierte: la inexistencia de calefacción en el hogar resta más peso al índice de la vivienda (-3,62) de lo que suma su posesión (0,22). Esto puede ser debido a que el bien 'calefacción' es menos usual en 1980-81

que en 1990-91, de forma que en el primer período, disponer de ella destaca en mayor medida mientras que, en el segundo es la 'no calefacción' la que juega un papel más destacado sobre la media de las viviendas.

b) La variable 'garaje' tiene un comportamiento semejante en ambos años: tener garaje representa un mayor peso absoluto en el índice que lo contrario, aunque la correlación con él es inferior en 1980-81 (0,41 frente a 0,58).

c) Con el 'teléfono', que tiene una correlación de 0,62 en 1980-81 y 0,69 en 1990-91, se produce un fenómeno similar al comentado para la calefacción: en 1980-81 tener teléfono pondera con mayor peso que su falta. En 1990-91, 'no teléfono' afecta negativamente en mayor medida, en comparación con la ponderación positiva a que da lugar su posesión. Nuevamente, la justificación se puede encontrar en las frecuencias relativas: en 1990-91 el 77 por ciento de las viviendas tenían teléfono mientras que este porcentaje no llegaba a un 51 por ciento en la década anterior.

La correlación entre índice y variables, junto con los pesos de las modalidades, permiten interpretar la nueva variable como un *Índice de Calidad de la Vivienda*. Este indicador hace posible ordenar la totalidad de viviendas en relación a su calidad, la cual representa un valor indicativo de la desviación (positiva o negativa) respecto a la calidad media que toma valor cero por construcción. Las modalidades comunes asociadas con la *Vivienda Media* en ambos períodos son las siguientes: una antigüedad del edificio entre 20 y 30 años, vivienda con luz, un baño (tal vez no completo), agua caliente, entre 61 y 90 metros cuadrados y sin garaje. La vivienda media en 1980-81 no posee teléfono ni calefacción, mientras que en 1990-91 dispone de ambos además de estar caracterizada por estar situadas en edificios con más de tres viviendas, sin ascensor, con servicios comunitario pero sin servicios como: aire acondicionado, jardín, piscina o zona deportiva. El tipo de energía que

se utiliza para cocinar y calentar el agua en la vivienda media es el butano, mientras que para la calefacción se combina este combustible con la energía eléctrica siendo los aparatos móviles el tipo de calefacción más asociado con la calidad de la vivienda media.

#### **2.4.4 Limitaciones de la variable índice de calidad**

Es importante hacer notar que los índices son interpretables en términos de un concepto de calidad restringido a las  $Q$  variables que intervienen en cada aplicación del ACM, con unos únicos pesos de las características físicas para toda vivienda a comparar. En este punto, se deben tener en cuenta dos limitaciones del Índice de Calidad hallado:

A) No se han considerado variables que pueden ser relevantes en un concepto más amplio de calidad y que se pueden agrupar en: i) otras características físicas: tipo de material utilizado en la construcción, estado de conservación, etc. y, ii) características del entorno: grado de contaminación, seguridad ciudadana, etc. La no disponibilidad de estas variables impide su utilización, limitando la información sintetizada en el índice.

B) No se recoge la posible heterogeneidad en el peso de las modalidades sobre la calidad para distintos segmentos de viviendas. Así, por ejemplo, es de suponer que la calefacción en Canarias debería tener menor peso sobre la calidad que en cualquier zona del norte de España. La consideración apropiada de esta circunstancia se podría resolver hallando un índice por cada agrupación de interés, aunque se haría imposible la comparación entre zonas. Dados los objetivos de este trabajo, se ha optado por calcular un único índice de calidad por período muestral. No obstante, dado que la variable calefacción es la más afectada por esta limitación, se decidió incluir en la modalidad 'con calefacción' la posesión de aparatos móviles con el fin de

atenuar la discrepancia entre zonas que puede distorsionar su comparación en términos de calidad.

## 2.5 Aplicaciones del índice de calidad

El ACM utilizado para sintetizar las características de la vivienda de la EPF de 1980-81 (EPF de 1990-91) ha permitido construir el índice de calidad de la vivienda *I80* (*I90*). Se trata de un indicador de carácter ordinal que tiene la cualidad de ser independiente del resto de indicadores obtenidos en el análisis, de forma que la consideración conjunta del total de indicadores permite sintetizar la totalidad de las características de la vivienda disponible en la encuesta. La aplicación de estos indicadores puede estar justificada en cualquier análisis multivariante que requiera la propiedad de independencia para su adecuada utilización. Así, en el Capítulo III se identifican los índices *I80* (*I90*) como determinantes del alquiler mensual pagado en 1980-81 (1990-91) por las viviendas en arrendamiento, lo que permite estimar el efecto de la calidad de la vivienda en el precio a través de un modelo de regresión.

Por su parte, el análisis de correspondencias aplicado a la muestra conjunta: EPF 80-90 ha proporcionado el índice de calidad *I8090* que, entre otras cosas, facilita la comparación del nivel de calidad de las viviendas ocupadas en 1990-91 en relación a 1980-81 ya que se reduce el número de variables a manejar, de ocho características a una única variable. Para hacer operativo este tipo de análisis descriptivo se crea una variable que identifica la muestra a la que pertenece cada vivienda, lo que permite controlar el posible efecto diferencial del nivel de calidad en ambos períodos. Seguidamente se realiza la comparación, tanto en media como en distribución por cuartiles, del nivel de calidad entre distintas agrupaciones de viviendas de la siguiente forma:

en una primera fase se identifican las características asociadas a los niveles de calidad de cada partición cuartílica para, en una segunda fase, comparar la calidad en 1990-91 respecto a 1980-81 para distintos tipos de vivienda.

### 2.5.1 Asociación entre las características de la vivienda y el nivel de calidad

Una vez construido el índice de calidad común *I8090*, se ordenan los hogares de menor a mayor calidad y se divide la distribución en cuartiles. Posteriormente, se analiza la distribución de frecuencias de las modalidades de cada partición cuartílica con objeto de asociar una *vivienda tipo* a cada nivel de calidad que vendrá dada por las modalidades más frecuentes de cada variable en cada partición del *I8090*. En el Cuadro 2.5 se muestra tanto la frecuencia relativa de las modalidades como su distribución en términos porcentuales en cada partición cuartílica. Como se puede observar, en la primera partición se encuentra la totalidad de hogares sin servicios higiénicos (4,8 por ciento del total de hogares y el 18,3 por ciento de los hogares que componen este 25 por ciento de viviendas de peor calidad); la totalidad de los que no tienen luz, y prácticamente el 100 por cien de los que no poseen agua corriente. De forma progresiva, en los siguientes cuartiles, se van registrando mayores porcentajes en las modalidades asociadas a un mayor nivel de calidad. Así, en el cuarto cuartil, el 97 por ciento de las viviendas tienen calefacción y/o teléfono, el 100 por cien tiene agua caliente; el 77,5 por ciento tiene menos de 20 años de antigüedad; el 76,3 por ciento tiene más de 90 metros cuadrados y el 63,1 por ciento tiene más de un baño. Resulta revelador comprobar la correspondencia de la medida ordinal 'Índice de Calidad', con las variables que representa y determinan su significado.

### **2.5.2 Comparación de la calidad de la vivienda en 1980-81 y 1990-91**

En este apartado se muestra el análisis descriptivo que permite comparar el nivel medio de la variable *I8090*, así como su distribución, según los períodos de referencia: 1980-81 y 1990-91. Así, en el Cuadro 2.6 se muestra esta información para el conjunto de viviendas ocupadas en cada período. Se observa como la calidad media de las viviendas ocupadas en España ha mejorado de forma significativa en la década de estudio. Los porcentajes de viviendas en cada cuartiles avalan esta conclusión ya que, su distribución se invierte de 1980-81 a 1990-91, ganando peso los dos cuartiles superiores en detrimento de los inferiores. En 1980-81, el 70,3 por ciento de los hogares se corresponde con la calidad representada por el primer y segundo cuartil mientras que, diez años después, la mayoría de los hogares (el 65 por ciento) se concentran en los dos últimos cuartiles, lo que corrobora la mejora en la calidad de las viviendas en su conjunto.

Este misma información se puede desagregar para distintos tipos de viviendas de tal forma que, se puede identificar tanto el cambio producido en esta década como la heterogeneidad existente en lo que al acondicionamiento de las viviendas se refiere. En concreto, las agrupaciones de viviendas que se consideran son: i) las Comunidades Autónomas (CCAA); ii) el tamaño de municipio; y iii) el régimen de tenencia y calificación legal de la vivienda.

#### **Comunidades Autónomas**

Así, en el Cuadro 2.7 se muestra la media y distribución cuartílica del índice de calidad *I8090* según las CCAA que componen el territorio nacional incluyendo a Ceuta y Melilla. La ordenación en que se presentan las dieciocho

CCAA corresponde al ranking de mayor a menor calidad en 1980-81. Coincidiendo con el resultado obtenido para el conjunto total de viviendas, en términos medios y para el período de referencia, la calidad de la vivienda ha mejorado de forma significativa en todas y cada una de las CCAA aunque a distintos niveles. En términos comparativos, las CCAA que han experimentado una mejora más destacada son las que presentaban una calidad media más baja en 1980-81: Andalucía, Ceuta-Melilla, Galicia, Castilla-La Mancha y Extremadura. En este sentido se puede afirmar que la desigualdad entre CCAA ha disminuido en la década de estudio, en lo referente al acondicionamiento de la vivienda. El ranking está encabezado por el País Vasco en el primer período y por Navarra en el segundo. Por otra parte, Madrid es la comunidad Autónoma que experimenta un menor cambio medio en su índice de calidad pasando, del segundo puesto en 1980-81, al octavo en 1990-91. El fenómeno contrario ocurre en Castilla-La Mancha que pasa, de ocupar la antepenúltima posición en 1980-81, precediendo a Galicia y Extremadura, a situarse en el séptimo lugar.

En lo que respecta a la distribución de la calidad según cuartiles, en la mayoría de las CCAA se observa, igualmente, un comportamiento de mejora similar al del conjunto de las viviendas. Destaca Navarra por ser la que tiene mayor porcentaje de viviendas en la última partición por cuartiles, 25 por ciento en 1980-81 y 54,6 por ciento en 1990-91. En Andalucía y Extremadura también se produce un cambio en el mismo sentido que en el total nacional, pero en menor medida que en la mayoría de las CCAA. Canarias y Ceuta-Melilla difieren del resto de comunidades por acumular tanto en 1980-81 como en 1990-91 más del 50 por ciento de las viviendas en la primera y segunda partición cuartílica. Puesto que el índice de calidad está determinado por unas ponderaciones únicas asociadas a cada característica,

ésto puede distorsionar la comparación de calidades entre CCAA en la medida en que existan características con distinta importancia según su localización.

### **Tamaño de municipio**

De forma similar al expuesto en el apartado anterior, en el Cuadro 2.8 se muestra el valor medio y la distribución por cuartiles del índice de calidad I8090 para distintos tamaño de municipio. Nuevamente se constata, tanto en media como en distribución, un cambio positivo en 1990-91 respecto a 1980-81 en cada una de las particiones consideradas. En ambos períodos se aprecia un mejor acondicionamiento de la vivienda en su conjunto a medida que el municipio es de mayor tamaño. También se evidencia una mayor desigualdad en 1990-91 que en la década anterior, tanto en media como en distribución. La mejora relativa más relevante se produce en los municipios con menos de 10.000 habitantes, ya que en 1980-81 el 65 por ciento de estas viviendas se correspondía con la calidad del primer cuartil, cuando en los demás municipios no se superaba el 48 por ciento, y en 1990-91 su proporción en el cuarto cuartil es similar a la de las demás particiones, y por tanto en torno al valor medio del 35 por ciento. Tratando de interpretar esta circunstancia, y teniendo en cuenta que INE, a través de las EPF, permite estimar un desalojo de 68.091 viviendas en este ámbito rural, se puede pensar que el tipo de viviendas que, con mayor probabilidad, han sido desocupadas son las de peor calidad.

### **Régimen de tenencia y calificación legal**

La comparación de la calidad por régimen de tenencia se muestra en el Cuadro 2.9. Tanto en la EPF de 1980-81 como en la de 1990-91 se obtiene una clara asociación entre las viviendas con calificación legal de protección oficial y

una 'buena calidad' de la vivienda. Este resultado es robusto en media y en distribución por cuartiles en el sentido de que, es el tipo de viviendas en cada período muestral presenta mejor calidad con independencia del régimen de tenencia, ya sea de propiedad, arrendamiento o cesión. Este hecho puede estar justificado en la medida en que se produzca un efecto sustitución del gasto en vivienda: el ahorro generado por el acceso a una vivienda más barata (protección oficial), se destina a un mejor acondicionamiento de la misma, con el resultado final de una ventaja comparativa respecto a otras viviendas. Si esto no ocurre en la cesión, puede ser debido a la temporalidad de la misma. Aunque este tipo de cuestiones sobrepasan los objetivos de este capítulo, se pospone al Capítulo IV de la tesis un análisis sobre la relación entre el valor del flujo de servicios de la vivienda, su calidad y el régimen de tenencia.

Otros comentarios que se pueden deducir del análisis del Cuadro 2.9 al comparar las muestras 1980-81 y 1990-91 son:

1.- Todas y cada una de las divisiones establecidas experimentan una mejora en el acondicionamiento de las viviendas en la década de estudio.

2. En ambos períodos se obtiene un comportamiento sistemático en la ordenación de los tipos de vivienda según su calidad, en dos sentidos:

a) Para todo régimen de tenencia, el ranking de viviendas queda establecido en orden descendente por: protección oficial, libre y sin calificación conocida.

b) Para toda calificación legal, la vivienda mejor acondicionada es la que se disfruta en régimen de propiedad. En la mayoría de los casos le sigue, por este orden, arrendamiento y cesión.

3.- La vivienda cedida, si es de protección oficial, supera en calidad al régimen de arrendamiento libre y sin calificación en ambos períodos, e incluso

al alquiler de protección oficial en 1980-81.

4.- En términos relativos a cada régimen de tenencia, la calificación legal de protección oficial es la experimenta la mejora más moderada. Por otro lado, también se obtiene que las viviendas en régimen de propiedad y cesión experimentan un cambio más importante que las de arrendamiento.

En cada período, la desviación típica de la media del *I8090* se puede considerar como una aproximación a la dispersión de la calidad entre las viviendas que componen cada muestra. En este sentido, se observa que en todos los regímenes de tenencia, salvo en arrendamiento libre, esta dispersión es menor en 1990-91 que en 1980-81. Este resultado también se observa en las distintas CCAA y los distintos tamaños de municipio, lo que corrobora una menor desigualdad en el segundo período con los tres criterios analizados.

## 2.6 Conclusiones

El objetivo del Capítulo II es diseñar algunas medidas objetivas que hagan posible comparar la totalidad del parque de viviendas de nuestro país en relación al nivel de acondicionamiento que proporcionan a sus ocupantes. Se busca la mejor síntesis de las características que definen la vivienda principal ocupada en la Encuesta de Presupuestos Familiares (EPF) de 1980-81 y 1990-91, de forma que se maximice la variabilidad observada en los respectivos conjuntos de viviendas. Para ello se aplica la técnica multivariante de Análisis de Correspondencias Múltiples que permite construir tres índices -*I80*, *I90* e *I8090*- como combinación lineal de las características disponibles en la EPF y cuya interpretación es la 'calidad de la vivienda' en la vivienda ocupada en 1980-81, en 1990-91 y en ambos períodos, respectivamente.

Para la construcción del *I8090* se consideran ambas EFP conjuntamente, por lo que es preciso prescindir de aquellas variables que están disponibles en la EFP de 1990-91 pero no en la de 1980-81. En concreto, las variables que se utilizan en los índices *I80* e *I8090* son ocho: antigüedad del edificio, tipo de servicios higiénicos, agua corriente, superficie construida, disponibilidad o no de luz eléctrica, calefacción, garaje y teléfono. Además de estas características, en la determinación del índice *I90* se utilizan diez variables más: el tipo de edificio, el tipo de combustible para cocinar, para calentar el agua, para la calefacción, disponibilidad o no de ascensor, aire acondicionado, jardín, piscina, zona deportiva y otros servicios comunitarios.

La pérdida de información que supone considerar la variable índice de calidad en lugar de las variables que sintetiza, es inferior al 27 por ciento en los tres análisis y, por construcción, dicha información está incorrelada con la calidad. Por otra parte, la correlación lineal media de las correspondientes variables originales y los índices de calidad hallados son de: 0,57 para *I80*, 0,58 para *I90* y 0,67 para *I8090*.

No obstante, no todas las variables tienen el mismo peso a la hora de determinar el índice de calidad. Así, en los tres análisis se obtiene que 'la falta servicios higiénico dentro de la vivienda', 'no tienen agua corriente' o 'no tener luz eléctrica' son las modalidades que en mayor medida penalizan la calidad de la vivienda, de tal forma que no hay ninguna característica que compense positivamente el efecto negativo de estas variables. Las variables 'superficie construida' y 'antigüedad del edificio' también tienen un comportamiento similar a la hora de determinar el índice de calidad de cada período. Así se observa que, la peor modalidad de las dos primeras variables ('menos de 60 metros cuadrados' y 'más de 50 años de antigüedad', respectivamente) reducen el nivel de calidad en mayor medida que lo aumenta la mejor

modalidad ('más de 130 metros cuadrados' y 'menos de 11 años'). Con el 'garaje' ocurre lo contrario, 'no tener garaje' penaliza menos a la calidad de la vivienda de lo que incrementaría su posesión.

Las variables 'calefacción' y 'teléfono' destacan por tener un efecto distinto sobre el índice de calidad definido para 1980-81 (*I80*) que para 1990-91 (*I90*). Mientras que en el primer período el efecto positivo derivado de disponer de estos bienes es superior al efecto negativo que supone la carencia de estos servicios, en 1990-91 se obtiene el resultado contrario. Este resultado es coherente si tenemos en cuenta que estas variables son las que experimentan un mayor cambio distribucional en el segundo período respecto al primero: el 89 por ciento de las viviendas ocupadas en 1990-91 tenían calefacción (el 77 por ciento tenían teléfono), mientras en 1980-81 este porcentaje era de un 38 por ciento (un 51 por ciento).

El índice *I80* (*I90*) es una variable referida únicamente a las viviendas de la EPF de 1980-81 (1990-91) ya que, este tipo de índices son variables ordinales que reflejan un nivel de calidad relativo a la vivienda media del período de referencia, de tal forma que el valor de la calidad media es cero. En los dos períodos analizados la vivienda de calidad media está caracterizada por: una antigüedad entre 20 y 30 años, vivienda con luz, un baño (tal vez no completo), agua caliente, entre 61 y 90 metros cuadrados y sin garaje. En 1980-81 no posee teléfono ni calefacción, mientras que en 1990-91 dispone de ambos servicios además de están situadas en edificios con más tres viviendas, sin ascensor, con servicios comunitario, sin aire acondicionado, sin jardín, ni piscina o zona deportiva, con cocina y calentador de agua de butano, y se utilizan aparatos eléctricos y móviles para caldear la vivienda.

Por su parte, el índice *I8090*, al considerar ambas encuestas conjuntamente, hace posible la comparación entre las viviendas ocupadas en ambos

períodos en términos del nivel de calidad. Así, se obtiene que la vivienda principal ocupada en 1990-91 posee mayor calidad que en 1980-81, tanto en media como en distribución. Este resultado se observa en todas y cada una de las Comunidades Autónomas, en los distintos tamaños de municipio considerados y en cada uno de los regímenes de tenencia y calificación legal analizados en la EPF. No obstante, se observa un comportamiento heterogéneo dentro de cada tipo de agrupación de viviendas.

En general se aprecia una mejora más apreciable en el conjunto de viviendas de peor calidad, y por tanto, una disminución de la desigualdad en 1990-91. En particular, Navarra destaca entre las demás Comunidades Autónomas, por tener la mayor concentración de viviendas de mejor calidad en 1990-91 mientras que, Ceuta-Melilla ocupa la peor situación. En ambos períodos de estudio se obtiene una asociación positiva entre el tamaño de municipio y el nivel de acondicionamiento de la vivienda. Lo mismo ocurre con la vivienda de calificación legal 'protección oficial', respecto al mercado libre, y con el régimen de tenencia en 'propiedad', respecto al arrendamiento y la cesión.



. Cuadros del capítulo II

Cuadro 2.1: Definición y frecuencia de las variables utilizadas en el Análisis de Correspondencias

VARIABLES: modalidades	NOTACIÓN	FRECUENCIA RELATIVA	
		EPF 1980-81 (% s/9.992.051)	EPF 1990-91 (% s/11.105.215)
<b>ANTIGÜEDAD DEL EDIFICIO</b>			
Más de 50 años	antig0	25,53	11,28
Entre 31 y 50 años	antig1	27,7	13,61
Entre 21 y 30 años	antig2	11,48	25,01
Entre 11 y 20 años	antig3	11,01	33,39
Menos de 11 años	antig4	24,28	16,71
<b>SERVICIOS HIGIÉNICOS</b>			
No tiene o compartido con otras viviendas	hig0	7,61	1,75
Un baño completo o menos [aseo(s) y/o retrete(s)]	hig1	77,33	73,3
Más de un baño completo	hig2	15,06	24,95
<b>SUPERFICIE CONSTRUIDA</b>			
Menos de 61 metros cuadrados	m2v0	12,7	9
Entre 61 y 90 metros cuadrados	m2v1	41,56	37,85
Entre 91 y 130 metros cuadrados	m2v2	33,45	37,36
Más de 130 metros cuadrados	m2v3	12,29	15,79
<b>AGUA</b>			
No tiene	agua0	3,84	0,36
Sólo agua fría	agua1	19,42	5,05
Agua fría y caliente	agua2	76,74	94,59
<b>TELÉFONO</b>			
No tiene	tel0	49,25	22,73
Sí tiene	tel1	50,75	77,27
<b>CALEFACCIÓN</b>			
No tiene	calf0	61,65	10,93
Individual, Colectiva y/o aparatos móviles	calf1	38,35	89,07
<b>GARAJE</b>			
No tiene	gar0	83,63	72,29
Sí tiene	gar1	16,37	27,71
<b>LUZ ELÉCTRICA</b>			
No tiene	luz0	0,9	0,15
Sí tiene	luz1	99,1	99,85
<b>TIPO DE EDIFICIO</b>			
Alojamiento fijo o fines no residenciales	tiped0		0,29
Edificio de una sola planta	tiped1		35,61
Edificio de dos plantas	tiped2		4,61
Edificio de 3 o más plantas	tiped3		59,49
<b>AIRE ACONDICIONADO</b>			
No tiene	aire0		97,63
Si tiene	aire1		2,37
<b>ASCENSOR</b>			
No tiene	asc0		70,32
Si tiene	asc1		29,68
<b>JARDÍN</b>			
No tiene	jar0		84,93
Comunitario o privado	jar1		15,07
<b>PISCINA</b>			
No tiene	pisc0		98,8
Si tiene	pisc1		1,2
<b>ZONA DEPORTIVA</b>			
No tiene	zdep0		98,95
Si tiene	zdep1		1,05
<b>OTROS SERVICIOS COMUNITARIOS</b>			
No tiene	serv0		52,75
Si tiene	serv1		47,25
<b>COMBUSTIBLE O ENERGÍA PARA CALENTAR AGUA</b>			
No tiene	cbagua0		5,41
Sólidos: leña, carbón	cbagua1		2,06
Butano	cbagua2		59,81
Energía eléctrica	cbagua3		16,16
Otros: gas ciudad, gas natural, gasóleo,...	cbagua4		16,56
<b>COMBUSTIBLE O ENERGÍA PARA CALEFACCIÓN</b>			
No tiene	cbcalf0		10,93
Sólidos: leña, carbón	cbcalf1		19,37
Butano	cbcalf2		13,09
Energía eléctrica	cbcalf3		40,58
Otros: gas ciudad, gas natural, gasóleo,...	cbcalf4		16,03
<b>COMBUSTIBLE O ENERGÍA PARA COCINAR</b>			
Sólidos: leña, carbón	cbcoc1		3,36
Butano	cbcoc2		76,82
Energía eléctrica	cbcoc3		6,84
Otros: gas ciudad, gas natural, gasóleo,...	cbcoc4		12,98

**Cuadro 2.2: Descomposición de la Inercia Proyectada**

FACTORES	EPF 1980-81		EPF 1990-91		CONJUNTO.	
	Inercia corregida	%	Inercia corregida	%	Inercia corregida	%
Primer factor	0,0542	86,17	0,1682	73,21	0,0561	87,79
Segundo factor	0,0076	12,08	0,0242	10,52	0,0076	11,89
Tercer factor						
Resto de factores	0,0011	1,75	0,0374	16,27	0,0002	0,31
Total	0,0629	100	0,2298	100	0,0639	100

**Cuadro 2.3: Correlación lineal de las variables y el primer factor**

EPF 1980-81		EPF 1990-91		CONJUNTO	
Servicios higiénicos	0,76	Otros servicios comunitarios	0,83	Agua	0,75
Agua	0,75	Ascensor	0,80	Servicios higiénicos	0,74
Teléfono	0,62	Aire acondicionado	0,80	Teléfono	0,63
Antigüedad del edificio	0,60	Zona deportiva	0,78	Calefacción	0,61
Calefacción	0,59	Tipo de edificio	0,73	Antigüedad del edificio	0,53
Garaje	0,41	Combustible para cocinar	0,72	Superficie construida	0,49
Superficie construida	0,40	Servicios higiénicos	0,72	Garaje	0,43
Luz eléctrica	0,26	Combustible para calentar agua	0,71	Luz eléctrica	0,26
		Agua	0,70		
		Teléfono	0,69		
		Antigüedad del edificio	0,68		
		Calefacción	0,65		
		Jardín	0,62		
		Garaje	0,58		
		Combustible para calefacción	0,57		
		Superficie Construida	0,49		
		Piscina	0,40		
		Luz eléctrica	0,33		

Cuadro 2.4: Ordenación de los pesos normalizados

EPF 1980-81		EPF 1990-91		CONJUNTO	
luz0	-10,00	agua0	-10,00	luz0	-10,00
agua0	-8,41	luz0	-8,72	agua0	-8,72
hig0	-8,09	hig0	-8,32	hig0	-7,89
agua1	-4,21	cbagua0	-7,16	agua1	-4,38
antig0	-3,25	agua1	-6,94	antig0	-2,89
m2v0	-2,59	cbcoc1	-4,97	m2v0	-2,75
tel0	-0,17	tiped0	-4,91	tel0	-2,38
calf0	-1,73	antig0	-3,66	calf0	-2,36
antig1	-1,49	calf0	-3,62	antig1	-0,76
gar0	-0,68	m2v0	-3,27	gar0	-0,66
antig2	-0,22	tel0	-3,27	hig1	-0,24
hig1	-0,07	tiped1	-2,93	m2v1	-0,19
luz1	0,09	antig1	-2,82	luz1	0,05
m2v1	0,36	serv0	-2,44	antig2	0,35
antig3	1,08	cbcalf1	-2,25	agua2	0,81
m2v2	1,43	antig2	-2,01	antig3	0,83
agua2	1,49	tiped2	-1,93	m2v2	1,11
m2v3	1,65	asc0	-1,82	calf1	1,26
tel1	2,29	cbcalf2	-1,35	tel1	1,31
antig4	2,67	gar0	-1,02	antig4	1,39
calf1	2,79	hig1	-0,98	m2v3	1,58
gar1	3,47	cbagua2	-0,89	gar1	2,32
hig2	4,34	cbagua1	-0,81	hig2	2,67
		antig3	-0,81		
		cbcoc2	-0,69		
		m2v1	-0,33		
		aire0	-0,08		
		zdep0	-0,05		
		pisc0	-0,04		
		jar0	-0,01		
		luz1	0,01		
		jar1	0,04		
		cbcalf3	0,15		
		calf1	0,22		
		m2v2	0,32		
		agua2	0,32		
		cbagua3	0,41		
		m2v3	0,81		
		tel1	1,14		
		antig4	2,14		
		tiped3	2,21		
		gar1	2,41		
		cbcalf4	2,65		
		serv1	3,16		
		aire1	3,29		
		hig2	3,41		
		cbagua4	3,65		
		cbcoc3	4,46		
		pisc1	4,69		
		asc1	4,69		
		cbcoc4	5,19		
		zdep1	5,49		

Cuadro 2.5: Distribución de los cuartiles del Índice de calidad 18090 según características de la vivienda					
VARIABLES: modalidades	FRECUENCIAS RELATIVAS EN %				
	Total	Particiones cuartiles del 18090			
		I	II	III	IV
<b>ANTIGÜEDAD DEL EDIFICIO</b>					
Más de 50 años	18,1	45,2	20,3	5,8	1,2
Entre 31 y 50 años	12,0	16,3	10,7	13,6	7,3
Entre 21 y 30 años	18,0	12,6	17,8	27,4	14
Entre 11 y 20 años	30,3	16,3	32,1	33,6	39,2
Menos de 11 años	21,6	9,6	19,1	19,6	38,3
<b>SERVICIOS HIGIÉNICOS</b>					
No tiene o compartido con otras viviendas	4,6	18,3	-	-	-
Un baño completo o menos [aseo(s) y/o retrete(s)]	75	80,3	95,4	87,3	36,9
Más de un baño completo	20,4	1,3	4,5	12,7	63,1
<b>SUPERFICIE CONSTRUIDA</b>					
Menos de 61 metros cuadrados	15,8	35,7	25,3	2,3	0,1
Entre 61 y 90 metros cuadrados	42,7	45,7	40,5	61,5	23,6
Entre 91 y 130 metros cuadrados	29,8	14,1	25,4	28,7	51,2
Más de 130 metros cuadrados	11,6	4,8	8,8	7,6	25,1
<b>AGUA</b>					
No tiene	2	8	0,1	-	-
Sólo agua fría	11,9	45	2,4	0,3	-
Agua fría y caliente	86,1	47	97,5	99,7	100
<b>TELÉFONO</b>					
No tiene	35,4	85,6	41,8	11,3	2,9
Sí tiene	64,6	14,4	58,2	88,7	97,1
<b>CALEFACCIÓN</b>					
No tiene	34,9	82,8	42,3	11,5	2,9
Individual, Colectiva y/o aparatos móviles	65,1	17,2	57,7	88,5	97,1
<b>GARAJE</b>					
No tiene	77,8	96,3	92,5	84,8	37,5
Sí tiene	22,2	3,7	7,5	15,2	62,5
<b>LUZ ELÉCTRICA</b>					
No tiene	0,5	2	-	-	-
Sí tiene	99,5	98	100	100	100

**Cuadro 2.6: Media y distribución en % del Índice de Calidad**

EPF	Media (Desviación Típica)	Particiones cuartílicas del I8090			
		I	II	III	IV
1980-81	-2,245 (0,004)	42,3	28,0	16,0	13,7
1990-91	2,217 (0,003)	9,7	22,3	33,0	35,0

**Cuadro 2.7: Media y distribución en % del Índice de Calidad por Comunidad Autónoma**

Comunidad Autónoma	EPF	Media (Desviación Típica)	Particiones cuartílicas del I8090			
			I	II	III	IV
País Vasco	1980-81	0,073 (0,013)	17,4	29,1	30,3	23,2
	1990-91	0,266 (0,008)	4,6	18,5	41,8	35,1
Madrid	1980-81	0,027 (0,013)	18,4	37,7	25,0	18,9
	1990-91	0,248 (0,015)	2,8	21,8	47,0	28,4
Navarra	1980-81	0,018 (0,025)	29,0	21,6	24,3	25,1
	1990-91	0,400 (0,015)	2,3	11,0	32,1	54,6
Aragón	1980-81	-0,091 (0,016)	30,8	27,9	21,8	19,5
	1990-91	0,281 (0,011)	5,8	16,3	36,5	41,4
Cataluña	1980-81	-0,100 (0,010)	35,3	32,9	15,7	16,1
	1990-91	0,321 (0,008)	3,6	16,9	39,4	40,1
Murcia	1980-81	-0,120 (0,023)	33,2	37,5	14,4	14,9
	1990-91	0,235 (0,020)	10,4	23,2	22,3	44,1
Valencia	1980-81	-0,134 (0,012)	39,0	27,5	18,9	14,6
	1990-91	0,294 (0,008)	5,7	20,5	32,3	41,5
La Rioja	1980-81	-0,182 (0,028)	46,3	19,3	20,8	13,6
	1990-91	0,329 (0,019)	6,4	18,4	27,5	47,7
Cantabria	1980-81	-0,236 (0,027)	38,6	33,5	14,9	13,0
	1990-91	0,181 (0,022)	8,4	24,8	33,2	33,6
Asturias	1980-81	-0,253 (0,024)	46,3	23,4	14,7	15,6
	1990-91	0,153 (0,024)	17,7	18,7	26,7	36,9
Balears	1980-81	-0,267 (0,030)	46,0	23,4	15,3	15,3
	1990-91	0,326 (0,019)	7,1	14,8	30,0	48,1
Castilla y León	1980-81	-0,355 (0,012)	48,5	21,4	15,3	14,8
	1990-91	0,237 (0,008)	10,4	19,2	28,7	41,7
Canarias	1980-81	-0,414 (0,021)	55,0	28,5	9,3	7,2
	1990-91	-0,093 (0,015)	31,0	37,2	20,1	11,7
Andalucía	1980-81	-0,449 (0,010)	56,9	27,2	10,1	5,8
	1990-91	0,130 (0,007)	13,9	30,7	28,1	27,3
Ceuta y Melilla	1980-81	-0,483 (0,031)	66,6	22,0	7,6	3,8
	1990-91	-0,151 (0,029)	32,6	41,2	17,8	8,4
Castilla la Mancha	1980-81	-0,511 (0,018)	58,2	21,7	10,6	9,5
	1990-91	0,256 (0,010)	9,0	21,9	27,9	41,2
Galicia	1980-81	-0,578 (0,020)	62,4	19,7	7,9	10,0
	1990-91	0,090 (0,014)	20,5	22,7	22,7	34,1
Extremadura	1980-81	-0,740 (0,023)	74,7	14,0	5,9	5,4
	1990-91	0,085 (0,017)	18,1	27,6	27,1	27,2

. Apéndice I: Ejemplos del  
Análisis de Correspondencias

Ejemplo 1: Análisis gráfico de una tabla de contingencias Se dispone de información de un total de 100 viviendas sobre dos variables: AGUA (disponibilidad de agua) y BAÑO (número de cuartos de baño completos). La primera presenta tres modalidades y la segunda cuatro, que se definen en la siguiente tabla de contingencia donde se muestra la frecuencia conjunta de cada cruce de modalidades:

AGUA\BAÑO	Ninguno	Menos de uno*	Uno	Más de uno	Total
No tiene	2	0	0	0	2
Sólo fría	1	20	12	0	33
Caliente	0	2	53	10	65
Total	3	22	65	10	100

\* Cuarto de aseo y/o retrete.

Para analizar la asociación entre BAÑO y AGUA se expresan las frecuencias de las modalidades de una de las variable en función de su total. Aquí se opta por analizar las modalidades BAÑO en relación a la variable AGUA, pero los resultados serían idénticos en el caso de la elección contraria. Así, la matriz de perfiles columna de la tabla anterior es:

$$P_j \equiv 100 * \begin{bmatrix} 2/3 & 0/22 & 0/65 & 0/10 \\ 1/3 & 20/22 & 12/65 & 0/10 \\ 0/3 & 2/22 & 53/65 & 10/10 \end{bmatrix}^T = \begin{bmatrix} 66 & 33 & 0 \\ 0 & 91 & 9 \\ 0 & 18 & 82 \\ 0 & 0 & 100 \end{bmatrix}$$

Esto muestra, por ejemplo, que el 66 por ciento de las viviendas sin baño corresponde a hogares que no disponen de agua corriente mientras que el 33 por ciento restante sólo tienen agua fría. Por su parte, la totalidad de viviendas con más de un cuarto de baño tiene agua caliente.

Cada punto perfil de la variable BAÑO tiene asociado un peso que informa de la proporción de viviendas que corresponde a cada modalidad,

$$c \equiv [ 0,03 \quad 0,22 \quad 0,65 \quad 0,10 ] ,$$

de manera que la matriz diagonal de pesos es

$$D_c \equiv \begin{bmatrix} 0,03 & 0 & 0 & 0 \\ 0 & 0,22 & 0 & 0 \\ 0 & 0 & 0,65 & 0 \\ 0 & 0 & 0 & 0,10 \end{bmatrix}.$$

Los pesos asignados a los perfiles permiten hallar el perfil columna medio  $r \equiv [r_1 \ r_2 \ r_3]^T$  como la suma de los perfiles columna con su correspondiente ponderación respecto al total de viviendas. Así, los componentes de  $r$  se aproximan de la siguiente forma:

$$r_1 = 0,03 * 66 + 0,22 * 0 + 0,65 * 0 + 0,10 * 0 = 1,98 \simeq 2$$

$$r_2 = 0,03 * 33 + 0,22 * 91 + 0,65 * 18 + 0,10 * 0 = 32,71 \simeq 33$$

$$r_3 = 0,03 * 0 + 0,22 * 9 + 0,65 * 82 + 0,10 * 100 = 65,28 \simeq 65$$

$$r \equiv [ 2 \quad 33 \quad 65 ]^T$$

Se observa que el perfil columna medio coincide con el peso de las modalidades fila: el 2 por ciento de las viviendas no tienen agua, el 33 por ciento sólo agua fría y el 65 por ciento dispone de agua caliente.

#### A. Representación gráfica:

La información de la tabla de contingencia del ejemplo 1 expresada en términos de frecuencias condicionadas a las columnas permite definir la nube de perfiles BAÑO  $N(J) = N(4)$  en relación a  $R^I = R^3$ , el espacio que determinan los tres puntos vértices de la variable AGUA en un espacio ortogonal (la tres modalidades son independientes entre sí)

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

En el Gráfico 2-1 se representan los perfiles BAÑO (de coordenadas  $P_J$ ) y el perfil medio (de coordenadas  $r$ ) en el espacio Euclídeo ordinario  $R^3$  definido

por los vértices AGUA, por lo que las distancias observadas son distancias euclídeas.

Por otra parte, las filas de  $P_J$  definen las coordenadas de los cuatro puntos perfiles en un espacio Euclídeo Ponderado donde los ejes están inversamente ponderados por la matriz diagonal  $D_r \equiv \text{diag}(r)$ . Así, la métrica del espacio que permite definir la distancia chi-cuadrado entre perfiles está determinada por

$$D_r^{-1} \equiv \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/33 & 0 \\ 0 & 0 & 1/65 \end{bmatrix}$$

En el Gráfico 2-2 se representan los perfiles en el espacio ponderado, de manera que al comparar con el gráfico anterior se aprecia la importancia de ponderar de forma distinta los puntos vértice. Ahora las distancias entre perfiles reflejan las similitudes entre modalidades BAÑO en términos relativos a las modalidades de AGUA y una vez consideradas las discrepancias de las frecuencias de esta última variable: sólo el 2 por ciento de las viviendas no tienen agua mientras que el 65 por ciento dispone de agua caliente.

Además, como se ha comentado, el conjunto de perfiles se localiza en el subespacio de dimensión  $(I - 1)$ , por lo que en este ejemplo los perfiles BAÑO se sitúan en el plano formado por la unión de los puntos vértices del Gráfico 2-2, como muestra el Gráfico 2-3.

Las coordenadas de los perfiles de las modalidades de la variable BAÑO en relación a los tres ejes de la variable AGUA, permiten hacer las siguientes deducciones: i) la modalidad "más de un cuarto de baño" se identifica con la disponibilidad de "agua caliente", mientras que "sin baño" se asocia con viviendas "sin agua" o "agua fría"; ii) la vivienda menos frecuente y por ello más alejada de la vivienda media, es la que no dispone de baño en su interior, frente a la vivienda de un baño que es la modalidad de mayor peso;



iii) las modalidades más similares entre sí son la disponibilidad de "un cuarto de baño" y "más de un baño" y es debido a su gran asociación con "agua caliente".

#### B. Inercia total y estadístico Chi-cuadrado:

Tanto la inercia total como el estadístico  $\chi^2$  permiten medir el grado de asociación entre las variables BAÑO y AGUA. La divergencia de cada perfil BAÑO con el perfil medio computada a través del cuadrado de la distancia chi-cuadrado es:

$$d^2(1, r) = \frac{1}{0,02}(0,66-0,02)^2 + \frac{1}{0,33}(0,33-0,33)^2 + \frac{1}{0,65}(0-0,65)^2 = 21,13$$

$$d^2(2, r) = \frac{1}{0,02}(0-0,02)^2 + \frac{1}{0,33}(0,91-0,33)^2 + \frac{1}{0,65}(0,09-0,65)^2 = 1,5219$$

$$d^2(3, r) = \frac{1}{0,02}(0-0,02)^2 + \frac{1}{0,33}(0,18-0,33)^2 + \frac{1}{0,65}(0,82-0,65)^2 = 0,13264$$

$$d^2(4, r) = \frac{1}{0,02}(0-0,02)^2 + \frac{1}{0,33}(0-0,33)^2 + \frac{1}{0,65}(1,00-0,65)^2 = 0,53846$$

Hallados estos valores, la inercia de la nube de modalidades BAÑO y el estadístico  $\chi^2$  son, respectivamente,

$$Inercia(N(J)) = 0,03*21,13+0,22*1,5219+0,65*0,13264+0,10*0,53846 = 1,1088$$

y

$$\chi^2 = 100 * 1,1088 = 110,88.$$

El valor de la  $\chi^2$  con seis grados de libertad que deja el 0,05 a su derecha es 0,872, lo que muestra que hay evidencia estadística en contra de la hipótesis de independencia entre las variables BAÑO y AGUA al nivel de confianza del 95 por ciento. La constatación de la asociación entre las variables apoya la interpretación dada a la distancia entre perfiles en términos de similitud de modalidades, para lo que resulta de gran ayuda la representación gráfica.

#### C. Análisis de perfiles centrados:

Se puede comprobar que:

a) Las coordenadas y pesos de los puntos de  $N(\tilde{J})$  en  $R^I = R^3$  son:

$$P_J - 1r^T = \begin{bmatrix} 64 & 0 & -65 \\ -2 & 58 & -56 \\ -2 & -15 & 17 \\ -2 & -33 & 35 \end{bmatrix}$$

$$c \equiv [ 0,03 \quad 0,22 \quad 0,65 \quad 0,10 ]$$

b) Las coordenadas del perfil medio de la nube  $N(\tilde{J})$  en  $R^3$  son  $\tilde{r} \equiv [ 0 \quad 0 \quad 0 ]^T$  ya que,

$$\tilde{r}_1 = 0,03*(66-1,98)+0,22*(0-1,98)+0,65*(0-1,98)+0,10*(0-1,98) = 0$$

$$\tilde{r}_2 = 0,03*(33-32,71)+0,22*(91-32,71)+0,65*(18-32,71)+0,10*(0-32,71) = 0$$

$$\tilde{r}_3 = 0,03*(0-65,28)+0,22*(9-65,28)+0,65*(82-65,28)+0,10*(1-65,28) = 0$$

c) Debido a que  $d^2(j, r) = d^2(\tilde{j}, \tilde{r})$  se obtiene que:

$$Inercia(N(\tilde{J})) = 1,1088 = Inercia(N(J))$$

$$\chi^2 = 110,88$$

Así pues, las representaciones gráficas de  $N(J)$  y  $N(\tilde{J})$  son idénticas. La ventaja de trabajar con la segunda opción es que el signo de la coordenada de cada perfil informa directamente de su posición relativa al valor medio, sin detrimento de la interpretación de la distancia entre puntos como asociaciones de perfiles ya que la posición relativa de los puntos en  $N(J)$  no ha variado y por tanto la distancia entre ellos es idéntica.

Gráfico 2-1: Espacio Euclideo ordinario

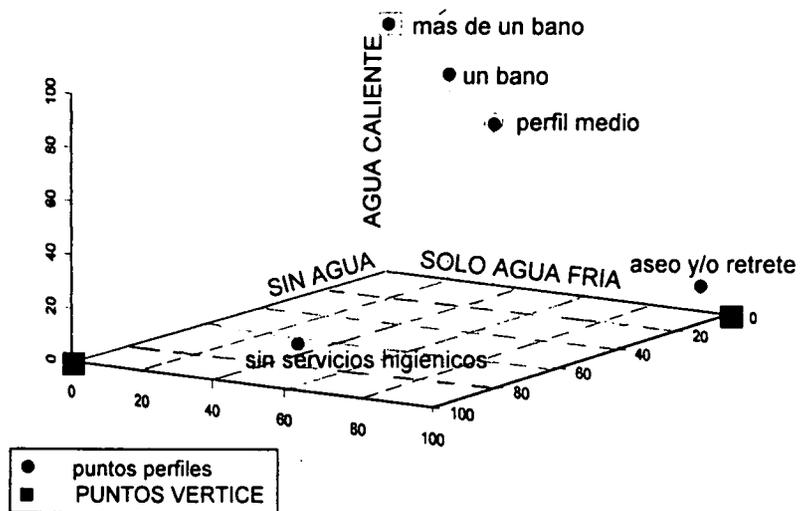


Gráfico 2-2: Espacio Euclideo ponderado

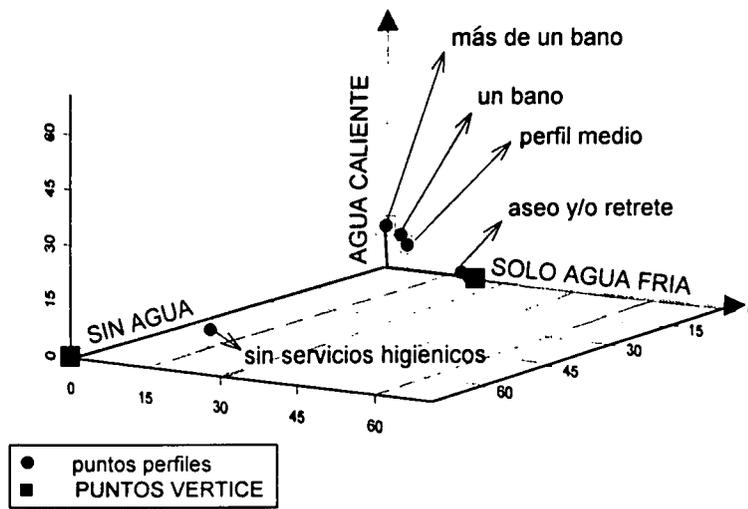
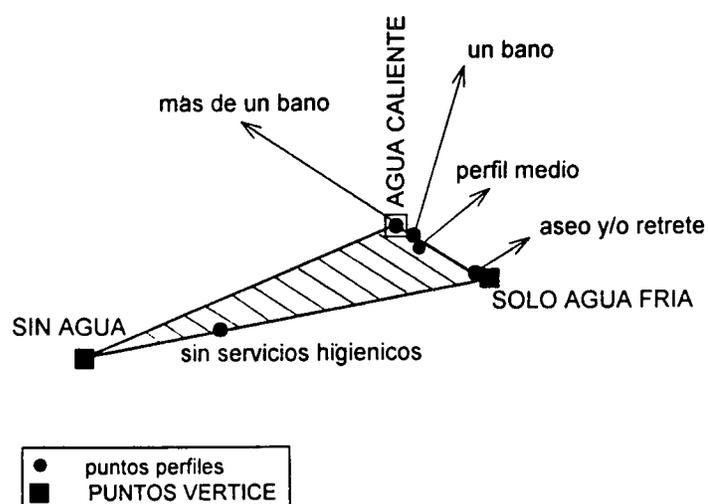


Gráfico 2-3: Plano del Espacio Euclideo ponderado



## Ejemplo 2: Búsqueda de subespacios Óptimos

Sea una Comunidad Autónoma (CA) de la que se extrae una muestra aleatoria simple de 2.536 viviendas y se consideran dos variables: ZONA (seis provincias dentro de la CA) y CALEFACCION (cuatro modalidades respecto a la disponibilidad de esta variable). El objetivo del análisis es conocer las posibles asociaciones de las distintas provincias en relación a la posesión del servicio calefacción de las viviendas ubicadas en cada una de

ellas. La información disponible aparece en la siguiente tabla de contingencia:

ZONA/CALEFACCION	No tiene	Aparatos móviles	Individual	Colectiva	Total
Provincia 1	138	74	14	10	236
Provincia 2	296	90	28	44	458
Provincia 3	340	130	24	58	552
Provincia 4	318	114	24	56	512
Provincia 5	244	94	22	82	442
Provincia 6	292	14	2	28	336
Total	1.628	516	114	2,78	2.536

Los perfiles columna y el perfil medio, en porcentajes, están recogidos en las filas de las matrices  $P_J$  y  $r^T$  respectivamente:

$$P_J = \begin{bmatrix} 8,48 & 18,18 & 20,88 & 19,53 & 14,99 & 17,94 \\ 14,34 & 17,44 & 25,19 & 22,09 & 18,22 & 2,71 \\ 12,28 & 24,56 & 21,05 & 21,05 & 19,30 & 1,75 \\ 3,60 & 15,83 & 20,86 & 20,14 & 29,5 & 10,07 \end{bmatrix}$$

$$r^T \equiv [ 9,30 \quad 18,06 \quad 21,77 \quad 20,19 \quad 17,43 \quad 13,25 ]$$

La caracterización de la nube de puntos centrada  $N(\tilde{J})$  viene dada por las matrices:  $P_J - 1r^T$  (coordenadas respecto a  $R^I = R^6$ ),  $D_c$  (pesos) y  $D_r^{-1}$  (métrica del espacio), que toman los siguientes valores:

$$P_J - 1r^T \equiv \begin{bmatrix} -0,83 & 0,12 & -0,88 & -0,66 & -2,44 & 4,69 \\ 5,04 & -0,62 & 3,43 & 1,90 & 0,79 & -10,54 \\ 2,97 & 6,50 & -0,71 & 0,86 & 1,87 & -11,49 \\ -5,71 & -2,23 & -0,90 & -0,05 & 12,07 & -3,18 \end{bmatrix}$$

$$D_c \equiv \begin{bmatrix} 64,20 & 0 & 0 & 0 \\ 0 & 20,34 & 0 & 0 \\ 0 & 0 & 4,50 & 0 \\ 0 & 0 & 0 & 10,96 \end{bmatrix}$$

$$D_r^{-1} \equiv \begin{bmatrix} 1/9,30 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/18,06 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/21,77 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/20,19 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/17,43 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/13,25 \end{bmatrix}$$

De esta forma la matriz cuadrada a diagonalizar (2.11) es:

$$E^T E = \begin{bmatrix} 136,85 & -162,31 & -77,27 & -60,55 \\ -162,31 & 241,69 & 101,20 & -1,30 \\ -77,27 & 101,2 & 60,80 & 10,17 \\ -60,55 & -1,30 & 10,17 & 141,76 \end{bmatrix}$$

Su descomposición en valores y vectores propios, tal que,

$$E^T E = V D_\lambda V^T, \text{ con } V^T V = Id$$

da como resultado:

$$D_\lambda \equiv \begin{bmatrix} 411,63 & 0 & 0 \\ 0 & 154,71 & 0 \\ 0 & 0 & 14,76 \end{bmatrix} \text{ y}$$

$$V \equiv \begin{bmatrix} -0,56 & -0,20 & -0,01 \\ 0,74 & -0,30 & 0,40 \\ 0,34 & -0,06 & -0,91 \\ 0,14 & 0,93 & 0,07 \end{bmatrix}$$

Este resultado indica que la base de la nube de puntos que componen los perfiles CALEFACCION en relación a la variable ZONA es de dimensión tres, el número de columnas de la matriz  $V$ . Es posible reproducir la estructura original de  $R^6$  en  $R^3$  sin pérdida de información puesto que es un simple cambio de espacio de referencia donde, ahora, es preciso dar una interpretación a los nuevos vectores  $V$ . Las coordenadas estandarizadas ( $M$ ) y principales ( $M_\alpha$ )

de los perfiles columna en  $R^3$ , halladas a través de la ecuación (2.8), son:

Modalidades de CALEFACCION	$M$			$M_\alpha$		
	Eje 1	Eje 2	Eje 3	Eje 1	Eje 2	Eje 3
No tiene	-0,070	-0,025	-0,002	-1,426	-0,313	-0,006
Sólo aparatos móviles	0,164	-0,066	0,088	3,329	-0,824	0,340
Calefacción individual	0,161	-0,026	-0,431	3,268	-0,323	-1,656
Calefacción colectiva	0,041	0,281	0,022	0,831	3,498	0,083

Salvo efectos en el redondeo a tres decimales, la divergencia entre ambos conjuntos de coordenadas supone un cambio de escala, pero la posición relativa de los perfiles CALEFACCION es la misma.

#### A. Ecuaciones de Transición

El número de las modalidades de la variable CALEFACCION es menor que el de la variable ZONA por lo que ha resultado práctico realizar el análisis de los perfiles columna. Para cada nueva dimensión de  $R^K = R^3$  el conjunto de perfiles CALEFACCION sirve para caracterizar cada factor de proyección  $k = 1, 2, 3$ , mientras que las posiciones del conjunto de modalidades ZONA se deben interpretar en relación a dichos factores. Sus coordenadas se calculan a través de la ecuación (2.9) de forma que, las coordenadas principales ( $N_\alpha$ ) de perfiles ZONA ( $i = 1, 2, \dots, 6$ ) en la nueva base  $R^3$  se hallan como puntos medios de los perfiles CALEFACCION ( $j = 1, 2, 3, 4$ ),  $n_{\alpha k}(i) = \sum_{j=1}^J \frac{f_{ij}}{f_i} m_k(j)$  para  $k = 1, 2, 3$ . Sus valores son:

Modalidades de ZONA	$N_\alpha$		
	Eje 1	Eje 2	Eje 3
Provincia 1	2,164	-2,514	0,220
Provincia 2	0,061	-0,388	-0,789
Provincia 3	0,666	-0,271	0,342
Provincia 4	0,492	-0,086	0,090
Provincia 5	1,172	2,287	0,052
Provincia 6	-4,990	-0,138	0,158

Como se verá en los siguientes apartados, para analizar la asociación entre las variables ZONA y CALEFACCION en el nuevo espacio de referencia, es necesario considerar medidas adicionales a las coordenadas que permiten interpretar correctamente los factores obtenidos.

#### B. Ayudas a la interpretación

A la hora de interpretar la posición de las modalidades CALEFACCION y ZONA en  $R^3$  se dispone de la siguiente información:

1. La inercia total  $IT = tr [E^T E] = tr [D_\lambda] = 581.1$  y su proyección sobre cada uno de los ejes que compone el espacio  $R^3$  es la siguiente:

	Inercia proyectada	% inercia	% acumulado
Eje 1	411,628	70,84	70,84
Eje 2	154,711	26,62	97,46
Eje 3	14,761	2,54	100
Total	581,100	100	

Así pues, la estructura que presentan los distintos perfiles CALEFACCION en relación a la ZONA en que se ubica la vivienda queda identificada, en más de un 97 por ciento, en el plano formado por los dos primeros factores de proyección. A su vez, se observa que la importancia relativa del segundo factor es considerablemente menor que la del primero.

2. Las contribuciones relativas expresadas en términos porcentuales son:

Modalidades CALEFACCION	Eje 1	Eje 2	Eje 3	Suma
No tiene	95,39*	4,61	0,00	100
Sólo aparatos móviles	93,31	5,72	0,97	100
Calefacción individual	78,95	0,77	20,28	100
Calefacción colectiva	5,34	94,61	0,05	100

$$* 95,39 = \left( \frac{(-1,426)^2}{(-1,426)^2 + (-0,313)^2 + (-0,006)^2} \right) * 100$$

El perfil 'calefacción colectiva' está mejor representado en el segundo factor que en el primero, mientras que a las demás modalidades les ocurre lo contrario.

3. La descomposición de la inercia de cada factor  $k$  según las modalidades proyectadas y su expresión en términos porcentuales respecto a la inercia del factor  $\lambda_k$  es<sup>22</sup>:

	Eje 1	%	Eje 2	%	Eje 3	%
No tiene	130,543*	31,72**	6,309	4,07	0,002	0,01
Sólo aparatos móviles	225,445	54,77	13,815	8,93	2,351	15,22
Calefacción individual	48,048	11,67	0,470	0,30	12,345	79,90
Calefacción colectiva	7,574	1,84	134,090	86,7	0,075	4,87
Suma	411,610	100	154,684	100	14,774	100

\*  $130,543 = (-1,426)^2 * 64,20$  ; \*\*  $31,72 = \frac{130,543}{411,628} * 100$

Las dos modalidades que reflejan "no instalación de calefacción" influyen en mayor medida en la orientación del primer factor de proyección y de forma más relevante "sólo aparatos móviles" que "no tiene". La clase "calefacción colectiva" determina de forma significativa el segundo factor, mientras que "calefacción individual" determina el tercero.

#### C. Aproximación a la nube de puntos

La aproximación óptima de la nube de puntos en  $R^3$  que contiene la información sobre la asociación entre CALEFACCION y ZONA del ejemplo 2-2, precisa prescindir de alguno, o algunos, de los tres factores de proyección. El factor con menor inercia proyectada es el que menor pérdida de información supone su eliminación. En este ejemplo es el tercer eje con una pérdida del 2.54 por ciento. En los Gráficos 2-4 y 2-5 se muestran, respectivamente, las proyecciones de las modalidades CALEFACCION (en coordenadas estándar.) y ZONA (coordenadas principales) en  $R^2$ , plano formado por los dos primeros factores de proyección.

En el Gráfico 2-4 se deben interpretar ambos factores. El primero muestra claramente la oposición entre disponer, o no, de algún sistema que permi-

<sup>22</sup> Contribuciones Absolutas (ecuación 2.12)

ta aclimatar la vivienda. El segundo factor muestra la oposición entre un sistema colectivo de calefacción frente a los demás. Los perfiles más similares entre sí son "calefacción individual" y "aparatos móviles" en cuanto a la ubicación por zonas. La modalidad más próxima al origen de coordenadas es "no calefacción", lo cual muestra que es la más influyente a la hora de determinar el perfil medio del total de viviendas consideradas<sup>23</sup>.

En el Gráfico 2-5 las ZONAS proyectadas sobre el primer eje permiten asociar la provincia "6" con viviendas que en promedio no disponen de menos medios de calefacción. La mayor similitud se da entre viviendas localizadas en las provincias "2", "3" y "4". En términos relativos a las demás zonas, la provincia "5" es la que presenta mayor asociación con 'calefacción colectiva'. Esto se deduce una vez visto que dicha modalidad de CALEFACCION es la que en mayor medida determina la orientación del segundo eje, mayor contribución absoluta.

---

<sup>23</sup>La proximidad de "calefacción colectiva" al origen no se interpreta porque se ha visto que la calidad de representación de esta modalidad no es muy buena para este factor.

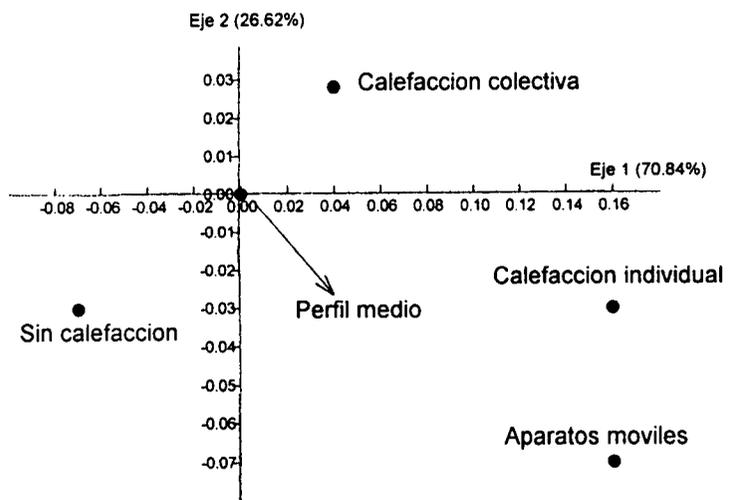
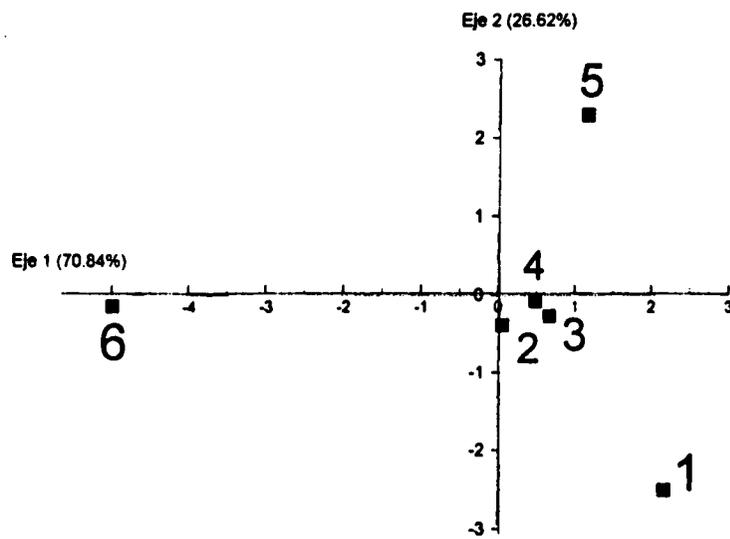
Gráfico 2-4: Proyección de las modalidades CALEFACCION en  $R^2$ 

Gráfico 2-5: Proyección de las modalidades ZONA en  $R^2$





**. Apéndice II: Deducción de las ecuaciones de transición**

La descomposición en valores y vectores singulares de la Matriz de Errores Estandarizados lleva a que  $E = UD_\alpha V^T$ , con  $U^T U = V^T V = I$ . Dicha matriz se puede escribir en función de las nubes de perfiles centrados  $N(\tilde{I})$ :  $E = D_r^{\frac{1}{2}}(P_I - 1c^T)D_c^{-\frac{1}{2}}$ , y de  $N(\tilde{J})$ :  $E^T = D_c^{\frac{1}{2}}(P_J - 1r^T)D_r^{-\frac{1}{2}}$ , lo que permite establecer las relaciones entre las matrices de coordenadas estandarizadas ( $N$  y  $M$ ) de los respectivos conjuntos de perfiles y de los vectores de proyección ( $V$  y  $U$ ) de la forma siguiente:

Para $N(\tilde{I})$ :	Para $N(\tilde{J})$ :
$U = EVD_\alpha^{-1}$	$V = E^TUD_\alpha^{-1}$
$U = D_r^{\frac{1}{2}}(P_I - 1c^T)D_c^{-\frac{1}{2}}VD_\alpha^{-1}$	$V = D_c^{\frac{1}{2}}(P_J - 1r^T)D_r^{-\frac{1}{2}}UD_\alpha^{-1}$
Premultiplicando por $D_r^{-\frac{1}{2}}$ ,	Premultiplicando por $D_c^{-\frac{1}{2}}$ ,
y dado que $N = D_r^{-\frac{1}{2}}U$	y dado que $M = D_c^{-\frac{1}{2}}V$
$N = (P_I - 1c^T)D_c^{-\frac{1}{2}}VD_\alpha^{-1}$	$M = (P_J - 1r^T)D_r^{-\frac{1}{2}}UD_\alpha^{-1}$

Para encontrar la relación entre  $N$  y  $M$ , se premultiplican las expresiones de  $U$  y  $V$  por  $(P_J - 1r^T)D_r^{-\frac{1}{2}}$  y  $(P_I - 1c^T)D_c^{-\frac{1}{2}}$  respectivamente y se posmultiplican ambas por  $D_\alpha^{-1}$ :

i) Para  $N(\tilde{I})$ :

$$(P_J - 1r^T)D_r^{-\frac{1}{2}}UD_\alpha^{-1} = (P_J - 1r^T)D_r^{-\frac{1}{2}}D_r^{\frac{1}{2}}(P_I - 1c^T)D_c^{-\frac{1}{2}}VD_\alpha^{-1}D_\alpha^{-1}$$

De forma que:

$$M = (P_J - 1r^T)ND_\alpha^{-1}$$

ii) Para  $N(\tilde{J})$ :

$$(P_I - 1c^T)D_c^{-\frac{1}{2}}VD_\alpha^{-1} = (P_I - 1c^T)D_c^{-\frac{1}{2}}D_c^{\frac{1}{2}}(P_J - 1r^T)D_r^{-\frac{1}{2}}UD_\alpha^{-1}D_\alpha^{-1}$$

De forma que:

$$N = (P_I - 1c^T)MD_\alpha^{-1}$$

En el siguiente paso se demuestra que  $r^T N = 0^T$  y que  $c^T M = 0$ , lo que permite simplificar las expresiones anteriores.

$$\begin{array}{l}
 r^T N = r^T (P_I - 1c^T) D_c^{-\frac{1}{2}} V D_\alpha^{-1} = \\
 r^T (D_r^{-1} F - 1c^T) D_c^{-\frac{1}{2}} V D_\alpha^{-1} = \\
 (1^T F - c^T) D_c^{-\frac{1}{2}} V D_\alpha^{-1} = \\
 (c^T - c^T) D_c^{-\frac{1}{2}} V D_\alpha^{-1} = \\
 0^T
 \end{array}
 \left|
 \begin{array}{l}
 c^T M = c^T (P_J - 1r^T) D_r^{-\frac{1}{2}} U D_\alpha^{-1} = \\
 c^T (D_c^{-1} F^T - 1r^T) D_r^{-\frac{1}{2}} U D_\alpha^{-1} = \\
 (1^T F^T - r^T) D_r^{-\frac{1}{2}} U D_\alpha^{-1} = \\
 (r^T - r^T) D_r^{-\frac{1}{2}} U D_\alpha^{-1} = \\
 0^T
 \end{array}
 \right.$$

Por lo tanto se deduce que las relaciones entre las coordenadas estandarizadas de los perfiles que componen  $N(\tilde{I})$  ( $N$ ) y las columnas ( $M$ ) son:

$$M = P_J N D_\alpha^{-1}$$

$$N = P_I M D_\alpha^{-1}$$

Teniendo en cuenta que la coordenadas principales de ambos conjuntos se denotan por  $N_\alpha = N D_\alpha$  para  $N(\tilde{I})$ , y por  $M_\alpha = M D_\alpha$  para  $N(\tilde{J})$ , se cumplen igualmente las relaciones:

$$M_\alpha = P_J N$$

$$N_\alpha = P_I M$$



. Apéndice III: Resultados de  
SAS

**RESULTADOS DEL PROGRAMA SAS AL APLICAR EL ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES  
A LOS DATOS SOBRE LA VIVIENDA DE LA ENCUESTA DE PRESUPUESTO FAMILIARES (EPF)**

(The SAS System: The Correspondence Analysis Procedure)

EPF de 1980-81		
Inertia Decomposition		
Singular Values	Principal Inertias	Percents
0,573	0,329	17,535
0,449	0,201	10,732
0,391	0,153	8,156
0,364	0,133	7,067
0,356	0,127	6,771
0,354	0,125	6,668
0,351	0,123	6,579
0,328	0,107	5,733
0,321	0,103	5,497
0,312	0,098	5,206
0,309	0,096	5,108
0,298	0,089	4,732
0,283	0,080	4,267
0,253	0,064	3,401
0,219	0,048	2,548
Sum	1,875	100

EPF de 1980-81 y 1990-91		
Inertia Decomposition		
Singular Values	Principal Inertias	Percents
0,576	0,332	17,719
0,448	0,201	10,723
0,391	0,153	8,135
0,362	0,131	6,980
0,357	0,127	6,778
0,353	0,125	6,654
0,350	0,123	6,547
0,328	0,108	5,738
0,320	0,102	5,461
0,314	0,099	5,271
0,304	0,092	4,917
0,299	0,089	4,763
0,285	0,081	4,318
0,254	0,065	3,447
0,219	0,048	2,548
Sum	1,875	100,000

EPF de 1990-91		
Inertia Decomposition		
Singular Values	Principal Inertias	Percents
0,666	0,443	23,448
0,450	0,202	10,716
0,391	0,153	8,075
0,361	0,130	6,906
0,341	0,117	6,171
0,337	0,113	5,998
0,332	0,110	5,841
0,327	0,107	5,675
0,319	0,102	5,399
0,310	0,096	5,096
0,298	0,089	4,712
0,279	0,078	4,119
0,263	0,069	3,661
0,222	0,049	2,604
0,151	0,023	1,202
0,084	0,007	0,377
Sum	1,889	100

Column Coordinates		
	Dim1	Dim2
antig0	-0,873	-0,391
antig1	-0,399	0,044
antig2	-0,060	0,473
antig3	0,290	0,463
antig4	0,716	-0,345
hig0	-2,172	-1,647
hig1	-0,020	0,415
hig2	1,164	-1,241
m2v0	-0,694	0,306
m2v1	0,095	0,350
m2v2	0,384	-0,699
m2v3	0,444	-1,281
agua0	-2,256	-2,613
agua1	-1,130	0,217
agua2	0,400	0,076
tel0	-0,633	0,022
tel1	0,615	-0,021
calf0	-0,465	0,102
calf1	0,748	-0,164
gar0	-0,182	0,177
gar1	0,931	-0,909
luz0	-2,683	-3,681
luz1	0,024	0,034

Column Coordinates		
	Dim1	Dim2
antig0	-1,028	-0,370
antig1	-0,272	0,212
antig2	0,124	0,549
antig3	0,295	0,188
antig4	0,496	-0,527
hig0	-2,801	-2,142
hig1	-0,087	0,432
hig2	0,949	-1,106
m2v0	-0,977	0,213
m2v1	-0,067	0,508
m2v2	0,396	-0,436
m2v3	0,561	-1,042
agua0	-3,104	-3,478
agua1	-1,560	0,103
agua2	0,289	0,068
tel0	-0,848	-0,010
tel1	0,465	0,005
calf0	-0,839	0,040
calf1	0,449	-0,022
gar0	-0,236	0,231
gar1	0,826	-0,808
luz0	-3,560	-4,738
luz1	0,018	0,024

Column Coordinates		Partial Contributions to Inertia for the Column Points			
	Dim1	Dim2	Dim1	Dim2	
antig0	-0,873	0,509	antig0	0,026	0,012
antig1	-0,673	0,177	antig1	0,018	0,002
antig2	-0,481	-0,481	antig2	0,017	0,024
antig3	-0,192	-0,637	antig3	0,004	0,056
antig4	0,510	0,286	antig4	0,013	0,006
hig0	-1,985	2,796	hig0	0,021	0,056
hig1	-0,235	-0,182	hig1	0,012	0,010
hig2	0,815	0,284	hig2	0,050	0,008
m2v0	-0,782	0,360	m2v0	0,016	0,005
m2v1	-0,078	-0,281	m2v1	0,001	0,012
m2v2	0,193	0,037	m2v2	0,004	0,000
m2v3	0,076	0,346	m2v3	0,000	0,008
agua0	-2,388	-2,388	agua0	0,006	0,008
agua1	-1,656	-1,656	agua1	0,041	0,057
agua2	0,077	-0,176	agua2	0,002	0,012
tel0	-0,781	-0,781	tel0	0,041	0,057
tel1	0,271	-0,123	tel1	0,017	0,005
calf0	-0,865	10,930	calf0	0,024	0,034
calf1	0,053	89,070	calf1	0,001	0,009
gar0	-0,244	-0,097	gar0	0,013	0,003
gar1	0,575	0,230	gar1	0,027	0,006
luz0	-2,081	3,425	luz0	0,002	0,007
luz1	0,004	-0,006	luz1	3,7E-06	1,4E-05
tiped0	-1,171	1,410	tiped0	0,001	0,002
tiped1	-0,699	-0,699	tiped1	0,052	0,072
tiped2	-0,462	-0,462	tiped2	0,003	0,004
tiped3	0,527	0,527	tiped3	0,049	0,068
aire0	-0,018	-0,018	aire0	9,8E-05	1,3E-04
aire1	0,786	0,088	aire1	0,004	7,5E-05
asc0	-0,435	-0,435	asc0	0,040	0,055
asc1	1,120	1,120	asc1	0,111	0,153
jar0	-0,002	-0,043	jar0	7,3E-07	0,001
jar1	0,010	0,246	jar1	4,2E-06	0,004
pisc0	-0,010	-0,006	pisc0	2,7E-05	1,7E-05
pisc1	1,119	0,752	pisc1	0,004	0,003
zdep0	-0,012	-0,004	zdep0	4,3E-05	7,7E-06
zdep1	1,311	0,472	zdep1	0,005	0,001
serv0	-0,582	0,097	serv0	0,053	0,002
serv1	0,754	-0,125	serv1	0,080	0,003
cbagua0	-1,710	-1,710	cbagua0	0,047	0,065
cbagua1	-0,193	0,488	cbagua1	2,3E-04	0,002
cbagua2	-0,213	-0,213	cbagua2	0,008	0,011
cbagua3	0,099	0,099	cbagua3	4,7E-04	0,001
cbagua4	0,871	-0,764	cbagua4	0,038	0,040
cbcalf1	-0,537	0,136	cbcalf1	0,017	0,001
cbcalf2	-0,323	-0,323	cbcalf2	0,004	0,006
cbcalf3	0,035	0,035	cbcalf3	1,5E-04	2,0E-04
cbcalf4	0,632	0,323	cbcalf4	0,019	0,007
cbcoc1	-1,187	1,082	cbcoc1	0,014	0,016
cbcoc2	-0,164	-0,164	cbcoc2	0,006	0,009
cbcoc3	1,065	1,065	cbcoc3	0,023	0,032
cbcoc4	1,240	-0,926	cbcoc4	0,060	0,046
Sum			Sum	1,000	1,000

Partial Contributions to Inertia for the Column Points		
	Dim1	Dim2
antig0	0,071	0,023
antig1	0,007	0,000
antig2	0,000	0,016
antig3	0,009	0,036
antig4	0,050	0,019
hig0	0,137	0,129
hig1	0,000	0,082
hig2	0,080	0,149
m2v0	0,042	0,013
m2v1	0,002	0,037
m2v2	0,012	0,065
m2v3	0,005	0,071
agua0	0,074	0,163
agua1	0,094	0,006
agua2	0,047	0,003
tel0	0,075	0,000
tel1	0,073	0,000
calf0	0,051	0,004
calf1	0,082	0,006
gar0	0,011	0,016
gar1	0,054	0,084
luz0	0,025	0,076
luz1	0,000	0,001
Sum	1,000	1,000

Partial Contributions to Inertia for the Column Points		
	Dim1	Dim2
antig0	0,072	0,015
antig1	0,003	0,003
antig2	0,001	0,034
antig3	0,010	0,007
antig4	0,020	0,037
hig0	0,135	0,131
hig1	0,002	0,087
hig2	0,069	0,155
m2v0	0,057	0,004
m2v1	0,001	0,069
m2v2	0,018	0,035
m2v3	0,014	0,078
agua0	0,073	0,152
agua1	0,109	0,001
agua2	0,027	0,002
tel0	0,096	0,000
tel1	0,053	0,000
calf0	0,092	0,000
calf1	0,049	0,000
gar0	0,016	0,026
gar1	0,057	0,090
luz0	0,024	0,072
luz1	0,000	0,000
Sum	1,000	1,000

## Capítulo 3

# Determinantes del alquiler de la vivienda

### **3.1 Introducción**

El objetivo de este capítulo es construir un modelo que explique los determinantes del alquiler mensual sufragado en el sector liberalizado de arrendamientos urbanos en España. Para este ejercicio se utilizan las Encuestas de Presupuestos Familiares (EPF) elaboradas por el Instituto Nacional de Estadística (INE) en 1980-81 y 1990-91. Esta base de datos tiene representatividad nacional y recoge una amplia información sobre aquellos factores que, a priori, pueden explicar el valor corriente de los servicios de vivienda.

Aunque la EPF recoge el alquiler realmente pagado por el inquilino en las viviendas en régimen de arrendamiento, es preciso recordar el impacto de la política de intervención pública en este sector que se concreta en dos tipos de actuaciones independientes: el sistema de Protección Oficial (PO) y el control de alquileres. El efecto directo de ambas políticas es la determinación de un alquiler por debajo del que imperaría en el mercado libre para viviendas similares.

La calificación legal de PO surge en España a finales de 1950 cuando el Estado comienza a desarrollar una serie de actuaciones encaminadas tanto a promocionar la construcción de nuevas viviendas (o la rehabilitación de ciertas unidades), como al establecimiento de programas asistenciales. Por su parte, el control de alquileres se remonta a 1920, año en que se congela el alquiler ya que se prohíbe legalmente su actualización y la prórroga del contrato pasa a ser forzosa. Esta situación perdura, prácticamente, hasta la Ley de Arrendamientos Urbanos (LAU) de 1964 que liberaliza el primer alquiler reflejado en todos los contratos celebrados con posterioridad a esta fecha. Aunque la LAU mantiene el carácter obligatorio de la prórroga, permite la adaptación del alquiler a través de cláusulas de actualización establecidas en el contrato. No obstante, este proceso de actualización se restringe a unos

topes máximos fijados por el gobierno a través de diversas disposiciones de Ley que surgen sin un criterio temporal definido. La liberalización total se produce con la entrada en vigor del Real Decreto-Ley de abril de 1985 sobre Medidas de Política Económica, el popularmente llamado Decreto Boyer. En los contratos celebrados con posterioridad a 1985, se deroga la prórroga forzosa además de permitir que tanto la fijación del primer alquiler como el sistema de su actualización responda al libre convenio de las partes. Así pues, la información relevante para este capítulo es el alquiler negociado entre propietario e inquilino que refleja las fuerzas del mercado en el momento en que se realiza el contrato, es decir, la información referida a los arrendamientos de calificación legal 'libre' en sentido estricto (*AL*), que son aquellos cuya contratación se efectuó con posterioridad a la LAU de 1964. En la EPF no se proporciona la fecha del contrato por lo que, a la hora de considerar el momento de la celebración del primer contrato de arrendamiento, se toma como referencia la variable 'año de ocupación'.

En este capítulo, se utiliza un modelo de regresión para explicar el alquiler mensual en el sector liberalizado de arrendamiento en función de: i) las características físicas de la vivienda, ii) sus características geográficas, que se refieren a la provincia y al tamaño del municipio donde están ubicadas, y iii) el año de ocupación. Se demuestra que el efecto de la ocupación en el alquiler se puede interpretar como una función de las cláusulas de actualización aplicadas en la renovación del contrato y la evolución de la inflación en el subsector de nueva contratación.

En relación al primer grupo de variables se proponen dos modelos alternativos: a) el método tradicional de regresión hedónica, y b) un nuevo método en el que las características físicas se sustituyen por el índice de calidad de la vivienda construido en el Capítulo II.

El método hedónico se utiliza para expresar el precio de un producto heterogéneo en función de sus características. Este procedimiento permite controlar los efectos del cambio de los atributos de un producto sobre el precio observado. Aparentemente, el primer estudio empírico que plantea esta relación se debe al economista agrario Waugh (1928). El propósito de su investigación era descubrir aquellas características físicas (color, tamaño, madurez, uniformidad,...) que causaban las subidas y bajadas de los precios diarios de venta al pormayor de ciertos productos (tomates, espárragos, pepinos,...). Posteriormente, y en un contexto distinto, Court (1939) diseña el llamado método de precios hedónicos. Ante el encargo por parte de la General Motors de un estudio para evaluar los efectos de los cambios en las características de los automóviles sobre el volumen total de ventas, Court mide la importancia relativa de estas variables para los consumidores a través de un modelo que relaciona los precios de los automóviles con sus características específicas. Sin embargo, la contribución de Court (convertir el problema de la calidad en una medida cuantitativa) no tiene una influencia inmediata en las investigaciones empíricas<sup>1</sup>. La regresión múltiple de precios hedónicos de Court fue definitivamente recuperada por Griliches (1961)<sup>2</sup> dos décadas después. Mientras que la noción de precio hedónico de Court se localiza en el lado de la demanda, Griliches lo conceptualiza como el resultado

---

<sup>1</sup>En Goodman (1998) se evalúa la aportación hedónica de Court (1939) y se conjetura sobre los motivos que lo dejan en desuso durante 20 años: falta de interés por parte de los economistas (fundamentalmente macroeconómicos) en la década de los 40 y 50, escasez de los oportunos datos de consumo temporal y falta de las herramientas de cálculo necesarias para realizar de forma adecuada este tipo de análisis.

<sup>2</sup>Este trabajo surge del encargo del *Price Statistics Review Committee* (conocido como comité Stigler) a la agencia estadística para examinar el método hedónico al considerarlo más objetivo que el tradicional método de *Bureau of Labor Statistics* a la hora de evaluar cambio de calidad en la medición de los Índices de Precios al Consumo oficiales.

de la interacción de la oferta y la demanda de las características. Se puede decir que su estudio abre el camino de la literatura moderna sobre precios hedónicos. El primer empleo del método hedónico para la estadística de precios con carácter oficial se produce en *U.S Census Bureau's 'Price Index of New One-Family Houses Sold'* en 1968<sup>3</sup>. Sin embargo su segundo empleo no se produce hasta que en 1986 el *U.S. Bureau of Economic Analysis* publica los índices de precios oficiales para ordenadores basados, en parte, en los resultados de regresiones múltiples hedónicas<sup>4</sup>. En Triplett (1990) se presenta una guía del método hedónico y se realiza un análisis de las razones por las que las oficinas estadísticas se resisten durante 30 años a su empleo, que aún sigue siendo poco frecuente. Entre estas razones destaca la argumentación mantenida durante los años 60 y principios de los 70 de que este tipo de construcciones empíricas carecía de relación con la teoría económica. Esta objeción se resuelve en Rosen (1974) donde se establecen los fundamentos microeconómicos de la regresión hedónica. Rosen proporciona un modelo de determinación del precio de un bien diferenciado e indivisible bajo condiciones de competencia perfecta donde el precio es el resultado de la interacción de la oferta y la demanda del conjunto de atributos que caracteriza a dicho bien. Aunque las funciones subyacentes de oferta y demanda de las características no pueden ser identificadas a partir de la información contenida en una regresión hedónica, la derivada parcial de esta ecuación de regresión puede ser interpretada como el precio marginal implícito de las características

---

<sup>3</sup>Este índice fue adoptado para los componentes de construcción de *National Income and Product Accounts (Survey of Current Business 54, N° 8 [August 1974] pp 18-27)*

<sup>4</sup>En la actualidad, esta metodología se utiliza con éxito en el *Bureau of Labor Statistics* para la elaboración del Índice de Precios de Consumo norteamericano de algunos tipos de bienes. Entre ellos figura la medición de la depreciación del stock de viviendas -Randolph (1988) y Moulton (1997).

prevalcientes en un mercado en equilibrio.

En Griliches (1971), Ball (1973) y Quigley (1979) se realiza una revisión de la literatura sobre el surgimiento de las funciones hedónicas en el campo de la economía urbana. Algunos trabajos empíricos sobre funciones hedónicas de la vivienda son: Linneman (1980) para las ciudades de Chicago y los Angeles, y Marks (1984) para Vancouver. Entre las aplicación empíricas realizadas en nuestro país se encuentran, por ejemplo, Peña y Ruíz-Castillo (1984) para el Área Metropolitana de Madrid y Saura (1995) para Murcia. La estimación hedónica para España que se realiza en este Capítulo es a nivel nacional.

En el modelo alternativo que se presenta, se sustituyen las características físicas de la vivienda por el Índice de Calidad construido en el capítulo anterior. El planteamiento es coherente con la hipótesis hedónica: el bien heterogéneo (en este caso la vivienda) es una agregación de características individuales. Por tanto, el precio marginal implícito de una unidad de calidad puede ser calculado como la derivada de la ecuación con respecto a la variable índice. Las ventajas encontradas en este segundo modelo frente al hedónico para los datos de 1980-81, justifican que la especificación del modelo de alquiler para 1990-91 se limite al modelo con Índice de Calidad.

El proceso de estimación de los modelos se completa con un análisis de diagnóstico que permite: i) detectar observaciones potencialmente influyentes; ii) medir su grado de influencia, es decir, sus efectos en la estimación, y iii) evaluar su significatividad como datos atípicos. Para ello se utiliza el método de Peña y Yohai (1995) aplicable en modelos lineales.

El resto del capítulo se estructura de la siguiente forma. En el segundo apartado se presenta la especificación del modelo. Representa un modelo general en relación al cual se hace explícita la interpretación del efecto de la

ocupación en el alquiler en términos de la tasa de inflación experimentada en el sector de nuevos arrendamientos y de la tasa de actualización aplicada en la renovación de contrato de arrendamiento. También en este apartado, se muestra que los datos de la EPF permiten determinar la especificación semi-logarítmica de dicho modelo. En el tercer y el cuarto apartado se estiman, respectivamente, los modelos hedónico y con índice de calidad para 1980-81, mientras que en el apartado quinto se estima el modelo con índice de calidad para 1990-91. En cada uno de los procesos de estimación se distinguen tres etapas, 1) el estudio de la forma más adecuada de incluir las variables explicativas, 2) el proceso de estimación de varios modelos previos junto con el resultado de depurar la muestra a través de la detección y eliminación de las observaciones atípicas influyentes, y 3) la interpretación de los resultados obtenidos en el modelo final. El sexto apartado se dedica a exponer las conclusiones más relevantes del capítulo.

### 3.2 Especificación del modelo

Consideremos el sector liberalizado del arrendamiento durante el período comprendido entre el año 1965 y el año  $T$ . Sea  $A_t$  el alquiler mensual, a precios corrientes del año  $t$ , de las viviendas de arrendamiento en el mercado libre ( $AL$ ) cuyo primer contrato se realiza en  $t$ . Se supone que el modelo de determinación de  $A_t$  es:

$$A_t = F(t, \chi_t, \xi_t) ,$$

donde  $\chi_t = [\chi_{1t}, \chi_{2t}]$  es el conjunto de atributos físicos ( $\chi_1$ ) y de localización ( $\chi_2$ ) que caracterizan al conjunto de viviendas de  $AL$  arrendadas por primera vez en el año  $t$ , y  $\xi_t$  es la perturbación aleatoria asociada al modelo. Se hace el supuesto de que la sucesión  $\{\xi_t\}$  es un ruido blanco, tal que la distribución

de  $\xi_t$  es normal con  $E(\xi_t) = 0$ ,  $Var(\xi_t) = \sigma_\xi^2$  y  $Cov(\xi_t, \xi_{t'}) = 0$  para todo  $t \neq t'$ .

Si se considera la misma distribución de viviendas  $\chi_t = \chi_{t'}$  en dos momentos  $t \neq t'$ , se tendrá que  $E[A_t/\chi_t] = E[A_{t'}/\chi_{t'}]$  únicamente en ausencia de inflación. En caso contrario, se tendrá que  $\partial F/\partial t \neq 0$ , lo que será achacable a un efecto inflacionista distinto de cero en el subsector de nuevos arrendamientos.

Sea  $\Delta_{t,t'}^l = \frac{E[A_t/\chi_t = \chi_l] - E[A_{t'}/\chi_{t'} = \chi_l]}{E[A_{t'}/\chi_{t'} = \chi_l]}$  la tasa media de la inflación experimentada entre el año  $t'$  y el año  $t$ , con  $1965 \leq t' < t$ , en el alquiler de nuevos arrendamientos para el conjunto de viviendas contratadas por primera vez en  $l$ . Para cada conjunto de viviendas  $\chi_l$ , es posible relacionar los alquileres esperados corrientes  $A_t$  y  $A_{t'}$  para cualquier año  $t = 1966, \dots, T$  de acuerdo con la siguiente expresión:

$$E[A_t/\chi_t = \chi_l] = E[A_{t'}/\chi_{t'} = \chi_l](1 + \Delta_{t,t'}^l), \quad (3.1)$$

es decir, en promedio, el alquiler esperado en los nuevos contratos en el año  $t$  para el conjunto de viviendas con características  $X_l$ , se puede expresar en función de su precio esperado en  $t' < t$  y el efecto medio de la inflación en el subsector de nuevos arrendamientos en el período correspondiente, que será  $(1 + \Delta_{t,t'}^l) = \prod_{s=t'}^{t-1} (1 + \Delta_{s+1,s}^l)$ , con  $t' = 1965, \dots, T - 1$ .

Sea  $\Lambda_t$  el subconjunto de tamaño  $N_t$  de las viviendas en  $AL$  ocupadas por primera vez en el año  $t$  que continúan ocupadas en el año  $T$ , y sea  $a_t$  su alquiler corriente en  $T$ , de tal forma que

$$a_t = f(t, X_t, \varepsilon_t) \quad (3.2)$$

siendo  $X_t = [X1_t, X2_t]$ , tal que  $X1_t \in \chi1_t$  y  $X2_t \in \chi2_t$ , el subconjunto de atributos que caracteriza a las viviendas en  $AL$  ocupadas desde el año  $t$  hasta el año  $T$ , y  $\varepsilon_t$  la perturbación aleatoria del modelo cuya sucesión  $\{\varepsilon_t\}$  es un

ruido blanco, de manera que la distribución de  $\varepsilon_t$  es normal con  $E(\varepsilon_t) = 0$ ,  $Var(\varepsilon_t) = \sigma_\varepsilon^2$  y  $Cov(\varepsilon_t, \varepsilon_{t'}) = 0$  para todo  $t \neq t'$ .

Se denota por  $a$  al conjunto de tamaño  $N$  de los alquileres en  $AL$  realmente pagados en  $T$ , es decir,  $a = \bigcup_{t=65}^T \{a_t\}$ , con  $N = \sum_{t=65}^T N_t$  y características  $X = [X1, X2]$  donde  $X1 = \bigcup_{t=65}^T \{X1_t\}$  y  $X2 = \bigcup_{t=65}^T \{X2_t\}$ . En la práctica, el alquiler  $a_{ti}$ , con  $i \in \Lambda_t$ , pagado en  $T$  por una vivienda  $X_{ti}$  en  $AL$  ocupada en  $t$ , es el resultado de la actualización del primer alquiler contratado ( $A_{ti}$ ) que se incrementa cada año en un porcentaje  $C_{i+1,t}^i$ , con  $t \leq l < T$ , como resultado de las cláusulas al respecto incluidas en el contrato de arrendamiento pactado en el momento  $t$ . En cuanto a la relación entre las funciones  $F(\cdot)$  y  $f(\cdot)$  que determinan los valores  $A_{ti}$  y  $a_{ti}$ , respectivamente, se tendrá que  $A_{Ti} = F(T, \chi_{ti} = X_{ti}, \xi_i) = f(t, X_{ti}, \varepsilon_i) = a_{ti}$ , si y sólo si las cláusulas de actualización aplicadas a la vivienda  $i$  entre  $t$  y  $T$  ( $C_{T,t}^i$ ) recogen exactamente el efecto inflacionista que experimenta dicha vivienda ( $\Delta_{T,t}^i$ ) para todo  $t = 1965, \dots, T$ .

Por otra parte, si se denota por  $C_{t,t'}^l = \frac{E[a_t/X_t=X_t] - E[a_{t'}/X_{t'}=X_t]}{E[a_{t'}/X_{t'}=X_t]}$  a la tasa media esperada entre  $t'$  y  $t$  ( $1965 \leq t' < t$ ) de las cláusulas de actualización aplicadas al conjunto de viviendas  $X_t$ , el alquiler medio esperado para el conjunto de viviendas  $i \in \Lambda_t$  es:

$$E[a_t] = E[A_t/\chi_t = X_t] \prod_t^{T-1} (1 + C_{t+1,t}^t) = E[A_t/\chi_t = X_t] (1 + C_{T,t}^t), \quad (3.3)$$

siendo  $(1 + C_{T,t}^t)$  el efecto medio, entre  $t$  y  $T$ , de las cláusulas de actualización en el conjunto de viviendas caracterizado por  $X_t$ .

Para analizar el efecto de la ocupación en el alquiler pagado en  $T$  para el conjunto de viviendas ocupadas en el año  $t$  ( $X_t$ ) cuando  $t \neq T$ , se define  $d_{T,t}$  como la diferencia entre el alquiler medio esperado en el caso de renovar el contrato desde el año  $t$  ( $E[a_t]$ ) y el que correspondería en el caso de acudir al

mercado de nuevos arrendamientos en el año  $T$  ( $E[A_T/\chi_T = X_t]$ ), de forma que

$$d_{T,t} = E[a_t] - E[A_T/\chi_T = X_t].$$

Según la ecuación (3.3) el valor esperado  $E[a_t] = E[A_t/\chi_t = X_t](1 + C_{T,t}^t)$  y según la ecuación (3.1),  $E[A_T/\chi_T = X_t] = E[A_t/\chi_t = X_t](1 + \Delta_{T,t}^t)$ . Por tanto,

$$d_{T,t} = E[A_t/\chi_t = X_t](C_{T,t}^t - \Delta_{T,t}^t)$$

es decir, el efecto medio esperado de ocupar una vivienda de  $AL$  en  $t$  en lugar de en  $T$ , con  $t < T$ , aunque depende del alquiler medio fijado en el primer contrato de arrendamiento en el sector de  $AL$ , se deriva de una diferencia entre la tasa de actualización media de los alquileres realmente aplicados en  $T$  de acuerdo con las cláusulas pactadas en los contratos celebrados desde el año  $t$ , y la tasa de inflación media experimentada, entre  $t$  y  $T$ , en el subsector de nuevos arrendamientos. En consecuencia, la expresión en términos relativos del efecto medio  $d_{T,t}$  para el conjunto de viviendas ocupadas en  $t$  es

$$D_{T,t} = \frac{E[a_t] - E[A_T/\chi_T = X_t]}{E[A_T/\chi_T = X_t]} = \frac{E[A_t/\chi_t = X_t](C_{T,t}^t - \Delta_{T,t}^t)}{E[A_T/\chi_T = X_t]}, \quad (3.4)$$

y, sólo en el caso en las cláusulas de actualización recojan la evolución inflación en su totalidad,  $C_{T,t}^t = \Delta_{T,t}^t$  para todo  $t$ , el efecto medio de la ocupación sería nulo, es decir, sería indiferente acudir al mercado de  $AL$  en  $T$  o renovar el contrato ya que  $E[a_t] = E[A_T/\chi_T = X_t]$ . En caso contrario, el año de ocupación será una variable explicativa del alquiler  $a$  realmente pagado en  $T$ .

Hay razones para pensar que el efecto medio de la ocupación en el alquiler sea negativo, lo que significaría que los propietarios estarían concediendo un descuento medio a aquellos inquilinos que renuevan su contrato de arrendamiento en relación al alquiler que aplicarían a un nuevo inquilino por las

mismas viviendas. Así, por ejemplo, en el trabajo de Börsch-Supan (1986) con datos de Estados Unidos y Alemania, se concluye<sup>5</sup>: '*we must...accept the discounts as a universal and independent phenomenon in rental housing markets*'. Esta ventaja para el inquilino que renova su contrato se debe al efecto denominado *tenure discount*, según el cual el propietario renuncia al ingreso adicional que le supondría el contrato con un nuevo inquilino en la medida en que este ingreso no compense los costes asociados tanto a la búsqueda de arrendatarios interesados, como a la incertidumbre sobre su comportamiento en relación al pago del alquiler y al mantenimiento de la vivienda en buenas condiciones. La hipótesis de que los propietarios conceden un *tenure discount* con objeto de mantener a los 'buenos' inquilinos se estudia también en Miron (1990), donde se presenta un modelo teórico en el que el descuento también depende del coste de traslado para el inquilino, y en Hubert (1995), donde se analizan los contratos de alquiler en un mercado con asimetría de información sobre el coste que producen los 'malos' inquilinos y la movilidad imperfecta.

En España, el trabajo realizado por Peña y Ruíz-Castillo (1984) para el Área Metropolitana de Madrid muestra evidencia empírica del efecto negativo de la ocupación en el alquiler y en referencia a  $T = 1974$ . En este capítulo se trabaja con modelos representativos del conjunto nacional para  $T = 1980-81$  y  $T = 1990 - 91$ , respectivamente.

### 3.2.1 Modelo empírico

Las dos EPF disponibles proporcionan información para una muestra de tamaño  $n < N$  sobre los cuatro aspectos siguientes: los alquileres de  $AL$ ,

---

<sup>5</sup>Este trabajo trata de diferenciar los efectos de la intervención gubernamental de los efectos de las fricciones intrínsecas al mercado de la vivienda.

$a = \bigcup_{t=65}^T \{a_t\}$ , pagados en  $T = 1980$  ( $T = 1990$ ); las características físicas de las viviendas,  $X1 = \bigcup_{t=65}^T \{X1_t\}$ ; las de localización,  $X2 = \bigcup_{t=65}^T \{X1_t\}$ , y el año en que fueron ocupadas,  $X3$ :

$$\{(a_i, X1_i, X2_i, X3_i), i = 1, \dots, n\}$$

Cuando sólo se dispone de información muestral se debe recurrir a la estimación de algún modelo, lo más sencillo posible y bajo determinados supuestos sobre el comportamiento poblacional. Una forma de estimar los determinantes del alquiler pagado en  $T$  a través de un único modelo que tenga en cuenta las posibles discrepancias entre la evolución de las cláusulas de actualización ( $C_{T,t}^t$ ) y el proceso inflacionista ( $\Delta_{T,t}^t$ ) es la siguiente. A la hora de considerar las variables explicativas del alquiler  $a$ , además de las variables  $X = [X1, X2]$  incluidas en los modelos (3.2) tal que  $X = \bigcup_{t=65}^T \{X_t\}$ , se puede incorporar un tercer conjunto de variables *dummy*,  $X3_{dummy} = [ocup65, ocup66, \dots, ocupT]$  que hagan referencia al año de ocupación, de manera que, para  $t = 1965, \dots, T$ ,  $ocup_t = 1$  si el año de ocupación es  $t$  y  $ocup_t = 0$  en otro caso. Una vez creadas las variables  $Xt = X * ocup_t$  para todo  $t$ , la expresión más general para el modelo del alquiler  $a$  que equivale a los  $T - 65 + 1$  modelos (3.2) es:

$$a = g(T, X, X65, \dots, XT, ocup65, \dots, ocupT - 1, \varepsilon), \quad (3.5)$$

donde se ha omitido la *dummy*  $ocupT$  (ocupación en  $T$ ), de forma que las variables  $Xt$  permiten medir el efecto adicional de  $X$  sobre el alquiler derivado de la ocupación en  $t$ , y se hace el supuesto de que la perturbación aleatoria  $\varepsilon$  sigue una distribución normal con media cero y varianza  $\sigma^2$ .

### 3.2.2 Determinación de la forma funcional

En el Apéndice I se determina la forma funcional de la función  $g$  de la ecuación (3.5) utilizando los datos observados en la EPF. Se comprueba el carácter semilogarítmico del modelo de alquiler libre en España, así, la relación entre la transformación logarítmica del alquiler ( $\ln a$ ) y las variables explicativas  $X = [X1, X2]$ ,  $Xt = [X1t, X2t]$  y  $X3_{dummy}$  viene dada por la expresión lineal

$$\ln a = \alpha + X\beta + \sum_{t=65}^{T-1} Xt\beta_t + \sum_{t=65}^{T-1} \gamma_t \text{ocupt} + \varepsilon. \quad (3.6)$$

Por tanto, la estimación del alquiler en  $T$  de una vivienda  $i \in \Lambda_t$  con características  $X_i = [X1_i, X2_i]$  y año de ocupación  $t = 1965, \dots, T$ , será igual a

$$\widehat{E}[a/X_i, \text{ocupt}_i = 1] = \exp(\widehat{\alpha'}) * \exp(X_i\beta) * \prod_{t=65}^{T-1} \exp([Xt]_i \widehat{\beta}_t) * \prod_{t=65}^{T-1} \exp(\widehat{\gamma}_t \text{ocupt}_i)$$

donde  $\exp(\widehat{\alpha'}) = \exp(\widehat{\alpha}) * \exp(\widehat{\sigma}^2/2)$ , y  $\widehat{\sigma}^2$  es la estimación de la varianza de la perturbación aleatoria  $\varepsilon$  del modelo (3.6). La interpretación del vector de parámetros estimados  $\widehat{\delta} = (\widehat{\beta}, \widehat{\delta}_{65}, \dots, \widehat{\delta}_{T-1})$ , donde  $\widehat{\delta}_t = (\widehat{\beta}_t, \widehat{\gamma}_t)$  hace referencia a los parámetros asociados a la *dummy ocupt* para  $t = 1965, \dots, T - 1$ , se debe hacer teniendo en cuenta el carácter, continuo o discreto, de las variables correspondientes. Así, la estimación media del cambio porcentual que se produce en el alquiler ante una variación unitaria de la  $k$ -ésima variable  $X$  viene dada por  $100 * \widehat{\delta}^k$  si la variable es continua y por  $100 * [\exp(\widehat{\delta}^k) - 1]$  si la variable es discreta.

En lo que se refiere a las variables *dummy* de ocupación, el efecto estimado en el alquiler corriente  $a$  de la ocupación en  $t$  (variable  $\text{ocupt} = 1$ , con  $t = 1965, \dots, T - 1$ ) frente a la ocupación en  $T$  ( $\text{ocup}T = 1$ ) para cada vivienda  $i \in \Lambda_t$ , con características  $[Xt]_i$ , es  $100 * [\exp([Xt]_i \widehat{\beta}_t + \widehat{\gamma}_t) - 1]$ , por lo que depende tanto de  $t$  como del vector de variables  $[Xt]_i$ . Es decir, según el

modelo (3.6) se pueden identificar dos componentes en el efecto medio de la ocupación en el alquiler: 1) un efecto medio que sólo depende del año de ocupación,  $\exp(\hat{\gamma}_t)$ , y 2) un efecto diferencial de las variables  $X$  explicativas del modelo en relación al año de ocupación de la vivienda,  $\exp(Xt\hat{\beta}_t)$ , que depende del valor alcanzado por las características de la vivienda y/o de su localización.

Una vez determinada la función  $g$  de la ecuación (3.5), resulta obvio deducir la función  $f$  de la ecuación (3.2) de forma que la estimación del alquiler medio de las viviendas ocupadas en  $t$  es:  $\hat{E}[a_t] = \exp(\hat{\alpha}') * \exp(X\hat{\beta} + Xt\hat{\beta}_t) * \exp(\hat{\gamma}_t)$  si  $t = 1965, \dots, T - 1$ , y  $\hat{E}[a_t] = \exp(\hat{\alpha}') * \exp(X\hat{\beta})$  si  $t = T$ . En consecuencia, dado que  $E(a/Xt, \text{ocupt} = 1) = E(a_t)$  para todo  $t = 1965, \dots, T$ , la estimación del alquiler en  $T$  de una vivienda  $i$  (con características  $X_i$ ) cumple la siguiente expresión en relación a las ecuaciones (3.5) y (3.2):

$$\hat{E}[a/X_i, \text{ocupt} = 1] = \hat{E}[a_t/X_i], \text{ para } t = 1965, \dots, T. \quad (3.7)$$

Las ventajas de estimar los determinantes del alquiler en  $T$  con la ecuación (3.5) en lugar de con las  $T - 65 + 1$  ecuaciones (3.2) son varias: i) la estimación se reduce a una única ecuación para obtener los mismos resultados; ii) es posible realizar contrastes individuales para cada variable  $X$  respecto a la igualdad de su efecto en distintos años de ocupación (Ho:  $\beta_t^k = \beta_{t'}^k$  para  $t \neq t'$ , siendo  $\beta_t^k$  el efecto en el alquiler de  $k$ -ésima variable  $X$  en el año  $t$ ). Nótese que el test de Chow aplicado a las correspondientes ecuaciones de (3.2) mostraría si existe evidencia estadística para la igualdad del conjunto de coeficientes estimados en dos períodos, pero no del coeficiente de cada variable; iii) la posibilidad de agrupar variables en la ecuación (3.5) puede aumentar considerablemente los grados de libertad, con lo que se mejora la precisión relativa de los parámetros estimados. En concreto, se puede observar que: i) si en la ecuación (3.2) se cumpliera que  $\partial f / \partial X_t = \partial f / \partial X_{t'}$ ,

para todo  $t \neq t'$ , entonces,  $\partial g/\partial X_t = \partial g/\partial X_{t'}$ , por lo que la ecuación (3.5) será igual a  $a = g(T, X, ocup_{65}, \dots, ocup_t, \dots, ocup_{T-1}, \varepsilon)$ , y ii) si en la ecuación (3.2) se cumpla que  $\partial f/\partial t = 0$ , es decir, que el alquiler en  $T$  sea independiente del momento de la ocupación de la vivienda, entonces, la ecuación (3.5) será  $a = g(T, X, \varepsilon)$ .

### 3.2.3 Estimación del proceso de actualización del alquiler

En este apartado se muestra como, en el caso de obtener que el año de ocupación es un determinante significativo del precio del arrendamiento, el modelo (3.6) permite estimar el *tenure discount*  $D_{T,t}$  de la ecuación (3.4), para  $t = 1965, \dots, T-1$ , lo que hace posible el análisis del proceso de actualización corriente en el mercado de *AL* de nuestro país.

Efectivamente, en primer lugar, nótese que en el momento  $T$  coincide el tipo de viviendas  $X_T = \chi_T$  por lo que  $a_T = A_T$  y, por tanto, la estimación del alquiler pagado en el mercado libre de nuevos arrendamientos por las viviendas del conjunto  $\Lambda_T$  viene determinado por  $\hat{E}[a/X, ocup_T = 1] = \hat{E}[a_T] = \hat{E}[A_T]$ .

Esta relación entre la estimación de  $a$ , condicionada a que  $t = T$ , y  $A_T$  hace posible identificar al modelo (3.6) como aquel que permite imputar a toda vivienda  $i$ , con características físicas y de localización  $X_i = [X_{1i}, X_{2i}]$ , la estimación del alquiler que se alcanzaría en el mercado a través de un nuevo contrato de arrendamiento realizado en  $T$ . Así, una vez estimados los parámetros  $\alpha$ ,  $\beta$ ,  $\beta_t$  y  $\gamma_t$  del modelo (3.6), para estimar un alquiler de mercado en  $T$  imputable a cada vivienda  $i$  ( $\hat{a}_i^{imp}$ ) se deben considerar las contribuciones estimadas ( $\hat{\beta}$ ) de las variables  $X$  al alquiler total, las características

$X_i$  y  $ocupT = 1$ , de forma que

$$\hat{a}_i^{imp} = \hat{E}[a/X_i, ocupT = 1, \hat{\alpha}', \hat{\beta}] = \hat{E}[A_T/\chi_T = X_i] \quad (3.8)$$

Así, la estimación corriente del alquiler que se debería pagar por una vivienda con características  $X_i$  en caso de acudir al mercado libre de arrendamientos en el año  $T$  es  $\hat{a}_i^{imp} = \exp(\hat{\alpha}' + X_i\hat{\beta})$ . Por su parte, la estimación del alquiler medio imputado al conjunto de viviendas ocupadas en el año  $t$  es  $\hat{E}[a_t^{imp}] = \hat{E}[a/X = X_t, ocupT = 1, \hat{\alpha}, \hat{\beta}]$ . Este valor representa una estimación objetiva del valor medio esperado en  $T$  para el conjunto de vivienda  $X_t$ ,  $\hat{E}[A_T/\chi_T = X_t]$ , mientras que  $\hat{E}[a/X, ocupt = 1] = \hat{E}[a_t]$  según la ecuación (3.7). Por lo tanto, el valor estimado del *tenure discount*  $D_{T,t}$  para todo  $t = 1965, \dots, T-1$ , es

$$\hat{D}_{T,t} = \frac{\hat{E}[a_t] - \hat{E}[a_t^{imp}]}{\hat{E}[a_t^{imp}]} = \frac{\hat{E}[a/X, ocupt = 1] - \hat{E}[a/X = X_t, ocupT = 1, \hat{\alpha}, \hat{\beta}]}{\hat{E}[a/X = X_t, ocupT = 1, \hat{\alpha}, \hat{\beta}]} \quad (3.9)$$

y, como indica la ecuación (3.4), sólo será cero si, en términos medios, las cláusulas de actualización recojen el efecto inflacionista del sector en su totalidad, en cuyo caso el año de ocupación no será un determinante del alquiler observado.

En los siguientes apartados del capítulo se procede a la estimación de los modelos de arrendamiento libre más plausibles en relación a los datos muestrales de las EPF de 1980-81 y 1990-91, respectivamente. En el proceso previo de análisis de los modelos alternativos se parte del modelo más sencillo, en el que se supone cierta la hipótesis de que no existe efecto *tenure discount*, es decir, en la ecuación (3.2) se cumpla que  $\partial f/\partial t = 0$ , para todo  $t \neq t'$ . Con el fin de contrastar esta hipótesis, posteriormente se incluyen las variables *dummy* de ocupación *ocupt*. Por último, la simplificación en el número de variables físicas que supone la utilización de la variable Índice de Calidad

construido en el Capítulo II, también permitirá estimar el modelo más general (3.5), donde se estima el efecto diferencial de las variables explicativas  $X$  en relación a la ocupación, es decir, se contrasta la hipótesis de que  $\partial f/\partial X_t = \partial f/\partial X_{t'}$  en la ecuación (3.2), para todo  $t \neq t'$ .

### 3.3 Modelo hedónico para 1980

El objetivo de este apartado es estimar el modelo (3.6) con el fin de considerar el efecto de los distintos factores que pueden explicar el alquiler de la vivienda en el sector liberalizado de arrendamiento con la información de la EPF de 1980-81. En primer lugar se describen los datos y las variables explicativas disponibles para, posteriormente, presentar el proceso de estimación que conduce a la especificación más razonable del modelo hedónico. En este proceso se aplica un procedimiento para detectar las posibles observaciones influyentes que permitirá determinar e interpretar los resultados del mejor modelo depurado de atípicos.

#### 3.3.1 Datos y variables disponibles

Las fuentes de información utilizadas en este capítulo son las EPF realizadas por el INE en 1980-81 y 1990-91 que se encuentran disponibles en la dirección de internet [www.eco.uc3m.es/epf.html](http://www.eco.uc3m.es/epf.html)<sup>6</sup>. Además, en el Capítulo I se analiza la calidad de la información referida a la vivienda y se aplican tres filtros con el objeto de mejorar su representatividad. Estos filtros afectan, respectivamente, a las variables: 1) 'régimen de tenencia', 2) 'año de construcción del edificio' y 'año de ocupación', y 3) 'alquiler'.

Así, aunque el número de viviendas en arrendamiento libre y ocupadas con

---

<sup>6</sup>Veáse documentación en Alonso-Colmenares (1994) y Arévalo (1995).

posterioridad a 1964 es de 2.181 en la EPF 1980-81, en 18 casos no se dispone de información sobre el alquiler sufragado por el inquilino y en 21 casos no se proporciona la variable 'antigüedad del edificio', variable relevante para la estimación del alquiler. Por tanto, se dispone de una muestra compuesta por 2.142 viviendas en *AL*, representativas de 867.627 viviendas al ser aplicado el correspondiente factor de elevación poblacional.

En un primer nivel exploratorio, se parte de la máxima discretización de las posibles variables explicativas con la pretensión de realizar un análisis que justifique agregaciones razonables de las mismas. En el Apéndice II.1 se presenta la lista de variables y su correspondiente discretización. Como se ha comentado anteriormente, dichas variables se pueden agrupar en los tres tipos siguientes. 1) El grupo *X1* hace referencia a siete características físicas de la vivienda: metros cuadrados construidos, instalación de servicios higiénicos, disponibilidad de garaje, teléfono, agua caliente, calefacción y antigüedad del edificio<sup>7</sup>. El coeficiente de correlación entre el alquiler y las dos únicas variables continuas del conjunto *X1*: 'número de metros cuadrados construidos' y 'antigüedad del edificio', es de 0,20 y -0,29, respectivamente. Las modalidades de estas variables que se definen en el anexo responden a su consideración por tramos. Los límites de estos tramos para la primera variable son: menos de 60 metros; entre 61 y 90; entre 91 y 130, y más de 130; y para 'antigüedad del edificio': menos de 11 años; entre 11 y 20; entre 21 y 30; entre 31 y 50, y más de 50. 2) El grupo *X2* incluye las variables de localización: las 50 provincias españolas más Ceuta y Melilla, y el tamaño del municipio medido por el número de habitantes de hecho. 3) Por último, la variable *X3* se refiere al período de ocupación de la vivienda

---

<sup>7</sup>La variable 'antigüedad del edificio' se define como:  $aconx = [1980 - (\text{año de construcción del edificio})]$ .

en régimen de *AL*. En el Apéndice II.1 se presentan, en primer lugar, las 16 variables *dummy* *ocupt* con  $t = 65, 66, \dots, 80^8$ . Como se verá más adelante, estas variables serán agregadas con el fin de simplificar el modelo sin alterar la estimación de los respectivos efectos individuales.

En la primera columna del apéndice se proporciona la distribución poblacional de frecuencias de las modalidades de cada variable. A continuación se presenta la relación entre el alquiler observado y cada una de las variables candidatas a explicar el modelo del alquiler. Para cada variable, se proporciona tanto el valor medio (columna 2) como los distintos cuartiles del alquiler (columnas 3 a 5). En la columna 6 se muestra el conjunto de agregaciones que se ha considerado en las primeras estimaciones del modelo, para lo que se ha seguido un criterio de similitud del alquiler observado. A priori, la información más objetiva que puede dar pie a considerar el efecto conjunto de, por ejemplo, dos provincias, es la similitud de los alquileres observados en las mismas.

En relación a este proceso se pueden hacer los siguientes comentarios:

1. Todas las variables  $X1$  muestran una correlación positiva con el alquiler observado. La agrupación de modalidades de las variables discretas busca la reducción del número de parámetros a estimar, teniendo en cuenta la significatividad de su efecto en los modelos estimados, lo cual será analizado más adelante.

2. Respecto a las variables  $X2$ , en una primera etapa se realizan cuatro agrupaciones del tamaño del municipio y cuatro de las 52 provincias, incluidas Ceuta y Melilla. En una segunda etapa se utilizan estas ocho agrupaciones para definir nueve variables *dummy* que permitirán estimar los efectos

---

<sup>8</sup>Tan solo se observan 11 viviendas en régimen *AL* cuya ocupación se produce en 1981, por lo que se tratan junto con las 156 viviendas ocupadas en 1980.

cruzados de ambas variables. Como se puede observar, el criterio de agregación de las variables  $X_2$  no responde a ninguna justificación económica, sino que se basa en la similitud de las distintas modalidades respecto al alquiler observado.

3. En cuanto a la variable  $X_3$ : 'número de años de ocupación de la vivienda ( $l$ )', siendo  $l = [1981 - \text{año de ocupación de la vivienda } (t)]$ , el coeficiente de correlación con el alquiler observado es de -0,41, lo que da una primera idea de la importancia de considerar esta variable en el modelo. En el apéndice se observa que existe una relación negativa con el alquiler que se hace más acusada cuando la ocupación tiene lugar a partir de mediados de los años setenta. En concreto, se observan dos posibles cambios en la relación 'alquiler-ocupación' en los años 1974 y 1977. Por ello, en el anexo se definen las siguientes seis variables que permitirán considerar la variable  $X_3$  por tramos de ocupación: i) tres variables continuas, 'acup6573', 'acup7477' y 'acup7880', que hace referencia al número de años ( $l = 1981 - t$ ) que lleva ocupada la vivienda si la ocupación se realizó entre 1965-73, 1974-77 o 1978-80, respectivamente, y ii) tres variables *dummy*, 'a6573', 'a7477' y 'a7880' con valor uno en los tramos correspondientes. Mientras que en el primer período se observa un alquiler medio de 3.116, este valor se eleva a 5.177 en el segundo y a 10.023 en el tercero. Estos datos medios pueden estar revelando la existencia de un *tenure discount*, o descuento concedido a los inquilinos que renuevan su contrato de arrendamiento, en función de  $t$ .

El análisis descriptivo de las variables independientes disponibles y su asociación con el alquiler, es insuficiente para concretar los determinantes del precio observado pero es necesario para obtener una primera aproximación razonable del modelo que se pretende estimar. En el Apéndice III.1 se plantean las formas alternativas de incluir dichas variables con el fin de

identificar aquella especificación que conduzca a la mejor estimación hedónica del modelo (3.6) para 1980. Por último, la agrupación de modalidades referenciadas en las dos últimas columnas del Apéndice II.1 indican la forma en que cada variable, discreta (columna 7) y continua (columna 8), entra en el modelo final estimado.

### 3.3.2 Proceso de estimación

Como se justifica en el Apéndice I, la variable dependiente es el logaritmo del alquiler corriente mensual,  $y = \ln(a)$ . En el proceso de estimación previo a la determinación del modelo final, se ha evitado hacer uso de los procedimientos automáticos de selección de variables que ofrecen distintos paquetes informáticos (métodos *forward*, *backward* o *stepwise*) por el riesgo de perder la relevancia teórica de los resultados. Estos métodos no son capaces de seleccionar entre las posibles modalidades de una variable de carácter cualitativo de forma que se mantenga un significado coherente. En el Apéndice III.1 se presentan y analizan los resultados obtenidos de la estimación Mínimos Cuadrados Ordinarios (MCO) con errores estándar robustos, de nueve modelos hedónicos. En la búsqueda del modelo más plausible también se realiza la etapa de diagnosis necesaria para completar el proceso de estimación. Con este objeto se plantea la depuración de las observaciones muestrales a través de la detección de observaciones anómalas que puedan estar distorsionando de forma significativa la estimación del modelo. Para ello se recurre al método de Peña y Yohai (1995) que permite detectar observaciones potencialmente influyentes en modelos lineales, mide sus efectos en la estimación de los parámetros y proporciona un test (basado en el estadístico  $t$ ) para evaluar la significatividad de las observaciones candidatas a ser atípicas.

El orden de presentación de los modelos estimados en el Apéndice III.1

responde al siguiente criterio. En primer lugar se muestra el tipo de modelo más sencillo donde sólo se estima el efecto de las variables  $X = [X1, X2]$  con varias consideraciones alternativas en su discretización, modelos 1 a 4. Posteriormente, en el modelo 5, se incluyen las variables  $X3_{dummy}$  (*ocup79*) con el objeto de estimar el *tenure discount*. Se obtiene que, salvo *ocup79*, el efecto de todas las variables *ocup79*, es significativamente negativo; muestra una tendencia negativa respecto al número de años de ocupación y un posible cambio estructural a partir de la ocupación en 1973. En relación a los resultados de los modelos anteriores se obtiene que: i) se mejora la bondad del ajuste ya que el coeficiente de determinación ajustado ( $R_{aj}^2$ ) pasa de un 0,38 del modelo anterior a un 0,51; ii) la estimación de los efectos de las variables  $X2$  apenas muestran variación, y iii) los efectos de todas las variables  $X1$  muestran una reducción sistemática, en especial las variables 'posesión de agua', 'calefacción', 'teléfono' y 'garaje', lo que está indicando la existencia de multicolinealidad entre las variables  $X1$  y  $X3$ .

A título ilustrativo, se presenta el alquiler medio (desviación típica) correspondiente a las cuatro modalidades de la variable 'agua' según los siguientes períodos de ocupación:

AGUA\OCUPACION	Entre 1965-73	Entre 1974-81
Sin agua corriente	1.940 (2.167)	1.950 (1.703)
Sólo agua fría	2.153 (1.592)	3.496 (2.698)
Agua caliente individual	3.903 (3.225)	7.967 (6.112)
Agua caliente colectiva	6.962 (4.588)	17.275 (12.372)

Al igual que ocurre con la variable 'agua', se podría comprobar que la relación del alquiler con el resto de variables  $X1$  también depende del período de ocupación.

La posibilidad de intentar un tratamiento cruzado del conjunto de variables  $X1$  y  $X3$  se realiza en un apartado posterior, donde se considera un

modelo alternativo al hedónico en el cual se sustituyen las 11 variables de características físicas del modelo 5 por un único indicador de la calidad. Así, el efecto cruzado de  $X1$  y  $X3$  se podrá tratar de una forma mucho más sencilla de lo que se podría hacerlo aquí.

**Detección de observaciones anómalas** Siguiendo con el modelo hedónico, la aplicación del método de Peña y Yohai (1995) para la depuración de observaciones anómalas permite detectar 90 atípicos influyentes en el modelo 5, lo que representa un 4,2 por ciento de la muestra disponible. Su distribución de frecuencias en relación al estadístico  $t$  es,

t-valor*	Frecuencia
$3 < /t/ \leq 4$	9
$4 < /t/ \leq 5$	31
$5 < /t/ \leq 6$	24
$6 < /t/ \leq 7$	11
$7 < /t/ \leq 8$	5
$8 < /t/ \leq 9$	4
$9 < /t/$	6
Total	90

\*/t/ valor absoluto del t-valor.

Cuando se eliminan estos atípicos en el modelo 6, se reduce la desviación típica estimada de todos los parámetros del modelo y se mejora la bondad del ajuste ya que el  $R_{aj}^2 = 0,61$ . Al igual que en el modelo 5, la estimación de los efectos de la ocupación en su máxima nivel de desagregación (dieciséis variables  $X3_{dummy}$ ) muestran una tendencia clara y sugieren algún tipo de agrupación. Así, en el Apéndice III.1 se presenta la estimación de los modelos 7 a 9 que permiten diferenciar dos tipos de efectos de la variable 'número de años de ocupación' según el período de ocupación: 1965-73 y 1974-80, respectivamente.

Seguidamente, en el Cuadro 3.1, se presentan las modalidades de las variables explicativas  $X_1$ ,  $X_2$  y  $X_3$  incluidas en el modelo hedónico más plausible hasta el momento para las viviendas de  $AL$  de la EPF de 1980-81, el modelo 9 del Apéndice III.1. En el Cuadro 3.2 se muestran los resultados de la estimación por MCO de dicho modelo antes y después de haber eliminado las observaciones anómalas. Se puede comprobar que la mejora observada con la eliminación de los atípicos en este modelo final es similar a la observada en los modelos anteriores: i) se incrementa la variabilidad explicada en relación al  $R_{aj}^2$ , en un 11,6 por ciento en este caso, ii) se estima una menor desviación típica para todos los parámetros, iii) las variables más afectadas por el sesgo en la estimación de sus efectos en el alquiler son: 'ocup71', 'ocup72', 'rloc4', 'rloc8', 'calf1' y 'gar1', y iv) el efecto de la variable garaje en el alquiler resulta significativo mientras que en los modelos sin depurar de atípicos no lo es.

### 3.3.3 Interpretación de los resultados

Como indican los valores alcanzados por el estadístico  $F$ , la consideración conjunta de las variables que intervienen en la determinación del alquiler resulta claramente significativa. En cuanto a los efectos individuales y en lo que respecta a las características físicas, en primer lugar se observa que la mayor reducción en el alquiler se produce cuando la vivienda no dispone de agua, una circunstancia que se da en el 3,85 por ciento de las viviendas de  $AL$ . Esta reducción se estima en un 49 por ciento respecto a la posesión de agua caliente central. La reducción equivalente por tener sólo agua fría es cercana al 31 por ciento, mientras que para el agua caliente individual es de un 12 por ciento. En cuanto a los servicios higiénicos, si la vivienda cuenta con menos de un baño completo (o dispone sólo de un baño completo), el alquiler estimado supondrá aproximadamente un 40 (o un 27) por ciento menos que

en la opción alternativa. Disponer de calefacción fija incrementa el alquiler en cerca de un 34 por ciento respecto a no tener ningún medio alternativo para caldear la vivienda, y en un 29 por ciento respecto a disponer sólo de aparatos móviles para el mismo fin.

El efecto positivo que supone el garaje es de un 13 por ciento y el del teléfono se aproxima al 10 por ciento. Por su parte, la superficie construida supone una elasticidad precio estimada cercana a 0,23, es decir, un incremento del 10 por ciento en los metros cuadrados implica un incremento medio de un 2,3 por ciento en el alquiler. La antigüedad del inmueble también rebaja el alquiler estimado en, aproximadamente, un 0,3 por ciento por año considerado.

En lo que respecta a la influencia de la localización, se puede decir, por ejemplo, que el alquiler mensual de una vivienda media en las ciudades de Madrid o Barcelona se estima un 144 por ciento superior al alquiler medio del mismo tipo de vivienda localizada en un municipio con menos de 10.000 habitantes en las provincias de Ávila, Soria, Albacete o Badajoz. Dicho incremento se reduciría a un 111 por ciento, si el tamaño de municipio en las provincias de Madrid o Barcelona estuviera comprendido entre 50.000 y 500.000 habitantes, o a un 87 por ciento si el municipio tuviera menos de 10.000 habitantes.

Por su parte, el efecto en el alquiler de la variable 'número de años de ocupación' es negativo y se estima un cambio estructural en el alquiler medio de la vivienda a partir de 1973. En el período de ocupación 1965-73, el descuento generado por la diferencia entre las cláusulas de actualización y la inflación en el sector, se estima en un 3,7 por ciento anual. Este descuento medio interanual se eleva a un 12,8 por ciento si la vivienda se ocupó entre 1974 y 1980-81. Los datos oficiales proporcionado por el Instituto Nacional

de Estadística (INE) respecto a la tasa de inflación interanual del alquiler (Apéndice IV.1) permiten constatar una evolución similar, incluyendo un cambio estructural en 1973, nótese que el modelo lo detecta un año después. Según el INE la tasa de variación interanual media en el alquiler es de un 5,8 por ciento en el período 1965-72 y de un 12,5 por ciento en el período 1973-80.

En el estudio elaborado por la Dirección General de Programación y Coordinación Económica sobre la situación del mercado de la vivienda en España a finales de los ochenta (MOPU, 1989, pp.79) se señala que el cambio que se observa a partir de 1973 se puede deber a dos sucesos que afectan al alquiler observado de forma combinada: 1) la gran elevación del nivel general de precios experimentada a partir de la primera crisis del petróleo en esa fecha, 2) La generalización paulatina de la inclusión de las cláusulas de actualización que se viene a completar en estos años. No obstante, y en lo que se refiere al régimen de *AL*, la similitud entre el *tenure discount* interanual estimado en el modelo y la evolución de la inflación según el INE, estaría indicando una escasa aplicación de las cláusulas de actualización en la práctica. En el Apéndice IV.2 se muestra la estimación media del *tenure discount*,  $\widehat{D}_{80,t}$  de la ecuación (3.9), para  $t = 1965, \dots, 1979$ . La estimación del mayor descuento, derivado de la ocupación en 1965 en lugar de en 1980, es de un 69,7 por ciento, mientras que la magnitud del mínimo *tenure discount*, ocupación en 1979, es de un 12,3 por ciento. El descuento medio correspondiente al período 1965-73 es de un 64,7 por ciento mientras que el equivalente al período 1974-79 es de un 34,6 por ciento.

El resultado obtenido sobre la relación negativa del *tenure discount* con el año de ocupación y la similitud del descuento interanual con la inflación, podría estar mostrado el siguiente patrón de comportamiento en el mercado

de *AL*. El alquiler medio fijado en los nuevos contratos de arrendamiento para  $t = 1965, \dots, 1980$  muestra un incremento interanual que viene a ser equivalente al proporcionado por el INE con carácter oficial<sup>9</sup>, sin embargo, en el proceso de renovación del arrendamiento apenas se hace uso de las cláusulas de actualización del precio. Así se justifica que la estimación del descuento interanual concedido al inquilino que renueva su contrato se aproxime a la tasa de inflación experimentada en el sector.

Por otra parte, tal y como se ha comentado anteriormente, hay evidencia de un efecto cruzado entre las variables físicas de la vivienda y el período de ocupación en relación al alquiler observado. En el siguiente apartado se propone un modelo en el que se sustituyen las 11 variables físicas por un indicador de la calidad de la vivienda, lo que hace factible estimar dicho efecto cruzado de una forma sencilla.

### 3.4 Modelo con índice de calidad para 1980

En este apartado se propone un modelo alternativo al hedónico para estimar los efectos de los determinantes del alquiler observado en 1980-81. Este modelo consiste en utilizar como variable explicativa el Índice de Calidad (IC) construido en el Capítulo II. La variable IC resume de forma ordinal la información disponible sobre las características físicas de la vivienda. Así, el alquiler estimado de una vivienda no se interpretará como la suma de los precios marginales de las distintas características, sino como la valoración conjunta de todas ellas.

---

<sup>9</sup>El hecho de que el modelo identifique con un año de retraso el cambio observado a partir de 1973 en la tasa de inflación del INE tendría su explicación si para la fijación anual del primer alquiler en  $t$  se tomara como referencia el último dato disponible del INE sobre la inflación interanual, inflación entre  $t - 1$  y  $t$ .

Las razones que justifican este nuevo método frente al hedónico son cinco:

1. Se evita un posible problema de multicolinealidad en la estimación de este tipo de modelos donde la información está altamente correlacionada (por ejemplo, 'metros cuadrados' con 'número de cuartos de baño').
2. Es posible estimar el efecto de variables que, debido a su reducida frecuencia, no tendría poder explicativo si se incluyeran de forma independiente en una regresión hedónica. Por ejemplo, en 1990-91 tan solo 6 de las 1026 viviendas de AL tienen piscina.
3. Facilita la estimación de los posibles efectos cruzados entre las variables explicativas que sintetiza dicho índice y otra(s) variable(s) explicativa(s).
4. Puede suponer una forma de depurar la información relevante a efectos de explicar el alquiler. Por construcción, la variable Índice de Calidad está incorrelada con el resto de indicadores obtenidos del Análisis de Correspondencias Múltiples. Aunque estos indicadores contribuyen a explicar la variabilidad total observada en el espacio de las características de la vivienda, en la medida que son independientes de la calidad de la vivienda es de esperar que no contribuyan a explicar su precio en el mercado. En caso contrario, estarían influyendo en el alquiler por razones que nada tienen que ver con dicha calidad, lo que justificaría su consideración de forma separada, y siempre factible, en el modelo<sup>10</sup>.

---

<sup>10</sup>En nuestro caso, el segundo indicador procedente del ACM que, junto con el IC explica prácticamente la totalidad de variabilidad, es 'la posesión de luz'. El porcentaje de viviendas en AL que disponen de luz en 1980-81 es el 99.62 por ciento, por lo que es una variable que, directamente, no ayuda a explicar el alquiler observado. No obstante, como se vio en el Capítulo II, la 'no posesión de luz' es la modalidad con mayor peso negativo

5. Al disminuir el volumen de datos a manejar, se facilita encontrar posibles similitudes entre distintas segmentaciones del parque residencial. Así, por ejemplo, es más fácil ordenar las distintas provincias españolas en relación al nivel medio del Índice de Calidad que en relación al conjunto combinado de las características físicas que éste representa.

### 3.4.1 Proceso de estimación

En el Apéndice III.2 se analiza la estimación de doce modelos en función del IC construido en el Capítulo II ( $X1$ ), la localización ( $X2$ ) y el año de ocupación ( $X3$ ). Como en el caso hedónico, en los primeros modelos se estima sólo el efecto del IC y las variables de localización para posteriormente estimar el efecto del *tenure discount*. Los efectos estimados de las variables de localización son similares a los del modelo hedónico y el IC tiene una incidencia significativa y positiva en la determinación del alquiler, lo que refleja que el mercado valora positivamente los niveles de servicios que dicho índice sintetiza. También, en analogía con los resultados hedónicos, la inclusión de las variables  $X3_{dummy}$  en el modelo 3 modifica significativamente la estimación del efecto del IC, no siendo así con las variables de localización. Por otra parte, la aplicación del análisis de robustez de Peña y Yohai (1995) con objeto de detectar observaciones atípicas en estos modelos, confirma la naturaleza atípica de los mismos 90 datos muestrales encontrados en el modelo hedónico.

De esta forma, salvo por la sustitución de las variables físicas de la vivienda por el IC, el modelo 7 del Apéndice III.2 es equivalente al modelo hedónico en el valor del IC, mientras que su posesión apenas incrementa su valor. De esta forma, 'la posesión de luz' es una variable que, debido al punto 2, no se incorpora en el modelo hedónico pero si estaría considerada en el modelo con IC.

9. En ambos se ha deducido la misma discretización de las variables  $X_2$  y  $X_3$  bajo el supuesto de que no existe un efecto diferencial de las variables  $X = [X_1, X_2]$  respecto a año de ocupación ( $t$ ), es decir, son estimaciones de un modelo en el que se acepta la hipótesis, referida a la expresión más general del modelo (3.5), de que  $\partial g / \partial X_t = \partial g / \partial X_{t'}$ , para todo  $t \neq t'$ . Sin embargo, se ha mostrado evidencia empírica de que esto no es cierto para las variables  $X_1$ .

Por este motivo, en los modelos 8 a 12 del Apéndice III.2 se trata de modelizar el efecto cruzado entre el IC y el año de ocupación. Para ello, el proceso se inicia con la estimación del posible efecto diferencial del IC en cada uno de años de ocupación, para posteriormente, a la vista de los resultados obtenidos, realizar las agrupaciones más aconsejables que conducen al modelo más razonable. Así, el modelo 12 se considera el más idóneo entre los analizados ya que, a diferencia del modelo 9 hedónico (o de su equivalente el modelo 7 con IC), se estima un efecto diferencial de las características físicas de la vivienda en el alquiler en dos períodos de ocupación, 1965-73 y 1974-80 recogido con las variables '*ind6573*' = IC si  $1965 \leq t \leq 1973$ , e '*ind7480*' = IC si  $1974 \leq t \leq 1980$ , respectivamente. El gran número de variables  $X_1$  implicadas hubiera complicado enormemente esta deducción en el modelo hedónico.

En definitiva, y en relación al efecto de la variable 'ocupación' en el alquiler, en el modelo con IC se constata que la revalorización estimada de la calidad de la vivienda, vía cláusulas de actualización, muestra dos ritmos distintos marcados por la frontera temporal de 1973. Este resultado implica que el supuesto asumido en el modelo hedónico referido a la ecuación (3.5) no es cierto, es decir,  $\partial g / \partial X_t \neq \partial g / \partial X_{t'}$  al menos para los períodos de ocupación  $t = 1965, \dots, 1973$  y  $t' = 1974, \dots, 1980$ .

Los resultados de la estimación del modelo 12, antes y después de eliminar las observaciones atípicas se muestran en el Cuadro 3.3. La depuración de anómalos en el modelo final se refleja en los siguientes puntos: i) el porcentaje de variabilidad explicada se incrementa en un 30 por ciento, lo que aproxima su valor al 60 por ciento; ii) se estima una menor desviación típica en todos los parámetros; iii) las variables más afectadas en la estimación de sus efectos sobre el alquiler son: 'ind6573', 'rloc4', 'rloc7', 'rloc8' y 'a6573', y iv) esta última variable *dummy* resulta significativamente distinta de cero tras la eliminación de los atípicos.

### 3.4.2 Interpretación de los resultados

Se puede comprobar la gran similitud del modelo del cuadro 3.3 con el correspondiente al modelo hedónico en cuanto a la bondad del ajuste, aún a pesar de haber reducido el número de parámetros estimados de 22 a 13. En ambos tipos de modelos, hedónico y con IC, los efectos estimados de las variables comunes de localización tampoco son significativamente distintos, y aunque los coeficientes estimados muestran ligeras variaciones, sus magnitudes relativas no se alteran. La forma en que se considera el IC en el segundo modelo permite estimar un efecto cruzado 'calidad-años de ocupación' en el alquiler observado. Así, se estima que la mejora en el nivel de calidad produce un incremento porcentual del alquiler superior si la ocupación se realiza con posterioridad a 1973. En concreto, la diferencia entre ambos efectos marginales se estima en 13 puntos porcentuales. Por tanto, a diferencia del modelo hedónico que no incorpora este efecto, el modelo con IC recoge un efecto interesante: el incremento porcentual en el alquiler debido a la ocupación post73 frente a pre73 es mayor en las viviendas que disponen de un mayor nivel de calidad.

Para ilustrar esta afirmación se propone el siguiente ejemplo. Sean  $\bar{a}_{\min,65}$ ,  $\bar{a}_{\max,65}$ ,  $\bar{a}_{\min,80}$ , y  $\bar{a}_{\max,80}$ , los respectivos alquileres medios de cuatro tipos de viviendas caracterizadas por dos tipos de calidad ( $IC_{\min} = 2,07$  e  $IC_{\max} = 2,74$ , que corresponden al primer y tercer cuartil del IC observado en las viviendas de *AL*) y dos años de ocupación distintos ( $t = 1965$  y  $t = 1980$ , es decir,  $acup6473 = 16$  y  $acup7480 = 1$  respectivamente). La estimación de  $\bar{a}_{\min,65}$  es  $\hat{a}_{\min,65} = \hat{E}(a/IC = 2,07, a6573 = 1, acup6473 = 16)$ , es decir, la estimación del alquiler esperado en 1980-81 del tipo de viviendas de nivel de calidad mínimo y ocupado en 1965. Según el modelo con IC, este valor es

$$\hat{a}_{\min,65} = \tilde{A} * \exp(0,700 * 2,07 + (-0,458 - 0,038 * 16)) = 1,472 * \tilde{A}$$

siendo  $\tilde{A}$  el valor medio estimado del alquiler controlado por las variables  $X_2$  de localización. En la siguiente tabla se presenta el alquiler estimado de los cuatro tipos de viviendas ( $\hat{a}_{\min,65}$ ,  $\hat{a}_{\max,65}$ ,  $\hat{a}_{\min,80}$ , y  $\hat{a}_{\max,80}$ ) según dos modelos con IC del Apéndice III.2: el modelo 7 donde, al igual que en el modelo hedónico, no se estima el efecto conjunto entre calidad y ocupación, y el modelo 12 en el que este efecto si se estima<sup>11</sup>. La columna 3 de la tabla muestra la estimación de la variación porcentual que supone la ocupación en 1965 respecto a la ocupación en 1980 (*tenure discount*,  $\hat{D}_{80,65}$ , de la ecuación (3.9)) para cada nivel de calidad. Las filas 3 y 6 muestran las respectivas variaciones porcentuales en el alquiler estimado derivadas del aumento de la

<sup>11</sup>Se considera el modelo 7 con IC, en lugar del hedónico, por la mayor facilidad para calcular el alquiler estimado condicionado a un nivel de calidad prefijado. No obstante, se puede comprobar que el resultado que se pretende ilustrar es idéntico en el modelo hedónico.

calidad  $Var(IC)$ .

Modelo 7	$t = 1965$	$t = 1980$	$\widehat{D}_{80,65}$
$IC_{\min}$	$\widehat{a}_{\min,65} = 1,256 * \widetilde{A}$	$\widehat{a}_{\min,80} = 4,340 * \widetilde{A}$	-71%
$IC_{\max}$	$\widehat{a}_{\max,65} = 2,110 * \widetilde{A}$	$\widehat{a}_{\max,80} = 7,289 * \widetilde{A}$	-71%
$Var(IC)$	680%	680%	
Modelo 12			
$IC_{\min}$	$\widehat{a}_{\min,65} = 1,472 * \widetilde{A}$	$\widehat{a}_{\min,80} = 4,888 * \widetilde{A}$	-70%
$IC_{\max}$	$\widehat{a}_{\max,65} = 2,352 * \widetilde{A}$	$\widehat{a}_{\max,80} = 8,524 * \widetilde{A}$	-72%
$Var(IC)$	587%	744%	

Como se puede observar, este ejemplo permite interpretar un efecto del *tenure discount* adicional al recogido con las variables  $X3$ : 'a6573', 'acup6573' y 'acup7480' en el modelo 12 con IC. Así, cuando se estima un único efecto marginal de la calidad (modelo 7 equivalente al modelo hedónico) la variación en el precio debida a la ocupación no depende del nivel de calidad. Sin embargo, cuando la revalorización de la calidad depende del período de ocupación (modelo 12) el *tenure discount* también depende del nivel de los servicios instalados en la vivienda. Así pues, además de la diferencia entre la evolución de la inflación y su reflejo en las cláusulas de actualización estimada con las variables  $X3$ , al incluir las variables 'ind6573' e 'ind7480' en el modelo 12 se detectan tres efectos del IC:

1. Tanto en el período de ocupación 1965-73 como 1974-80, se observa que el alquiler es superior para la vivienda de mayor calidad.
2. El incremento de los alquileres atribuible a la mejora de la calidad es inferior en el primer período que en el segundo.
3. A medida que la calidad de la vivienda es mayor el *tenure discount* es mayor, es decir, hay mayor diferencia entre el precio del mercado de nuevos arrendamiento y el que procede de la renovación del contrato. Esto es debido a que el incremento en el alquiler debido a la ocupación posterior a 1973 aumenta, Posiblemente esta evolución en los precios este influida por

dos factores: 1) la evolución de la inflación interanual del alquiler que, según los datos oficiales del INE comentados anteriormente, se incrementa considerablemente a partir de 1973, y 2) una escasa aplicación de las cláusulas de actualización de los alquileres en la renovación de los contratos de arrendamiento, que por otra parte están sometidos a prorroga forzosa. En estas condiciones, el descuento a favor del inquilino (*tenure discount*) será sensiblemente superior en las viviendas de mayor calidad donde la justificación de un incremento interanual del precio vía mercado es mayor.

En el modelo hedónico no se aprecia más que el primero de los tres efectos: la mejora en cada de las características de la vivienda conlleva un mayor alquiler. Por otra parte, la estimación del descuento interanual en el alquiler para el inquilino que renueva su contrato de arrendamiento es similar en ambos modelos. Así, en el modelo con IC se estima un descuento de un 3,8 por ciento anual si la ocupación tiene lugar entre 1965-73 y un 13,2 por ciento si es posterior. Se recuerda que los respectivos descuentos en el modelo hedónico son de 3,7 y 12,8 mientras que la inflación media interanual es de un 5,82 entre 1965-72 y de un 12,46 por ciento entre 1973-80, según el INE.

Como se ha comentado anteriormente, no se constata una utilización generalizada de las cláusulas de actualización por lo que el incremento del alquiler pagado en  $T$  cuando  $t$  aumenta, y en consecuencia la relación negativa del *tenure discount* y  $t$ , se puede deber a dos razones: 1) la evolución creciente del alquiler medio fijado en los nuevos arrendamientos derivada de un efecto inflacionista, y 2) la mejora de la calidad media de la vivienda en  $t$ , así, por ejemplo, el IC medio (desviación típica) en 1965 es de 2,2 (0,52) mientras que en 1980 es de 2,5 (0,44). Más adelante se comprobará esta misma tendencia en relación a la calidad de la vivienda de la EPF de 1990-91 por lo que, se debe hacer notar la dificultad de aproximar los niveles de los servicios de

la vivienda en el caso de no disponer de las variables IC construidas en el Capítulo II.

Por último, en el Apéndice IV.2 se muestra la estimación de los descuentos sobre el alquiler de mercado en 1980-81 (estimación del *tenure discount*  $D_{80,t}$  de la ecuación (3.9)), para  $t = 1965, \dots, 1979$ , según el modelo hedónico (columna 1) y según el modelo con IC para el nivel de calidad medio  $\overline{IC} = 2,38$  (columna 2). Aunque estos descuentos son ligeramente superiores en el segundo modelo para todo  $t$ , su valor es similar. Así, el descuento medio estimado en el período 1965-73 según el modelo con IC es de un 66,2 por ciento, frente a un 64,7 del hedónico, y de un 35,3 por ciento, frente a un 34,6 en el período 1974-79.

A modo de síntesis se puede decir que hay dos razones que justifican la superioridad del modelo con IC frente al hedónico: 1) con el primero se reduce considerablemente el número de parámetros estimados sin perjuicio de la bondad de ajuste del modelo, y 2) ha sido factible la estimación del efecto cruzado entre el año de ocupación y el conjunto de características de la vivienda de una forma sencilla. Así, el modelo con IC ha permitido demostrar que la hipótesis  $\partial g/\partial X_t = \partial g/\partial X_{t'}$  para  $t \neq t'$  referente a la ecuación (3.5) es cierta para los tramos de ocupación 1965-1973 y 1974-1980. Por tanto, además de detectar el denominado '*tenure discount*' recogido con las variables  $X_3$ , en el modelo con IC se ha identificado un efecto diferencial de la variable IC que depende del período en que se ocupó la vivienda.

A la vista del análisis comparativo de los modelos hedónico y con IC para 1980-81, a partir de ahora y en lo referente a la EPF de 1990-91 se procede a exponer únicamente el modelo con IC. Es de esperar que en 1990-91 las ventajas sean aún mayores puesto que el número de variables que el IC sintetiza es mayor que en 1980-81, y por tanto, se potencian las razones

expuestas sobre la superioridad de este modelo frente al hedónico.

### **3.5 Modelo con índice de calidad para 1990**

Como en los modelos para 1980-81, la presentación del análisis realizado sobre el modelo de alquiler en *AL* para 1990-91 se inicia con una descripción de los datos disponibles. Posteriormente se presenta el proceso de estimación previo a la determinación del modelo final y la interpretación de los resultados obtenidos.

#### **3.5.1 Datos y variables disponibles**

La EPF de 1990-91 proporciona una muestra de 1.061 de hogares cuyas viviendas corresponden al régimen de *AL*. De esta muestra inicial se eliminan 26 hogares: 19 que no proporcionan información sobre el alquiler y 7 que desconocen el año de construcción. Así, el número de observaciones muestrales disponibles es de 1.035, representativas de 590.948 viviendas al aplicar el factor de elevación poblacional proporcionado por el INE.

En lo que respecta a las posibles variables explicativas del alquiler, tanto la localización ( $X_2$ ) como el año de ocupación ( $X_3$ ) presentan el mismo grado de detalle que en la EPF de 1980-81, mientras que la información sobre las variables  $X_1$  es bastante más completa. Así, además de contar con las ocho características físicas de la muestra anterior (metros cuadrados construidos, instalación de servicios higiénicos, disponibilidad de garaje, luz, teléfono, agua caliente, calefacción, antigüedad del edificio), en la EPF de 1990-91 se dispone de diez variables adicionales: tipo de edificio, aire acondicionado, ascensor, jardín, piscina, zona deportiva, otros servicios comunitarios, tipo de energía utilizada para calentar el agua, para la calefacción y para cocinar.

En el Capítulo II se construye el IC que sintetiza la calidad del conjunto de viviendas ocupadas en 1990-91, utilizando la totalidad de estas variables.

Al igual que se hizo con la base de datos de 1980-81, en el Apéndice II.2 se presenta la relación entre el alquiler observado en 1990-91 y cada una de las variables disponibles. El grado de discretización de las variables  $X1$  corresponde a la división de modalidades utilizadas para construir el IC. Por su parte, la discretización de las variables  $X2$  y  $X3$  es la máxima posible y su agrupación (columna 6) responde a la similitud de los alquileres observados. Los comentarios que surgen de este análisis descriptivo se pueden concretar en los siguientes puntos:

1. Todas las variables  $X1$ , salvo 'servicios higiénicos' y 'calefacción', muestran una correlación positiva con el alquiler observado. Las excepciones que presentan estas dos variables son las siguientes: la modalidad 'no disponer de servicios higiénicos' supone un alquiler medio (10.250 pesetas) superior al de 'servicios higiénicos compartidos con otras viviendas' (3.430 pesetas) y 'uno o dos aseos, o uno o dos cuartos retrete' (6.935 pesetas). Como se verá posteriormente, 3 de las 22 observaciones de la modalidad 'no disponer de servicios higiénicos' son datos atípicos en el modelo estimado. El alquiler medio de estas 3 observaciones es de 23.175 pesetas y el del resto de viviendas de esa modalidad, tras eliminar los atípicos, es de 5.041. Por otra parte, las modalidades 'no disponer de calefacción' y 'disponer de sólo aparatos móviles' tienen un alquiler medio de 18.163 y 13.338 pesetas, respectivamente. Esta diferencia no es muy relevante y menos aún si se tiene en cuenta que los aparatos móviles no tienen por qué formar parte del alquiler de la vivienda.

2. En cuanto a la localización de la vivienda (variables  $X2$ ), en primer lugar se realizan cuatro agrupaciones tanto de las 52 provincias como del

tamaño del municipio. Como era de esperar, esta última variable muestra una relación positiva con el alquiler. En segundo lugar se definen cinco variables *dummy* que hacen referencia al efecto cruzado 'provincia-tamaño de municipio' que responde a la similitud de los alquileres observados.

3. Respecto a la variable  $X_3$ , se observa claramente una tendencia negativa entre el alquiler medio observado y el 'número de años de ocupación de la vivienda', con un coeficiente de correlación entre ambas variables de -0,45. Sin embargo, tanto el alquiler medio como los cuartiles analizados no muestran una tendencia constante en todo el período de ocupación comprendido entre 1965 y 1990<sup>12</sup>. Así, se observa un crecimiento sistemático del alquiler en tres subperíodos: 1965-75, 1976-82 y 1983-90 cuyos alquileres medios (desviación típica) son 6.796 (15.017), 17.404 (18.997) y 23.575 (15.017) respectivamente. En cada uno de estos subperíodos se encuentran las siguientes particularidades:

- (a) Entre 1965 y 1975 el alquiler medio anual experimenta un incremento de un 148 por ciento (pasando de 3.288 a 8.170 pesetas). En este período destaca una gran estabilidad entre 1968 y 1973 en torno a un nivel medio de 6.471.
- (b) Entre 1976 y 1982 se produce un incremento más atenuado (un 29 por ciento) al pasar de un alquiler medio de 11.893 a 15.383. Sin embargo, se observa que: i) en el subperíodo 1976 y 1979, con un alquiler medio de 19.253, se produce un incremento sistemático que eleva el alquiler medio en un 146 por ciento en cuatro años, ii) entre 1980 y 1982 se observa un freno al crecimiento del subperíodo anterior acompañado

---

<sup>12</sup>Únicamente se observan 3 viviendas ocupadas en 1991 que se consideran junto con las 82 ocupadas en 1990.

de una reducción del nivel alcanzado: el alquiler medio en estos tres años es de 14.876.

- (c) En los años comprendidos entre 1983-90, entre los que destaca 1990 por ser el año con el menor alquiler medio (20.340 pesetas), también se observa cierta estabilidad caracterizada ahora por un alquiler medio aproximado de 23.575 pesetas mensuales. Por tanto, el crecimiento medio respecto al período anterior (1976-82) es de un 35 por ciento y de un 61 por ciento respecto a 1965-75.

Este análisis descriptivo puede orientar una primera aproximación al comportamiento no lineal del efecto del año de ocupación en el alquiler.

En el siguiente apartado se presenta el proceso de estimación de una serie de modelos previos que conduce a la determinación del modelo más convincente de los analizados. Las variables discretas y continuas del mismo se presentan, respectivamente, en las columnas 7 y 8 del Apéndice II.2.

### 3.5.2 Proceso de estimación

Los resultados obtenidos en el proceso de estimación de los modelos con IC para 1990-91 se presentan en el Apéndice III.3. La definición de las variables utilizadas se encuentra en el Anexo I.2. En todos los modelos se aplica el factor de elevación poblacional y se utiliza el método de MCO con errores estándar robustos. El orden de los modelos responde a un proceso secuencial similar al realizado para los datos de la EPF de 1980-81: 1) primero se estiman únicamente los efectos del IC y las variables de localización, esto es, se estima el modelo (3.5) bajo el supuesto de que la ocupación no tiene efecto en el alquiler ; 2) posteriormente, se añade la estimación del efecto medio del *tenure discount* sólo en función del año de ocupación, es decir, bajo el

supuesto de que  $\partial g/\partial X_t = \partial g/\partial X_{t'}$  para  $t \neq t'$ , siendo  $X_t = [X_{1t}, X_{2t}]$ , y 3) finalmente se relaja el supuesto anterior, con lo que se estimará el efecto diferencial del *tenure discount* en relación a las variables explicativas  $X = [X_1, X_2]$  que proceda.

En la primera etapa, al igual que en 1980-81, se vuelve a revelar la capacidad sintética del IC a la hora de explicar el alquiler en el régimen de *AL* y la conveniencia de considerar conjuntamente las variables  $X_2$ : tamaño del municipio y provincia. Cuando se incluye el año de ocupación (variables  $X_3$ ) en el modelo se obtienen los siguientes resultados: i) una reducción significativa del efecto marginal estimado del IC sin que los efectos de las variables  $X_2$  se vean sensiblemente alterados, y ii) un incremento considerable de la capacidad explicativa del modelo, pues el valor del  $R_{aj}^2$  pasa de 0,27 a 0,51. Por tanto, al igual que en el modelo para 1980-81, se constata la importancia de la ocupación, y su correlación con la calidad de la vivienda, a la hora de explicar el alquiler observado.

En esta segunda etapa de estimación, aún sin efectos cruzados entre  $X_1$  y  $X_3$ , la eliminación de 53 observaciones atípicas detectadas con el método de Peña y Yohai (1995) hace que la variabilidad explicada por el modelo se eleve a un 60 por ciento. Dichos atípicos representan un 5,2 por ciento de las observaciones muestrales disponibles y su distribución de frecuencias en

relación al estadístico  $t$  es,

t-valor*	Frecuencia
$3 < t \leq 4$	18
$4 < t \leq 5$	15
$5 < t \leq 6$	9
$6 < t \leq 7$	4
$7 < t \leq 8$	2
$8 < t \leq 9$	4
$9 < t$	1
Total	53

\* $t$ / valor absoluto del t-valor.

La depuración del modelo permite corregir los sesgos en la estimación que, como se analiza en el Apéndice III.3, en 1990-91 afectaban de forma significativa a los efectos de la ocupación en el alquiler.

Ya en la tercera etapa, el análisis llevado a cabo para deducir la mejor forma de incluir las variables  $X3$ , junto con la estimación de los efectos cruzados del año de ocupación ( $t$ ) con el IC, permite identificar tres períodos que determinan la influencia de  $t$  en el alquiler observado en 1990-91. Así, al igual que en 1980-81, en el modelo final se estiman dos componentes del *tenure discount*: 1) un efecto medio que depende del año de ocupación con un comportamiento distinto en los períodos: 1965-75, 1976-82 y 1983-90, respectivamente, y 2) un efecto diferencial del IC en cada período. En el Cuadro 3.4 se muestra la definición de las variables que forman parte del modelo y en el Cuadro 3.5 se presenta los resultados de la estimación, antes y después de la eliminación de las observaciones atípicas.

Como se puede observar, la eliminación de los atípicos produce los siguientes resultados en el modelo final: i) el porcentaje de variabilidad explicada por el modelo se incrementa en nueve unidades porcentuales, ii) la desviación típica estimada es menor en todos los parámetros, iii) los efectos estimados

de las variables de localización se incrementan de forma que mantienen su magnitud relativa, y iv) el *tenure discount* es el efecto estimado más afectado, en particular para la ocupación en el período 1983-90.

### 3.5.3 Interpretación de los resultados

El modelo estimado para 1990-91 explica algo más del 60 por ciento de la variabilidad observada en los alquileres de las viviendas en *AL*, y el valor alcanzado por el estadístico *F* muestra la significatividad conjunta de las variables consideradas. En cuanto a los determinantes del alquiler se pueden hacer los siguientes comentarios.

La localización de la vivienda tiene un efecto que depende tanto de la provincia como del tamaño del municipio, lo que permite los dos tipos de comparaciones. Así, en la provincia de Madrid, por ejemplo, el alquiler esperado de una vivienda se incrementa en un 17,3 por ciento si el municipio tiene más de 50.000 habitantes según el modelo de 1990-91. Por otra parte, si se considera como referencia el alquiler medio en municipios con menos de 10.000 habitantes de, por ejemplo, Ceuta, la localización en Sevilla o Valencia incrementa el alquiler medio en un 46,7 por ciento mientras que si la provincia es Madrid el incremento estimado es de un 83,4 por ciento.

De forma similar al modelo de la década anterior, existe un descuento significativo concedido por los propietarios a los inquilinos según el año de ocupación. Esta ventaja para los antiguos inquilinos se debe a dos componentes interrelacionados del denominado '*tenure discount*': un efecto medio del año de ocupación y un efecto diferencial del nivel de calidad que depende del período temporal en que se celebró el primer contrato de arrendamiento.

El efecto medio de cada año de ocupación se refleja tanto en un efecto marginal distinto como en un efecto fijo que afecta al alquiler medio estimado

en los períodos: 1965-75, 1976-82 y 1983-90. En cuanto al primero tipo de efectos, se observa que los inquilinos que han renovado su contrato de arrendamiento desde antes de 1976 obtienen un descuento estimado de un 7 por ciento por año de ocupación, un 4 por ciento si el período es entre 1976-82, y un 2,6 por ciento entre 1983-90. Como se puede observar, el orden de magnitud de estos efectos interanuales, a diferencia de lo que ocurría en 1980-81<sup>13</sup>, se reduce en el tiempo.

Si se analizan los datos oficiales del INE sobre la evolución de la inflación interanual del alquiler (Apéndice IV.1) se puede observar un comportamiento homogéneo en tres períodos: entre 1965-72 el incremento medio interanual del alquiler es de 5,82 por ciento; entre 1973-1982 es de un 12,51 por ciento, y entre 1983-90 este incremento se reduce a un 7,21 por ciento. Por tanto, a la vista de los efectos de la ocupación estimados con las dos EPF, se obtiene que el primer cambio estructural es recogido por ambos modelos, con un año de retraso en el modelo para 1980-81 y con tres años de retraso en el modelo para 1990-91, mientras que el cambio producido en 1983 es recogido en este mismo año por el segundo de los modelos. En cuanto a los efectos interanuales, en el modelo de 1990-91, a diferencia del de 1980-81, se refleja un uso más generalizado de las cláusulas de actualización del alquiler, lo que permite interpretar con mayor nitidez el *tenure discount* como el resultado de un descuento concedido por el propietario con objeto de retener a los buenos inquilinos, de tal forma que, se estima un mayor descuento interanual a los inquilinos más antiguos. Por su parte, los incrementos marginales en el alquiler derivado del nivel de calidad de la vivienda se estiman en: un 11,7 por ciento en el período 1965-75; un 9,4 por ciento en 1976-82, y un 6,5 por

---

<sup>13</sup>Se recuerda que en el modelo de 1980-81 se obtenía un descuento interanual aproximado de un 4 por ciento en 1965-73 y cercano al 13 por ciento en 1974-80.

ciento en 1983-90.

Por consiguiente, en el modelo para 1990-91 se han estimado dos factores derivados de la ocupación que afectan conjuntamente al nivel medio del alquiler: 1) la duración del arrendamiento, de forma que el efecto marginal estimado es mayor para los períodos de ocupación más alejados, y 2) la calidad de la vivienda, de forma que a medida que se consideran mayores niveles de calidad, el descuento debido al período de ocupación se reduce (esto es debido a que los efectos marginales son mayores para la ocupación anterior).

Para ilustrar el orden de magnitud de este efecto conjunto se presenta un ejemplo donde se estima la variación en el alquiler medio cuando la ocupación es anterior a 1990. Para ello se consideran los siguientes años de ocupación: 1966, 1976, y 1986. En la tabla siguiente se presenta la estimación del efecto porcentual medio sobre el alquiler derivado de ocupar una vivienda en  $t$  frente al que correspondería si la ocupación se realiza en  $T = 1990$ , estimación del *tenure discount*  $D_{90,t}$  de la ecuación (3.9). Como ya se ha comentado, en el modelo estimado este efecto depende de la calidad de la vivienda por lo que en este ejemplo se han fijado dos niveles del IC ( $IC_{\min} = 12,59$  y  $IC_{\max} = 17,81$ ) que corresponden al primer y tercer cuartil del IC observado en las viviendas de *AL*. En la primera columna de la tabla se muestran, respectivamente, el alquiler medio estimado según el modelo presentado para un nivel de calidad  $IC$ , donde  $\tilde{A}$  es el valor medio estimado del alquiler controlado por las variables  $X2$  de localización. En las columnas 2 y 3 se muestran los correspondientes efectos  $\widehat{D}_{90,t}$  expresados en términos porcentuales cuando el nivel de calidad observado es  $IC_{\min}$ , y  $IC_{\max}$ , respectivamente.

Año $t$	Alquiler medio estimado para $IC$	$\widehat{D}_{90,t}^{IC_{\min}}$	$\widehat{D}_{90,t}^{IC_{\max}}$
1966	$\tilde{A} \exp[0,117 * IC - 0,747 - 0,070 * 25]$	-84,57	-78,38
1976	$\tilde{A} \exp[0,094 * IC - 0,631 - 0,040 * 15]$	-56,40	-49,19
1986	$\tilde{A} \exp[0,065 * IC - 0,026 * 5]$	-9,75	-9,75

Como se puede observar, el *tenure discount* es considerablemente mayor cuando aumenta el número de años que lleva ocupada la vivienda<sup>14</sup>. Además, el descuento en el alquiler para los antiguos inquilinos es mayor en las viviendas de peor calidad, es decir, la revalorización del precio vía cláusulas de actualización es mayor cuando la calidad de la vivienda también es mayor. Este resultado, aunque es contrario al obtenido en 1980, es lógico en cuanto que la demanda de viviendas de mayor calidad aumenta en el tiempo. Para mostrar la constatación de esta tendencia, en la siguiente tabla se presenta el número de viviendas en *AL* y la calidad media (desviación típica) del índice de calidad según el período de ocupación: Pre80 (con año de ocupación  $t = 1965, \dots, 1980$ ) y Post80 (con  $t = 1981, \dots, 1990$ ), respectivamente. Como medida de la calidad se utiliza el índice de calidad I8090 construido en el Capítulo II que permite comparar la calidad de las viviendas de la EPF de 1980-81 y de 1990-91 con un criterio homogéneo.

Número y nivel de calidad medio ( $\overline{I8090}$ ) para las viviendas en <i>AL</i>						
	Pre80			Post80		
	N	$\overline{I8090}$	(d.t)	N	$\overline{I8090}$	(d.t)
EPF 1980-81	867.627	-0,364	(0,581)	-	-	-
EPF 1990-91	276.402	0,041	(0,460)	314.546	0,063	(0,429)

Nótese, que las viviendas Pre80 de la EPF de 1990-91, corresponden al subconjunto de las viviendas de la EPF de 1980-81 que renuevan su contrato de arrendamiento durante diez años más. Representan el 32 por ciento del total de las viviendas en *AL* ocupadas en 1980-81 y el 47 por ciento de las ocupadas en 1990-91. En la tabla se comprueba que: 1) las viviendas de más

<sup>14</sup>En el Cuadro 3 del Apéndice IV.2 se presenta la estimación del *tenure discount*,  $D_{90,t}$  con  $t = 1965, \dots, 1989$ , para el nivel de calidad medio observado  $\overline{IC} = 15,46$ . Para este nivel de calidad se estima un descuento que va desde un 2,53 por ciento, si  $t = 1989$ , a un 82,18 por ciento si  $t = 1965$ .

reciente ocupación disponen de un mayor nivel de calidad medio, y 2) que el subconjunto de viviendas Pre80 que se mantiene ocupado en 1990-91 posee una calidad media superior a la del total de viviendas en *AL* ocupadas en 1980-81. En definitiva, estas viviendas son las de mejor calidad respecto a 1980-81 pero las de peor calidad respecto a 1990-91, en consecuencia son las viviendas que proporcionan un mayor *tenure discount* en ambos periodos.

Por otra parte, es posible estimar el descuento medio que se obtiene por estas viviendas en función del año de ocupación,  $t < 1980$ , respecto al alquiler que debieran pagar si la ocupación hubiera sido en 1980 – 81, en dos situaciones distintas: 1) en relación al precio de mercado de 1980-81 (a través de los modelos hedónico y con IC estimados con la EPF de 1980-81), y 2) en relación al precio de mercado de 1990-91 (modelo de la EPF de 1990-91). En el Cuadro 2 del Apéndice IV.2 se proporcionan estas variaciones estimadas en el precio medio para cada  $t = 1965, \dots, 1979$ <sup>15</sup>. En las columna 1 y 2 se muestra el *tenure discount* estimado en la primera situación según los modelos hedónico y con IC, respectivamente, mientras que en la columna 3 se muestra dicho descuento según el modelo de 1990-91. Como era de esperar, las ventajas para los inquilinos del efecto *tenure discount* debidas a la ocupación anterior a 1980 son mayores en 1980-81 que diez años después.

### 3.6 Conclusiones

En este capítulo se han construido los modelos que permiten estimar los determinantes del alquiler de mercado en el régimen de arrendamiento libre (*AL*) en España (ocupación posterior a la Ley de Arrendamientos Urbanos

---

<sup>15</sup>Estos *tenure discount* se calculan para el nivel de calidad medio correspondiente al respectivo grupo de viviendas de referencia.

de 1964) a precios corrientes de 1980 y 1990, respectivamente. La base de datos utilizada es la Encuesta de Presupuestos Familiares (EPF) realizadas por el Instituto Nacional de Estadística (INE) en 1980-81 y 1990-91.

El estudio se inicia con un análisis descriptivo y gráfico sobre las correlaciones entre las variables independientes disponibles y la variable alquiler que permite especificar un modelo lineal donde la variable dependiente es la transformación logarítmica del alquiler mensual, y las variables explicativas son de tres tipos: 1) características físicas de la vivienda, 2) variables de localización y 3) número de años que la vivienda lleva ocupada, una variable cuyo efecto en el alquiler, *tenure discount*, refleja una diferencia entre las cláusulas de actualización y la evolución de la inflación en el subsector de nueva contratación. En la fase de diagnóstico, correspondiente al proceso de estimación de los modelos provisionales, se aplica el método de Peña y Yohai (1995) para detectar observaciones atípicas influyentes, lo que permite corregir los sesgos de estimación provocados por dichas observaciones.

Para los datos de 1980-81 se presentan dos modelos alternativos en relación a las características físicas de la vivienda. El primero es el modelo hedónico tradicional, que consiste en considerar de forma individual tales características, mientras que en el segundo dicho conjunto de variables se sustituye por el Índice de Calidad (IC) que fue construido en el Capítulo anterior. La principal diferencia es que con el método hedónico se obtiene la estimación marginal de cada atributo mientras que con el modelo con IC se obtiene una estimación conjunta de todas ellas. Los resultados del modelo hedónico muestran que las reducciones en el alquiler debidas a no tener agua caliente, o tener este servicio de forma individual, son del 31 y 12 por ciento respecto al suministro colectivo. Este descuento se estima en un 49 por ciento si no se dispone de agua en el interior de la vivienda. Respecto al tipo

de vivienda más frecuente que dispone de 'un cuarto de baño completo', el alquiler se reduce en un 17 por ciento, aproximadamente, si las instalaciones higiénicas son menores, y se eleva en un 36,6 por ciento si éstas son mayores. También la 'calefacción fija' incrementa en mayor medida el alquiler que la disponibilidad de 'sólo aparatos móviles', un 34 frente a un 3,6 por ciento, respectivamente. Los servicios que proporcionan el garaje y el teléfono elevan en menor medida el coste del arrendamiento: un 13 y un 9,6 por ciento, respectivamente. Por otra parte, se estima un incremento porcentual medio en el alquiler de un 0,23 por ciento ante incrementos porcentuales unitarios de la superficie construida. Por último, la antigüedad del edificio reduce el precio mensual de la vivienda arrendada en un 0,3 por ciento por año considerado.

Cuando se sustituye el conjunto de características físicas de la vivienda por el IC se detectan las siguientes ventajas frente al modelo hedónico: i) se reduce considerablemente el número de parámetros a estimar, mientras que la bondad de ajuste y los efectos independientes de las variables de localización y ocupación son similares en ambos modelos, y ii) la estimación del efecto cruzado entre la ocupación y la calidad de la vivienda se lleva a cabo de forma sencilla. En el modelo hedónico hubiera sido difícil estimar este efecto debido al gran número de variables físicas involucradas. Estos resultados justifican que el modelo con IC sea el método aplicado con los datos de la EPF de 1990-91, donde cabría esperar que las ventajas sean aún más importantes puesto que el número de características físicas disponibles es mayor. Seguidamente se comentan los resultados más relevantes.

Tanto en 1980-81 como en 1990-91 la estimación del efecto en el alquiler derivado del incremento en la calidad es positivo y depende del período de ocupación. Así, en el modelo de la EPF de 1980-81 se estima un efecto marginal diferencial de la calidad para la ocupación entre los años 1965-73 y

1974-80 del 7 y el 8,3 por ciento, respectivamente. En el modelo de la EPF de 1990-91 este efecto es de un 11,7 por ciento para el período de ocupación 1965-75, un 9,4 por ciento para 1976-83 y un 6,5 por ciento para 1983-90.

Por su parte, el efecto estimado de la ocupación (*tenure discount*) muestra una tendencia negativa y depende tanto del número de años que lleve arrendada la vivienda como del nivel de calidad del inmueble. Mientras que en el modelo de 1980-81 apenas se constata la utilización de las cláusulas de actualización, en el modelo de 1990-91 su uso es generalizado. No obstante, en ambos casos se estima una ventaja clara para los inquilinos que renuevan su contrato de arrendamiento frente a los que acuden a la nueva contratación de viviendas similares: el incremento del alquiler vía cláusulas de actualización, a que se ven sometidos los primeros no recoge en su totalidad el efecto inflacionista que deben soportar los segundos. A expensas del nivel de calidad, el *tenure discount* estimado en ambas muestras es equivalente en el sentido de que es mayor cuantos más años lleve ocupada la vivienda en *AL*. De esta forma, si se considera una vivienda en *AL* de calidad media, el descuento medio que se estima en 1980-81 es de un 66 por ciento si la ocupación se realizó en el período 1965-73 y de un 35 por ciento si tuvo lugar en 1974-79. Los correspondientes descuentos respecto al nuevo arrendamiento en 1990-91 se estiman en un 74, 46, y un 9,6 por ciento en los períodos de ocupación 1965-75, 1976-82 y 1983-90, respectivamente.

La comparación entre los resultados de la EPF de 1980-81 y la EPF de 1990-91 respecto a la revalorización del alquiler en función de la calidad nos lleva a unos resultados que, a primera vista, parecen contradictorios: según la primera muestra, las viviendas de mayor calidad son las que tienen un mayor descuento derivado de la ocupación mientras los datos de la EPF de 1990-91 muestran que son las viviendas más revalorizadas. Por otra parte se

se constata que la vivienda de calidad es el tipo de viviendas más demandado con el paso del tiempo: en términos medios, el subconjunto de viviendas ocupadas antes de 1981 que se mantienen ocupadas en 1990-91 son las de mejor calidad respecto al stock de viviendas en AL de 1980-81 y las de peor calidad respecto al stock de 1990-91. Ambas consideraciones permiten estimar que este subconjunto de viviendas es el que presenta un mayor *tenure discount* en ambos periodos.

En cuanto a la influencia de la localización en el alquiler, el efecto cruzado entre la provincia y el tamaño del municipio resulta estadísticamente significativo en las dos EPF analizadas. Así por ejemplo, el alquiler mensual medio de una vivienda de arrendamiento libre en 1980-81 en Madrid capital, se estima entre un 127 y un 144 por ciento más que si se ubica en una población con menos de 10.000 habitantes en la provincia de Badajoz. Este incremento de precio supone un 145 por ciento si la vivienda arrendada corresponde a 1990-91.

La estimación de los determinantes del alquiler en 1980-81 y 1990-91 hace posible la valoración del coste corriente de los servicios de la vivienda del stock residencial en otros regímenes de tenencia. Una de las aplicaciones de este ejercicio podría ser de utilidad para el INE a la hora de incluir la vivienda en la elaboración del índice de precios de consumo (IPC) en nuestro país. Como es sabido, para estimar el flujo de los servicios de la vivienda distinta del arrendamiento en la elaboración del sistema del IPC con base en 1976 y 1983, el INE emplea la autoevaluación del alquiler por parte del ocupante, propietario o beneficiario de una vivienda en cesión que proporcionan las EPF. La evaluación del coste corriente de los servicios de vivienda a través de los modelos estimados en este capítulo, representa una alternativa objetiva a dicha imputación que, cuando menos, merece considerarse a los efectos de

evaluar el posible sesgo en la medición de la inflación. Por este motivo, en el capítulo siguiente se plantea realiza la evaluación del flujo corriente del stock residencial y su comparación con el valor imputado por los hogares de la EPF en régimen distinto del arrendamiento.



. Cuadros del capítulo III

Cuadro 3.1: Definición de las variables del modelo hedónico para 1980-81.

Variable	Notación <sup>16</sup>	Definición
Servicios higiénicos	hig1	Menos de un cuarto de baño completo
	hig2	Un baño completo, con o sin cuarto retrete
	hig3*	Más de un baño completo y cuarto retrete
Agua	agua1	No tiene
	agua2	Sólo agua fría
	agua3	Agua caliente individual
	agua4*	Agua caliente central
Calefacción	calf1	No tiene
	calf2	Sólo aparatos móviles
	calf3*	Instalación fija: individual o colectiva
Garaje	gar1	No tiene
Teléfono	tel	Si tiene teléfono
Superficie construida	lm2v	Logaritmo de los metros cuadrados
Antigüedad del edificio	aconx	Número de años desde su construcción
Localización	rloc1*-rloc9	Tamaño de municipio y provincia
Año de ocupación ( <i>t</i> )	a6573	=1 si $1965 \leq t \leq 1972$ ; =0 en otro caso
	acup6573	Número de años si $1965 \leq t \leq 1973$
	acup7480	Número de años si $1974 \leq t \leq 1981$

<sup>16</sup> Las variables señaladas con el superíndice \* hacen referencia a las modalidades omitidas.

Cuadro 3.2: Modelos hedónicos, con y sin atípicos, para 1980-81

Variables	Con atípicos	$F_{22,2119}=79$	Sin atípicos	$F_{22,2029}=99$
	Nºobs.: 2.142	$R_{aj}^2=0,514$	Nºobs.: 2.052	$R_{aj}^2=0,604$
	Coefficiente	t-valor	Coefficiente	t-valor
constante	8,156	24,793	8,305	26,747
hig1	-0,495	-6,155	-0,502	-7,063
hig2	-0,313	-5,494	-0,312	-6,088
agua1	-0,708	-4,078	-0,681	-4,308
agua2	-0,358	-2,912	-0,370	-3,354
agua3	-0,174	-1,530	-0,131	-1,275
calf1	-0,344	-5,628	-0,292	-5,284
calf2	-0,282	-4,187	-0,257	-4,270
gar1	-0,070	-1,065	-0,116	-2,167
tel	0,096	2,292	0,092	2,472
lm2v	0,235	4,198	0,227	4,194
aconx	-0,003	-4,547	-0,003	-4,789
rloc2	0,247	2,841	0,213	2,878
rloc3	0,416	4,330	0,460	5,908
rloc4	0,715	5,604	0,625	5,241
rloc5	0,534	6,098	0,521	6,885
rloc6	0,761	7,251	0,733	8,377
rloc7	0,762	8,611	0,702	9,252
rloc8	0,833	9,403	0,747	9,766
rloc9	0,935	9,560	0,892	10,262
a6573	-0,674	-5,046	-0,730	-6,440
acup6573	-0,037	-3,302	-0,037	-3,931
acup7480	-0,116	-10,896	-0,128	-13,297

Estimación robusta,  $y=\ln(a)$ , con factor de elevación poblacional.

Cuadro 3.3: Modelo con IC, con y sin atípicos, para 1980-81

Variables	Con atípicos	$F_{13,2128}=125$	Sin atípicos	$F_{13,2038}=155$
	Nºobs.: 2.142	$R^2_{aj}=0,505$	Nºobs.: 2.052	$R^2_{aj}=0,596$
	Coficiente	t-value	Coficiente	t-valor
constante	6.,399	49,739	6,556	59,159
ind6573	0,644	10,837	0,700	14,790
ind7480	0,838	19,120	0,830	21,320
loc2	0,187	2,229	0,151	2,178
loc3	0,372	3,982	0,421	5,737
loc4	0,635	5,026	0,556	4,692
loc5	0,463	5,409	0,439	6,089
loc6	0,709	6,937	0,678	8,006
loc7	0,730	8,581	0,667	9,420
loc8	0,799	9,292	0,706	9,688
loc9	0,877	9,492	0,820	10,245
a6573	-0.241	-1,031	-0,458	-2,351
acup6573	-0.039	-3,372	-0,038	-3,905
acup7480	-0,120	-11,363	-0,132	-13,629

Estimación robusta,  $y=\ln(a)$ , con factor de elevación poblacional.

Cuadro 3.4: Definición de las variables del modelo con IC para 1990-91

Variable	Notación	Definición
IC de la vivienda	ind6575	IC si $1965 \leq \text{año de ocupación } (t) \leq 1975$
	ind7682	IC si $1976 \leq t \leq 1982$
	ind8390	IC si $1983 \leq t \leq 1991$
Localización	rloc1-rloc5	Tamaño de municipio y provincia
Año de ocupación	a6575	=1 si $1966 \leq t \leq 1990$ ; =0 en otro caso
	a7682	=1 si $1966 \leq t \leq 1990$ ; =0 en otro caso
	acup6575	Número de años si $1965 \leq t \leq 1975$
	acup7682	Número de años si $1976 \leq t \leq 1982$
	acup8390	Número de años si $1983 \leq t \leq 1991$

Cuadro 3.5: Modelo con IC, con y sin atípicos, para 1990-91

Variables	Con atípicos	$F_{12,1022}=59$	Sin atípicos	$F_{12,949}=74$
	Nºobs.: 1035	$R^2_{aj}=0,512$	Nºobs.: 982	$R^2_{aj}=0,602$
	Coefficiente	t-valor	Coefficiente	t-valor
constante	8,105	32.095	8,323	37,407
ind6575	0,119	7.954	0,117	8,782
ind7682	0,096	4.767	0,094	4,835
ind8390	0,072	8.713	0,065	8,258
loc2	0,292	1.329	0,383	1,985
loc3	0,525	2.427	0,607	3,205
loc4	0,681	2.948	0,806	3,999
loc5	0,777	3.504	0,898	4,634
a6575	-0,322	-0.734	-0,747	-1,859
acup6575	-0,079	-4.034	-0,070	-3,663
a7682	-0,441	-1.119	-0,631	-1,645
acup7682	-0,032	-1.212	-0,040	-1,559
acup8390	0,001	0.081	-0,026	-1,849

Estimación robusta,  $y=\ln(a)$ , con factor de elevación poblacional.



**. Apéndice I: Determinación de  
la forma funcional del alquiler**

Sea  $a$  el conjunto de alquileres en  $AL$  ocupados en el año  $T$  desde el año  $t = 1965, \dots, T$  y sea  $z$  el conjunto de dichos alquileres que se observan en la EPF de 1980-81 (equivalentemente de la EPF de 1990-91). Por tanto,  $z$  es una muestra aleatoria de la variable  $a$  que permitirá determinar la forma funcional del alquiler en  $AL$  de nuestro país. La determinación de la forma funcional  $g$  de la ecuación (3.5) requiere un análisis descriptivo y gráfico de la variable dependiente  $z$  del modelo. Con el fin de simplificar la exposición, y sin pérdida de generalidad, este apartado se limita a los datos correspondientes a la EPF de 1980-81, por lo que  $T = 1980$  y al supuesto de que  $\partial g / \partial X_t = \partial g / \partial X_{t'}$ , para todo  $t \neq t'$ . Así, en las Figuras 1, 2 y 3 se observa que: 1) el alquiler mensual observado en el régimen de  $AL$  muestra una distribución Chi-cuadrado, 2) si se denota  $e$  al error asociado a cualquiera de las especificaciones lineales  $h(z)$  previa del modelo (3.5), se observa que el gráfico de  $e_i = h(z_i)$  (*Residuals*) muestra heteroscedasticidad y una estructura curva y, 3) la distribución de  $e$  (*Density:Residuals*) es claramente asimétrica.

Estos resultados sugieren la transformación logarítmica de la variable dependiente:  $y = \ln z$ . De esta manera, los problemas anteriores se evitan y la forma funcional del modelo se puede expresar como la siguiente relación lineal entre  $y$  y las variables explicativas: características físicas de la vivienda  $X_1$ , localización  $X_2$  y año de ocupación  $X_3$   $dummy = [ocup65, ocup66, \dots, ocupT]$  con  $ocup_t = 1$  si el año de ocupación es  $t$  y  $ocup_t = 0$  en caso contrario para  $t = 65, \dots, 80$ ,

$$y = \alpha + \sum_{j=1}^2 \beta_j X_j + \sum_{t=65}^{79} \gamma_t ocup_t + \varepsilon$$

Dada la expresión de la variable dependiente  $y$ , el alquiler de las viviendas en  $AL$  se obtiene a partir de la ecuación

$$z = \exp(y) = \exp \alpha * \prod_{j=1}^2 \exp(\beta_j X_j) * \prod_{t=65}^{79} \exp(\gamma_t ocup_t) * \exp(\varepsilon),$$

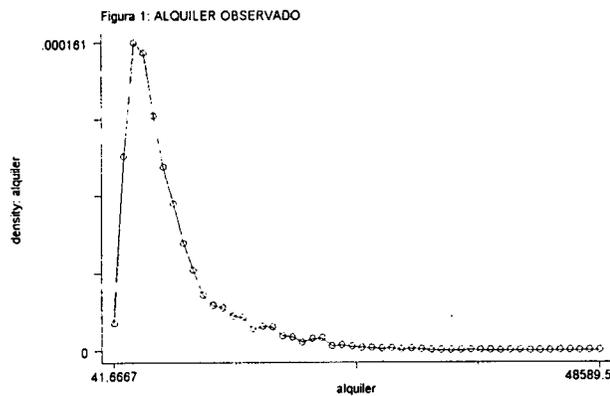
quedando así determinada la función  $g$  de la ecuación (3.5). Bajo el supuesto de que  $\varepsilon$  tiene una distribución normal con media cero y varianza  $\sigma^2$ , la variable aleatoria  $\exp(\varepsilon)$  sigue una distribución lognormal cuyos momentos son:

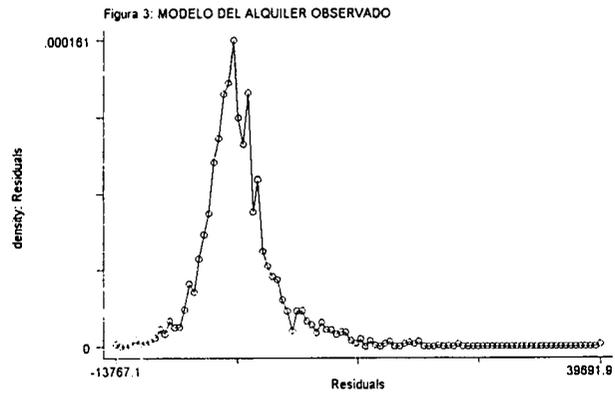
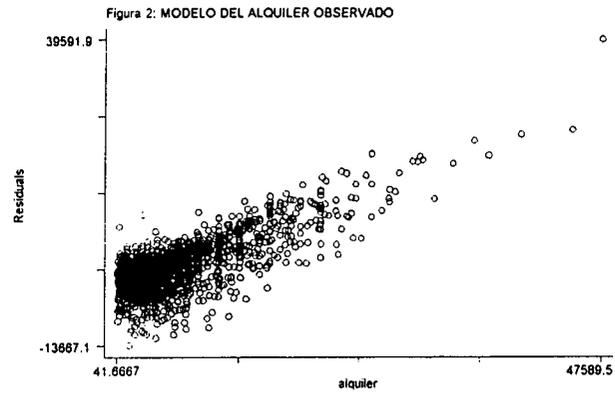
$$\begin{aligned} E(\exp(\varepsilon)) &= \exp(\sigma^2/2) > 1 \\ \text{Var}(\exp(\varepsilon)) &= \exp \sigma^2 (\exp \sigma^2 - 1). \end{aligned}$$

Por tanto, el alquiler estimado en  $T = 1980$  de un tipo de vivienda  $i$  con vector de características físicas y de localización  $X_i = [X1_i, X2_i]$  y vector de ocupación  $X3_{dummyi} = [ocup65_i, ocup66_i, \dots, ocupT_i]$ , se puede obtener a través de la expresión

$$\hat{E}(z/X_i, ocupT_i = 1) = \exp \hat{\alpha}' * \prod_{j=1}^2 \exp(\hat{\beta}_j X_{j_i}) * \prod_{t=65}^{79} \exp \hat{\gamma}_t ocupT_i$$

siendo  $\exp \hat{\alpha}' = \exp \hat{\alpha} * \exp(\hat{\sigma}^2/2)$ .





. Apéndice II: Relación entre  
las variables explicativas  
disponibles y el alquiler

La descripción de las columnas es,

% porcentaje de viviendas a nivel poblacional;

**MEDIA:** alquiler medio observado; **Q25, Q50 y Q75:** el primer, segundo y tercer cuartil del alquiler;

**VA:** variables auxiliares creadas en el proceso de especificación del modelo;

**VD y VC:** variables, discretas y continua respectivamente, del modelo final.

## II. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1980-81

Tamaño muestral: 2.142; Tamaño poblacional: 867.627.

	%	MEDIA	Q75	Q50	Q25	VA	VD	VC
<b>TOTAL</b>	<b>100</b>	<b>5.629</b>	<b>7.000</b>	<b>3.809</b>	<b>2.000</b>			
<b>X1: CARACTERÍSTICAS FÍSICAS DE LA VIVIENDA</b>								
<b>Variables discretas</b>								
<b>Instalaciones higiénicas</b>								
No tiene	4,93	1.842	2.500	1.500	705		hig1	
Compartido con otras viviendas	1,45	2.386	2.600	1.500	1.000		hig1	
Uno o dos cuartos retrete	3,39	2.206	2.500	1.800	1.000		hig1	
Cuarto(s) de aseo, o aseo y retrete	8,95	2.508	3.000	2.000	1.200		hig1	
Un cuarto de baño	68,55	5.552	6.870	4.000	2.474		hig2	
Un baño y un retrete	0,50	4.241	11.500	2.500	375		hig2	
Un baño y uno o dos aseos	7,45	9.386	12.690	6.830	4.320		hig3	
Dos baños	3,46	11.951	17.096	11.467	6.000		hig3	
Dos o más baños y uno o dos aseos	0,59	22.032	44.730	12.000	9.097		hig3	
Uno o más baños, aseo y retrete	0,20	20.029	33.000	20.029	7.058		hig3	
Tres o cuatro baños	0,54	17.870	21.712	17.000	11.000		hig3	
<b>Agua corriente</b>								
No tiene	3,85	1.945	3.000	1.382	600		agua1	
Sólo agua fría	23,82	2.819	3.500	2.030	1.300		agua2	
Agua caliente individual	68,15	6.297	8.000	4.800	2.735		agua3	
Agua caliente central	4,18	14.145	19.529	11.200	5.200		agua4	
<b>Calentamiento</b>								
No tiene	66,32	4.332	5.613	3.200	1.793		calf1	
Algún aparato móvil	19,88	5.713	7.000	4.000	2.381		calf2	
Individual con aparatos fijos	7,60	10.340	14.778	8.000	4.316		calf3	
Central	6,20	13.455	18.724	10.296	5.618		calf3	
<b>Teléfono</b>								
No tiene	63,15	4.517	6.000	3.200	1.900		tel	
Si tiene	36,85	7.535	10.000	5.000	2.750		tel	
<b>Garaje</b>								
No tiene	88,30	5.177	6.387	3.500	2.000		gar1	
Tiene pero fuera del edificio	6,75	7.462	10.000	6.000	2.500		gar1	
Tiene en el propio edificio	4,95	11.204	16.601	7.000	3.500		gar1	
<b>Variables continuas</b>								
<b>Número de metros cuadrados construidos (se considera en logaritmos)</b>								
Menos de 60	18,72	3.801	5.000	2.690	1.500	lm2v1		lm2v
Entre 61 y 90	46,67	5.423	6.769	3.756	2.000	lm2v2		
Entre 91 y 130	27,94	6.360	8.000	4.500	2.735	lm2v3		
Más de 130	6,66	9.144	12.000	5.750	2.500	lm2v4		
<b>Número de años de antigüedad del edificio</b>								
10 años o menos	18,24	9.967	13.290	7.500	4.719	aconx1		aconx
Entre 11 y 20 años	28,64	6.016	7.500	5.000	2.835	aconx2		
Entre 21 y 30 años	10,59	5.202	5.400	3.500	2.000	aconx3		
Entre 31 y 50 años	15,98	4.299	5.409	3.000	1.750	aconx4		
Más de 50 años	26,55	3.202	4.096	2.227	1.135	aconx5		

# 1. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1980-81

Tamaño muestral: 2.142; Tamaño poblacional: 867.627.

		MEDIA	Q75	Q50	Q25	VA	VD	VC
X2: LOCALIZACIÓN		5.629	7.000	3.809	2.000			
<b>VARIABLES DISCRETAS</b>								
<b>Tamaño de municipio</b>								
Menos de 10.000 habitantes	22,26	3.381	4.925	2.500	1.500	rtmun1		
10.001-50.000 habitantes	20,72	4.154	5.000	3.000	1.500	rtmun2		
50.001-500.000 habitantes	34,78	6.093	7.800	4.719	2.542	rtmun3		
Más de 500.000 habitantes	22,24	8.528	11.165	5.500	2.937	rtmun4		
<b>Provincia</b>								
		5.629	7.000	3.809	2.000			
Albacete	0,58	3.348	5.500	2.000	1.000	rpro1		
Alicante	3,30	3.156	4.000	2.500	1.200	rpro1		
Almería	1,30	3.183	4.000	2.000	1.000	rpro1		
Ávila	0,43	2.655	3.118	2.490	650	rpro1		
Badajoz	1,03	2.829	5.000	2.227	1.500	rpro1		
Cáceres	1,44	2.776	3.000	2.000	1.000	rpro1		
Ciudad Real	1,21	2.793	3.400	2.000	1.000	rpro1		
Córdoba	1,07	5.397	5.000	2.075	600	rpro1		
Jaén	1,27	2.883	4.416	2.100	1.300	rpro1		
Soria	0,27	3.093	5.274	2.000	1.000	rpro1		
Teruel	0,34	2.350	4.000	1.500	1.000	rpro1		
Burgos	0,96	3.851	5.000	2.300	1.000	rpro2		
Cádiz	2,66	3.787	4.003	2.500	1.395	rpro2		
Cuenca	0,28	3.174	4.000	2.835	2.000	rpro2		
Huesca	0,69	3.974	5.000	3.000	2.000	rpro2		
León	2,22	4.204	5.300	2.000	1.355	rpro2		
La Rioja	0,72	4.117	5.700	3.000	1.600	rpro2		
Navarra	0,70	4.056	5.934	2.502	1.000	rpro2		
Palencia	0,61	4.723	7.000	2.525	1.003	rpro2		
Salamanca	1,08	3.894	6.000	2.500	1.200	rpro2		
Segovia	0,49	4.491	5.000	3.309	1.124	rpro2		
Tarragona	1,95	4.479	6.370	3.000	1.500	rpro2		
Toledo	0,91	3.903	6.000	2.500	1.459	rpro2		
Valencia	2,92	4.081	5.000	2.000	1.500	rpro2		
Vizcaya	0,68	4.296	6.000	2.800	1.380	rpro2		
Ceuta	0,30	4.497	6.500	3.075	1.600	rpro2		
Castellón	0,69	4.993	6.500	4.000	2.800	rpro3		
La Coruña	3,30	5.767	6.000	3.746	2.500	rpro3		
Gerona	1,98	4.125	4.860	3.000	2.300	rpro3		
Granada	1,46	4.601	6.000	3.500	2.000	rpro3		
Guadalajara	0,16	4.455	5.000	3.800	2.647	rpro3		
Guipuzcoa	1,43	5.587	8.000	3.600	2.000	rpro3		
Huelva	1,21	4.729	5.500	3.500	2.000	rpro3		
Lérida	1,74	5.233	6.879	4.000	2.500	rpro3		
Lugo	1,01	4.251	5.437	4.000	1.125	rpro3		
Málaga	2,41	4.161	4.682	3.637	2.150	rpro3		
Orense	0,69	5.972	10.000	4.000	2.000	rpro3		
Asturias	4,32	5.497	8.000	2.800	1.500	rpro3		
Cantabria	1,16	6.263	7.000	3.000	1.841	rpro3		
Sevilla	2,46	5.359	5.588	3.500	2.000	rpro3		
Valladolid	0,88	5.247	8.000	5.000	2.000	rpro3		
Zamora	0,25	5.053	5.000	4.500	1.350	rpro3		
Zaragoza	1,65	5.601	8.000	5.000	200	rpro3		
Melilla	0,37	5.393	10.000	3.200	1.712	rpro3		
Alava	0,30	7.640	12.000	6.337	5.250	rpro4		
Baleares	4,97	6.170	5.753	4.500	2.870	rpro4		
Barcelona	18,21	7.113	8.300	5.100	3.200	rpro4		
Madrid	13,41	8.100	10.296	5.294	2.800	rpro4		
Murcia	1,39	6.568	7.500	6.000	3.175	rpro4		
Las Palmas	0,78	6.384	10.000	4.500	3.000	rpro4		
Pontevedra	2,52	6.172	8.259	5.000	3.000	rpro4		
Santa Cruz Tenerife	1,85	6.066	7.746	5.023	2.735	rpro4		

## EFFECTOS CRUZADOS 'TAMAÑO DE MUNICIPIO- PROVINCIA' (rtmun, rpro)

	%	MEDIA	Q75	Q50	Q25	
rtmun1, rpro1	4,71	2.089	2.500	1.500	1.000	rlloc1
rtmun1, rpro2	5,53	3.096	4.500	2.000	1.250	rlloc2
rtmun2, rpro1	3,85	2.803	4.000	2.500	1.000	rlloc2
rtmun2, rpro2	3,59	2.980	3.500	1.937	1.500	rlloc2
rtmun1, rpro3	6,83	3.621	4.500	3.000	2.000	rlloc3
rtmun2, rpro3	6,17	3.828	4.335	2.700	1.500	rlloc3
rtmun1, rpro4	5,19	4.544	6.000	4.000	2.300	rlloc4
rtmun3, rpro1	3,66	5.281	6.000	3.500	2.000	rlloc5
rtmun3, rpro2	6,38	5.193	6.500	3.495	1.764	rlloc5
rtmun4, rpro2	1,68	5.583	8.000	3.000	1.200	rlloc5
rtmun2, rpro4	7,11	5.762	8.000	4.000	3.000	rlloc6
rtmun3, rpro3	11,33	6.554	9.000	4.682	2.500	rlloc7
rtmun4, rpro3	2,83	6.358	9.083	4.474	2.900	rlloc7
rtmun3, rpro4	13,40	6.353	7.790	5.100	3.371	rlloc8
rtmun4, rpro4	17,73	9.154	11.749	5.772	3.190	rlloc9

### 1. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1980-81

Tamaño muestral: 2.142; Tamaño poblacional: 867.627.

X3: AÑO DE OCUPACIÓN DE LA VIVIENDA	%	MEDIA	Q75	Q50	Q25	VA	VD	VC
		5.629	7.000	3.809	2.000			
<b>Variables discretas</b>								
Ocupación en 1980-81	7,70	9.397	13.290	8.050	4.000	<b>ocup80</b>	0	
En 1979	9,78	10.926	15.000	8.500	4.760	<b>ocup79</b>	0	
En 1978	7,59	9.497	12.648	6.000	3.000	<b>ocup78</b>	0	
En 1977	6,41	5.939	8.500	5.500	2.470	<b>ocup77</b>	0	
En 1976	6,55	5.939	6.910	4.418	3.000	<b>ocup76</b>	0	
En 1975	8,70	4.980	6.000	4.224	2.500	<b>ocup75</b>	0	
En 1974	9,97	4.359	6.000	3.500	2.000	<b>ocup74</b>	0	
En 1973	7,23	3.685	5.200	3.000	1.750	<b>ocup73</b>	1	
En 1972	7,50	3.884	5.500	3.000	1.600	<b>ocup72</b>	1	
En 1971	3,95	2.979	4.000	2.489	1.250	<b>ocup71</b>	1	
En 1970	7,48	4.001	5.000	3.500	1.900	<b>ocup70</b>	1	
En 1969	3,86	3.461	3.981	2.735	1.500	<b>ocup69</b>	1	
En 1968	4,56	2.893	3.200	2.200	1.096	<b>ocup68</b>	1	
En 1967	2,30	2.604	3.715	2.000	1.225	<b>ocup67</b>	1	
En 1966	2,40	2.566	2.800	1.500	1.000	<b>ocup66</b>	1	
En 1965	4,03	2.924	4.500	2.200	1.150	<b>ocup65</b>	1	
Ocupación entre 1965 y 1973	43,30	3416	4474	2647	1500	<b>a6573</b>		<b>acup6573</b>
Entre 1974 y 1977	31,64	5177	6626	4100	2500	<b>a7477</b>		
Entre 1978 y 1980-81	25,07	10023	14000	7500	4000	<b>a7880</b>		
Entre 1974 y 1980-81	56,70	7320	9886	5000	2912	<b>a7480</b>		
<b>Variables continuas</b>								
Número de años que lleva ocupada la vivienda si se ocupó entre 1965 y 1973						<b>acup6573</b>		<b>acup6573</b>
Entre 1974 y 1977						<b>acup7477</b>		
Entre 1978 y 1980-81						<b>acup7880</b>		
Entre 1974 y 1980-81						<b>acup7480</b>		<b>acup7480</b>

La descripción de las columnas es,

% porcentaje de viviendas a nivel poblacional;

**MEDIA:** alquiler medio observado; **Q25, Q50 y Q75:** el primer, segundo y tercer cuartil del alquiler;

**VA:** variables auxiliares creadas en el proceso de especificación del modelo;

**VD y VC:** variables, discretas y continua respectivamente, del modelo final.

## 2. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1990-91

Tamaño muestral: 1.035; Tamaño poblacional: 590.948.

	%	MEDIA	Q75	Q50	Q25	VA	VD
<b>TOTAL</b>	<b>100,00</b>	<b>17.145</b>	<b>25.000</b>	<b>12.650</b>	<b>5.500</b>		
<b>X1: CARACTERÍSTICAS FÍSICAS DE LA VIVIENDA</b>							
<b>Variables discretas</b>							
<b>Instalaciones higiénicas</b>							
No tiene	2,33	10.250	16.072	2.092	1.240		
Compartido con otras viviendas	0,62	3.430	5.500	500	500		
Uno o dos aseos, o uno o dos cuartos retrete	5,25	6.935	9.000	5.093	3.311		
Un baño, o retrete con uno o dos aseos	74,00	15.470	20.972	11.667	5.250		
Un baño y uno o dos aseos o un baño y un retrete	5,00	21.986	29.000	20.500	15.000		
Un baño, aseo y retrete, o dos baños y aseo(s) o retrete(s)	11,90	31.182	38.000	28.000	15.000		
Dos baños, aseos y retrete o más	0,90	29.131	38.000	13.268	13.268		
<b>Agua corriente</b>							
No tiene	0,45	2.274	1.250	1.240	333		
Sólo agua fría	8,11	6.477	8.000	4.100	2.067		
Agua caliente individual	85,16	17.397	25.000	13.200	6.000		
Agua caliente central	6,28	28.565	35.000	27.969	13.268		
<b>Calefacción</b>							
No tiene	13,71	18.163	26.000	14.000	5.100		
Algún aparato móvil	64,22	13.338	17.975	10.000	5.000		
Individual con aparatos fijos	15,60	26.070	37.311	20.000	12.000		
Central	6,47	31.263	38.749	31.460	16.702		
<b>Teléfono</b>							
No tiene	35,62	14.037	20.000	11.000	5.000		
Un teléfono	48,01	17.137	26.500	12.000	5.000		
Dos teléfonos	14,83	22.542	28.620	15.000	9.350		
Tres o más teléfonos	1,54	37.373	42.000	33.800	29.111		
<b>Garaje</b>							
No tiene	81,15	15.437	22.000	11.500	5.000		
Si tiene	18,85	24.499	31.460	17.025	10.000		
<b>Luz</b>							
No tiene	0,12	2.298	8.000	333	333		
Si tiene	99,88	17.163	25.000	12.800	5.500		
<b>Tipo de edificio</b>							
Otros fines distintos de los residenciales	0,15	9.180	16.600	16.600	1.500		
Alojamiento fijo	0,28	13.288	20.000	13.250	6.500		
Edificio de una sola vivienda	18,66	10.484	14.202	8.000	4.000		
Edificio de dos viviendas	6,36	12.527	17.000	11.116	5.000		
Edificio de tres o más viviendas	74,55	19.238	28.000	15.000	6.340		
<b>Aire acondicionado (AIREAC)</b>							
No tiene	98,18	16.997	24.880	12.322	5.500		
Sistema privado	1,74	23.154	32.000	28.435	7.500		
Sistema colectivo	0,08	70.000	70.000	70.000	70.000		
<b>Ascensor (ASC)</b>							
No tiene	73,36	12.823	17.000	9.600	4.500		
Si tiene	26,64	29.047	37.485	27.019	16.000		
<b>Jardín (JAR)</b>							
No tiene	91,45	16.940	24.570	12.000	5.500		
Si tiene	8,55	19.336	32.000	17.025	5.500		
<b>Piscina (PISC)</b>							
No tiene	98,13	17.042	25.000	12.160	5.400		
Si tiene	1,87	22.553	35.000	17.025	13.268		
<b>Zona Deportiva (ZDEP)</b>							
No tiene	98,91	17.129	25.000	12.300	5.500		
Si tiene	1,09	18.645	28.809	17.025	13.268		
<b>Otros servicios comunitarios (OSERC)</b>							
No tiene	51,10	11.624	15.200	8.514	4.000		
Si tiene	48,90	22.915	32.000	18.850	9.500		

## 2. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1990-91

Tamaño muestral: 1.035; Tamaño poblacional: 590.948.

	%	MEDIA	Q75	Q50	Q25	VA
<b>TOTAL</b>	<b>100,00</b>	<b>17.145</b>	<b>25.000</b>	<b>12.650</b>	<b>5.500</b>	
<b>X1: CARACTERÍSTICAS FÍSICAS DE LA VIVIENDA</b>						
<b>Variables discretas</b>						
<b>Combustible o energía para calentar el agua (COMBA)</b>						
No tiene o sólo posee agua fría	8,56	6.256	8.000	4.100	2.000	
Leña	0,6	4.303	4.450	4.450	2.208	
Butano	56,58	16.077	22.000	12.600	6.500	
Energía eléctrica	18,41	15.036	20.417	11.250	4.000	
Carbón	1,2	14.537	16.000	15.281	3.250	
Otros combustibles gaseosos o sólidos	0,18	15.488	28.809	15.488	2.167	
Propano	1,72	23.645	34.000	17.025	13.268	
Gas ciudad o gas natural	8,61	30.776	39.048	28.000	13.200	
Fuel-oil, gasóleo u otros combustibles líquidos	4,14	35.274	42.000	34.000	26.000	
<b>Combustible o energía para la calefacción (COMBCF)</b>						
No tiene calefacción	13,71	18.163	26.000	14.000	5.100	
Leña u otros combustibles sólidos	8,3	10.620	12.000	8.000	4.000	
Butano	20,58	12.276	17.000	10.000	5.000	
Carbón	4,5	17.182	22.900	15.000	4.919	
Energía eléctrica	41,57	15.945	23.000	12.300	5.500	
Otros combustibles gaseosos o líquidos	0,85	35.134	70.676	25.000	5.133	
Gasóleo	3,63	26.291	34.000	26.000	16.500	
Propano	2,06	21.439	32.500	17.025	13.268	
Gas ciudad o gas natural	2,96	47.172	52.500	39.048	28.435	
Fuel-oil	1,83	41.171	46.628	35.000	31.700	
<b>Combustible o energía para cocinar (COMBACOC )</b>						
Leña	0,54	3.339	5.000	2.083	1.000	
Carbón	0,89	9.696	15.000	8.615	2.500	
Butano	77,79	15.112	20.417	11.000	5.093	
Energía eléctrica	7,41	25.924	37.645	23.500	11.161	
Propano	3,99	18.539	26.725	17.025	5.000	
Gas ciudad o gas natural	9,37	27.991	37.311	18.301	12.000	
<b>Variables continuas</b>						
<b>Número de metros cuadrados construidos</b>						
Menos de 60	18,86	14.275	16.783	8.950	4.000	
Entre 61 y 90	38,87	16.482	24.317	13.000	5.500	
Entre 91 y 130	30,06	17.543	26.000	15.000	6.200	
Más de 130	12,21	22.709	30.316	16.495	6.680	
<b>Número de años de antigüedad del edificio</b>						
Antes de 1801	0,88	9.597	13.273	10.000	4.450	
Entre 1801 y 1899	6,91	12.334	20.000	5.000	3.293	
Entre 1900 y 1935	14,32	14.170	18.000	9.300	4.000	
Entre 1937 y 1945	6,34	11.655	15.000	6.000	3.750	
Entre 1946 y 1955	8,17	14.517	22.000	8.000	4.129	
Entre 1956 y 1965	21,95	15.059	20.000	12.000	6.680	
Entre 1966 y 1975	28,81	20.158	28.000	15.000	7.200	
Entre 1976 y 1980	8,28	25.545	32.000	24.700	13.268	
Entre 1981 y 1985	3,73	23.250	32.500	21.890	15.000	
Después de 1985	0,62	26.060	31.000	20.417	18.000	

## 2. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1990-91

Tamaño muestral: 1.035; Tamaño poblacional: 590.948.

### X2: LOCALIZACIÓN

Variables discretas

Tamaño de municipio	MEDIA	Q75	Q50	Q25	VA	VD
Menos de 10.000 habitantes	16,28	11.723	16.000	9.000	4.000	rtmun1
10.001-50.000 habitantes	23,64	14.725	20.000	12.800	5.500	rtmun2
50.001-500.000 habitantes	35,85	17.875	26.000	14.600	5.920	rtmun3
Más de 500.000 habitantes	24,24	22.068	33.800	15.000	6.000	rtmun4
<b>Provincia</b>						
Alava	0,06	5.000	5.000	5.000	5.000	rpro1
Badajoz	0,35	9.428	15.000	10.000	2.000	rpro1
Huesca	0,08	4.628	6.800	6.800	2.032	rpro1
Jaén	0,99	9.055	18.000	5.000	2.532	rpro1
Murcia	2,31	9.633	16.667	8.000	2.000	rpro1
Navarra	0,13	6.389	11.498	6.389	1.281	rpro1
Teruel	0,18	7.803	12.000	8.000	3.000	rpro1
Ceuta	0,29	5.013	8.500	3.250	1.110	rpro1
Albacete	0,38	14.050	12.000	7.583	3.500	rpro2
Almería	0,95	15.236	18.000	10.000	5.000	rpro2
Ávila	0,19	13.424	11.000	5.000	4.853	rpro2
Burgos	0,57	16.312	22.900	6.839	4.382	rpro2
Cáceres	0,82	11.164	12.000	9.000	5.000	rpro2
Cádiz	3,30	12.377	15.000	6.507	3.500	rpro2
Huelva	0,21	13.927	28.000	7.136	6.250	rpro2
León	1,89	14.655	16.000	13.000	5.000	rpro2
Lérida	1,15	13.227	19.647	10.000	5.500	rpro2
La Rioja	0,52	13.254	24.125	7.240	1.700	rpro2
Palencia	0,42	14.351	19.354	9.428	3.600	rpro2
Salamanca	0,76	15.183	25.000	10.000	1.000	rpro2
Sevilla	2,22	12.570	17.000	7.000	2.500	rpro2
Toledo	0,53	16.181	27.000	7.500	5.400	rpro2
Melilla	0,99	11.120	15.000	8.000	4.730	rpro2
Alicante	2,29	16.329	26.000	16.667	4.000	rpro3
Baleares	5,88	14.281	20.000	10.833	5.000	rpro3
Barcelona	19,41	16.948	18.637	12.000	5.507	rpro3
Castellón	0,57	20.874	30.000	9.000	6.386	rpro3
Córdoba	0,78	19.022	22.850	12.000	9.600	rpro3
Cuenca	0,38	13.239	15.000	10.367	9.231	rpro3
Gerona	1,98	17.318	24.000	15.000	5.129	rpro3
Guadalajara	0,25	15.970	21.932	12.000	7.000	rpro3
Guipuzcoa	1,52	16.220	25.000	12.650	5.417	rpro3
Lugo	0,41	16.642	16.500	15.000	10.417	rpro3
Málaga	1,39	18.394	23.000	13.333	8.937	rpro3
Cantabria	1,14	17.347	23.000	12.000	6.000	rpro3
Segovia	0,42	13.162	16.000	10.000	7.820	rpro3
Soria	0,23	18.265	34.500	10.000	6.000	rpro3
Tarragona	2,28	15.319	17.000	13.000	7.765	rpro3
Valencia	2,98	13.345	21.890	13.229	4.000	rpro3
Valladolid	1,19	10.550	15.000	10.000	7.000	rpro3
Vizcaya	1,53	17.105	25.000	17.500	9.500	rpro3
Zamora	0,76	12.743	18.000	9.167	6.000	rpro3
Zaragoza	0,75	20.395	29.200	12.000	4.200	rpro3
Ciudad Real	0,67	21.952	32.500	20.000	17.000	rpro4
La Coruña	2,27	18.746	28.000	17.000	6.000	rpro4
Granada	1,56	19.631	32.000	16.000	6.040	rpro4
Madrid	15,86	21.530	34.000	15.500	6.000	rpro4
Orense	1,39	17.534	20.000	17.206	13.000	rpro4
Asturias	3,26	20.979	30.000	21.500	9.136	rpro4
Las Palmas	3,94	21.440	33.000	16.000	5.200	rpro4
Pontevedra	2,47	18.229	28.000	17.000	8.000	rpro4
Santa Cruz Tenerife	3,16	25.137	33.750	23.000	11.000	rpro4

### EFFECTOS CRUZADOS 'TAMAÑO DE MUNICIPIO- PROVINCIA' (rtmun, rpro)

	%	MEDIA	Q75	Q50	Q25	
rtmun1, rpro1	0,80	6.351	8.000	5.000	2.000	rloc1
rtmun2, rpro1	1,53	7.169	10.000	5.000	2.500	rloc1
rtmun2, rpro2	2,62	9.410	10.000	5.133	2.600	rloc2
rtmun1, rpro2	4,41	9.983	15.000	8.000	5.000	rloc2
rtmun3, rpro1	2,05	11.099	17.400	10.000	2.000	rloc2
rtmun1, rpro3	7,52	11.324	15.000	9.000	4.000	rloc2
rtmun3, rpro2	6,87	15.964	22.660	11.000	4.450	rloc3
rtmun3, rpro3	14,21	14.088	18.251	10.666	5.000	rloc3
rtmun2, rpro3	13,62	16.169	23.000	13.273	7.000	rloc3
rtmun2, rpro4	5,87	15.706	21.500	15.000	7.250	rloc3
rtmun1, rpro4	3,54	15.959	25.000	15.000	5.200	rloc3
rtmun4, rpro2	1,01	20.996	32.000	16.250	2.332	rloc4
rtmun4, rpro3	10,79	22.218	28.435	15.000	7.100	rloc4
rtmun4, rpro4	12,44	22.024	34.100	15.140	5.750	rloc4
rtmun3, rpro4	12,71	24.233	34.000	24.000	11.000	rloc5

## 2. VIVIENDAS DE ARRENDAMIENTO LIBRE SEGÚN LA EPF DE 1990-91

Tamaño muestral: 1.035; Tamaño poblacional: 590.948.

X3: AÑOS DE OCUPACIÓN DE LA VIVIENDA	MEDIA	Q75	Q50	Q25	VA	VD	VC
En 1965	2,48	3.288	3.719	2.100	870	ocup65	
En 1966	0,92	2.991	3.250	2.550	2.268	ocup66	
En 1967	0,65	4.557	5.400	3.078	1.187	ocup67	
En 1968	2,14	6.897	9.136	5.920	2.092	ocup68	
En 1969	1,52	5.723	8.350	2.694	1.993	ocup69	
En 1970	5,44	6.343	9.000	5.000	2.000	ocup70	
En 1971	2,06	6.396	8.200	7.100	4.000	ocup71	
En 1972	3,05	6.825	10.426	5.000	2.500	ocup72	
En 1973	3,2	10.130	15.281	12.000	3.700	ocup73	
En 1974	2,36	6.883	11.000	4.000	3.311	ocup74	
En 1975	5,39	8.170	10.400	6.000	4.100	ocup75	
En 1976	3,83	11.893	12.821	7.200	6.000	ocup76	
En 1977	3,85	15.538	28.000	11.000	5.000	ocup77	
En 1978	3,69	22.746	33.000	18.500	9.000	ocup78	
En 1979	2,96	29.251	18.637	13.000	10.000	ocup79	
En 1980	3,25	16.042	27.019	10.942	5.417	ocup80	
En 1981	3,65	13.343	16.500	13.268	7.500	ocup81	
En 1982	3,58	15.383	23.500	13.000	5.200	ocup82	
En 1983	4,17	22.737	37.311	20.972	8.900	ocup83	
En 1984	4,68	22.550	28.110	18.301	12.000	ocup84	
En 1985	5,44	22.480	28.898	18.125	12.000	ocup85	
En 1986	4,53	25.097	32.000	17.500	13.000	ocup86	
En 1987	5,59	26.098	30.800	24.000	16.000	ocup87	
En 1988	6,93	22.610	30.000	19.942	11.680	ocup88	
En 1989	7,18	26.935	35.000	25.200	16.667	ocup89	
Ocupación en 1990-91	7,47	20.340	30.000	16.500	6.839	ocup90	
<b>Variables continuas</b>							
Número de años de ocupación según el período de ocupación							
Entre 1965 y 1975	29,20	6.796	9.500	5.000	2.645	a6575	acup6575
Entre 1976 y 1982	24,80	17.404	20.000	11.667	6.500	a7682	acup7682
Entre 1983 y 1990	46,00	23.575	32.000	20.972	12.000	a8390	acup8390

**. Apéndice III: Estimación de  
los modelos de alquiler**

En este apéndice se analiza el proceso de estimación de los modelos del alquiler en relación a los datos muestrales de las Encuestas de Presupuestos Familiares de 1980-81 y 1990-91. El método de estimación es Mínimos Cuadrados Ordinarios con errores estándar robustos y se aplica el factor de elevación proporcionado por el Instituto Nacional de Estadística. Para el primer período se presenta dos modelos alternativos: un modelo hedónico y un modelo donde se sustituyen las características físicas del hogar por el índice de Calidad (IC) construido en el Capítulo II. Para el período 1990-91 sólo se muestra el modelo con IC.

#### 1. Modelos hedónicos para 1980

En este apartado se analiza la estimación de nueve estimaciones previas a la identificación del modelo hedónico más apropiado para 1980-81. Los resultados de los modelos estimados se muestran en el Cuadro 1 de este apéndice y la definición de todas las variables se encuentra en el Apéndice II.1. En el primer modelo se incluyen solamente las características físicas ( $X1$ ) y las de localización ( $X2$ ). En cuanto a las primeras, se consideran todas las variables disponibles en su versión más discretizada. Para intentar captar el posible efecto no lineal en el alquiler, las variables continuas 'metros cuadrados construidos' (en logaritmos) y 'antigüedad del edificio' mantienen su carácter continuo dentro de los cuatro y cinco tramos que se han distinguido, respectivamente. Las variables  $X2$  de este primer modelo son las cuatro modalidades del 'tamaño de municipio' y los cuatro grupos de provincias propuestos en el Apéndice II.1.

Como resultado de la estimación del primer modelo se obtiene que: i) se explica el 40 por ciento de la variabilidad total según el coeficiente de determinación ajustado ( $R_{aj}^2$ ); ii) no resultan significativos los efectos en el alquiler de las siguientes variables  $X1$ : 'metros cuadrados', 'garaje' y 'teléfono'

no'; iii) todas las variables significativas muestran el signo esperado salvo la variable 'antigüedad del edificio' <sup>17</sup>, y iv) en lo que respecta a las variables  $X2$ , el tipo de provincia 'rpro2' y el tamaño de municipio 'rmun2' no son significativamente distintas de las modalidades omitidas ('rpro1' y 'rmun1', respectivamente).

En el segundo modelo se evalúa el efecto cruzado de estas dos últimas variables en el alquiler de forma independiente al resto de variables ( $X1$ ). Para ello, se mantienen las mismas variables del primer modelo con la única variación de sustituir las variables 'tamaño del municipio' y 'provincia' por su consideración conjunta, las variables rloc1, rloc2,...,rloc9 definidas en el Apéndice II.1. Esto supone estimar ocho parámetros de localización en lugar de seis. Todas las variables  $X2$  son ahora significativamente distintas a la variable omitida ('rloc1'). Tanto la bondad de ajuste del modelo ( $R_{aj}^2 = 40,32$ ) como el valor estimado de todas las variables  $X1$ , son similares al primer modelo. Estos resultados aconsejan mantener la especificación conjunta para las variables  $X2$  en los siguientes modelos.

En el tercer y cuarto modelo se procede a considerar, de forma sucesiva, la forma lineal de las variables 'antigüedad del edificio' y 'metros cuadrados construidos'. Con ello se obtiene que sus efectos estimados en el alquiler sean significativos y del signo esperado. El resto de variables presenta una estimación similar a la de los modelos anteriores, por lo que, por el momento, se da por concluida la determinación de las variables  $X1$ ,  $X2$ .

En el paso siguiente se estudia la forma más apropiada de incluir la variable  $X3$ : 'Año de ocupación de la vivienda'. Como se ha expuesto anteriormente, esta variable permite estimar la diferencia media entre el alquiler que

---

<sup>17</sup>De los cinco tramos considerados, sólo los tres primeros son significativos, muestran un efecto similar y con el signo negativo esperado.

pagarían en 1980 dos inquilino de una vivienda de *AL* cuya única diferencia es la fecha de celebración del contrato. Para estimar el efecto correspondiente a cada año de ocupación, en el modelo 5 se incluyen 15 variables *dummy* ( $X3_{dummy}$ ) '*ocup<sub>t</sub>*', con  $t = 1965, \dots, 1979$ , de forma que se omite la ocupación en 1980-81, '*ocup80*'.

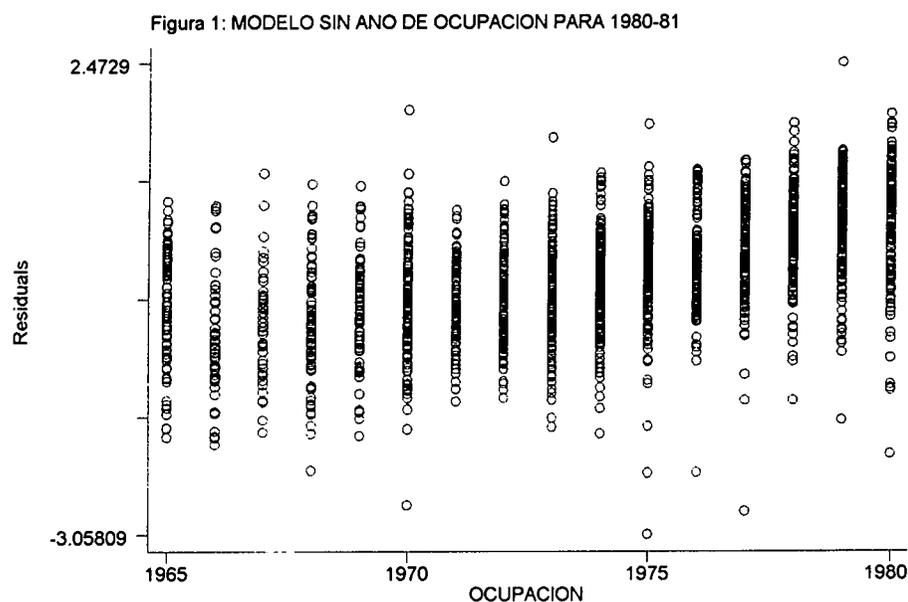
La inclusión de las variables  $X3_{dummy}$  reduce la raíz del error cuadrático medio (*RECM*) del modelo 4 en un 12,8 por ciento, se obtiene un  $R_{aj}^2 = 0,51$ , frente a un  $R_{aj}^2 = 0,38$  del modelo anterior, y no se experimentan cambios significativos en la estimación de las variables  $X2$ . El efecto en el alquiler de todas las variables  $X3_{dummy}$ , salvo el de '*ocup79*', resulta significativo y del signo esperado. Respecto a los efectos de las variables  $X1$  se observan los siguientes cambios: i) la variable 'posesión de teléfono' presenta un efecto positivo y significativo cuando no era así en los modelos anteriores, y ii) se reduce de forma sistemática el efecto estimado para el resto de variables.

La aplicación del método de Peña y Yohai (1995) para la depuración de observaciones anómalas permite detectar 90 atípicos influyentes en el modelo 5. En el modelo 6 se muestran los resultados equivalentes al modelo 5 después de prescindir de estos anómalos. Así, se observa que: i) el coeficiente de determinación ajustado es de 0,61, lo que supone un incremento de la variabilidad explicada de un 15 por ciento respecto al modelo anterior; ii) se reduce en un 20 por ciento la raíz del error cuadrático medio (*RECM*) asociada al modelo; iii) se estima una menor desviación típica para todos los parámetros; iv) en lo que respecta al sesgo que provocan los atípicos en los efectos estimados, los más significativo entre las variables  $X3$  se produce en las variables '*ocup71*' y '*ocup72*'; entre las variables  $X2$  destaca '*rloc4*' y '*rloc8*' y en las variables  $X1$  destaca '*calf1*' y '*gar1*', y v) el efecto de la variable garaje en el alquiler resulta significativo mientras que en los cinco

modelos anteriores no lo era.

En el modelo 6, aunque los efectos de las variables  $X3_{dummy}$  muestran una tendencia positiva respecto a  $t$ , no se estiman diferencias apreciables que justifiquen este nivel de desagregación, lo que hace pensar en la conveniencia de algún tipo de agrupación. Para determinar la forma de agregar estas variables se recurre a un análisis gráfico que se completa con un análisis descriptivo. En la Figura 1 se presenta el gráfico que relaciona los residuos del modelo 4, donde sólo se incluyen las variables  $X1$  y  $X2$ , con la variable 'año de ocupación'. Se observa una relación positiva entre ambas variables que se hace más evidente a partir de 1973, lo que sugiere el tratamiento no lineal de esta variable. Además, como se muestra en el Apéndice II.1, el alquiler medio de las viviendas en  $AL$  es de 3.416, si se ocupa entre 1965 y 1973, frente a 7.320 si se ocupa después. También la correlación del alquiler con la variable 'número de años ocupación' es distinta en ambos períodos: -0.12 en el primer período y -0.34 en el segundo. En el apéndice también se observa una cierta similitud en el nivel alcanzado por el alquiler cuando la vivienda es ocupada en 1978, 1979 y 1980 (10.023 pesetas de media), lo que

podría sugerir la existencia de un efecto similar en estos años.



A la vista de estos resultados, se estima el modelo 7 del Cuadro III.1. En él, junto a las variables  $X_1$  y  $X_2$  del modelo 6, se añaden las cinco variables en relación a la variable  $X_3$  referenciadas en el Apéndice II.1; éstas son: dos variables *dummy* ('a6573' y 'a7477'), cuya estimación recoge el efecto relativo a la ocupación posterior a 1977 (la variable omitida es 'a7880'); y tres variables continuas ('acup6573', 'acup7477' y 'acup7880') que representan el número de años que lleva ocupada la vivienda por los mismos inquilinos según el correspondiente período de ocupación con valor cero en el resto. Tanto la variable 'a6573' como las que recogen los tres efectos marginales, resultan significativas y con signo negativo, lo que muestra un descuento en el alquiler medio a favor del inquilino que renueva su contrato frente al de nueva contratación. La no significatividad de la *dummy* 'a7477' muestra un efecto similar al de la variable omitida, por lo que se elimina en el modelo 8. En el modelo 9, simplemente se sustituyen las variables 'acup7477' y 'acup7880' del modelo anterior por la variable 'acup7480' debido a la similitud de sus efectos

estimados, lo que permite reducir un parámetro sin perjuicio del ajuste del modelo.

Si bien la inclusión de las variables  $X3$  en su máximo nivel de desagregación y la depuración de atípicos en el modelo 6 mejoran la estimación de los modelos anterior, se puede justificar la superioridad del modelo 9 frente al 6 porque requiere la estimación de doce parámetros menos y los resultados obtenidos son equivalentes.

## 2. Modelos con índice de calidad para 1980

En este apartado se presenta la estimación de doce modelos donde se sustituyen las características físicas utilizadas en el modelo hedónico por el IC construido en el Capítulo II<sup>18</sup>. Los resultados de estas estimaciones se presentan en el Cuadro 2. En el primero modelo se estima sólo el efecto del IC y las variables de localización: tamaño de municipio y provincia en su versión más desagregada. Se obtiene que ambos tipos de variables explican el 35 por ciento de la variabilidad total del alquiler observado. Los efectos de las variables de localización son similares a los del primer modelo hedónico estimado, mientras que el IC tiene una incidencia significativa y positiva en la determinación del alquiler, lo que refleja que el mercado valora positivamente los niveles de servicios que dicho índice sintetiza. En el segundo modelo se sustituyen las variables 'tamaño de municipio' y 'provincia' por su consideración cruzada, lo que apenas eleva el valor del  $R_{aj}^2$ . La significatividad de la nuevas variables  $X2$  consideradas hace que se mantengan en el resto de los modelos estimados.

Al igual que se hizo en el modelo hedónico 5, en el modelo 3 con IC se incorporan las variables  $X3_{dummy}$  al modelo anterior de tal forma que se

---

<sup>18</sup>Por conveniencia se aplica un cambio de escala al IC obtenido directamente del ACM que consiste en sumarle 2,5 para que todos sus valores sean positivos.

puede estimar el efecto de cada año de ocupación ( $t = 65, \dots, 79$ ) en referencia al año 1980-81. El resultado es similar al observado en el caso hedónico. El efecto estimado de la ocupación muestra una tendencia negativa sobre el alquiler que eleva el  $R_{aj}^2$  del modelo 2 de 0,35 a 0,50.

Se observa que, aunque apenas varía el efecto estimado de las variables  $X2$ , la incorporación de  $X3$  (del modelo 2 al 3) modifica significativamente el efecto de  $X1$  (variable IC) lo que indica un problema de multicolinealidad que será tratado más adelante al estimar los efectos cruzados de ambos tipos de variables en el modelo.

Al someter el modelo 3 al método de Peña y Yohai (1995) para la estimación de los posibles datos muestrales anómalos se vuelven a detectar las mismas 90 observaciones atípicas del modelo hedónico. El modelo 4 del Cuadro 2 muestra el resultado de eliminar estas observaciones del modelo 3. Las principales diferencias entre ambos modelos son similares a las del modelo hedónico: i) el coeficiente de determinación ajustado se incrementa aproximadamente en un 16 por ciento, lo que significa que dicho modelo sin anómalos explica un 60 por ciento de la variabilidad total observada en el alquiler; ii) la RECM del modelo se reduce en un 20 por ciento; iii) se estima una menor desviación típica para todos los parámetros; iv) el efecto marginal estimado del IC en el alquiler apenas se incrementa un 1,8 por ciento, y v) las variables de localización '*rloc4*' y '*rloc8*' son las más afectadas, mientras que los efectos negativos de las *dummy* de ocupación que experimentan un mayor sesgo son '*ocup68*', '*ocup71*' y '*ocup72*'.

Los resultados obtenidos y la justificación para agregar  $X3$  en los modelos siguientes, del modelo 5 al 7, son prácticamente idénticos a los de los respectivos modelos hedónicos (modelos 7 a 9). Por tanto, hasta el momento y al igual que en el modelo hedónico, la agregación más razonable de  $X3$

estaría compuesta por la *dummy* 'a6573' y las variables que recogen el número de años de ocupación de la vivienda en los tramos de ocupación 1965-73 ('acup6573') y 1974-80 ('acup7480'), respectivamente. Se puede observar que aunque la bondad del ajuste del modelo con IC es ligeramente inferior a la del modelo hedónico directamente comparable ( $R_{aj}^2$ , de 0,59 frente a 0,60), se requiere la estimación de diez parámetros menos. Por lo demás, ambos modelos (el 9 hedónico y el 7 con IC) son equivalente en el sentido de que no se ha tratado el efecto cruzado entre las variables  $X1$  y  $X3$  o, lo que es lo mismo, en la ecuación (3.5) se supone que  $\partial g/\partial X_t = \partial g/\partial X_{t'}$ , para  $t \neq t'$  y  $X_t = [X1t, X2t]$  para todo  $t$ , cuando se ha comprobado que esta igualdad no se cumple en  $X1$ .

Para obtener una estimación del efecto diferencial de  $X1$  según  $t$ , en los modelos 8 a 12 del Cuadro 2 se relaja este supuesto con el fin de llegar al mejor modelo para 1980-81. En los cinco modelos se mantiene la misma especificación de las variables  $X2$ . En el modelo 8 se estima un posible efecto cruzado 'indt' distinto para cada año de ocupación  $t = 1965, \dots, 1980$ . Para ello se considera la variable IC y quince variables 'indl', tal que  $indl = IC$  si  $l = t$ , e  $indl = 0$  en caso contrario, siendo  $l = 1966, \dots, 1980$  el año de ocupación. De esta forma se estima el efecto marginal adicional al recogido con el IC, atribuible a un efecto diferencial de la calidad en cada año de ocupación. Los resultados muestran que hasta 1973 no existe un efecto adicional significativamente distinto a la variable de referencia, 'ind65'. Sin embargo, a partir de esta fecha, todos los efectos marginales adicionales de las variables 'indt' son significativamente distintos y, además, muestran una tendencia creciente en  $t$ .

En el modelo 9 se estiman los efectos marginales medios del IC según se ocupe antes o después de 1973 (variables 'ind6573' = IC si  $1965 \leq$  año

de ocupación  $\leq 1973$ ; = 0 en otro caso, y '*ind7480*' = IC si  $1974 \leq$  año de ocupación  $\leq 1980$ ; = 0 en otro caso) junto con los respectivos efectos adicionales del IC sobre '*ind7480*' cuando la ocupación es posterior a 1973 (variables '*ind74*' a '*ind79*'). Como resultado se obtiene que todas las variables, salvo '*ind79*', resultan significativas. En el modelo 10 se añaden a las variables anteriores las variables  $X3_{dummy}$ , lo que ocasiona que los efectos de las seis variables '*indt*' del modelo 9 dejan de ser significativas. Por este motivo, en el modelo 11 se mantienen únicamente las variables  $X1$ : '*ind6573*' y '*ind7480*' junto a las quince variables  $X3_{dummy}$  y las ocho variables  $X2$  de localización. Como se puede comprobar, los efectos estimados de las variables  $X3$ , aunque son distintos a los del modelo 4, nuevamente aconsejan la agrupación que conducía al modelo 7<sup>19</sup> y que se presenta en el modelo 12. En este proceso cabe destacar que en los modelos 8 a 12 las variables de control  $X2$  siguen mostrando un efecto estimado similar al de los modelos anteriores, manteniéndose también la bondad del ajuste ( $R_{aj}^2 = 0,60$ ). Estos resultados, permiten considerar al modelo 12 del Cuadro 2 como el más idóneo entre los modelos con IC analizados. A diferencia del modelo 7 con IC, en el modelo 12 se estima el efecto diferencial del IC en dos tramos de ocupación (1965-73 y 1974-80) y no se alteran significativamente los efectos estimados del resto de variables (salvo el de la *dummy* '*a6573*', que se reduce considerablemente, ya que parte del efecto que recogía esta variable viene ahora recogido por la diferencia entre la estimación de '*ind6573*' y '*ind7480*').

### 3. Modelos con índice de calidad para 1990

En este apartado se analiza el proceso de estimación de nueve modelos de alquiler para el régimen de *AL* en 1990-91. Los resultados de la estimación

---

<sup>19</sup>No se considera necesario presentar las estimaciones equivalentes a los modelos hedónicos, modelos 5 a 7 del Cuadro III.1, para comprobar esta deducción.

se presentan en el Cuadro 3 En el primer modelo únicamente se incluye el  $IC^{20}$  y el mayor nivel de desagregación de las dos variables de localización: provincias y tamaño del municipio. Como resultado se obtiene que: i) el modelo explica un 27 por ciento de la variabilidad total observada en el alquiler según el valor del  $R_{aj}^2$ ; ii) el nivel de calidad recogido con la variable IC tiene un efecto positivo y significativo sobre el precio del arrendamiento, y iii) mientras que los efectos de las provincias resultan significativos y de las magnitudes esperadas en relación a los datos observados en el Apéndice II.2, el efecto del tamaño de municipio no resulta significativo.

En el segundo modelo se mantiene la variable IC y se sustituyen las variables provincia y tamaño de municipio por su consideración conjunta. Esto supone la estimación de dos parámetros menos que en el modelo anterior, respecto al cual apenas mejoran los resultados obtenidos en cuanto al porcentaje de variabilidad explicada y el efecto del IC en el alquiler estimado. En lo que respecta a las variables de localización, se obtiene que las modalidades incluidas tienen efectos significativos y acorde con el análisis descriptivo mostrado en el Apéndice II.2.

Con el objeto de estimar el efecto independiente de cada uno de los años de ocupación de la vivienda, se estima el modelo 3 donde, manteniendo las mismas variables  $X1$  y  $X2$  del modelo 2, se añade una *dummy* (variable *ocupt*) para cada año ( $t$ ) desde 1965 a 1989. Por tanto, la interpretación del efecto estimado es en relación a la variable omitida *ocup90*: ocupación en 1990-91. Se debe advertir que, a diferencia del modelo equivalente estimado con datos de la EPF de 1980-81, aquí se deben estimar 10 parámetros más con la mitad de las observaciones, aproximadamente, lo cual puede influir en el

---

<sup>20</sup>Se aplica un cambio de escala en la variable IC que ha consistido en sumarle 16 al índice obtenido directamente del ACM.

valor y la significatividad individual de los efectos que se pretenden estimar. Los resultados del modelo 3 muestran que la inclusión de las 25 variables de ocupación reduce el error cuadrático medio del modelo anterior en un 18 por ciento y eleva el  $R^2$  ajustado de 0,27 a 0,51. A la vista de la estimación de los efectos individuales de cada año de ocupación en referencia a la ocupación en 1990 se obtiene que: i) la ocupación en cada uno de los años comprendidos entre 1965 y 1977 presenta un significativo efecto negativo sobre el precio, de forma que se estima una tendencia clara: el descuento en el alquiler es mayor cuanto mayor es el número de años que lleva ocupada la vivienda; ii) la ocupación entre los años 1978 y 1986 no resulta significativa y, iii) la ocupación en los tres años inmediatamente anteriores a la realización de la encuesta supone un alquiler medio estimado significativamente superior a la ocupación en 1990.

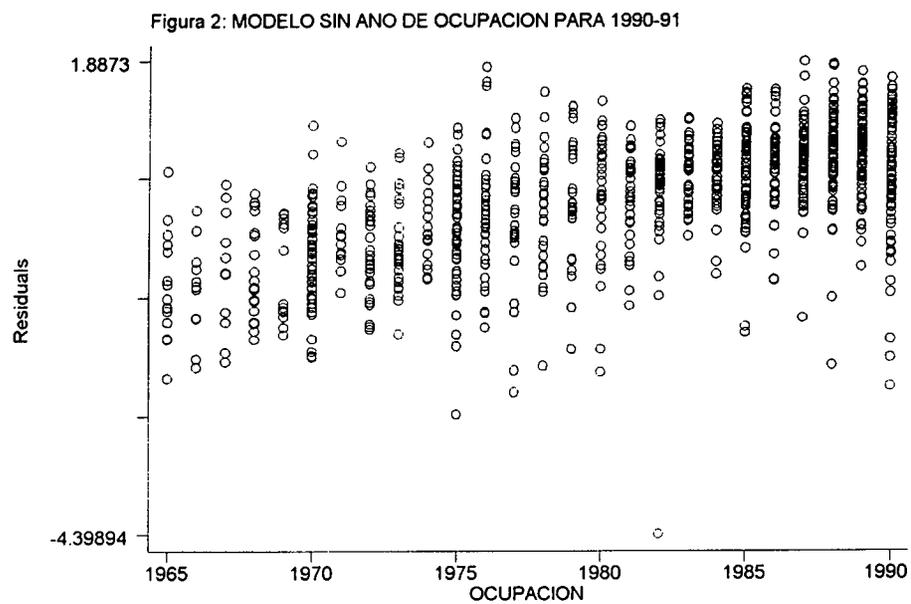
El método de Peña y Yohai (1995) permite detectar 53 observaciones que pueden estar distorsionando de forma significativa los efectos estimados de los determinantes del alquiler en el modelo 3. En el modelo 4 se muestran los resultados de haber eliminado los datos atípicos del modelo anterior. Se observa una menor desviación típica en todos los parámetros estimados, salvo en la variable 'ocup69', y una mejora considerable de la bondad del ajuste en relación al modelo sin depurar: el  $R_{aj}^2$  pasa de 0,510 a 0,596. La estimación del efecto del IC apenas disminuye su valor mientras que los cuatro parámetros de localización experimentan un ligero incremento sin alterar su magnitud relativa. En cuanto al año de ocupación, se obtiene que 13 de los 53 atípicos encontrados pertenecen a 1990. Una vez eliminadas estas observaciones, y a diferencia de lo obtenido en el modelo 3, el efecto de la ocupación entre 1987 y 1989 deja de ser significativamente distinto de la ocupación en 1990. También, de forma más clara que en el modelo anterior, en el modelo 4 se pueden

observar los tres períodos identificados en el análisis descriptivo respecto a su relación con el alquiler observado: 1965-75, 1976-82 y 1983-1990. De forma complementaria, en la Figura 2 se presenta un análisis gráfico de los residuos del modelo 2 (donde no se incluyen las variables  $X3$ ) en relación a la variable 'año de ocupación de la vivienda'. En este gráfico se revela la conveniencia de incluir tres variables de ocupación para explicar el efecto marginal de esta variable en cada tramo de ocupación. Así, en el modelo 5 se estima el efecto conjunto de la ocupación en estos períodos a través de dos variables *dummy* ('a6575' y 'a7682' con valor uno en 1965-75 y 1976-82 respectivamente) y tres variables continuas ('acup6575', 'acup7682' y 'acup8390') que representan el número de años que lleva ocupada la vivienda en régimen de *AL* cuando fue ocupada en el respectivo período y valor cero en otro caso. En este modelo no se obtienen cambios sustanciales respecto al anterior. El resultado de la estimación de las variables 'a6575' y 'a7682' muestra que los efectos fijos de la ocupación en los respectivos períodos no resultan ser significativamente distintos de la ocupación posterior. Consecuentemente, en el modelo 6 se eliminan estas dos *dummy*. Además de la misma estimación de los efectos de las variables  $X1$  y  $X2$  que en el modelo anterior, se obtiene una relación positiva entre el descuento interanual estimado de los años de ocupación y el alquiler con un comportamiento diferencial en los tres períodos de ocupación citados.

Finalmente, surge la cuestión, ya resulta en el modelo para 1980-81, sobre la posibilidad de que exista un efecto cruzado entre las variables  $X1$  y  $X3$  que aumente la capacidad para explicar el alquiler observado en la EPF de 1990-91. Junto a las variables  $X2$  de localización, en el modelo 7 se estima el efecto cruzado entre el IC y cada año de ocupación ( $t$ ) a través de 25 variables  $indl = IC$  si  $l = t$  e  $indl = 0$  en caso contrario, con  $t = 1966, \dots, 1990$  y la

variable IC. En relación al modelo anterior, el valor del  $R_{aj}^2$  se reduce de un 0,59 a un 0,54. Sin duda ésto se debe a que en este modelo no se han incluido las variables  $X3$  que permiten estimar el efecto medio del *tenure discount* para los inquilinos que renuevan su contrato de arrendamiento. Por su parte, los parámetros estimados de las variables  $X2$  apenas muestran variación. En cuanto a la estimación del efecto cruzado 'índice de calidad-ocupación', se observa una tendencia positiva del efecto marginal del IC respecto al año de ocupación y un comportamiento similar en los tres tramos de ocupación identificados anteriormente, lo que está indicando que, al igual que se obtuvo en los modelos de 1980-81, en 1990-91 el *tenure discount* tiene un componente diferencial según la calidad de la vivienda. A la vista de estos resultados, en el modelo 8 se estiman los efectos medios del efecto cruzado del IC en cada tramo de ocupación (variables '*ind6575*', '*ind7682*' y '*ind8390*' con valor: = IC si el año de ocupación ( $t$ ) pertenece al correspondiente período de ocupación; = 0 en otro caso) y las 25 variables de ocupación  $X3_{dummy}$ : '*ocupl*' con valor uno si  $l = t$ , para  $t = 1965, \dots, 1989$ . Los resultados de este modelo en relación al anterior son: i) el porcentaje de variabilidad explicada se eleva hasta superar el 60 por ciento; ii) el efecto estimado de las variables de localización apenas muestra variación; iii) el efecto marginal medio del IC es significativamente distinto en los tramos de ocupación considerados, y iv) la parte del *tenure discount* estimado de forma individual para cada año de ocupación sugiere de nuevo la agrupación para los mismos períodos: 1965-75, 1976-82 y 1983-90. Así, al igual que se hizo en el modelo 5, en el modelo 9 se sustituyen las 25 variables  $X3_{dummy}$  por las variables '*a6575*', '*a7682*', '*acup6575*', '*acup7682*' y '*acup8390*' con el resultado de que aquí ninguno de los efectos resulta estadísticamente irrelevante. Por tanto, con este modelo

se da por concluido el proceso de estimación del alquiler en *AL* para 1990-91.



MODELOS ALTERNATIVOS CONSIDERADOS PARA LA ESPECIFICACION DEL MODELO DE ALQUILER EN 1980-81 Y 1990-91

Variable dependiente: Inquiliner mensual

Estimacion por Minimos Cuadrados Ordinarios con errores estandar robustos

Se aplica el factor de elevacion poblacional

CUADRO 1. RESULTADOS DEL MODELO HEDONICO PARA LA EPF DE 1980-81

Modelo	Nº observaciones	F(24, 2117)	Prob > F	R2-ajustado	RMSE	Variables	Coef.	Std Err	Robust	Variables	Coef.	Std Err	Robust
Modelo 1	2142	50.18	0.0000	0.3986	0.7524	constante	8.670	0.540	16.067	constante	8.479	0.543	15.602
						h1g1	-0.552	0.092	-6.025	h1g1	-0.548	0.091	-6.045
						h1g2	-0.322	0.065	-4.937	h1g2	-0.320	0.065	-4.893
						agual1	-0.442	0.172	-3.480	agual1	-0.432	0.173	-3.459
						agual2	-0.201	0.118	-1.689	agual2	-0.207	0.117	-1.770
						agual3	-0.386	0.070	-5.484	agual3	-0.376	0.069	-5.410
						ca11z	-0.320	0.077	-4.176	ca11z	-0.316	0.077	-4.106
						ca12z	-0.096	0.070	-1.375	ca12z	-0.107	0.069	-1.540
						gar1	-0.027	0.045	-0.610	gar1	-0.035	0.045	-0.780
						gar2	0.032	0.129	0.247	gar2	0.039	0.130	0.303
						ln2v1	0.064	0.116	0.548	ln2v1	0.073	0.117	0.625
						ln2v2	0.076	0.109	0.689	ln2v2	0.086	0.109	0.763
						ln2v3	0.066	0.099	0.668	ln2v3	0.065	0.109	0.596
						ln2v4	0.074	0.106	0.682	ln2v4	0.082	0.099	0.825
						acon1	-0.004	0.001	-3.204	acon1	-0.003	0.001	-5.408
						acon2	-0.004	0.001	-3.204	acon2	-0.003	0.001	-5.408
						acon3	-0.004	0.001	-3.204	acon3	-0.003	0.001	-5.408
						acon4	-0.002	0.001	-1.178	acon4	-0.003	0.001	-2.986
						acon5	-0.002	0.001	-1.178	acon5	-0.003	0.001	-2.986
						acon6	-0.002	0.001	-1.178	acon6	-0.003	0.001	-2.986
						acon7	-0.002	0.001	-1.178	acon7	-0.003	0.001	-2.986
						acon8	-0.002	0.001	-1.178	acon8	-0.003	0.001	-2.986
						acon9	-0.002	0.001	-1.178	acon9	-0.003	0.001	-2.986
						acon10	-0.002	0.001	-1.178	acon10	-0.003	0.001	-2.986
						acon11	-0.002	0.001	-1.178	acon11	-0.003	0.001	-2.986
						acon12	-0.002	0.001	-1.178	acon12	-0.003	0.001	-2.986
						acon13	-0.002	0.001	-1.178	acon13	-0.003	0.001	-2.986
						acon14	-0.002	0.001	-1.178	acon14	-0.003	0.001	-2.986
						acon15	-0.002	0.001	-1.178	acon15	-0.003	0.001	-2.986
						acon16	-0.002	0.001	-1.178	acon16	-0.003	0.001	-2.986
						acon17	-0.002	0.001	-1.178	acon17	-0.003	0.001	-2.986
						acon18	-0.002	0.001	-1.178	acon18	-0.003	0.001	-2.986
						acon19	-0.002	0.001	-1.178	acon19	-0.003	0.001	-2.986
						acon20	-0.002	0.001	-1.178	acon20	-0.003	0.001	-2.986
						acon21	-0.002	0.001	-1.178	acon21	-0.003	0.001	-2.986
						acon22	-0.002	0.001	-1.178	acon22	-0.003	0.001	-2.986
						acon23	-0.002	0.001	-1.178	acon23	-0.003	0.001	-2.986
						acon24	-0.002	0.001	-1.178	acon24	-0.003	0.001	-2.986
						acon25	-0.002	0.001	-1.178	acon25	-0.003	0.001	-2.986
						acon26	-0.002	0.001	-1.178	acon26	-0.003	0.001	-2.986
						acon27	-0.002	0.001	-1.178	acon27	-0.003	0.001	-2.986
						acon28	-0.002	0.001	-1.178	acon28	-0.003	0.001	-2.986
						acon29	-0.002	0.001	-1.178	acon29	-0.003	0.001	-2.986
						acon30	-0.002	0.001	-1.178	acon30	-0.003	0.001	-2.986
						acon31	-0.002	0.001	-1.178	acon31	-0.003	0.001	-2.986
						acon32	-0.002	0.001	-1.178	acon32	-0.003	0.001	-2.986
						acon33	-0.002	0.001	-1.178	acon33	-0.003	0.001	-2.986
						acon34	-0.002	0.001	-1.178	acon34	-0.003	0.001	-2.986
						acon35	-0.002	0.001	-1.178	acon35	-0.003	0.001	-2.986
						acon36	-0.002	0.001	-1.178	acon36	-0.003	0.001	-2.986
						acon37	-0.002	0.001	-1.178	acon37	-0.003	0.001	-2.986
						acon38	-0.002	0.001	-1.178	acon38	-0.003	0.001	-2.986
						acon39	-0.002	0.001	-1.178	acon39	-0.003	0.001	-2.986
						acon40	-0.002	0.001	-1.178	acon40	-0.003	0.001	-2.986
						acon41	-0.002	0.001	-1.178	acon41	-0.003	0.001	-2.986
						acon42	-0.002	0.001	-1.178	acon42	-0.003	0.001	-2.986
						acon43	-0.002	0.001	-1.178	acon43	-0.003	0.001	-2.986
						acon44	-0.002	0.001	-1.178	acon44	-0.003	0.001	-2.986
						acon45	-0.002	0.001	-1.178	acon45	-0.003	0.001	-2.986
						acon46	-0.002	0.001	-1.178	acon46	-0.003	0.001	-2.986
						acon47	-0.002	0.001	-1.178	acon47	-0.003	0.001	-2.986
						acon48	-0.002	0.001	-1.178	acon48	-0.003	0.001	-2.986
						acon49	-0.002	0.001	-1.178	acon49	-0.003	0.001	-2.986
						acon50	-0.002	0.001	-1.178	acon50	-0.003	0.001	-2.986
						acon51	-0.002	0.001	-1.178	acon51	-0.003	0.001	-2.986
						acon52	-0.002	0.001	-1.178	acon52	-0.003	0.001	-2.986
						acon53	-0.002	0.001	-1.178	acon53	-0.003	0.001	-2.986
						acon54	-0.002	0.001	-1.178	acon54	-0.003	0.001	-2.986
						acon55	-0.002	0.001	-1.178	acon55	-0.003	0.001	-2.986
						acon56	-0.002	0.001	-1.178	acon56	-0.003	0.001	-2.986
						acon57	-0.002	0.001	-1.178	acon57	-0.003	0.001	-2.986
						acon58	-0.002	0.001	-1.178	acon58	-0.003	0.001	-2.986
						acon59	-0.002	0.001	-1.178	acon59	-0.003	0.001	-2.986
						acon60	-0.002	0.001	-1.178	acon60	-0.003	0.001	-2.986
						acon61	-0.002	0.001	-1.178	acon61	-0.003	0.001	-2.986
						acon62	-0.002	0.001	-1.178	acon62	-0.003	0.001	-2.986
						acon63	-0.002	0.001	-1.178	acon63	-0.003	0.001	-2.986
						acon64	-0.002	0.001	-1.178	acon64	-0.003	0.001	-2.986
						acon65	-0.002	0.001	-1.178	acon65	-0.003	0.001	-2.986
						acon66	-0.002	0.001	-1.178	acon66	-0.003	0.001	-2.986
						acon67	-0.002	0.001	-1.178	acon67	-0.003	0.001	-2.986
						acon68	-0.002	0.001	-1.178	acon68	-0.003	0.001	-2.986
						acon69	-0.002	0.001	-1.178	acon69	-0.003	0.001	-2.986
						acon70	-0.002	0.001	-1.178	acon70	-0.003	0.001	-2.986
						acon71	-0.002	0.001	-1.178	acon71	-0.003	0.001	-2.986
						acon72	-0.002	0.001	-1.178	acon72	-0.003	0.001	-2.986
						acon73	-0.002	0.001	-1.178	acon73	-0.003	0.001	-2.986
						acon74	-0.002	0.001	-1.178	acon74	-0.003	0.001	-2.986
						acon75	-0.002	0.001	-1.178	acon75	-0.003	0.001	-2.986
						acon76	-0.002	0.001	-1.178	acon76	-0.003	0.001	-2.986
						acon77	-0.002	0.001	-1.178	acon77	-0.003	0.001	-2.986
						acon78	-0.002	0.001	-1.178	acon78	-0.003	0.001	-2.986
						acon79	-0.002	0.001	-1.178	acon79	-0.003	0.001	-2.986
						acon80	-0.002	0.001	-1.178	acon80	-0.003	0.001	-2.986
						acon81	-0.002	0.001	-1.178	acon81	-0.003	0.001	-2.986
						acon82	-0.002	0.001	-1.178	acon82	-0.003	0.001	-2.986
						acon83	-0.002	0.001	-1.178	acon83	-0.003	0.001	-2.986
						acon84	-0.002	0.001	-1.178	acon84	-0.003	0.001	-2.986
						acon85	-0.002	0.001	-1.178	acon85	-0.003	0.001	-2.986
</													



Cuadro 3. RESULTADOS DEL MODELO CON INDICE DE CALIDAD PARA LA EPF DE 1990-91

Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6	Modelo 7	Modelo 8	Modelo 9
Nº observaciones = 1035 F(1, 1027) = 33.05 Prob > F = 0.0000 R2-ajustado = 0.2720 RMSE = 0.9127	Nº observaciones = 1035 F(5, 1029) = 53.20 Prob > F = 0.0000 R2-ajustado = 0.2741 RMSE = 0.9114	Nº observaciones = 1035 F(30,2119) = 30.65 Prob > F = 0.0000 R2-ajustado = 0.3103 RMSE = 0.7485	Nº observaciones = 982 F(30, 951) = 40.32 Prob > F = 0.0000 R2-ajustado = 0.5980 RMSE = 0.6297	Nº observaciones = 982 F(10, 971) = 81.05 Prob > F = 0.0000 R2-ajustado = 0.5980 RMSE = 0.6314	Nº observaciones = 982 F( 8, 973) = 98.05 Prob > F = 0.0000 R2-ajustado = 0.5945 RMSE = 0.6310	Nº observaciones = 982 F(30, 951) = 31.33 Prob > F = 0.0000 R2-ajustado = 0.5466 RMSE = 0.6672	Nº observaciones = 982 F(32, 949) = 40.16 Prob > F = 0.0000 R2-ajustado = 0.5804 RMSE = 0.6232	Nº observaciones = 982 F(12, 949) = 74.73 Prob > F = 0.0000 R2-ajustado = 0.6025 RMSE = 0.6247
Variables constantes Coef. 7.009 StdErr 0.114	Variables constantes Coef. 6.983 StdErr 0.110	Variables constantes Coef. 7.649 StdErr 0.088	Variables constantes Coef. 7.917 StdErr 0.083	Variables constantes Coef. 8.015 StdErr 0.084	Variables constantes Coef. 8.013 StdErr 0.084	Variables constantes Coef. 7.438 StdErr 0.117	Variables constantes Coef. 8.205 StdErr 0.117	Variables constantes Coef. 8.323 StdErr 0.117
Robust t 34.935 11.875	Robust t 28.925 12.073	Robust t 28.925 12.073	Robust t 33.995 11.757	Robust t 35.898 11.282	Robust t 36.549 11.174	Robust t 34.387	Robust t 34.694	Robust t 37.407
IC	IC	IC	IC	IC	IC	IC	IC	IC
rmun2 -0.054 rmun3 0.032 rmun4 -0.005 pmo2 0.334 pmo3 0.503 pmo4 0.697	fioc2 0.419 fioc3 0.575 fioc4 0.685 fioc5 0.949	fioc2 0.274 fioc3 0.518 fioc4 0.683 fioc5 0.772	fioc2 0.351 fioc3 0.585 fioc4 0.783 fioc5 0.878	fioc2 0.383 fioc3 0.605 fioc4 0.805 fioc5 0.903	fioc2 0.382 fioc3 0.605 fioc4 0.807 fioc5 0.904	fioc2 0.381 fioc3 0.610 fioc4 0.779 fioc5 0.916	fioc2 0.356 fioc3 0.590 fioc4 0.788 fioc5 0.876	fioc2 0.383 fioc3 0.607 fioc4 0.806 fioc5 0.898
occup5 -1.665 occup6 -1.530 occup7 -1.267 occup8 -1.019 occup9 -1.195 occup10 -1.006 occup11 -0.897 occup12 -1.082 occup13 -0.857 occup14 -0.887 occup15 -0.771 occup16 -0.405 occup17 -0.361 occup18 -0.054 occup19 -0.174 occup20 -0.227 occup21 -0.286 occup22 -0.087 occup23 0.206 occup24 0.119 occup25 0.178 occup26 0.409 occup27 0.403 occup28 0.519	occup5 -1.665 occup6 -1.530 occup7 -1.267 occup8 -1.019 occup9 -1.195 occup10 -1.006 occup11 -0.897 occup12 -1.082 occup13 -0.857 occup14 -0.887 occup15 -0.771 occup16 -0.405 occup17 -0.361 occup18 -0.054 occup19 -0.174 occup20 -0.227 occup21 -0.286 occup22 -0.087 occup23 0.206 occup24 0.119 occup25 0.178 occup26 0.409 occup27 0.403 occup28 0.519	occup5 -1.665 occup6 -1.530 occup7 -1.267 occup8 -1.019 occup9 -1.195 occup10 -1.006 occup11 -0.897 occup12 -1.082 occup13 -0.857 occup14 -0.887 occup15 -0.771 occup16 -0.405 occup17 -0.361 occup18 -0.054 occup19 -0.174 occup20 -0.227 occup21 -0.286 occup22 -0.087 occup23 0.206 occup24 0.119 occup25 0.178 occup26 0.409 occup27 0.403 occup28 0.519	occup5 -1.626 occup6 -1.657 occup7 -1.527 occup8 -1.311 occup9 -1.464 occup10 -1.273 occup11 -0.908 occup12 -1.100 occup13 -0.944 occup14 -1.046 occup15 -1.059 occup16 -0.694 occup17 -0.680 occup18 -0.385 occup19 -0.305 occup20 -0.503 occup21 -0.518 occup22 -0.510 occup23 -0.384 occup24 -0.144 occup25 -0.013 occup26 0.033 occup27 0.153 occup28 0.050	occup5 -1.025 occup6 -0.775 occup7 -0.620 occup8 -0.469 occup9 -0.317 occup10 -0.166 occup11 -0.015 occup12 0.139 occup13 0.289 occup14 0.439 occup15 0.589 occup16 0.739 occup17 0.889 occup18 1.039 occup19 1.189 occup20 1.339 occup21 1.489 occup22 1.639 occup23 1.789 occup24 1.939 occup25 2.089 occup26 2.239 occup27 2.389 occup28 2.539 occup29 2.689 occup30 2.839	occup5 -1.068 occup6 -1.311 occup7 -1.554 occup8 -1.797 occup9 -2.040 occup10 -2.283 occup11 -2.526 occup12 -2.769 occup13 -3.012 occup14 -3.255 occup15 -3.498 occup16 -3.741 occup17 -3.984 occup18 -4.227 occup19 -4.470 occup20 -4.713 occup21 -4.956 occup22 -5.199 occup23 -5.442 occup24 -5.685 occup25 -5.928 occup26 -6.171 occup27 -6.414 occup28 -6.657 occup29 -6.900 occup30 -7.143	occup5 -0.005 occup6 -0.001 occup7 0.017 occup8 0.038 occup9 0.059 occup10 0.080 occup11 0.101 occup12 0.122 occup13 0.143 occup14 0.164 occup15 0.185 occup16 0.206 occup17 0.227 occup18 0.248 occup19 0.269 occup20 0.290 occup21 0.311 occup22 0.332 occup23 0.353 occup24 0.374 occup25 0.395 occup26 0.416 occup27 0.437 occup28 0.458 occup29 0.479 occup30 0.500	occup5 -2.379 occup6 -2.398 occup7 -2.417 occup8 -2.436 occup9 -2.455 occup10 -2.474 occup11 -2.493 occup12 -2.512 occup13 -2.531 occup14 -2.550 occup15 -2.569 occup16 -2.588 occup17 -2.607 occup18 -2.626 occup19 -2.645 occup20 -2.664 occup21 -2.683 occup22 -2.702 occup23 -2.721 occup24 -2.740 occup25 -2.759 occup26 -2.778 occup27 -2.797 occup28 -2.816 occup29 -2.835 occup30 -2.854	occup5 -0.747 occup6 -0.631 occup7 -0.515 occup8 -0.400 occup9 -0.284 occup10 -0.168 occup11 -0.052 occup12 0.064 occup13 0.180 occup14 0.296 occup15 0.412 occup16 0.528 occup17 0.644 occup18 0.760 occup19 0.876 occup20 0.992 occup21 1.108 occup22 1.224 occup23 1.340 occup24 1.456 occup25 1.572 occup26 1.688 occup27 1.804 occup28 1.920 occup29 2.036 occup30 2.152

. Apéndice IV: Datos oficiales  
sobre la inflación y estimación  
del *tenure discount*

## 1. Datos oficiales sobre la inflación en el sector

El Instituto Nacional de Estadística (INE) proporciona información sobre la evolución del índice de precios de consumo (IPC) de una gran variedad de bienes y servicios con bases 1976, 1983 y 1992. Como es sabido, la información que sirve de base para establecer estos sistemas de IPC en España son las Encuestas de Presupuestos Familiares de 1973-74, 1980-81 y 1990-91, respectivamente. En el Cuadro 1 se presenta la tasa de variación interanual media del IPC,  $\Delta_{t+1,t}^{INE}$ , para el período comprendido entre 1966 y 1989, correspondiente a las viviendas en alquiler (Rúbrica 33) con base 1992.

Cuadro 1. Tasa de variación anual del alquiler

Año t	$\Delta_{t+1,t}^{INE}$	Año t	$\Delta_{t+1,t}^{INE}$
1966	7,85	1978	12,62
1967	8,87	1979	13,52
1968	6,80	1980	12,06
1969	1,34	1981	13,28
1970	4,87	1982	12,12
1971	5,91	1983	8,58
1972	5,10	1984	5,61
1973	10,27	1985	6,66
1974	12,85	1986	7,52
1975	13,98	1987	5,24
1976	11,04	1988	6,53
1977	13,37	1989	9,11

Fuente: Instituto Nacional de Estadística.

## 2. Estimación del *tenure discount*

En este apartado se presenta la estimación de los descuentos medios en el alquiler relativos a la nueva contratación en el año  $T = 1980$  ( $T = 1990$ ) que se derivan de la ocupación en  $t = 1965, \dots, T-1$ . En el Cuadro 2 se muestra la estimación del *tenure discount* ( $\widehat{D}_{T,t}$  de la ecuación (3.9)) según los modelos hedónico y con IC de 1980-81,  $\widehat{D}_{80,t}^H$  y  $\widehat{D}_{80,t}^{\overline{IC}}$ , para el nivel de calidad medio observado en las viviendas de  $AL$ ,  $\overline{IC} = 2,38$ . En la columna 3 se muestran los equivalentes descuentos porcentuales, relativos a la ocupación en 1980, para el subconjunto de viviendas ocupadas en 1980-81 que se mantienen ocupadas en 1990-91,  $\widehat{D}_{90/80,t}^{\overline{IC}}$ . Este valor se calcula según el modelo estimado en 1990-91 para un  $\overline{IC} = 14,85$  que corresponde al IC medio observado en dicho subconjunto de viviendas. En el Cuadro 3 se presenta la estimación del *tenure discount* según el modelo de 1990-91,  $\widehat{D}_{90,t}^{\overline{IC}}$ , para el nivel de calidad medio observado  $\overline{IC} = 15,46$ .

Cuadro 2. Estimación del  $D_{80,t}$  y  $D_{90/80,t}$ 

Año $t$	$\widehat{D}_{80,t}^H$	$\widehat{D}_{80,t}^{IC}$	$\widehat{D}_{90/80,t}^{IC}$
1965	-69,73	-71,08	-68,38
1966	-68,59	-69,97	-66,10
1967	-67,40	-68,81	-63,65
1968	-66,17	-67,61	-61,02
1969	-64,89	-66,37	-58,20
1970	-63,57	-65,07	-55,18
1971	-62,19	-63,73	-51,93
1972	-60,76	-62,33	-48,46
1973	-59,28	-60,89	-44,73
1974	-53,59	-54,60	-40,73
1975	-47,26	-48,22	-36,45
1976	-40,06	-40,93	-14,65
1977	-31,88	-32,62	-11,20
1978	-22,58	-23,14	-7,61
1979	-12,01	-12,33	-3,88

Cuadro 3. Estimación del  $D_{90,t}$ 

Año $t$	$\widehat{D}_{90}^{IC}$	Año $t$	$\widehat{D}_{90}^{IC}$
1965	-82,18	1978	-48,66
1966	-80,89	1979	-46,59
1967	-79,51	1980	-44,43
1968	-78,03	1981	-42,19
1969	-76,44	1982	-39,85
1970	-74,74	1983	-16,43
1971	-72,91	1984	-14,26
1972	-70,95	1985	-12,03
1973	-68,85	1986	-9,75
1974	-66,60	1987	-7,40
1975	-64,18	1988	-5,00
1976	-52,57	1989	-2,53
1977	-50,66		

## Capítulo 4

# Valoración del flujo de servicios de la vivienda

## 4.1 Introducción

Entre las motivaciones de esta tesis está el interés por considerar la importancia de la vivienda tanto en la medición del Índice de Precios de Consumo (IPC), como en su repercusión en el análisis de la desigualdad en España. Ambas consideraciones requieren la adjudicación de un valor al flujo de los servicios generados por un bien duradero tan complejo como es la vivienda, en el caso en que ésta esté en algún régimen distinto del arrendamiento. El objetivo de este capítulo es la imputación de un alquiler de mercado a la totalidad de las viviendas que compone el parque residencial en nuestro país, a precios corrientes de 1980 y 1990 respectivamente.

Para ello se utiliza la información sobre la vivienda que proporcionan las Encuestas de Presupuestos Familiares (EPF) con representatividad del total del parque residencial en 1980-81 y 1990-91 respectivamente. Estas encuestas, además de informar sobre las características físicas de la vivienda, el año de ocupación y la localización geográfica, disponen de la siguiente información sobre la valoración del flujo de servicios que se pretende estimar:

- 1) Si la vivienda está en régimen de propiedad o cesión, la EPF proporciona la autoestimación del encuestado sobre el alquiler de mercado que, según él, debería de pagar en 1980-81 (1990-91) por la vivienda que habita.
- 2) Si la vivienda pertenece al régimen de arrendamiento, la EPF proporciona el alquiler realmente pagado por el inquilino. Sin embargo, este valor difiere de la valoración corriente de mercado en 1980-81 (1990-91, respectivamente) en los siguientes casos: i) cuando la vivienda tiene la calificación legal (CL) de protección oficial (PO) el precio está por debajo del que imperaría en el mercado libre para una vivienda similar; ii) si la vivienda es de renta libre pero fue ocupada con anterioridad a la entrada en vigor de la Ley de Arrendamientos Urbanos (LAU) de 1964, se trata de una vivienda de 'renta

antigua' cuyo alquiler está afectado por la intervención gubernamental; iii) si la vivienda en arrendamiento tiene la CL de libre y fue ocupada después de la LAU de 1964, es decir, es de arrendamiento libre en sentido estricto (*AL*), el alquiler pagado estaría rebajado en la medida en que el propietario concede un *tenure discount* a los inquilinos en función del número de años que llevan ocupando la misma vivienda. En este punto se remite al Capítulo anterior donde se identifica el 'año de ocupación' como un determinante del alquiler, tanto en 1980-81 como en 1990-91. En consecuencia, la EPF proporciona el valor de mercado del flujo de los servicios de la vivienda en sentido estricto únicamente en las viviendas de *AL* ocupadas por primera vez en 1980-81 y en 1990-91 respectivamente. No obstante, la estimación de los efectos de los determinantes del alquiler en *AL* hace posible la estimación de un alquiler imputado a cualquier tipo de vivienda.

La imputación que se realiza en este capítulo es una adjudicación objetiva que procede de la estimación de un modelo econométrico que, al igual que en el capítulo anterior, trata de explicar los datos observados en la EPF. En el Capítulo III se identifican tres tipos de variables  $x$  que determinan el alquiler observado en las vivienda en *AL* (el índice calidad (*IC*) de la vivienda, la localización y el año de ocupación) y se realiza la estimación de sus efectos sobre el alquiler por mínimos cuadrados ordinarios (MCO). Sin embargo, este tipo de modelos sólo se puede utilizar para la imputación de un alquiler al resto de viviendas distintas del *AL* bajo el supuesto de independencia entre las decisiones del régimen de tenencia ('*AL*' frente a '*Otros*') y las características de la vivienda. En caso contrario, la estimación MCO de los efectos de  $x$  en el alquiler estará sesgada.

En este capítulo se plantea la necesidad de contrastar este supuesto a través de un modelo que permite corregir el sesgo de selección muestral en los

parámetros estimados cuando ambas cuestiones se deciden de forma simultánea. Para ello se considera la decisión de tenencia, *AL* u *Otros*, como un factor endógeno en la decisión del hogar sobre las características de la vivienda. Este planteamiento requiere abordar el análisis de los determinantes ( $z$ ) del régimen *AL* frente a *Otros*. En dicho análisis se comprueba la asociación de cuatro tipos de variables  $z$  con el régimen de tenencia: i) características físicas de la vivienda, ii) variables de localización: tamaño de municipio y provincia, iii) el año de ocupación, y iv) la condición de emigrante del sustentador principal de hogar<sup>1</sup>.

En consecuencia, en la especificación del modelo de este capítulo se tiene en cuenta que los datos observados sobre el alquiler de mercado están condicionados al régimen de *AL*. Los parámetros del modelo se estiman de forma robusta por máxima verosimilitud y como valores iniciales se consideran los resultados estimados en el modelo de selección de Heckman (1979) en dos etapas con el fin de buscar estimadores eficientes además de consistentes.

Una vez completada la fase de estimación, se procede a evaluar el efecto sobre el modelo final de considerar determinadas viviendas no incluidas en el modelo a través de un remuestreo por *bootstrap*. Se trata de las viviendas en arrendamiento sin calificación legal (CL), libre o PO, a las que se aplica un análisis de regresión logística que permite estimar su condición de '*AL*' frente a '*Otros*'. El objetivo último de este ejercicio es conocer el carácter, aleatorio o sesgado, de este tipo de viviendas sin CL. Los resultados obtenidos permiten llegar a la conclusión de que, para un nivel de significatividad del 95 por ciento, no hay evidencia en contra de su carácter aleatorio: la inclusión de estas observaciones en el modelo no introduce sesgo en la estimación de

---

<sup>1</sup>En principio, las variables  $x$  y  $z$ , a diferencia del alquiler de mercado, están disponibles para todo el parque residencia que compone la EPF.

los parámetros estimados.

Los modelos especificados para 1980-81 y 1990-91, proporcionan un mecanismo objetivo que, además de considerar la heterogeneidad del bien duradero valorado, permite imputar un valor, a precios corrientes de 1980 y 1990 respectivamente, al flujo de servicios de la vivienda a todo el parque residencial.

El análisis del alquiler imputado se completa con: i) una consideración pormenorizada de distintos tipos de viviendas según la disponibilidad de servicios en la vivienda, lo que permite comprobar una asociación entre la calidad del inmueble y el régimen de tenencia similar en ambas EPF, ii) una comparación, en términos medios, entre el alquiler imputado y el alquiler, real o autoestimado, que proporcionan las EPF de 1980-81 y 1990-91 para los distintos regímenes de tenencia, lo que hará posible, entre otras cosas, aproximar el grado de infra o sobrevaloración de la autoestimación que realiza el ocupante de la vivienda en régimen de propiedad o cesión.

Por otra parte, la imputación realizada para el conjunto de viviendas de la EPF tiene un interés específico, al menos, en cuatro posibles tipos de estudio: 1) una estimación del subsidio concedido a las viviendas de 'renta antigua' y de PO en referencia a lo que sería una valoración de mercado, 2) una evaluación del impacto distributivo de la política gubernamental en materia de vivienda en relación al tipo de hogares beneficiados: nivel y tipo de ingresos percibidos, edad del sustentador principal, nivel de educación, situación profesional, número de miembros del hogar,... 3) un análisis sobre desigualdad y bienestar de los hogares españoles en 1980-81 y 1990-91 respecto al gasto y calidad del bien vivienda, y 4) una estimación del efecto que llevaría consigo considerar el alquiler imputado a la hora de incluir la vivienda en régimen distinto al arrendamiento en la medición del índice de

precios de consumo (IPC) con base 1983 y 1992. Como es sabido, ambos indicadores fueron elaborados por el Instituto Nacional de Estadística (INE) en relación a la EPF de 1980-81 y 1990-91 respectivamente, aunque con criterios distintos a la hora de considerar la vivienda en propiedad y cesión. Efectivamente, mientras que en el IPC con base 1983 se considera el alquiler autoestimado por el ocupante de este tipo de viviendas, en el IPC con base 1992 no se utiliza esta información disponible en la EPF.

El trabajo que se presenta en este capítulo se estructura de la forma siguiente. En el segundo apartado se especifica el modelo de corrección del sesgo de selección. En el tercer y cuarto apartado se obtiene la valoración del parque de viviendas de la EPF de 1980-81 y 1990-91 respectivamente. En ambos casos se siguen las mismas etapas: i) se describen los datos disponibles y las variables consideradas; ii) se muestran los resultados de la estimación del modelo, así como una estimación de la influencia de aquellas viviendas en arrendamiento que no se incluyen en el modelo debido a la falta de información sobre su calificación legal; iii) se realiza la comparación entre el alquiler medio, real y autoestimación, observado en la EPF y el correspondiente alquiler imputado para distintos regímenes de tenencia. Por último, bajo el título de posibles extensiones, en el apartado seis se muestran distintas líneas abiertas de investigación para explotar las aportaciones de esta tesis.

## **4.2 Especificación del modelo**

El objetivo de este capítulo es determinar el valor del flujo de servicios de la totalidad de viviendas que compone el parque residencial español. En el Capítulo III se construyó un modelo que explicaba el alquiler ( $a$ ) observado en el sector de arrendamientos liberalizado ( $AL$ ) en la Encuesta de Presupuestos

Familiares (EPF). Se estimó que el modelo del alquiler  $a$  de los servicios la vivienda en  $AL$  en nuestro país seguía la siguiente expresión lineal

$$y = x\beta + \varepsilon, \quad (4.1)$$

siendo  $y = \ln(a)$  el logaritmo del alquiler,  $x$  el conjunto de explicativas y  $\varepsilon$  la perturbación aleatoria asociada al modelo.

La información sobre  $a$  tan solo está disponible para aquellas viviendas en régimen de  $AL$ . Si se denota por  $y_i$  al valor de  $y$  de la vivienda  $i$ , siendo  $i = 1, 2, \dots, N$ , y  $N$  el número total de viviendas que compone el parque residencial, la muestra estará dividida en dos grupos de viviendas: 'censuradas' si no se observa  $y_i$ , es decir, si  $i \notin AL$  y 'no censuradas' en caso contrario. Sea  $I$  la variable dicotómica que identifica el régimen de  $AL$  frente a los demás regímenes ('Otros' si  $i \notin AL$ ) a través de la siguiente expresión

$$I_i = 1 \iff i \in AL \iff z_i^* = z_i\gamma + u_i \geq 0$$

$$I_i = 0 \iff i \notin AL \iff z_i^* = z_i\gamma + u_i < 0$$

donde  $z$  es el conjunto de variables que permite clasificar a la vivienda  $i$  por régimen de tenencia: ' $AL$ ' u ' $Otros$ ',  $z^*$  es la variable latente que determina la selección muestral ( $y$  se observa sólo cuando  $z^* \geq 0$ ), y  $u$  la perturbación aleatoria asociada a dicha selección.

En estas circunstancias, la función de densidad condicionada que se puede observar es:

$$f(y, z^*/z^* \geq 0) = \frac{f(y, z^*)}{\Pr(z^* \geq 0)},$$

siendo  $f(y, z^*)$  la función de densidad conjunta de las variables aleatorias  $y$  y  $z^*$ , y  $\Pr(z^* \geq 0)$  la probabilidad de pertenecer al conjunto de viviendas  $AL$ .

En consecuencia, el modelo (4.1) estimado en el Capítulo III no es válido para estimar el alquiler  $a$  de la totalidad de viviendas muestrales en el caso

en que las decisiones sobre tenencia y características de la viviendas sean independientes. Por tanto, el truncamiento selectivo de la variable  $y$  obliga a reformular el modelo (4.1) a través de un mecanismo que corrija el posible sesgo producido por las particularidades muestrales de la información disponible cuando el proceso de decisión, sobre tenencia y características, se realiza simultáneamente.. De esta forma, la especificación del modelo que permitirá imputar un alquiler de mercado a toda vivienda  $i = 1, \dots, N$  se puede expresar de la forma siguiente:

1.- Modelo de regresión

$$y_i = x_i\beta + \varepsilon_i, \text{ observado si } i \in AL$$

donde  $\varepsilon$  se distribuye como una normal  $\varepsilon \sim N(0, \sigma)$ .

2.- Mecanismo de selección

$$z_i^* = z_i\gamma + u_i,$$

que, bajo el supuesto de que  $u$  se distribuye como una normal  $u \sim N(0, 1)$ , corresponde al modelo Probit siguiente

$$I_i = 1 \iff z_i^* \geq 0; \text{ y } I_i = 0 \text{ en otro caso}$$

de forma que

$$\Pr(i \in AL) = \Pr(I_i = 1) = \Phi(z_i\gamma), \text{ y}$$

$$\Pr(i \notin AL) = \Pr(I_i = 0) = 1 - \Phi(z_i\gamma)$$

donde  $\Phi$  es la función de distribución de la normal estándar.

En este contexto, donde  $\rho$  es el coeficiente de correlación entre las perturbaciones aleatorias  $u$  y  $\varepsilon$ , la función log-verosimilitud para cada observación muestral  $i$  (con factor de elevación poblacional  $w_i$ ) es

$$l_i = \begin{cases} w_i \ln \Phi\left(\frac{z_i\gamma + (y_i - x_i\beta)\rho/\sigma}{\sqrt{1-\rho^2}}\right) - \frac{w_i}{2} \left(\frac{y_i - x_i\beta}{\sigma}\right)^2 - w_i \ln(\sqrt{2\pi}\sigma), & \text{si } i \in AL \\ w_i \ln \Phi(-z_i\gamma) & \text{si } i \notin AL \end{cases}$$

Dado que las  $N$  observaciones son independientes, la función a maximizar es

$$L(\beta, \gamma, \sigma, \rho/y, x, z) = \sum_{i=1}^N l_i. \quad (4.2)$$

Como este modelo no es lineal en los parámetros, en el proceso de estimación se utilizarán como valores iniciales, las estimaciones consistentes procedentes del método de Heckman (1979) en dos etapas. Con ello se simplifica el procedimiento computacional a la hora de conseguir parámetros consistentes y asintóticamente eficientes en la estimación de (4.2) por Máxima Verosimilitud (MV). Por otra parte, para asegurar la identificación del modelo a través de restricciones de exclusión se deben seleccionar variables 'z' que no pertenezcan al conjunto de variables 'x'.

De esta forma, la estimación de los parámetros  $\beta$ , junto con el vector de características  $x_i$ , permite aproximar el alquiler corriente de mercado a toda vivienda  $i$ .

### 4.3 Valoración del parque de viviendas a precios de 1980

En este apartado se analiza sucesivamente: i) la información disponible en la EPF de 1980-81; ii) los resultados de la estimación del modelo (4.2); iii) el efecto que produce en la estimación de los parámetros la inclusión de las viviendas en arrendamiento sin información sobre la calificación legal; y iii) la comparación del alquiler corriente imputado a toda las viviendas de la EPF con la autoestimación realizada por sus ocupantes.

En el Cuadro 4.1 se presenta la distribución de frecuencias del parque de viviendas principales en 1980-81 según el régimen de tenencia: propiedad, arrendamiento o cesión. A su vez, este régimen aparece clasificado según la

calificación legal: libre, protección oficial (PO) o desconocida, de tal forma que en el caso particular del arrendamiento libre se muestra la distinción entre *AL* en sentido estricto (ocupación anterior a la Ley de arrendamientos Urbanos de 1964) y de renta antigua (ocupación anterior a 1965). En la primera columna aparece el número total de observaciones muestrales disponibles en la EPF de 1980-81. Hay 13 viviendas en arrendamiento no incluidas en la desagregación por CL que están identificadas en la EPF como de arrendamiento libre pero sin año de ocupación, lo que impide su posterior clasificación entre *AL* o renta antigua. En las columnas 2 y 3 se muestran, respectivamente, la representación población de cada tipo de vivienda y su peso relativo respecto al total.

Como ya se había constatado en el Capítulo I, el régimen de vivienda más frecuente en nuestro país es la propiedad y la calificación legal de libre. La falta de información en relación a la CL, que supone un 11 por ciento entre las viviendas en propiedad y un 9,7 por ciento en el arrendamiento, representa un 9,9 por ciento en la EPF de 1980-81. Más adelante se comentarán las implicaciones de incluir estas observaciones en la estimación de los parámetros del modelo.

### 4.3.1 Datos y variables

Las variables  $x$  que mejor explican el alquiler en el régimen de *AL* fueron identificadas en el Capítulo III<sup>2</sup>. Por su parte, la identificación de las variables  $z$  que permiten la clasificación de viviendas en *AL* y *Otros*, ha requerido el análisis descriptivo previo del Apéndice I.1. donde se muestra la asociación existente entre un conjunto de posibles variables  $z$  y el régimen de tenencia:

---

<sup>2</sup>El análisis descriptivo que ayuda a determinar las variables  $x$  se muestra en el Apéndice II.1 del Capítulo III.

*AL* u *Otros*. Cada variable aparece al máximo nivel de desagregación posible<sup>3</sup> con objeto de proceder a la agrupación de modalidades más coherente con los datos observados. Estas variables hacen referencia a los cuatro aspectos siguientes: i) la condición de migrante de sustentador principal, ii) las características de la vivienda, iii) el año de ocupación, y iv) la localización.

En las columnas 1, 2 y 3 se muestra la distribución porcentual de viviendas en *AL*, en *Otros* y la distribución total, respectivamente. La información disponible sobre las posibles variables explicativas condiciona el tamaño muestral de las viviendas que se incluyen en el modelo, de tal forma que en la EPF de 1980-81 se dispone de una muestra compuesta por 2.052 viviendas en *AL*<sup>4</sup> y 19.470 en *Otros*<sup>5</sup>, representativas de 834.046 y 8.211.540 viviendas al aplicar el correspondiente factor de elevación poblacional. En la columna 4 se presenta la proporción de viviendas que representa cada modalidad en la variable correspondiente mientras que, en la columna 5 (columna 6) se presenta el porcentaje de viviendas en *AL* (*Otros*) respecto a cada modalidad. La comparación de estos porcentajes con la proporción media de viviendas en *AL* (un 9,22 por ciento) ayuda a identificar tanto el conjunto como la forma más apropiada de incluir las variables  $z$  en el modelo. Así, por ejem-

---

<sup>3</sup>Su descripción se encuentra en los documentos Alonso-Colmenares et al. (1994) y los datos están disponibles en la página web <<http://www.eco.uc3m.es/epf80-81.html>>.

<sup>4</sup>Como se vió en el Capítulo III, en un principio hay 2.181 viviendas de arrendamiento libre de las que fueron eliminadas 21 por no disponer de año de construcción; 18 por no proporcionar información sobre el alquiler, y 90 se identifican como atípicos en el modelo que explica el alquiler observado.

<sup>5</sup>En principio hay 21.790 viviendas, de las cuales 548 son de arrendamiento pero no es posible su clasificación entre *AL* o *Otros* (535 no proporcionan la calificación legal: libre frente a protección oficial y en 13 no se facilita el año de ocupación); y en otros 1.772 casos no se dispone de variables relevantes en el modelo (hay 1.728 viviendas sin 'año de ocupación'; 36 sin año de construcción y 8 sin información en ambas variables).

plo, la proporción de viviendas en *AL* en el grupo de viviendas caracterizado porque el sustentador principal inmigró al municipio después de 1975 es de un 29 por ciento, lo que muestra una asociación superior a la media entre este tipo de viviendas y el *AL*. Este análisis descriptivo se completa con un proceso de estimación en el que se prueba con distintos modelos alternativos para determinar la elección más apropiado de las variables  $z$ . Respecto a este proceso interactivo se pueden hacer las siguientes comentarios:

1. La 'condición de migrante del sustentador principal' es la variable que en mayor medida ayuda a discriminar entre el régimen de *AL* y *Otros*. Además, es independiente del nivel de alquiler por lo que se trata de una variable instrumental que permite la identificación del modelo 4.2.

2. Entre las variables  $z$  no se incluye el IC que sintetiza las características físicas de la vivienda debido a que, a diferencia de su efecto en el alquiler, su asociación con el régimen de tenencia no es lineal. Así, se observa, por ejemplo, que las modalidades extremas de la variable servicios higiénicos, 'no disponibilidad de servicios' y 'disponer de uno o más cuarto de baño completo', son las más asociadas al régimen *Otros*, mientras que la modalidad 'un baño no completo' es la más asociada al *AL*. El resto de características  $z$  influyentes en el régimen de tenencia son: disponibilidad de agua caliente, garaje dentro del edificio, número de metros cuadrados construidos, disponibilidad de teléfono y antigüedad del edificio. Por su parte, la disponibilidad de calefacción no ayuda a discriminar por régimen de tenencia.

3. La variable 'año de ocupación' muestra una relación positiva con el *AL*.

4. El efecto de la localización sobre el régimen de tenencia se considera a través de las variables tamaño de municipio y la provincia. El régimen de *AL* se asocia en mayor medida con municipios de mayor población de

forma que se pueden definir los tramos: i) menos de 10.000 habitantes, ii) entre 10.001 y 50.000, y iii) más de 50.000 habitantes. Las variables definidas sobre la provincia responden a la agrupación propuesta en el Apéndice I.1 en relación a la similitud de los datos observados. En la columna 7 del apéndice se muestra la notación de las variables incluidas en el modelo final.

En el Cuadro 4.2 se definen las variables  $x$  y  $z$  del modelo (4.2) cuyos resultados se presentan en el siguiente apartado.

### 4.3.2 Resultados de la estimación

Los resultados del modelo con corrección del sesgo de selección (4.2) para 1980-81 se muestran en el Cuadro 4.3. En este tipo de modelos la significatividad del parámetro  $\rho$  indica la importancia de considerar el modelo de selección muestral frente a su omisión. En este caso, no hay evidencia para rechazar la hipótesis nula  $H_0 : \rho = 0$ .

Como es sabido, los efectos de las variables  $z$  no son directamente interpretables en términos de probabilidades estimadas de pertenecer al régimen *AL* u *Otros*. Sin embargo, el signo positivo en el parámetro estimado revela una asociación positiva entre la variable correspondiente y el régimen *AL* mientras que, un signo negativo indica una mayor asociación con el régimen *Otros*. Así, los factores con mayor probabilidad estimada de pertenecer al *AL* son: que el sustentador principal resida en el municipio desde hace menos de 5 años; la antigüedad del edificio superior a 11 años y en mayor medida si supera los 30 años; la ocupación más reciente; el mayor tamaño de municipio donde se encuentra la vivienda y su localización, por ejemplo, en Barcelona, Madrid o Baleares frente a su ubicación en provincias como Zamora o Cuenca.

Por otra parte, la consideración de las trece variables ' $x$ ' que intervienen

en el análisis para explicar el precio de alquiler, resulta claramente significativa, como indican los valores alcanzados por el estadístico  $\chi^2$ . En lo que respecta a la interpretación económica de los efectos de las variables  $x$ , se remite al modelo de alquiler estimado en el capítulo anterior. Con objeto de comparar aquellos resultados con los obtenidos en este capítulo, en el Cuadro 4.4 se proporciona el efecto porcentual medio de cada variable sobre el alquiler según el modelo sin corrección del sesgo de selección (columna 1) y según el modelo corregido (columna 2), así como el intervalo de variación de este último para un nivel de confianza del 95 por ciento (columna 3). Como se puede comprobar, las magnitudes medias son muy similares en los dos modelos y en ningún caso el efecto medio estimado experimenta una variación superior al 8,6 por ciento. Las variables más afectadas por la corrección son: 'loc2' (cuyo efecto pasa de un 16,31 a un 14,42), 'loc5' (de un 55,19 a un 58,69 por ciento), 'loc9' (de un 127,02 a un 134,98 por ciento) y 'acup6573' (de un 3,77 a un 4,04 por ciento). Como se puede observar en la columna 3 del Cuadro 4.4, en todas las variables se cumple que, su efecto estimado en el primer modelo está dentro del intervalo de confianza estimado en el modelo corregido.

### 4.3.3 Influencia de las observaciones sin CL

En las EPF hay hogares en régimen de arrendamiento que no proporcionan información sobre su calificación legal (CL): Libre o PO. La CL en el régimen de propiedad resulta irrelevante a la hora de estimar el modelo (4.2), tanto la propiedad libre como en PO pertenecen al conjunto de viviendas censuradas, grupo 'Otros'. Sin embargo, en el régimen de arrendamiento no ocurre lo mismo, así, mientras que la calificación libre de renta antigua y la PO formarían parte de las viviendas censuradas 'Otros', sólo la calificación

libre con ocupación posterior a 1964 está en el grupo de 'AL'.

En este apartado se cuestiona la repercusión que pudiera tener en la estimación del modelo (4.2) la consideración de aquellas viviendas en régimen de arrendamiento que no se incluyeron en el modelo por falta de información sobre la CL. En el Apéndice II se desarrolla con todo detalle un análisis que permite estimar la influencia de este tipo de observaciones en el modelo (4.2), tanto para 1980-81 como para 1990-91. El contenido del apéndice se estructura en dos apartados. En el apartado I se define una regla de clasificación discriminante que permite estimar la CL de toda vivienda en arrendamiento en relación a la probabilidad (estimada en un análisis de regresión logístico) de pertenecer al régimen 'libre', frente a la alternativa de 'PO'. Esta regla, a diferencia de la clasificación: 'libre', 'PO' que se derivaría directamente de la estimación logística, se define sobre un umbral de probabilidad considerado óptimo porque maximiza el número de asignaciones 'libre' a la vez que se penaliza en mayor medida el error que supone asignar la CL 'libre' cuando la vivienda es de 'PO', que el error contrario. Haciendo uso de las estimaciones: libre,  $\widehat{L}$ , y PO,  $\widehat{PO}$  para las viviendas en arrendamiento sin CL, en el apartado II del apéndice se plantea la posible influencia de estas observaciones en el modelo (4.2). Para ello se procede en tres etapas: i) identificar la clasificación de estas viviendas sin CL en los regímenes de tenencia que requiere el modelo, es decir, identificar el régimen libre en sentido estricto,  $\widehat{AL}$ <sup>6</sup>, y el resto de regímenes,  $\widehat{Otros}$ <sup>7</sup>, ii) eliminar las observaciones con calificación  $\widehat{AL}$  que, al igual que se comprobó en el Capítulo III con ciertas viviendas

---

<sup>6</sup>Viviendas en arrendamiento sin CL con régimen estimado  $\widehat{L}$  que fueron ocupadas después de 1964.

<sup>7</sup>Viviendas en arrendamiento sin CL con régimen estimado  $\widehat{L}$  que fueron ocupadas antes de 1965 (renta antigua), más las de régimen estimado de  $\widehat{PO}$ .

en régimen *AL*, tienen un comportamiento atípico en el modelo (4.1)<sup>8</sup>, y iii) medir, en términos de eficiencia y sesgo, el efecto de incluir las viviendas  $\widehat{AL}$  y  $\widehat{Otros}$  en el modelo. Con el objeto de analizar cada uno de estos efectos de forma independiente, se aplicará la metodología *bootstrap* de tal forma que se seleccionan muestras aleatorias con reemplazamiento y se reestiman el modelo (4.2) en tres procesos distintos según el número y tipo de observaciones consideradas: con o sin información estimada de la CL. El análisis de los resultados *bootstrap* permitirá dar respuesta a dos cuestiones:

- 1<sup>a</sup>.- ¿Cuál es la mejora de eficiencia en el modelo, es decir, la reducción en la desviación típica de los parámetros estimados, debido únicamente a incrementar las observaciones con CL en un número igual al que representan las viviendas en arrendamiento sin CL?
- 2<sup>a</sup>.- ¿Qué sesgo provocaría en la estimación de los parámetros del modelo el hecho de incluir las viviendas en arrendamiento con CL estimada?

Se puede observar que, la respuesta a la segunda pregunta puede suponer una evidencia empírica en contra del carácter aleatorio de las viviendas en arrendamiento sin CL en lo que al modelo (4.2) se refiere. Es decir, en función de la significatividad del sesgo obtenido, se puede detectar la existencia de un posible sesgo muestral en el modelo estimado puesto que no incluye este tipo de viviendas.

En lo que respecta a la EPF de 1980-81, el número de viviendas en arrendamiento sin CL es de 535 unidades muestrales, lo que representa un 2,23

---

<sup>8</sup>Se aplicará la metodología de Peña y Yohai (1995) a los modelos de alquiler del Capítulo III donde, además de las viviendas *AL*, se incluyen las viviendas  $\widehat{AL}$  con el fin de eliminar aquellas observaciones que puedan introducir sesgo en el modelo debido a su condición de atípicos.

por ciento del total de viviendas y un 9,7 de las viviendas en arrendamientos (véase Cuadro 4.1). El análisis realizado sobre estas viviendas en el Apéndice II<sup>9</sup> se puede concretar en los siguientes resultados en relación al modelo (4.2):

- 1.- El mayor número de observaciones recuperadas que se pueden considerar válidas para ser incluidas en el modelo (4.2) son 508; en 203 se estima la CL de  $\widehat{AL}$  y en 305 de  $\widehat{Otros}$ . Aunque en un principio se estiman 217 viviendas en  $\widehat{AL}$ , 14 de ellas se identifican como observaciones atípicos en el modelo de alquiler (4.1), por lo que no se considerarán en el análisis.
- 2.- La mejora de eficiencia en el modelo (4.2) debida únicamente a un incremento del tamaño muestral de 508 observaciones, seleccionadas aleatoriamente entre las viviendas con información sobre la CL, se estima en un valor medio de un 3,91 por ciento para el conjunto de los parámetros estimados, con un valor máximo de un 5,26 por ciento que corresponde a la reducción de la desviación típica del parámetro asociado a la variable 'ind6573'.
- 3.- No hay evidencia de un sesgo significativo en los efectos estimados de las variables en el modelo (coeficientes de variables) por incluir las 508 viviendas en arrendamiento con CL estimada para un nivel de confianza del 95 por ciento.

Con estos resultados se puede concluir que si bien es verdad que en principio la condición de CL estimada haría desaconsejable la inclusión de estas

---

<sup>9</sup>Al aplicar la regla de clasificación a las 535 arrendamientos sin CL se estima, con un error estimado del 32 por ciento, que 217 son de 'AL', 122 de 'renta antigua' y 183 de 'PO'. En los 13 casos restantes no es posible asignar una CL debido a la falta de información sobre variables explicativas del modelo discriminante (en 5 hogares no se proporciona el 'año de ocupación', en 7 el 'año de construcción' y en 1 ninguno de los dos).

observaciones en el modelo, no hay evidencia de que, en caso de hacerlo, se produzca un sesgo significativo en la estimación del efecto de ninguna de las variables en el alquiler.

Para ilustrar la magnitud de estos efectos se remite al Cuadro 4.4 donde, junto a la estimación del efecto marginal de cada variable  $x$  según el modelo de referencia con información de la CL (columnas 2 y 3), se muestra el correspondiente efecto cuando se incluyen las viviendas con CL estimada (columna 4). Como se puede observar, el sesgo, o efecto diferencial, en ningún caso supera los límites establecidos en el intervalo de confianza (columna 3) para estimar el efecto en el alquiler de cada variables para un nivel de confianza del 95 por ciento. En este sentido, se puede decir que no hay evidencia en contra del carácter aleatorio de las viviendas en arrendamiento que no proporcionan información sobre la CL en las EPF de 1980-81.

#### **4.3.4 Comparación del alquiler medio observado e imputado**

Este apartado se compara el alquiler observado en la EPF, real o autoestimación, con el alquiler imputable vía estimación de los efectos de las variables  $x$  en el alquiler según el modelo (4.2). Los resultados de dicha estimación hacen posible la imputación de un alquiler corriente (a precios de 1980) a toda vivienda con información sobre las variables ' $x$ '. En este sentido, sólo las variables 'año de ocupación', y 'antigüedad del edificio' presentan falta de información en 1.755 y 73 viviendas de la EPF<sup>10</sup> de 1980-81 respectivamente.

La primera variable no supone problema. A la hora de imputar un alquiler

---

<sup>10</sup>Para mayor detalle se remite al segundo filtro aplicado a los datos disponibles en las EPF y definido en el Capítulo I. En él se muestra el tipo de viviendas afectadas y se realiza una corrección en relación a ciertas incoherencias detectadas entre ambas variables.

de mercado, se toma 1980 como 'año de ocupación' ya que es la fecha en que están valorados los alquileres observados en arrendamiento y, también, es el año de referencia para la pregunta de la EPF sobre alquiler autoestimado en caso de propiedad o cesión. En consecuencia, a la hora de computar el efecto del índice de calidad (IC), el cual depende del período de ocupación estimado con las variables 'ind6573' e 'ind7480' respectivamente, se considera únicamente la segunda variable por recoger el efecto asociado a la ocupación en el año en que se realiza la encuesta.

En cuanto a la antigüedad del edificio, variable que condiciona el IC, en 1980-81 se asigna un 'año de construcción' a aquellas viviendas que no proporcionan tal información. El criterio seguido es asignar el año medio correspondiente a las viviendas de igual régimen de tenencia<sup>11</sup>. Con esta aproximación se consigue disponer de una estimación del valor corriente del flujo de los servicios de la vivienda para todo parque residencial de la EPF. Se trata de un alquiler imputado que representa una estimación objetiva del valor que el mercado libre asignaría a cada vivienda en 1980 en función de la calidad de los servicios que la compone y su localización: tamaño de municipio y provincia.

La información que se muestra en el Cuadro 4.5 hace referencia al alquiler medio observado en la EPF 1980-81 y a la correspondiente imputación de mercado estimada a precios corrientes. El conjunto de viviendas presentado está restringido a la disponibilidad de información sobre el alquiler obser-

---

<sup>11</sup>La asignación del 'año de construcción' es de: 1932 a propiedad libre (13 casos); 1966 a propiedad de Protección Oficial (3 casos); 1924 a propiedad sin calificación legal (3 casos); 1910 a arrendamiento libre con ocupación previa a la 1965 (7 casos); 1937 a arrendamiento libre con ocupación posterior a 1964 (21 casos); 1928 a arrendamiento de Protección Oficial (13 casos); 1927 a arrendamiento sin calificación legal (8 casos) y 1931 a algún tipo de cesión (5 casos).

vado, real en caso de arrendamiento, o autoestimación en otro régimen<sup>12</sup>. Esto permite que la comparación, alquiler observado-alquiler imputado, sea relevante ya que está referida al mismo conjunto de viviendas. En la primera columna del cuadro se presenta la distribución porcentual de las viviendas según el régimen de tenencia y la CL conocida (un total de 21.556 observaciones muestrales representativas de 8.991.308 a nivel poblacional). En la columna 2 se muestra el alquiler medio observado (desviación típica, DT) mientras que en la columna 3 se muestran los respectivos valores imputados. En la columna 4 se presenta la diferencia entre el alquiler medio observado e imputado expresada en términos porcentuales respecto al valor imputado. Obsérvese que, en términos estimados, el signo negativo en el régimen de propiedad (cesión) muestra que el propietario (el beneficiario de la cesión) infravalora el alquiler que debería de pagar en el mercado libre en 1980 por la vivienda que ocupa. Sin embargo, el signo negativo en el arrendamiento de renta antigua y PO representa el descuento medio que el inquilino obtiene por no arrendar su vivienda en el mercado libre en 1980; mientras que, en el caso de *AL* representa la estimación del denominado *tenure discount*, descuento medio derivado de la renovación del contrato en lugar de acudir al mercado de nuevos arrendamientos.

Como era de esperar, el inquilino que habita una vivienda de renta antigua, y en menor medida si es de PO, realiza unos pagos corrientes muy por debajo de lo que se estima que correspondería en el mercado libre; los descuentos medios suponen un 78 y un 63 por ciento respectivamente. Por su parte, la estimación del *tenure discount* es considerablemente menor, un

---

<sup>12</sup> Hay 63 viviendas que no muestran información sobre la variable alquiler: 2 casos en propiedad de PO, 18 casos en *AL*, 30 en arrendamiento de renta antigua, 2 en arrendamientos de PO y 11 en régimen de cesión.

42 por ciento. También la autoestimación media de los propietarios, tanto en régimen libre como de PO, y de los ocupantes de una vivienda en cesión está por debajo de la valoración de mercado. La mayor infravaloración se obtiene entre los propietarios de viviendas de PO, donde, por otra parte, el valor medio observado es superior al de la propiedad libre y, a su vez, éste superior al de la vivienda cedida. En este punto cabría hacerse la siguiente pregunta: ¿qué influencia puede tener la posible asociación entre niveles de calidad de la vivienda y el régimen de tenencia en este dato medio?, es decir, ¿qué papel juega la calidad en la diferencia observada en los precios por régimen de tenencia?

Con el objeto de realizar la comparación 'alquiler observado-alquiler imputado' según distintos niveles de calidad y posibilidades de tenencia, en los Cuadros 4.6 a 4.9 se presenta la misma información sobre el alquiler medio observado e imputado del cuadro anterior, pero ahora desagregada para cada una de las cuatro particiones de la distribución del IC definidas por los valores cuartiles<sup>13</sup>. En la primera columna se muestra, a diferencia del cuadro anterior, el porcentaje de viviendas de cada régimen de tenencia que se corresponde con la partición por cuartiles representada en cada cuadro, lo que permite analizar la asociación entre cada nivel de calidad y la tenencia. Así, por ejemplo, en el Cuadro 4.6 se observa que el arrendamiento de renta antigua es el régimen de tenencia que presenta el mayor porcentaje de viviendas con peor calidad (un 41,49 por ciento de este tipo de viviendas se corresponde con el primer cuartil) mientras que el régimen de propiedad de PO ocuparía la situación opuesta (sólo un 5,41 por ciento de estas viviendas

---

<sup>13</sup>Los cuartiles están determinados en relación a la distribución del IC del parque compuesto por la totalidad de viviendas principales de la EPF de 1980-81 (véase el Capítulo II).

está en el primer cuartil). El análisis conjunto de los Cuadros 4.6 a 4.9 indica, por ejemplo, que la vivienda en propiedad adquirida en el mercado libre muestra de calidad más equitativamente distribuida por cuartiles, mientras que la PO resulta ser el régimen de tenencia con una mayor concentración de viviendas en el tercer y cuarto cuartil, aproximadamente el 76 por ciento del total. La PO también tiene esta característica en el arrendamiento, aunque en menor medida puesto que la correspondiente acumulación de viviendas de mayor calidad es del 54,70 por ciento frente al 37,89 por ciento del arrendamiento libre en sentido estricto<sup>14</sup> y al 28,45 por ciento si es de renta antigua. Esto puede ser debido a que, como se sabe, existe el requerimiento de unas condiciones mínimas para que una vivienda sea considerada en el sistema de PO<sup>15</sup>, las cuales no tienen porqué cumplir las viviendas adquiridas en el mercado libre.

Por su parte, la vivienda arrendada de renta antigua se identifica con el régimen de peor acondicionamiento, el 71,55 por ciento de este tipo de viviendas dispone de un índice de calidad inferior a la mediana de la totalidad del parque residencial.

Resulta curioso que los precios medios autoestimados en la EPF guardan una relación positiva con los niveles de calidad que sintetiza el IC de tal forma que, mientras el coeficiente de correlación del IC con el alquiler pagado en AL es de 0,51, la correlación entre el IC y el alquiler autoestimado es de 0,55. Con independencia de que la vivienda sea de propiedad libre, de PO o cedida, el precio medio obtenido directamente de la correspondiente pregunta de la EPF es significativamente mayor en la medida en que se incrementa la calidad media que representa cada partición de la distribución del IC.

---

<sup>14</sup> AL, ocupación posterior a la Ley de Arrendamientos Urbanos de 1964.

<sup>15</sup> Por ejemplo, estas condiciones aparecen recogidas en el Real Decreto Ley 31/1978 cuando se define la vivienda de Protección Oficial en el período que nos ocupa.

Por su parte, la autoestimación del ocupante de la vivienda en propiedad libre declara un valor medio menor que el de la propiedad en PO sólo para los niveles más bajos de calidad; en el tercer y cuarto cuartil se observa la situación contraria.

En cuanto a los alquileres medios imputados, a diferencia de los alquileres proporcionados directamente en la EPF, están relacionados positivamente con la variable IC por construcción. Su comparación con el alquiler observado en los Cuadros 4.6 a 4.9, y en lo que a la propiedad y cesión se refiere, pondrá de relieve el nivel de infra o sobrevaloración con que el entrevistado estima el alquiler medio de mercado que debería pagar por una vivienda similar según las correspondientes particiones del IC. A este respecto se observa que la propiedad libre infravalora el valor del flujo de servicios de la vivienda que ocupan de forma similar en los distintos cuartiles, en torno a un 13 por ciento de media. Sin embargo, en la propiedad de PO se ajusta en mayor medida el valor de mercado cuando la calidad es menor (la infravalora en el primer cuartil es de un 7 por ciento mientras que oscila entre un 21 y un 25 por ciento en los cuartiles superiores). En el régimen de cesión se observan unos niveles de infravaloración menos sistemáticas, del orden de un 16 por ciento en los dos primeros niveles de calidad frente a un 23 y un 14 por ciento en el tercer y cuarto cuartil.

En lo que respecta a los regímenes de arrendamiento, el descuento definido como la diferencia media entre el alquiler que realmente se paga y el que se debería pagar en caso de un nuevo arrendamiento en 1980, mantiene el orden de magnitud en las tres alternativas posible: i) en la renta antigua se obtiene un descuento superior el 75 por ciento en todos los niveles de calidad (con un valor máximo de un 80 por ciento en el primer cuartil), ii) el descuento en el arrendamiento de un PO oscila entre un 58 y un 66 por ciento, y iii)

en el *AL* se estima un descuento medio del 42 por ciento. Obsérvese que el descuento de este último régimen se identifica con el efecto del denominado *tenure discount* o rebaja en el alquiler medio pagado en el mercado libre debido a la ocupación anterior a 1980. En el modelo especificado para 1980-81 se estima una relación negativa entre dicho descuento y el nivel de calidad, *ceteris paribus*. En este apartado, se comprueba esta relación para los niveles medios de calidad que representan los cuartiles 2, 3 y 4 (con descuentos aproximados de un 49, 44 y 33 por ciento respectivamente), aunque no para el del primer cuartil que presenta un *tenure discount* medio de un 43 por ciento. Por consiguiente, este resultado se debe a dos factores diferenciales en el *AL* según el nivel de calidad: i) la distribución de las variables de localización, 'tamaño de municipio' y 'provincia', que también determinan el alquiler de mercado imputado, y ii) la distribución del año de ocupación. De hecho, en la EPF de 1980-81 se puede comprobar, por ejemplo, que: i) el 34 por ciento de las viviendas en *AL* del primer cuartil están ubicadas en municipios de menos de 10.000 habitantes, mientras que este porcentaje en los cuartiles segundo, tercer y cuarto es de un 20, un 16 y un 9 por ciento respectivamente, y ii) el 46 por ciento de las viviendas en *AL* del primer cuartil fueron ocupadas antes de 1974, mientras que dicho porcentaje se reduce a un 43 por ciento en los cuartiles intermedios y a un 39 por ciento en las viviendas de mejor calidad.

#### 4.4 Valoración del parque de viviendas a precios de 1990

Como en el apartado correspondiente a 1980-81, comenzamos por analizar la información disponible en la EPF de 1990-91. Posteriormente se presenta el

resultado de estimar el modelo (4.2) corregido por sesgo de selección muestral para, finalmente, realizar la comparación entre el alquiler medio de mercado imputable en 1990 y los distintos alquileres medios, reales o autoestimados, proporcionados por la EPF de 1990-91 según el régimen de tenencia. En el Cuadro 4.10 se muestra la distribución de frecuencias del parque de viviendas principales según el régimen de tenencia y su correspondiente clasificación por CL. La distribución de las observaciones muestrales disponibles aparece en la primera columna mientras que la distribución de la representación poblacional y el porcentaje de viviendas de cada régimen sobre el total aparecen en la segunda y tercera columna respectivamente.

Al comparar la distribución de las viviendas de la EPF de 1990-91 con la distribución equivalente de la EPF de 1980-81 (Cuadro 4.1), se observa que: i) el régimen de arrendamiento tiene un menor peso relativo (un 15 por ciento frente a un 22,92), y ii) se incrementa el peso de la falta de información sobre la CL (un 15,62 por ciento frente a un 9,96). Como ya se ha comentado, la falta de información sobre la CL en la propiedad no afecta a la estimación del modelo (4.2) puesto que la propiedad, con independencia de la CL, forma parte del tipo de viviendas censuradas, es decir, del régimen 'Otros'. Sin embargo, en el arrendamiento es imprescindible conocer la CL para asignar al régimen 'AL' las viviendas con calificación 'libre' que, además, se ocuparon después de la entrada en vigor de la LAU de 1964, y al régimen 'Otros' el resto. Tal y como se hizo en la valoración del parque de viviendas en 1980-81, en 1990-91 tampoco se incluirán, en un principio, las viviendas en arrendamiento sin CL conocida (914 casos). No obstante, en el Apéndice II.2 se estima su CL, libre o PO, para posteriormente analizar el efecto de este tipo de viviendas en el modelo, efecto que será evaluado, de forma similar a como se hizo en 1980-81, en términos de eficiencia y sesgo en la estimación de los

parámetros. Seguidamente se concreta la disponibilidad de las observaciones y variables  $x$  e  $z$  que permitirán estimar el modelo corregido de sesgo de selección para 1990-91.

#### 4.4.1 Datos y variables

En el Capítulo III se identifica el conjunto de variables  $x$  del modelo que explica el alquiler pagado en el régimen de  $AL$ <sup>16</sup> en 1990-91. Por otra parte, en el Apéndice I.2 se muestra una serie de posibles variables  $z$ <sup>17</sup> y su asociación con la clasificación del régimen de tenencia  $AL$  y *Otros*. Al igual que en 1980-81, la falta de información sobre algunas de las posibles variables explicativas condiciona el tamaño muestral en 1990-91 de forma que para estimar el modelo (4.2) se dispone de 982 viviendas en  $AL$ <sup>18</sup>, representativas de 562.013 viviendas al aplicar el factor poblacional del INE, y 17.5800 en *Otros*<sup>19</sup>, representativas de 9.325.481 viviendas. El análisis descriptivo que se presenta en el Apéndice I.2 permite agrupar las modalidades de cada variable de forma coherente con los datos observados en la EPF de 1990-91 y

---

<sup>16</sup>Se remite al Apéndice II.2 del Capítulo III para analizar la relación de cada una de estas variables con el alquiler observado.

<sup>17</sup>Cada variable aparece en el mayor grado de desagregación disponible, cuya descripción aparece documentada en Arévalo et al.(1995). Los datos están disponibles en la página web <<http://www.eco.uc3m.es/epf90-91.html/>>.

<sup>18</sup>Aunque en un principio se dispone de 1.061 viviendas en  $AL$ , es preciso eliminar 79 de ellas: 19 no tienen información sobre el alquiler, 7 no proporcionan el año de construcción (variable que determina el valor del IC) y 53 se identifican con observaciones atípicas en el modelo de alquiler estimado en el Capítulo III.

<sup>19</sup>En principio habría 19.180 viviendas en el régimen '*Otros*', sin embargo, se eliminan 1.600 por no disponer de alguna de las variables explicativas del modelo (4.2): hay 1.557 viviendas sin 'año de ocupación', 30 sin 'año de construcción', y 13 sin 'metros cuadrados construidos'.

sirve de ayuda en el proceso interactivo que supone estimar varios modelos alternativos hasta identificar las variables  $z$  más apropiadas en el modelo (4.2) final elegido. La notación de estas variables aparece en la última columna del apéndice. Al igual que en 1980-81, entre las variables  $z$  para 1990-91 se considera, además de las características de la vivienda, localización y año de ocupación, la variable instrumental 'condición de migrante del sustentador principal'. En el Cuadro 4.11 se define tanto el conjunto de variables  $x$  como el de las variables  $z$  que se utilizan para estimar el modelo (4.2) cuyos resultados se muestran en el Cuadro 4.12 y se comentan en el siguiente apartado.

#### 4.4.2 Resultados de la estimación

La estimación del parámetro  $\rho$  en el modelo con corrección del sesgo de selección (4.2) para 1990-91, al igual que en el modelo de 1980-81, muestra un signo positivo aunque resulta significativo para un nivel de confianza inferior al convencional del 95 por ciento. Las variables más asociadas con el régimen de *AL* son: fecha de residencia del sustentador principal en el municipio posterior a 1985, menos de 61 metros cuadrados construidos, reciente ocupación de la vivienda, ubicación de la vivienda en municipios con más de 10.000 habitantes y en provincias como Madrid, Barcelona, Baleares o Valencia. Por su parte, las variables asociadas positivamente con el régimen '*Otros*' son: el año de construcción del edificio, disponer de un baño completo o más, ubicación de la vivienda en edificios de una sola planta, garaje dentro del edificio, teléfono y jardín.

El efecto que la variación en las variables  $x$  produce en el alquiler de 1990-91 fue estimado en Capítulo III y corregido por el sesgo de selección muestral en este capítulo. En las columnas 1 y 2 del Cuadro 4.13 se muestran los

respectivos efectos marginales medios expresados en términos porcentuales sobre el alquiler. Como se puede observar, la variable más afectada por la corrección es 'acup8390' cuyo efecto negativo medio se incrementa en un 18,16 por ciento. En menor medida, también se ven afectadas las variables 'rloc4', que experimenta un incremento de su efecto medio de un 7,37 por ciento, 'rloc5', un 6,35 por ciento, y 'acup7682', un 6,59 por ciento. Las tres modalidades del IC son las únicas variables que muestra una reducción de su efecto medio en el alquiler tras la corrección del sesgo de selección muestral, en cualquier caso las magnitudes relativas de los efectos de las distintas variables  $x$  no se ven alteradas.

En las columnas (1) y (2) del Cuadro 4.13 se muestran las respectivas estimaciones de la variación porcentual del alquiler ante variaciones unitarias de cada una variable independiente, mientras que en la columna (3) se muestra el intervalo de confianza de dicho efecto según el modelo corregido.

Tal y como se procedió en el apartado correspondiente de 1980-81, también en el modelo de 1990-91 se plantea la posible influencia que pudieran tener las viviendas en arrendamiento sin CL. Así, ante la hipótesis implícitamente asumida en el modelo del carácter aleatorio de este tipo de viviendas, en el siguiente apartado se muestra que no hay evidencias en contra de dicha hipótesis.

#### **4.4.3 Influencia de las observaciones sin CL**

En la EPF de 1990-91 hay 914 observaciones muestrales en régimen de arrendamiento cuya CL se desconoce (véase Cuadro 4.10); representan a más de medio millón de viviendas a nivel poblacional, lo que supone un 31,6 por ciento de las viviendas en arrendamiento y un 4,74 por ciento del total del parque residencial.

En el Apéndice II se define un regla de clasificación discriminante que, de forma similar al ejercicio realizado en 1980-81, permiten estimar la CL:  $\widehat{L}$  o  $\widehat{PO}$ , y por tanto el régimen:  $\widehat{AL}$  u  $\widehat{Otros}$ , de este tipo de viviendas<sup>20</sup>. También como en 1980-81, se utiliza la técnica *bootstrap* para seleccionar muestras aleatorias con reemplazamiento que permiten reestimar el modelo (4.2) y evaluar los efectos de incluir las viviendas en arrendamiento sin CL en dicho modelo.

Los resultados obtenidos en el Apéndice II en relación al modelo para 1990-91 se pueden resumir en los siguientes puntos:

- 1.- El número máximo de viviendas de arrendamiento sin CL que pueden ser recuperadas para estimar el modelo (4.2), son 858. La clasificación estimada del régimen de tenencia de estas viviendas es de: 397 viviendas en  $\widehat{AL}$  y 461 en  $\widehat{Otros}$ . En el cómputo de estas viviendas no se consideran 21 viviendas, para las que se había estima un régimen de  $\widehat{AL}$ , porque se identifican como observaciones atípicas en el modelo de alquiler (4.1) al aplicar el método de Peña y Yohai, 1995.
- 2.- Cuando se incrementa de forma aleatoria el tamaño muestral en 858 observaciones se experimenta una reducción media de un 17,33 por ciento en la desviación típica de los parámetros estimados. Las variables  $x$  son las más afectadas en la ganancia de eficiencia (con una reducción media del 20,61 por ciento) y en particular las siguientes variables de ocupa-

---

<sup>20</sup>Esta regla estima que, de las 914 viviendas en arrendamiento sin calificación legal se estima que 418 son de  $\widehat{AL}$  y 461 del régimen  $\widehat{Otros}$  (64  $\widehat{L}$  de renta antigua y 397 de  $\widehat{PO}$ ), con un error estimado de un 31 por ciento. Hay 35 viviendas a las que no se puede estimar la CL por falta de información en alguna de las variables explicativas del modelo discriminante (11 aparecen sin 'alquiler', 21 sin el 'año de construcción del edificio', 1 sin 'alquiler ni 'año de construcción' y 2 sin 'metros cuadrados construido').

ción: 'acup8390' con una reducción del 22,38 por ciento, 'acup7682', un 21,80 por ciento y 'a6575', un 21,62 por ciento.

- 3.- No hay evidencia de sesgo en la estimación del efecto medio en ninguna de las variables  $x$  y  $z$  cuando se incluyen las 858 viviendas en arrendamiento sin CL para un nivel de confianza del 95 por ciento.

En lo que se refiere al sesgo en el efecto de las variables  $x$  sobre el alquiler, en el Cuadro 4.13 se muestra la estimación porcentual media de dichos efectos respecto al alquiler según los tres modelos estimados en esta tesis para 1990-91. Así, en la columna 1 se muestra el efecto medio según el modelo sin corrección del sesgo de selección muestral estimado en el Capítulo III [modelo (1)]; en las columnas 2 y 3 se muestra, respectivamente, el efecto medio e intervalo de confianza según el modelo corregido [modelo (2)]; y en la columna 4 se presenta el correspondiente valor estimado cuando, además de corregir el sesgo de selección muestral, se incluyen las observaciones con CL estimada [modelo (3)]. Como se puede observar, este último modelo presenta mayores divergencias con el modelo (2) de referencia que el modelo (1). Además, como se comentó anteriormente, las variables más afectadas son: 'acup8390', 'acup7682', y 'a6575', sin embargo, los valores medios estimados en el modelo (3) se encuentran en los intervalos de confianza identificados en el modelo de referencia para un nivel de significación del 95 por ciento.

En consecuencia, aunque la información sobre la CL mejoraría la estimación del modelo en términos de eficiencia, se puede concluir que, al igual que en 1980-81, en 1990-91 no hay evidencia en contra del carácter aleatorio de las viviendas en arrendamiento sin CL en lo que al modelo estimado se refiere.

#### 4.4.4 Comparación del alquiler medio observado e imputado

El modelo corregido de sesgo de selección permite imputar un alquiler corriente de mercado a toda vivienda con información sobre el nivel de calidad de la vivienda, *IC*, y su localización ya que en todos los casos se considera como 'año de ocupación' 1990. En este sentido, en la EPF de 1990-91 hay 289 hogares que no proporcionan el 'año de construcción del edificio' y 67 que no proporcionan los 'metros cuadrados construidos'. En ambos casos se asigna el valor medio que presentan estas variables en las viviendas con el mismo régimen de tenencia<sup>21</sup>, con ello es posible disponer de un valor estimado del *IC* que hace posible imputar un alquiler corriente a la totalidad del parque residencial representado en la EPF de 1990-91. En este apartado, al igual que en el apartado equivalente para los datos 1980-81, se realiza la comparación de las imputaciones vía estimación del modelo (4.2) con los alquileres, real o procedente de la autoestimación del encuestado, que directamente proporciona la EPF de 1990-91. Así, en el Cuadro 4.14 se muestran ambos alquileres medios según el régimen de tenencia y CL conocida para el conjunto de viviendas con información al respecto<sup>22</sup>. En la primera columna

---

<sup>21</sup>En concreto, se asignan los siguientes valores medios: i) para el 'año de construcción': 1963 a propiedad libre (3 casos); 1951 a *AL* (7 casos); 1925 a arrendamiento de renta antigua (16 casos); 1964 a arrendamiento de PO (11 casos); 1946 a arrendamiento sin CL (22 casos); 1960 a la vivienda en cesión (230 casos), y ii) para 'metros cuadrados construidos' se asigna: 115 a propiedad libre (7 casos); 93 a propiedad de PO (1 caso); 103 a la propiedad sin CL (2 casos); 90 al arrendamiento de renta antigua (1 caso); 87 al arrendamiento sin CL (2 casos); 92 al arrendamiento sin CL (2 casos), y 97 a algún tipo de cesión (52 casos).

<sup>22</sup>No se consideran 75 viviendas por no disponer de la variable alquiler en la EPF de 1990-91: 3 casos en propiedad de PO, 19 casos en *AL*, 11 en arrendamiento de renta

del cuadro se muestra la distribución porcentual de las viviendas (17.909 observaciones muestrales representativas de 9.500.423 a nivel poblacional); en la segunda columna aparece el alquiler medio (desviación típica, DT) proporcionado directamente en la EPF; en la tercera columna se muestra el alquiler medio (desviación típica, DT) imputado mientras que en la cuarta columna representa la diferencia entre dichos valores medios expresada en relación al valor imputado.

En términos medios se puede observar que, a diferencia de lo que ocurre en 1980-81, tanto los propietarios como los beneficiarios de las viviendas en cesión, sobrevaloran el alquiler de mercado que deberían pagar en 1990-91 por la vivienda que ocupan. No obstante, en ambos periodos se mantiene la ordenación de los distintos regímenes de tenencia según la diferencia porcentual entre alquiler imputado y observado. Así, la ordenación del arrendamiento de mayor a menor descuento en el pago del alquiler realmente pagado respecto al estimado en el mercado es: renta antigua, PO y AL. Por su parte, los regímenes de propiedad de PO, cesión y propiedad libre son los que presentan, de menor a mayor valor, una sobrevaloración de dicho alquiler de mercado.

También en 1990-91. la propiedad de PO destaca por ser el tipo de viviendas que presenta un mayor alquiler medio tanto autoestimado como imputado, lo que, como se verá más adelante, se puede justificar por una diferenciación en la calidad a favor a este tipo de viviendas.

Con objeto de analizar la relación entre el nivel de calidad y el tipo de tenencia, tal y como se hizo en 1980-81, en los Cuadros 4.15 a 4.18 se presenta la misma información que en el cuadro anterior de forma desagregada para las cuatro particiones según los cuartiles de la variable IC<sup>23</sup>, donde, antigua, 29 en arrendamientos de PO y 12 en arrendamiento sin CL y 1 en régimen de cesión.

<sup>23</sup>Estos cuartiles están definidos en relación a la distribución de la variable IC del total

ahora, la primera columna muestra el porcentaje de las viviendas de cada régimen de tenencia que representa del cuartil correspondiente. Aunque el IC que se construye en 1990-91 sintetiza un mayor número de variables que el construido en 1980-81<sup>24</sup>, el análisis pormenorizado de la distribución de la calidad muestra unos resultados muy similares en ambos períodos: i) el tipo de vivienda con peor calidad es el arrendamiento de renta antigua, con un 69 por ciento de viviendas acumuladas en el primer y segundo cuartil, seguido de la cesión, con un 63 por ciento en 1990-91 (un 64 por ciento en 1980-81); ii) el tipo de viviendas con mejor calidad es la propiedad de PO con menos de un 10 por ciento de las viviendas en el primer cuartil en 1990-91 (un 5,4 por ciento en 1980-81); iii) la propiedad libre junto con el régimen de arrendamiento libre, muestran las distribuciones más uniforme por cuartiles. En relación a los niveles de calidad propios de cada período, el arrendamiento mejora su posición relativa al resto de regímenes de tenencia en 1990-91 respecto a 1980-81 mientras que la propiedad y la cesión empeoran, está última en menor medida. No obstante, la calidad media de todos los regímenes de tenencia en 1990-91 es superior a la observada en 1980-81 tal y como indica el IC que permite comparar el nivel de servicios de ambos períodos<sup>25</sup>.

En lo que respecta a los alquileres observados en 1990-91, la autoestimación media de los ocupantes de las viviendas en propiedad y cesión muestra una correlación positiva con los niveles de calidad considerados en los cuartiles. Esto indica que los encuestados estarían dispuestos a pagar un mayor precio por la mejora de calidad que sintetiza el IC. Este resultado, que tam-

---

de viviendas de la EPF de 1990-91 (véase Capítulo II).

<sup>24</sup>Se remite al Capítulo III para tener en cuenta, tanto las variables que intervienen en la construcción de cada IC, como su peso relativo a la hora de determinar el valor del índice.

<sup>25</sup>Veáse Capítulo II.

bién se obtuvo en 1980-81, da consistencia al significado y utilidad de los respectivos IC construidos en el Capítulo III.

En 1990-91, a diferencia de 1980-81, se obtiene que hay una generalizada sobrestimación media del alquiler valorado por parte de los propietarios y los beneficiarios de viviendas en régimen de cesión en las distintas particiones por cuartiles. A su vez, la mayor sobrevaloración se produce en las viviendas en propiedad libre donde de forma destacada se sobrestima el precio de las viviendas en la medida en que estas disponen de mayor nivel de calidad llegando a suponer un 33 por ciento en el cuarto cuartil, sólo suponía un 10 por ciento en el primer cuartil. La mayor sobrestimación en la cesión es de un 18 por ciento, observado en las viviendas del primer cuartil, y de un 10 por ciento en la propiedad de PO, lo que se da en el segundo cuartil.

La justificación de por qué la autoestimación del ocupante de una viviendas en 1980-81 está por debajo del valor corriente imputado, mientras que en 1990-91 está por encima, puede venir de los resultados obtenidos en el Capítulo III. Efectivamente, es lógico pensar que la referencia de que dispone el ocupante para estimar el alquiler que debería pagar en el mercado libre por su vivienda, la encuentre en la comparación con el alquiler real observado en viviendas con características similares. Pues bien, con independencia de que el régimen de arrendamiento libre se reduce en algo más de un 32 por ciento en 1990-91 respecto a 1980-81<sup>26</sup>, en el capítulo III se muestra evidencia empírica de que las cláusulas de actualización del alquiler apenas se utilizaban en 1980, mientras que su uso se generaliza en 1990. Así, mientras que el descuento estimado por año de ocupación en el AL es de un 13 por ciento en 1980-81, apenas supone un 3 por ciento en 1990-91<sup>27</sup>. Por tanto, en la

---

<sup>26</sup>En la EPF de 1980-81 hay 888.945 (Cuadro 1) viviendas en AL, mientras que en la EPF de 1990-91 hay 601.970 (Cuadro 10).

<sup>27</sup>Ambos descuentos interanuales por ocupación (estimación del *tenure discount*) se re-

medida en la que los alquileres observados en el *AL* sirven de referencia para la autoestimación del alquiler en 1980 y 1990 respectivamente (suponiendo el primer contrato de arrendamiento y, por tanto, sin efecto *tenure discount*), resulta más razonable el resultado obtenido: a principios de los ochenta este valor se subestima y a principios de los noventa se sobrestima.

Por último, en el régimen de arrendamiento en 1990-91 se obtiene que, con independencia del nivel de calidad, el descuento debido a la diferencia entre el alquiler realmente pagado y el imputado de mercado, es mayor en las viviendas de renta antigua que en las de PO y en éstas menor que en el *AL*, nuevamente un resultado que coincide con el obtenido en 1980-81. También en 1990-91, el descuento medio en la PO es considerablemente menor en las viviendas de mayor calidad, es del orden de un 42 por ciento en las viviendas del cuarto cuartil, mientras que en las del primer cuartil es de un 70 por ciento.

## 4.5 Conclusiones

El objetivo de este Capítulo es imputar un alquiler de mercado a las viviendas que componen el parque residencial de nuestro país en 1980-81 y 1990-91, respectivamente. Este alquiler representa una valoración del flujo de servicios generados por la vivienda a precios de 1980 y 1990 respectivamente.

La clave para realizar dicha imputación es la estimación de un modelo que revele los efectos de los factores  $x$  que explican el alquiler de mercado observado. Con la estimación de estos efectos será posible la adjudicación de un valor de mercado a toda vivienda con información sobre  $x$ .

---

fieren a los períodos más próximos al año de valoración del alquiler: ocupación entre 1974 y 1980 para el *AL* de 1980-81 y entre 1983 y 1990 para el *AL* de 1990-91.

Las Encuestas de Presupuestos Familiares (EPF) proporcionan las variables  $x$  para todas las viviendas que componen la muestra disponible. Sin embargo, sólo dispone de información sobre el alquiler de mercado en las viviendas en régimen de arrendamiento libre en sentido estricto ( $AL$ ), es decir, cuya ocupación se realizó con posterioridad a la entrada en vigor de la Ley de Arrendamientos Urbanos de 1964 que pone fin a la congelación del alquiler a que estaba sometido el arrendamiento desde 1920.

El modelo que se estima en este capítulo tiene en cuenta que el alquiler que se observa está condicionado al tipo de viviendas en  $AL$ , de tal forma que se evita el sesgo de selección muestral que se podría cometer en el caso de que exista relación entre el régimen de tenencia (' $AL$ ', frente a ' $Otros$ '), y el gasto en vivienda. Para considerar esta circunstancia en el modelo es preciso determinar los factores que determinan ambas variables.

En el Capítulo III se identificaron los determinantes  $x$  que explican el alquiler observado en las viviendas en  $AL$ : i) el índice de calidad de la vivienda (IC), ii) la localización en referencia al tamaño de municipio y la provincia, y iii) el año de ocupación. Por su parte, en este capítulo se identifican los siguientes factores  $z$  en relación a la clasificación de los regímenes ' $AL$ ' y ' $Otros$ ': i) las características de la vivienda, ii) el tamaño de municipio, ii) la provincia, y iii) la condición de emigrante del sustentador principal del hogar.

El modelo especificado para imputar un alquiler sin sesgo de selección, se estima por máxima verosimilitud de tal forma que proporciona estimaciones robustas de los parámetros de  $x$  y  $z$  de forma conjunta. No se consideran las observaciones atípicas influyentes en la estimación del modelo de alquiler que fueron detectadas al aplicar el método de Peña y Yohai, 1995 en el capítulo anterior. Tampoco se incluyen las viviendas en arrendamiento sin calificación

legal (CL), un 2,23 por ciento del total de viviendas en 1980-81 y un 4,74 por ciento en 1990-91. No obstante se comprueba que no hay evidencia empírica de sesgo, para un nivel de confianza del 95 por ciento, en ninguno de los parámetros estimados en el caso de incluir estas últimas viviendas en el modelo una vez se estima su CL a través de un modelo de regresión logística que permite adjudicar la clasificación discriminante: libre frente a protección oficial, y por tanto hace posible la clasificación: 'AL' frente a 'Otros'. En consecuencia, y en lo que al modelo de alquiler se refiere, se obtiene que no hay evidencia estadística en contra el carácter aleatorio del tipo de hogares en arrendamiento de la EPF que no proporciona la CL.

Los resultados obtenidos en la estimación del modelo muestran, tanto en 1980-81 como en 1990-91, una asociación entre el *AL* y los siguiente factores *z*: una mayor antigüedad del edificio, menor número de metros cuadrados construidos, disponer de menos de un baño completo dentro de la vivienda, no tener teléfono, no tener garaje; la localización de la vivienda en municipios de más de 10.000 habitantes y pertenecer a provincias como: Madrid, Barcelona, Tarragona, Baleares, Ceuta o Melilla; la ocupación más reciente, y la condición de emigrante del sustentador principal. Por su parte, la corrección del sesgo de selección que se aplica en este capítulo respecto a su omisión, afecta en el mismo sentido a cada tipo de variable *x* explicativa del alquiler en 1980-81 que en 1990-91: los efectos positivos estimados del IC sobre el alquiler experimentan una reducción que en ningún caso supera el 5 por ciento, mientras que el efecto marginal del año de ocupación y el correspondiente a las variables de localización experimentan un incremento que en ningún caso supera el 9 por ciento, salvo el efecto marginal de la ocupación posterior a 1983 que aumenta un 18 por ciento. No obstante, en ambos periodos se mantienen las magnitudes relativas de los efectos estimados en los tres tipos

de variables.

Una vez realizada la estimación corregida del modelo se asigna un valor al flujo de servicios, alquiler imputado, a toda vivienda de la EPF de 1980-81 y 1990-91. Dicho valor representa una estimación del alquiler que correspondería pagar en el mercado de nuevos arrendamientos en 1981 y 1991 respectivamente, por una vivienda con las mismas características y localización.

Este alquiler imputado se compara, en términos medios, con el observado en las EPF, es decir, tanto con el alquiler real en las viviendas en arrendamiento: *AL*, de renta antigua y de *PO*, como con la autoestimación proporcionada por la EPF en los regímenes de propiedad libre, propiedad de *PO* y cesión. En ambas EPF se obtiene que los inquilinos más beneficiados por la diferencia entre el alquiler pagado en arrendamiento y el estimado en el mercado son los que ocupan las viviendas de renta antigua, se estima un descuento porcentual de un 78 por ciento en 1980-81 y un 81 por ciento en 1990-91. Los correspondientes descuentos medios en el arrendamiento de *PO* son de un 66 y un 59 por ciento respectivamente, mientras que las estimaciones medias del *tenure discount* en el régimen de *AL* son de un 42 por ciento en el primer período y de un 34 por ciento en 1990-91.

En lo que respecta a los regímenes distintos del arrendamiento, se obtiene un resultado contrario en ambas EPF; el hogar infraestima el valor medio de mercado en 1980-81 mientras que en 1990-91 lo sobrestima. En cualquier caso, tanto los alquileres medios procedentes de la autoestimación, como los alquileres medios imputados, son superiores en la propiedad de *PO* que en la propiedad libre y, en esta última, superiores a los de la vivienda cedida.

La comparación similar entre el alquiler medio observado e imputado por régimen de tenencia también se realiza de forma desagregada para distintos

niveles de la calidad de la vivienda. Para establecer los niveles se consideran las cuatro particiones definidas sobre los cuartiles identificados sobre la distribución del IC en el total de viviendas de cada EPF, lo que permite analizar la relación existente entre calidad y tipo de tenencia. Así, en ambos períodos se comprueba que: i) el conjunto de viviendas asociado con un mayor nivel de calidad es la propiedad de PO, más del 75 por ciento de estas viviendas tiene una calidad superior al valor mediano del IC en 1980-81, más de un 76 por ciento en 1990-91; por su parte, la calidad de la vivienda en el régimen de arrendamiento de PO también es superior a la del arrendamiento libre; ii) el tipo de vivienda con peor calidad es el arrendamiento de renta antigua con un 72 por ciento acumulado en el primer y segundo cuartil en 1980-81, un 82 por ciento en 1990-91; el orden inmediato superior lo ocupa el régimen de arrendamiento libre y la cesión, con un porcentaje de viviendas cercano al 62 por ciento en 1980-81 y en torno al 67 por ciento en 1990-91; iv) el régimen de propiedad libre es el que muestra una distribución más equitativa entre los cuatro cuartiles del IC, con una representatividad superior al 21 por ciento en cada nivel de calidad en los dos periodos analizados.

A la vista de los resultados obtenidos en este capítulo, cabe destacar una gran coherencia por parte de los hogares de la EPF a la hora de discriminar por precio los distintos niveles de calidad de las viviendas que en esta tesis se logran sintetizar con la variable ordinal IC. Tanto los propietarios como los ocupantes de las viviendas en cesión, en 1980-81 y en 1990-91, revelan su percepción de mejora en la calidad a través de una mayor autoestimación media del alquiler que deberían pagar en el mercado.

El IC construido en el Capítulo II que permite comparar la calidad las viviendas de ambas EPF, revela una mejora en el nivel medio de calidad en todos y cada uno de los regímenes de tenencia en 1990-91 respecto a 1980-81.

De forma similar, los IC construido en 1980-81 y 1990-91 con las correspondientes características de la viviendas disponibles en las respectivas EPF, permite identificar la misma asociación entre el nivel de calidad medio y el tipo de tenencia. Así, en ambos períodos la ordenación de los regímenes de tenencia de mayor a menor calidad media es: propiedad de PO, propiedad libre, arrendamiento de PO, cesión, arrendamiento libre y por último arrendamiento de renta antigua. Finalmente, se puede añadir que no hubiera sido fácil llegar a este tipo de resultados si no se dispusiera de indicadores de calidad de la vivienda como los construidos en esta tesis.

. Cuadros del capítulo IV

Cuadro 4.1: Distribución del parque residencial en 1980-81

Régimen de tenencia	Muestral	Poblacional	%
PROPIEDAD	16.427	6.928.150	69,11
Propiedad libre	9.307	4.104.814	40,95
Propiedad de PO	5.316	2.048.206	20,43
Propiedad sin CL	1.804	775.130	7,73
ARRENDAMIENTO	5.484	2.297.105	22,92
Arrendamiento libre ( <i>AL</i> )	2.181	888.945	8,87
Arrendamiento libre de renta antigua	968	405.290	4,04
Arrendamiento de PO	1.787	773.164	7,71
Arrendamiento sin CL	535	223.795	2,23
CESIÓN (gratuita o semigratuita)	2.060	798.911	7,97
TOTAL	23.971	10.024.166	100

Cuadro 4.2: Definición de las variables del modelo para 1980-81

Variables x	Notación <sup>28</sup>	Definición
IC de la vivienda	ind6573	IC si $1965 \leq \text{año de ocupación } (t) \leq 1973$
	ind7480	IC si $1974 \leq t \leq 1981$
Localización	rloc1*-rloc9	(Véase Apéndice II.1 del Capítulo III)
Año de ocupación	a6573	=1 si $1965 \leq t \leq 1973$ ; =0 en otro caso
	acup6573	Número de años si $1965 \leq t \leq 1973$
	acup7480	Número de años si $1974 \leq t \leq 1981$
Variables z	Notación	Definición
Condición de migrante	migz	=1 si año de traslado $\geq 1975$ ; =0 en otro caso
Servicios higiénicos	higz1	No tiene
	higz2*	Menos de un baño completo
	higz3	Más de un baño completo
Agua caliente	aguaz	=1 si tiene; =0 en otro caso
Garaje en el edificio	garz	=1 si tiene; =0 en otro caso
Superficie construida	m2v	Metros cuadrados
Teléfono	telz	=1 si tiene; =0 en otro caso
Antigüedad del edificio	aconz1*	=1 si es $\leq 10$ años; =0 en otro caso
	aconz2	=1 si es entre 11 y 30 años; =0 en otro caso
	aconz3	=1 si es $\geq 31$ años
Año de ocupación (t)	ocupz	=0 si $t \leq 1964$ ; = t en otro caso
Tamaño de municipio	tmunz1*	Menos de 10.000 habitantes
	tmunz2	Entre 10.001 y 50.000 habitantes
	tmunz3	Más de 50.000 habitantes
Provincia	provz1*-provz4	(Véase Apéndice I.1)

<sup>28</sup>Las variables señaladas con el superíndice \* hacen referencia a las modalidades omitidas.

Cuadro 4.3: Resultados del modelo estimado\* para 1980-81

Variables $x$	Coef.	t-valor	Variables $z$	Coef.	t-valor
ind6573	0,665	12,116	migz	0.447	6.156
ind7480	0,792	15,617	higz1	-0.383	-4.074
loc2	0,163	2,329	higz3	-0.154	-2.261
loc3	0,431	5,825	aguaz	-0.198	-3.454
loc4	0,574	4,858	garz	-0.231	-3.222
loc5	0,462	6,232	m2v	-0.003	-4.719
loc6	0,702	8,143	telz	-0.398	-9.129
loc7	0,685	9,520	aconz2	0.673	12.527
loc8	0,729	9,830	aconz3	1.352	23.262
loc9	0,854	10,215	ocupz	0.071	15.579
a6573	-0,474	-2,434	tmunz2	0.144	2.433
acup6573	-0,040	-4,060	tmunz3	0.344	6.174
acup7480	-0,136	-13,213	provz2	0.194	3.521
			provz3	0.480	8.504
			provz4	0.795	12.381
constante	6,579	57,998	constante	-140.748	-15.695

$\hat{\rho}(\text{Std.err}) = 0,105 (0,080)$ ;  $\hat{\sigma}(\text{Std.err}) = 0,570(0,015)$

Número de observaciones = 21.522;

$\chi^2(13) = 1.944$ ;  $\text{Prob} > \chi^2 = 0.000$

Log-Likelihood = -2.608.742; Wald test ( $H_0: \rho=0$ ):  $\chi^2(1) = 1,67$

\*Estimación robusta con factor de elevación poblacional.

Cuadro 4.4: Efecto (%) de las variables  $x$  en el alquiler de 1980

Variabes	(1)	(2)	[Intervalo de (2) al 95%]	(3)
ind6573	69,80	66,50	[ 55,74 ; 77,26]	65,80
ind7480	83,01	79,19	[ 69,25 ; 89,13]	79,40
loc2	16,31	17,70	[ 2,62 ; 35,00]	14,42
loc3	52,30	53,81	[ 33,07 ; 77,79]	50,12
loc4	74,35	77,53	[ 40,84 ; 123,79]	73,03
loc5	55,19	58,69	[ 37,24 ; 83,49]	54,15
loc6	97,08	101,74	[ 70,39 ; 138,87]	95,23
loc7	94,78	98,40	[ 72,30 ; 128,45]	87,57
loc8	102,64	107,36	[ 79,30 ; 139,81]	109,95
loc9	127,02	134,98	[ 99,45 ; 176,83]	127,17
a6573	-36,38	-37,78	[ -57,53 ; -8,83]	-33,38
acup6573	-3,77	-4,04	[ -5,99 ; -2,09]	-4,33
acup7480	-13,16	-13,59	[ -15,60 ; -11,57]	-13,37

(1) modelo sin corrección del sesgo de selección; (2) modelo corregido;  
(3) modelo corregido y observaciones recuperadas.

Cuadro 5.5: Alquiler medio a precios corrientes de 1980-81

Régimen de tenencia	% <sup>(1)</sup>	Observado Medio (D.T)	Imputado Medio (D.T)	Diferencia % <sup>(2)</sup>
Propiedad libre	45,65	8.631 (7.519)	9.915 (5.903)	-12,95
Propiedad de PO	22,77	9.613 (5.891)	12.672 (4.711)	-24,14
Arrend. libre	9,77	5.620 (5.753)	9.683 (4.904)	-41,96
Arrend. renta antigua	4,37	2.046 (2.475)	9.229 (4.698)	-77,83
Arrend. de PO	8,59	4.273 (4.403)	11.665 (4.416)	-63,37
Cesión	8,84	6.902 (5.703)	8.352 (4.614)	-17,36

(1) Distribución porcentual de 8.991.308 viviendas. (2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.6: Alquiler medio a precios corrientes de 1980-81

Cuartil I		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	28,02	3.656 (3.246)	4.233 (1.667)	-13,63
Propieda de PO	5,41	5.391 (3.412)	5.792 (1.751)	-6,92
Arrend. libre	31,68	2.949 (2.627)	5.175 (1.905)	-43,01
Arrend. renta antigua	41,49	1.080 (1.278)	5.397 (2.081)	-79,98
Arrend. de PO	15,05	2.330 (2.774)	6.518 (1.789)	-64,25
Cesión	36,14	3.677 (3.221)	4.388 (1.854)	-16,20

(1) Porcentaje de viviendas del cuartil I en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.7: Alquiler medio a precios corrientes de 1980-81

Cuartil II		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	24,95	6.825 (4.384)	7.679 (2.160)	-11,12
Propieda de PO	18,97	6.928 (4.587)	8.813 (2.117)	-21,39
Arrend. libre	30,43	4.500 (3.373)	8.771 (2.037)	-48,70
Arrend. renta antigua	30,06	2,258 (2.645)	9.683 (2.162)	-76,68
Arrend. de PO	30,25	3,195 (3.190)	9.399 (2.016)	-66,01
Cesión	27,84	6.605 (3.958)	7.862 (1.998)	-15,99

(1) Porcentaje de viviendas del cuartil II en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.8: Alquiler medio a precios corrientes de 1980-81

Cuartil III		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	22.44	9.780 (5.419)	11.273 (2.769)	-13,24
Propieda de PO	32.80	8.667 (4.560)	11.469 (2.539)	-24,43
Arrend. libre	23,07	6.760 (5.097)	11.967 (2.615)	-43,51
Arrend. renta antigua	18,35	2.583 (2.239)	12.215 (2.525)	-78,85
Arrend. de PO	33,59	4.236 (3.924)	12.574 (2.476)	-66,31
Cesión	19,98	8.322 (4.971)	10.809 (2.453)	-23,01

(1) Porcentaje de viviendas del cuartil III en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.9: Alquiler medio a precios corrientes de 1980-81

Cuartil IV		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	24.59	15.083 (9.864)	17.420 (5.190)	-13,41
Propieda de PO	42.82	12.061 (6.499)	16.172 (4.310)	-25,42
Arrend. libre	14,82	11.855 (9.292)	17.638 (4.337)	-32,79
Arrend. renta antigua	10,10	4.407 (3.762)	18.193 (4.224)	-75,77
Arrend. de PO	21,11	7.260 (5.842)	17.134 (4.131)	-57,63
Cesión	16.04	12.911 (7.709)	15.071 (4.700)	-14,33

(1) Porcentaje de viviendas del cuartil IV en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.10: Distribución del parque residencial en 1990-91

Régimen de tenencia	Muestral	Poblacional	%
PROPIEDAD	16.623	8.789.287	77,79
Propiedad libre	9.132	4.883.659	43,22
Propiedad de PO	5.222	2.676.098	23,69
Propiedad sin CL	2.269	1.229.530	10,88
ARRENDAMIENTO	2.975	1.694.184	15,00
Arrendamiento libre ( <i>AL</i> )	1.061	601.970	5,33
Arrendamiento libre de renta antigua	229	129.968	1,15
Arrendamiento de PO	771	426.875	3,78
Arrendamiento sin CL	914	535.371	4,74
CESIÓN (gratuita o semigratuita)	1.557	815.038	7,21
TOTAL	21.155	11.298.509	100

Cuadro 4.11: Definición de las variables del modelo para 1990-91

Variables x	Notación	Definición
IC de la vivienda	ind6575	IC si $1965 \leq \text{año de ocupación } (t) \leq 1975$
	ind7682	IC si $1976 \leq t \leq 1982$
	ind8390	IC si $1983 \leq t \leq 1991$
Localización	rloc1*-rloc5	(Véase Apéndice II.2 del Capítulo III)
Año de ocupación (t)	a6575	=1 si $1965 \leq t \leq 1975$ ; =0 en otro caso
	a7682	=1 si $1976 \leq t \leq 1982$ ; =0 en otro caso
	acup6575	Número de años si $1965 \leq t \leq 1975$
	acup7682	Número de años si $1976 \leq t \leq 1982$
	acup8390	Número de años si $1983 \leq t \leq 1991$
Variables z	Notación	Definición
Condición de migrante	migz	=1 si año de traslado $\geq 1986$ ; =0 en otro caso
Servicios higiénicos	higz1*	Menos de un baño completo
	higz2	Un baño completo
	higz3	Más de un baño completo
Garaje en el edificio	garz	=1 si tiene; =0 en otro caso
Superficie construida	m2v1	=1 si metros cuadrados $\leq 60$ ; =0 en otro caso
Teléfono	telz	=1 si tiene; =0 en otro caso
Tipo de edificio	tipedz	=1 si es una sola planta; =0 en otro caso
Jardín	jarz	=1 si tiene; =0 en otro caso
Antigüedad del edificio	aconz	Año de construcción
Año de ocupación (t)	ocupz	=0 si $t \leq 1964$ ; = t en otro caso
Tamaño de municipio	tmunz1*	Menos de 10.000 habitantes
	tmunz2	Más de 10.000 habitantes
Provincia	provz1*-provz4	(Véase Apéndice II.2)

Cuadro 4.12: Resultados del modelo estimado<sup>(1)</sup> para 1990-91

VARIABLES $x$	Coef.	t-valor	VARIABLES $z$	Coef.	t-valor
ind6575	0,114	8,645	migz	0,374	4,688
ind7682	0,090	4,568	higz2	-0,256	-1,947
ind8390	0,061	7,329	higz3	-0,401	-2,792
loc2	0,394	2,044	garz	-0,109	-1,903
loc3	0,632	3,300	m2v1	0,244	3,147
loc4	0,846	4,174	telz	-0,501	-9,004
loc5	0,934	4,723	tipedz	-0,396	-5,907
a6575	-0,773	-1,923	jarz	-0,157	-3,121
acup6575	-0,071	-3,732	aconz	-0,011	-10,794
a7682	-0,633	-1,665	ocupz	0,028	7,530
acup7682	-0,042	-1,662	tmunz2	0,396	6,613
acup8390	-0,030	-2,194	provz2	0,733	11,578
			provz3	1,259	15,519
			provz4	0,157	2,477
constante	8,264	36,220	constante	-36,098	-4,814

$\hat{\rho}(\text{Std. err}) = 0,114 (0,093)$ ;  $\hat{\sigma}(\text{Std. err}) = 0,623 (0,023)$

Número de observaciones = 18.562;

$\chi^2(12) = 899$ ;  $\text{Prob} > \chi^2 = 0,000$

Log-Likelihood = -2.210.958; Wald test ( $H_0: \rho=0$ ):  $\chi^2(1) = 1,48$

(1) Estimación robusta con factor de elevación poblacional.

Cuadro 4.13: Efecto (%) de las variables  $x$  en el alquiler de 1990

Variabes	(1)	(2)	[Intervalo de (2) al 95%]	(3)
ind6575	11.71	11.37	[ 8,79 ; 13,94]	11.26
ind7682	9.39	9.03	[ 5,16 ; 12,91]	8.57
ind8390	6.46	6.14	[ 4,50 ; 7,78]	6.81
loc2	46.65	48.28	[ 1,63 ; 116,35]	41.71
loc3	83.40	88.21	[29,27 ; 174,01]	67.84
loc4	123.90	133.02	[56,63 ; 246,68]	110.98
loc5	145.36	154.58	[72,75 ; 275,17]	117.98
a7682	-52.63	-53.82	[ -78,99 ; 1,52]	-54.67
acup7682	-6.98	-7.09	[ -10,81 ; -3,37]	-6.52
a7682	-46.81	-46.93	[-74,82 ; 11,87]	-27.48
acup7682	-3.96	-4.22	[ -9,20 ; 0,76]	-5.44
acup8390	-2.56	-3.03	[ -5,74 ; -0,32]	-4.48

(1) modelo sin corrección del sesgo de selección; (2) modelo corregido;  
 (4) modelo corregido y observaciones recuperadas.

Cuadro 4.14: Alquiler medio a precios corrientes de 1990-91

Régimen de tenencia	% <sup>(1)</sup>	Observado Medio (D.T)	Imputado Medio (D.T)	Diferencia % <sup>(2)</sup>
Propiedad libre	51,40	30.743 (27.618)	24.105 (11.790)	27,54
Propiedad de PO	28,15	32.597 (20.376)	30.666 (10.667)	6,30
Arrend. libre	6,25	17.079 (15.885)	25.831 (10.340)	-33,88
Arrend. renta antigua	1,29	4.434 ( 5.320)	23.611 ( 8.245)	-81,22
Arrend. de PO	4,33	11.590 (13.107)	28.134 ( 8.936)	-58,81
Cesión	8,57	24.933 (21.310)	22.755 ( 9.927)	9,57

(1) Distribución porcentual de 9.500.423 viviendas. (2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.15: Alquiler medio a precios corrientes de 1990-91

Cuartil I		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	30,82	17.087 (11.242)	15.532 (3.675)	10,02
Propiedad de PO	6,51	20.017 (11.507)	18.560 (4.443)	7,85
Arrend. libre	29,03	9.380 (16.797)	16.798 (3996)	-44,16
Arrend. renta antigua	40,49	2.434 (3.315)	16.672 (3.606)	-85,40
Arrend. de PO	9,69	5.255 (5.950)	17.777 (4.304)	-70,44
Cesión	35,19	18.131 (19.935)	15.331 (4.320)	18,26

(1) Porcentaje de viviendas del cuartil I en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.16: Alquiler medio a precios corrientes de 1990-91

Cuartil II		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	26,78	25.150 (14.891)	19.507 (4.362)	28,93
Propiedad de PO	17,15	23.685 (13.678)	21.552 (4.934)	9,90
Arrend. libre	28,58	14.040 (11.494)	22.925 (4.250)	-38,76
Arrend. renta antigua	28,39	4.302 (5.008)	24.058 (4.312)	-82,12
Arrend. de PO	31,22	7.651 (8.322)	22.655 (4.605)	-66,23
Cesión	27,85	23.339 (16.575)	20.830 (4.601)	12,05

(1) Porcentaje de viviendas del cuartil II en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.17: Alquiler medio a precios corrientes de 1990-91

Cuartil III		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	21,25	33.201 (19.770)	25.129 (5.295)	32,12
Propiedad de PO	34,98	29.663 (16.493)	27.392 (5.453)	8,29
Arrend. libre	21,75	19.638 (15.581)	27.408 (4.801)	-28,35
Arrend. renta antigua	20,87	6.278 (6.445)	28.601 (3.777)	-78,05
Arrend. de PO	35,17	9.456 (9.774)	28.463 (5.053)	-66,78
Cesión	21,56	27.013 (17.503)	25.919 (5.542)	4,22

(1) Porcentaje de viviendas del cuartil III en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.

Cuadro 4.18: Alquiler medio a precios corrientes de 1990-91

Cuartil IV		Observado	Imputado	Diferencia
Régimen de tenencia	% <sup>(1)</sup>	Medio (D.T)	Medio (D.T)	% <sup>(2)</sup>
Propiedad libre	21,15	55.244 (42.797)	41.385 (12.314)	33,49
Propiedad de PO	41,36	40.754 (23.324)	39.120 (9.937)	4,18
Arrend. libre	20,64	29.417 (20.545)	40.895 (9.518)	-28,07
Arrend. renta antigua	10,25	8.942 (6.482)	39.626 (5.999)	-77,43
Arrend. de PO	23,92	22.431 (17.727)	38.994 (7.787)	-42,48
Cesión	15,40	40.441 (27.631)	38.764 (10.646)	4,32

(1) Porcentaje de viviendas del cuartil IV en cada régimen de tenencia.

(2) Diferencia porcentual entre el alquiler medio observado e imputado respecto al valor imputado.



. Apéndice I: Relación entre el  
régimen de tenencia y las  
variables explicativas

**RELACIÓN ENTRE EL RÉGIMEN DE TENENCIA: 'ALQUILER LIBRE' Y 'OTROS' Y POSIBLES VARIABLES EXPLICATIVAS SEGÚN LA ENCUESTA DE PRESUPUESTOS FAMILIARES DE 1980-81**

Descripción de la información mostrada en las columnas:

- (1) Distribución de las 834.046 viviendas en régimen de arrendamiento libre (AL).
- (2) Distribución de las 8.211.540 viviendas en régimen distinto al arrendamiento libre (Otros).
- (3) Distribución de las 9.045.586 viviendas del total (Total).
- (4) Porcentaje de viviendas en AL en la modalidad de cada variable.
- (5) Porcentaje de viviendas en el régimen Otros en la modalidad de cada variable.
- (6 y 7) Notación de las variables continuas y descretas que formarán parte del modelo final.

VARIABLES	Distribución de las viviendas según el régimen de tenencia:			Relación de cada variable con el régimen de tenencia:		Notación de las variables: Discretas Continuas	
	AL	Otros	Total	AL	Otros		
				Media: 9,22	Media: 90,78		
<b>CONDICION DE MIGRANTE DEL SUSTENTADOR PRINCIPAL</b>							
Reside en el mismo municipio desde 31-12-1975	86,67	96,67	95,74	8,35	91,65	migrz 0 1	
Se trasladó después del 31-12-1975	13,33	3,33	4,26	28,89	71,11		
<b>CARACTERISTICAS DE LA VIVIENDA</b>							
<b>Instalaciones higiénicas</b>							
No tiene	4,93	6,75	6,59	6,91	93,09	higz1 higz2 higz2 higz2 higz3 higz3 higz3 higz3 higz3 higz3	
Compartido con otras viviendas	1,45	0,38	0,48	27,76	72,24		
Uno o dos cuartos retrete	3,35	2,05	2,17	14,21	85,79		
Cuarto (s) de aseo o aseo y retrete	9,10	6,64	6,87	12,22	87,78		
Un cuarto de baño	68,38	67,68	67,75	9,31	90,69		
Un baño y un retrete	0,37	0,50	0,49	7,01	92,99		
Un baño y uno o dos aseos	7,59	8,81	8,70	8,04	91,96		
Dos baños	3,45	5,61	5,41	5,88	94,12		
Dos o más baños y uno o dos aseos	0,61	0,77	0,75	7,46	92,54		
Uno o más baños, aseo y retrete	0,21	0,13	0,14	13,48	86,52		
Tres o cuatro baños	0,56	0,66	0,65	7,94	92,06	higz3	
<b>Agua corriente</b>							
No tiene	3,79	3,45	3,48	10,04	89,96	aguaz 0 0 1 1	
Sólo agua fría	24,01	18,00	18,56	11,93	88,07		
Agua caliente individual	68,10	72,90	72,45	8,67	91,33		
Agua caliente central	4,11	5,65	5,51	6,87	93,13		
<b>Calefacción</b>							
No tiene	65,92	60,58	61,07	9,95	90,05	garz 0 0 1	
Sólo aparatos móviles	19,98	21,43	21,29	8,65	91,35		
Calefacción fija individual	7,73	9,10	8,98	7,94	92,06		
Calefacción fija central	6,37	8,89	8,66	6,78	93,22		
<b>Garaje</b>							
No tiene	88,61	82,78	83,31	9,81	90,19	m2v 0 0 1	
Tiene pero fuera del edificio	6,39	6,32	6,33	9,31	90,69		
Tiene en el propio edificio	4,99	10,90	10,35	4,45	95,55		
<b>Metros cuadrados</b>							
menos de 60	18,61	11,23	11,91	14,40	85,60	telz 0 1	
entre 61 y 90	46,50	41,38	41,85	10,24	89,76		
entre 91 y 130	28,28	34,33	33,77	7,72	92,28		
más de 130	6,61	13,05	12,46	4,89	95,11		
<b>Teléfono</b>							
No tiene	62,89	45,98	47,54	12,20	87,80	aconz1 aconz2 aconz2 aconz3 aconz3	
Si tiene	37,11	54,02	52,46	6,52	93,48		
<b>Antigüedad del edificio</b>							
10 años o menos	12,42	21,17	20,37	5,62	94,38		
Entre 11 y 20 años	28,09	29,16	29,06	8,91	91,09		
Entre 21 y 30 años	12,90	13,80	13,71	8,67	91,33		
Entre 31 y 40 años	8,47	5,80	6,05	12,92	87,08		
Entre 41 y 50 años	7,46	4,30	4,59	14,98	85,02		
Más de 50 años	30,65	25,77	26,22	10,78	89,22		
<b>AÑO DE OCUPACIÓN (t)</b>							
Antes de 1965	0,00	39,69	36,03	0,00	100,00	ocupz 0 t t t	
Entre 1965 y 1970	17,06	16,15	16,23	9,69	90,31		
Entre 1971 y 1975	35,97	24,30	25,38	13,07	86,93		
Entre 1976 y 1980-81	46,98	19,86	22,36	19,37	80,63		

**LOCALIZACIÓN**  
**Tamaño de municipio**  
 menos 2.000 habitantes  
 2.001-10.000 habitantes  
 10.001-50.000 habitantes  
 50.001-500.000 habitantes  
 mas de 500.000 habitantes

**Provincia**

Alava  
 Castellón  
 Córdoba  
 Cuenca  
 Guadalajara  
 Murcia  
 Navarra  
 Orense  
 Las Palmas  
 Valencia  
 Vizcaya  
 Zamora  
 Albacete  
 Avila  
 Badajoz  
 Burgos  
 Ciudad Real  
 Granada  
 Guipuzcoa  
 Jaén  
 La Rioja  
 Lugo  
 Málaga  
 Salamanca  
 Cantabria  
 Sevilla  
 Soría  
 Teruel  
 Toledo  
 Valladolid  
 Zaragoza  
 Alicante  
 Almería  
 Cáceres  
 Cádiz  
 La Coruña  
 Huelva  
 Huesca  
 Madrid  
 Asturias  
 Palencia  
 Pontevedra  
 Santa Cruz Tenerife  
 Segovia  
 Baleares  
 Barcelona  
 Gerona  
 León  
 Lérida  
 Tarragona  
 Ceuta  
 Melilla

	Distribución de las viviendas según el régimen de tenencia:			Relación de cada variable con el régimen de tenencia:		
	AL	Otros	Total	AL	Otros	
				Media: 9,22	Media: 90,78	
	7,59	11,49	11,13	6,29	93,71	tmunz1
	14,62	18,82	18,44	7,31	92,69	tmunz1
	20,49	20,22	20,25	9,33	90,67	tmunz2
	34,75	29,76	30,22	10,60	89,40	tmunz3
	22,56	19,71	19,97	10,41	89,59	tmunz3
	0,31	0,71	0,67	4,23	95,77	provz1
	0,67	1,27	1,21	5,07	94,93	provz1
	1,02	1,80	1,73	5,45	94,55	provz1
	0,30	0,66	0,63	4,33	95,67	provz1
	0,16	0,47	0,44	3,42	96,58	provz1
	1,38	2,64	2,53	5,04	94,96	provz1
	0,69	1,35	1,29	4,97	95,03	provz1
	0,66	1,32	1,26	4,81	95,19	provz1
	0,79	1,62	1,54	4,69	95,31	provz1
	2,92	6,01	5,72	4,70	95,30	provz1
	0,67	3,35	3,10	1,98	98,02	provz1
	0,25	0,79	0,74	3,09	96,91	provz1
	0,58	0,92	0,89	6,01	93,99	provz2
	0,42	0,51	0,50	7,66	92,34	provz2
	1,07	1,58	1,53	6,43	93,57	provz2
	0,95	1,07	1,06	8,26	91,74	provz2
	1,23	1,26	1,25	9,04	90,96	provz2
	1,52	1,93	1,89	7,40	92,60	provz2
	1,29	1,63	1,60	7,45	92,55	provz2
	1,32	1,78	1,74	6,99	93,01	provz2
	0,71	0,74	0,74	8,86	91,14	provz2
	0,92	1,05	1,04	8,22	91,78	provz2
	2,45	2,67	2,65	8,52	91,48	provz2
	1,08	1,10	1,10	9,12	90,88	provz2
	1,12	1,30	1,28	8,06	91,94	provz2
	2,51	3,72	3,61	6,43	93,57	provz2
	0,22	0,30	0,29	6,95	93,05	provz2
	0,35	0,51	0,49	6,59	93,41	provz2
	0,82	1,19	1,16	6,52	93,48	provz2
	0,88	1,28	1,25	6,54	93,46	provz2
	1,67	2,43	2,36	6,52	93,48	provz2
	3,24	3,11	3,12	9,58	90,42	provz3
	1,13	0,99	1,00	10,40	89,60	provz3
	1,49	1,22	1,24	11,02	88,98	provz3
	2,56	2,18	2,21	10,66	89,34	provz3
	3,43	2,84	2,89	10,94	89,06	provz3
	1,26	1,05	1,07	10,85	89,15	provz3
	0,72	0,61	0,62	10,70	89,30	provz3
	13,73	12,55	12,66	10,00	90,00	provz3
	3,82	3,10	3,17	11,13	88,87	provz3
	0,64	0,48	0,50	11,81	88,19	provz3
	2,46	1,92	1,97	11,54	88,46	provz3
	1,81	1,62	1,64	10,19	89,81	provz3
	0,51	0,42	0,43	11,05	88,95	provz3
	5,06	1,72	2,02	23,07	76,93	provz4
	18,59	12,13	12,73	13,47	86,53	provz4
	1,93	1,22	1,28	13,82	86,18	provz4
	2,27	1,35	1,43	14,61	85,39	provz4
	1,81	0,97	1,04	16,03	83,97	provz4
	1,95	1,33	1,39	12,96	87,04	provz4
	0,29	0,14	0,15	18,02	81,98	provz4
	0,36	0,11	0,13	24,75	75,25	provz4

**Notación de las variables:**  
 Discretas Continuas

**RELACIÓN ENTRE EL RÉGIMEN DE TENENCIA: 'ALQUILER LIBRE' Y 'OTROS' Y POSIBLES VARIABLES EXPLICATIVAS  
SEGÚN LA ENCUESTA DE PRESUPUESTOS FAMILIARES DE 1990-91**

Descripción de la información mostrada en las columnas:

- (1) Distribución de las 562.013 viviendas en régimen de arrendamiento libre (AL).
- (2) Distribución de las 9.325.481 viviendas en régimen distinto al arrendamiento libre (Otros).
- (3) Distribución de las 9.887.494 viviendas del total (Total).
- (4) Porcentaje de viviendas en AL en la modalidad de cada variable.
- (5) Porcentaje de viviendas en el régimen Otros en la modalidad de cada variable.
- (6 y 7) Notación de las variables continuas y discretas que formarán parte del modelo final.

VARIABLES	Distribución de las viviendas según el régimen de tenencia:			Relación de cada variable con el régimen de tenencia:		Notación de las variables: Discretas Continuas
	AL	Otros	Total	AL	Otros	
				Media: 5,68	Media: 94,32	
<b>CONDICION DE MIGRANTE DEL SUSTENTADOR PRINCIPAL</b>						
Reside en el mismo municipio desde antes de 1-3-1986	87,07	96,39	95,86	5,16	94,84	migrz 0
Se trasladó despues del 1-3-1986	12,93	3,61	4,14	17,76	82,24	1
<b>CARACTERISTICAS DE LA VIVIENDA</b>						
<b>Instalaciones higiénicas</b>						
No tiene	2,18	1,59	1,62	7,63	92,37	hlgz1
Compartido con otras viviendas	0,63	0,03	0,06	58,02	41,98	hlgz1
Uno o dos aseos, o uno o dos cuartos retrete	4,59	2,58	2,69	9,69	90,31	hlgz1
Un baño, o retrete con uno o dos aseos	74,13	69,42	69,69	6,05	93,95	hlgz2
Un baño y uno o dos aseos o un baño y un retrete	5,23	6,22	6,17	4,82	95,18	hlgz3
Un baño, aseo y retrete, o dos baños y aseo(s) o retrete(s)	12,29	18,65	18,29	3,82	96,18	hlgz3
Dos baños, aseos y retrete o más	0,95	1,51	1,48	3,65	96,35	hlgz3
<b>Agua corriente</b>						
No tiene	0,47	0,34	0,35	7,80	92,20	aguaz 0
Sólo agua fría	8,00	4,51	4,71	9,65	90,35	0
Agua caliente individual	85,08	87,95	87,78	5,51	94,49	1
Agua caliente central	6,45	7,20	7,16	5,12	94,88	1
<b>Calefacción</b>						
No tiene	13,20	10,54	10,70	7,02	92,98	caffz 0
Algún aparato móvil	64,09	61,24	61,41	5,93	94,07	0
Individual con aparatos fijos	16,10	19,40	19,21	4,76	95,24	1
Central	6,61	8,82	8,69	4,32	95,68	1
<b>Garaje</b>						
No tiene	80,34	70,50	71,06	6,43	93,57	garz 0
Si tiene	19,66	29,50	28,94	3,86	96,14	1
<b>Número de metros cuadrados construidos</b>						
Menos de 60	17,85	7,91	8,48	11,97	88,03	m2v1 1
Entre 61 y 90	39,35	37,17	37,30	6,00	94,00	0
Entre 91 y 130	30,27	38,19	37,74	4,56	95,44	0
Más de 130	12,53	16,72	16,48	4,32	95,68	0
<b>Teléfono</b>						
No tiene	34,55	20,20	21,02	9,34	90,66	telz 0
Un teléfono	49,06	61,29	60,60	4,60	95,40	1
Dos teléfonos	14,78	16,78	16,67	5,04	94,96	1
Tres o más teléfonos	1,62	1,72	1,72	5,35	94,65	1
<b>Tipo de edificio</b>						
Edificio de una sola vivienda	18,13	37,34	36,25	2,84	97,16	tipez 1
Edificio de dos viviendas	6,33	4,06	4,19	8,59	91,41	0
Edificio de tres o más viviendas	75,09	58,39	59,34	7,19	92,81	0
Otros fines distintos de los residenciales	0,16	0,10	0,10	9,10	90,90	0
Alojamiento fijo	0,29	0,11	0,12	13,71	86,29	0
<b>Jardín</b>						
No tiene	91,45	84,07	84,49	6,15	93,85	jarz 0
Si tiene	8,55	15,93	15,51	3,14	96,86	1
<b>Aire acondicionado</b>						
No tiene	98,09	97,43	97,47	5,72	94,28	aconz
Sistema privado	1,83	2,37	2,34	4,46	95,54	
Sistema colectivo	0,08	0,20	0,19	2,37	97,63	
<b>Ascensor</b>						
No tiene	72,28	69,33	69,50	5,91	94,09	
Si tiene	27,72	30,67	30,50	5,17	94,83	
<b>Piscina</b>						
No tiene	98,14	98,76	98,72	5,65	94,35	
Si tiene	1,86	1,24	1,28	8,28	91,72	
<b>Zona Deportiva</b>						
No tiene	98,86	98,93	98,92	5,68	94,32	
Si tiene	1,14	1,07	1,08	6,04	93,96	
<b>Otros servicios comunitarios</b>						
No tiene	50,99	51,80	51,75	5,60	94,40	
Si tiene	49,01	48,20	48,25	5,77	94,23	
<b>Antigüedad del edificio</b>						
10 años o menos	7,78	17,91	1,08	2,55	97,45	
Entre 11 y 20 años	28,00	34,79	48,25	4,63	95,37	
Entre 21 y 40 años	27,06	24,80	34,41	6,17	33,83	
Entre 41 y 50 años	8,82	9,57	3,47	5,26	94,74	
Entre 51 y 60 años	6,28	3,30	0,00	10,30	89,70	
Más de 60 años	22,06	9,63	36,57	12,13	87,87	
<b>AÑO DE OCUPACIÓN (t)</b>						
Antes de 1965	0,00	20,19	19	0,00	100,00	ocupz 0
Entre 1965 y 1974	23,89	28,27	28	4,84	95,16	t
Entre 1975 y 1984	39,41	36,40	37	6,13	93,87	t
Entre 1985 y 1990-91	36,70	15,14	16	12,75	87,25	t

**LOCALIZACIÓN**

**Tamaño de municipio**  
 Menos de 10.000 habitantes  
 10.001-50.000 habitantes  
 50.001-100.000 habitantes  
 100.001-500.000 habitantes  
 Más de 500.000 habitantes

**Provincia**  
 Alava  
 Albacete  
 Avila  
 Badajoz  
 Castellón  
 Córdoba  
 Huelva  
 Huesca  
 Lugo  
 Málaga  
 Navarra  
 Salamanca  
 Teruel  
 Toledo  
 Vizcaya  
 Zaragoza  
 Alicante  
 Almería  
 Burgos  
 Cáceres  
 Ciudad Real  
 La Coruña  
 Cuenca  
 Granada  
 Guadalajara  
 Guipuzcoa  
 Jaén  
 La Rioja  
 Murcia  
 Palencia  
 Cantabria  
 Segovia  
 Sevilla  
 Soria  
 Valencia  
 Valladolid  
 Barcelona  
 Cadiz  
 Gerona  
 León  
 Lérida  
 Madrid  
 Orense  
 Asturias  
 Pontevedra  
 Santa Cruz Tenerife  
 Tarragona  
 Zamora  
 Ceuta  
 Baleares  
 Las Palmas  
 Melilla

	Distribución de las viviendas según el régimen de tenencia:			Relación de cada variable con el régimen de tenencia:		
	AL	Otros	Total	AL	Otros	
				Media: 5,68	Media: 94,32	
	15,98	27,15	26,51	3,43	96,57	tmunz1
	23,86	22,19	22,29	6,09	93,91	tmunz2
	9,51	8,67	8,72	6,20	93,80	tmunz2
	26,80	22,98	23,20	6,57	93,43	tmunz2
	23,85	19,01	19,29	7,03	92,97	tmunz2
	0,06	0,78	0,74	0,46	99,54	provz1
	0,40	0,97	0,93	2,40	97,60	provz1
	0,20	0,57	0,55	2,09	97,91	provz1
	0,37	1,71	1,63	1,29	98,71	provz1
	0,60	1,36	1,32	2,57	97,43	provz1
	0,82	2,04	1,97	2,36	97,64	provz1
	0,22	1,21	1,16	1,10	98,90	provz1
	0,09	0,57	0,54	0,90	99,10	provz1
	0,43	1,08	1,05	2,36	97,64	provz1
	1,36	3,03	2,94	2,63	97,37	provz1
	0,07	1,44	1,37	0,29	99,71	provz1
	0,49	1,03	1,00	2,80	97,20	provz1
	0,18	0,45	0,44	2,40	97,60	provz1
	0,56	1,29	1,25	2,55	97,45	provz1
	1,31	3,16	3,05	2,44	97,56	provz1
	0,79	2,57	2,47	1,81	98,19	provz1
	2,41	3,35	3,29	4,15	95,85	provz2
	1,00	1,10	1,09	5,20	94,80	provz2
	0,60	1,07	1,04	3,26	96,74	provz2
	0,86	1,17	1,16	4,22	95,78	provz2
	0,70	1,27	1,24	3,22	96,78	provz2
	2,38	2,66	2,64	5,12	94,88	provz2
	0,40	0,66	0,65	3,50	96,50	provz2
	1,44	2,07	2,04	4,01	95,99	provz2
	0,27	0,44	0,43	3,55	96,45	provz2
	1,59	1,82	1,81	5,01	94,99	provz2
	1,04	1,69	1,66	3,57	96,43	provz2
	0,51	0,70	0,69	4,26	95,74	provz2
	1,98	2,44	2,41	4,66	95,34	provz2
	0,41	0,48	0,48	4,90	95,10	provz2
	0,96	1,35	1,33	4,12	95,88	provz2
	0,42	0,43	0,43	5,50	94,50	provz2
	2,33	3,52	3,46	3,84	96,16	provz2
	0,23	0,31	0,30	4,39	95,61	provz2
	3,14	5,79	5,64	3,16	96,84	provz2
	1,10	1,38	1,36	4,58	95,42	provz2
	20,41	11,34	11,85	9,79	90,21	provz3
	3,47	2,03	2,11	9,35	90,65	provz3
	2,08	1,22	1,27	9,32	90,68	provz3
	1,99	1,37	1,40	8,07	91,93	provz3
	1,21	0,95	0,96	7,16	92,84	provz3
	15,04	12,69	12,83	6,67	93,33	provz3
	1,46	1,26	1,27	6,55	93,45	provz3
	3,29	3,02	3,04	6,15	93,85	provz3
	2,27	2,04	2,06	6,27	93,73	provz3
	3,13	1,73	1,81	9,86	90,14	provz3
	2,35	1,43	1,48	9,02	90,98	provz3
	0,75	0,69	0,69	6,14	93,86	provz3
	0,23	0,15	0,16	8,18	91,82	provz3
	6,19	1,56	1,83	19,26	80,74	provz4
	3,57	1,48	1,60	12,68	87,32	provz4
	0,85	0,07	0,12	41,27	58,73	provz4

Notación de las variables:  
 Discretas Continuas



. Apéndice II: Análisis sobre  
las variables en arrendamiento  
sin CL

Como se sabe, la diferencia fundamental que lleva consigo el arrendamiento en el mercado libre, frente al considerado en algún sistema de protección oficial (PO), es el pago de un mayor alquiler en viviendas de similares características. En la Encuesta de Presupuestos Familiares (EPF) de 1980-81 hay 535 viviendas en régimen de arrendamiento que no proporcionan información sobre la calificación legal (CL): libre o protección oficial (PO). Este tipo de viviendas se eleva a 914 en la EPF de 1990-91, lo que representa un 31 por ciento del total de viviendas en arrendamiento de dicha muestra; esta relación era inferior al 10 por ciento en 1980-81.

Los objetivos de este apéndice son dos: 1) estimar la CL de las viviendas en arrendamientos de la EPF a través de un análisis de regresión logística, y 2) estimar los efectos que tendría en el modelo de alquiler (4.2) la inclusión de este tipo de viviendas con la CL estimada.

Los resultados obtenidos se muestran en dos apartados que hacen referencia a cada uno de los objetivos planteados. Antes de mostrar dichos resultados se procede a detallar el análisis realizado en cada apartado.

### I. ESTIMACIÓN DE LA CL

En este apartado se estiman los determinantes de la probabilidad de que una vivienda en arrendamiento haya sido contratada en el mercado libre frente a la alternativa de pertenecer al sistema de PO<sup>29</sup>. Para ello se especifica un modelo de regresión logística que será estimado por máxima verosimilitud (Greene, 1999). La CL es la variable dicotómica:  $Y = 1$ , si el arrendamiento es libre e  $Y = 0$  si es de PO, de forma que: i) el valor medio de  $Y$  representa la proporción de viviendas libres de la muestra, y ii) la distribución condicional de la variable  $Y$  (respecto a las variables explicativas ' $X$ ' de la

---

<sup>29</sup>En Arévalo (2000) también se realiza la estimación de la CL en las viviendas de propiedad que no proporcionan este dato.

CL) corresponde a una distribución binomial con probabilidad dada por la media condicional  $\pi(X) = E(Y/X) = P[Y = 1]$ . El modelo utilizado para estimar la probabilidad de arrendamiento libre considera la función logística:  $\pi(X) = \left[ 1 + e^{-\left(\beta_0 + \sum_{m=1}^M \beta_m X_m\right)} \right]^{-1}$ , con  $\beta^T = (\beta_0, \beta_1, \dots, \beta_M)$  el vector de parámetros, de forma que la contribución de cada vivienda  $i$  (con vector de características  $X_i$  y CL  $Y_i$ ) a la función máxima verosimilitud es:

$$\zeta(X_i) = \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i} ,$$

siendo  $i = 1, 2, \dots, I$  e  $I$  el número de viviendas en arrendamiento con CL conocida. Bajo el supuesto de que las  $I$  observaciones son independientes, la función de verosimilitud se define como el producto de dichas contribuciones. Para facilitar la resolución matemática, se define la función objetivo a maximizar como el logaritmo de esta función y se expresa como:

$$L(\beta) = \ln \left[ \prod_{i=1}^N \zeta(X_i) \right] = \sum_{i=1}^n (Y_i \ln [\pi(X_i)] + (1 - Y_i) \ln [1 - \pi(X_i)])$$

siendo  $\beta^T = (\beta_0, \beta_1, \dots, \beta_M)$  el vector de parámetros que permite estimar la probabilidad  $\pi(X_i)$ .

En relación a las probabilidades estimadas para las viviendas con CL, se establecen las reglas de clasificación discriminante (RCD) que permiten estimar la CL libre ( $\widehat{L}$ ), o PO ( $\widehat{PO}$ ), de toda vivienda en arrendamiento con información sobre las variables 'X'. Cualquier RCD dependen de un umbral,  $P$ , que representa la probabilidad de corte, o de referencia para discriminar a toda viviendas  $i$  entre:  $\widehat{PO}$ , si  $\pi(X_i) \leq P$ , y  $\widehat{L}$ , si  $\pi(X_i) > P$ . A su vez, cualquier RCD lleva asociado un error en la clasificación. Los dos tipos de errores posibles son: tipo I, asignación libre cuando el precio del arrendamiento es protegido, y tipo II, el caso contrario.

Mientras que la RCD que surge directamente de la estimación logística está basada en la minimización de la suma de ambos tipos de errores, la RCD que se definen en este trabajo, responde a un criterio que permite penalizar en mayor medida al error de tipo I que al de tipo II<sup>30</sup>. Efectivamente, en este trabajo, a diferencia de lo que supondría asignar las estimaciones:  $\hat{L}$  y  $\widehat{PO}$  imputables a las viviendas sin CL en función del modelo estimado, se evalúa la capacidad del modelo para discriminar en función de la probabilidad umbral. Para ello se considera la denominada curva *receiver operating characteristics* (curva ROC) que muestra la relación existente entre la sensibilidad, o proporción de viviendas libres adjudicadas correctamente, y la proporción del error tipo I asociado a los distintos niveles de la probabilidad umbral, véase Jovell (1995). Esta relación se puede presentar en un gráfico, de dimensiones cuadradas, donde la sensibilidad aparece en el eje de ordenadas y el error de tipo I en eje de abscisas, de forma que si la curva ROC coincide con la línea bisectriz del gráfico estaría mostrando que el modelo en cuestión no tiene capacidad de discriminación. En otro caso, el punto de la curva ROC cuya pendiente coincide con la pendiente de la bisectriz permite identificar a la probabilidad umbral óptima ( $P^*$ ) para la que se maximiza la sensibilidad del modelo a la vez que se minimiza el error de tipo I.

En este trabajo, se halla la probabilidad  $P^*$  que delimita las dos zonas de clasificación: i) la zona de 'PO':  $[0, P^*]$  de amplitud  $P^*$ , y ii) la zona de 'libre':  $(P^*, 1]$  de amplitud  $[1 - P^*]$ , y, en relación a ellas, se definen cuatro RCD.

En Arévalo (2000) se definen distintas RCD en relación al valor de los

---

<sup>30</sup>Este criterio es coherente con el objetivo último de estimar un modelo de alquiler de mercado con el mayor número de viviendas disponibles en la EPF; será preferible prescindir de viviendas en arrendamiento libre que incluir en el análisis viviendas con el precio protegido.

umbrales,  $P_{\widehat{PO}} \leq P^*$  y  $P_{\widehat{L}} < P^*$ , definidos para asignar  $\widehat{L}$  y  $\widehat{PO}$  en las respectivas zonas, por tanto, en el tipo de viviendas  $i$  cuyo probabilidad estimada  $\widehat{\pi}(X_i)$  sea  $P_{\widehat{PO}} < \widehat{\pi}(X_i) < P_{\widehat{L}}$  no se puede estimar la CL. En consecuencia, la diferencia entre las posibles RCD así definidas, estriba en su capacidad de clasificación en relación al *trade-off*: número de viviendas clasificadas y error asociado. La RCD que se presenta en este trabajo es la más inmediata y menos estricta, será denotada por  $RCD(P^*)$  y corresponde a la regla que asigna la CL de  $\widehat{PO}$  a toda vivienda  $i$  con probabilidad estimada  $\widehat{\pi}(x_i) \leq P^*$ , o de  $\widehat{L}$  si  $\widehat{\pi}(x_i) > P^*$ . Por tanto, la  $RCD(P^*)$  permite estimar la CL de todas las viviendas en arrendamiento de la EPF que no proporcionan su CL.

Cuando se realiza el análisis descrito al conjunto de viviendas en arrendamiento de la EPF de 1980-81 y 1990-91 se obtienen los siguientes resultados que hacen referencia a: (1) la identificación de las variables explicativas de la CL libre, frente a PO, en el modelo de regresión logística, (2) resultados obtenidos en la estimación del modelo, y (3) resultados obtenidos al estimar la CL de las viviendas en arrendamiento según la  $RCD(P^*)$ .

#### (1) Variables explicativas de la CL

Una vez realizado el análisis descriptivo pertinente, las variables  $X$  de las viviendas en arrendamiento que se identifican como factores explicativos de la CL libre, frente a PO, se muestran en el Cuadro I, de forma que en la primera y última columna aparece la notación de las variables del modelo estimado para 1980-81 y 1990-91 respectivamente. Estas variables son: el alquiler pagado mensualmente, un indicador de la calidad de la vivienda, el año de ocupación, el tamaño del municipio y la comunidad autónoma en la que se localiza.

La variable indicador de calidad se construye expresamente para sinteti-

zar el nivel de calidad de las viviendas de arrendamiento en 1980-81, *ICA80*, y 1990-91, *ICA90*, respectivamente. La técnica utilizada es el análisis de correspondencias múltiples expuesto en el Capítulo II de esta tesis, de tal forma que el criterio para seleccionar las características de la vivienda que tendrán un peso en el indicador de calidad, se consideran aquellas que presentan un mayor grado de divergencia según el régimen de tenencia (libre o PO) en cada EPF<sup>31</sup>. Estas variables son: 'servicios higiénicos', 'luz', 'metros cuadrado', 'agua', 'calefacción' y 'año de ocupación' en ambas EPF, además de 'garaje', y 'teléfono', en 1980-81, así como 'tipo de edificio', 'ascensor', 'piscina', 'otros servicios comunitarios', tipo de combustible para calentar: 'el agua', 'la calefacción' y 'la cocina', en 1990-91. Los principales resultados en relación a la construcción del *ICA80* son: i) esta variable explica el 72 por ciento de la variabilidad total observada en las 8 variables que sintetiza; ii) la correlación media entre estas variables y el índice es de 0,55 destacando su relación con 'agua' y 'servicios higiénicos' (coeficientes de correlación de 0,79 y 0,78, respectivamente); y iii) las modalidades con pesos extremos son: 'no disponer de luz eléctrica', 'no tener servicios higiénicos dentro de la vivienda' y 'no tener agua corriente' ó 'sólo agua fría', en la parte negativa, y 'más de dos baños', 'agua caliente' y 'calefacción' en la parte positiva. En lo que respecta a la construcción del *ICA90* se obtiene que: i) explica el 67 por ciento de las 13 variables que intervienen en su cálculo; ii) la correlación media de estas variables con el índice es de un 0,52 de forma que las variables con mayor coeficiente de correlación son: 'tipo de combustión para calentar el agua' (0,79), 'ascensor' (0,66) y 'otros servicios comunitarios' (0,64); y iii) las modalidades extremas con un peso negativo en el índice son: 'no disponer

---

<sup>31</sup>El análisis completo sobre el desarrollo de la construcción de estos índices están a disposición del interesado que lo solicite.

de luz', 'no tener agua', 'sistema de leña para cocinar' y 'no tener servicios higiénicos dentro de la vivienda', y con un peso positivo: 'calefacción por 'gas ciudad' o 'propano', 'disponer de piscina' y 'agua caliente' y 'más de un baño'.

En relación a las variables discretas que intervienen en el modelo de CL (año de ocupación, tamaño de municipio y comunidad autónoma), se ha optado por una codificación que permite interpretar los coeficientes como desviaciones a un valor medio, en lugar de tomar como referencia un individuo tipo definido por el conjunto de características omitidas (Hosmer y Lemeshow (1989), página 51). La ventaja de esta alternativa es que a través de una parametrización de los coeficientes, que consiste en que la suma de los coeficientes de las distintas características de cada variable discreta sea cero, es posible obtener el efecto sobre la variable dependiente de la característica omitida.

La disponibilidad de información sobre las variables condiciona el tamaño muestral, de tal forma que en el modelo de 1980-81 se dispone de 4.839<sup>32</sup> viviendas con CL que representan a 2.021.865 viviendas a nivel poblacional. Por su parte, en el modelo de 1990-91 se dispone de una muestra de 1.967<sup>33</sup> observaciones representativas de 1.111.299 viviendas con información sobre la CL. El porcentaje de viviendas de calificación libre es de un 63,53 por

---

<sup>32</sup>En un principio se contabilizan 4.949 viviendas en arrendamiento con CL, sin embargo, en 110 viviendas se detecta falta de información en las siguientes variables explicativas: alquiler, en 49 casos, año de construcción (necesaria para hallar el ICA80) en 40 casos, alquiler y año de construcción en 1 caso, y año de ocupación en 20.

<sup>33</sup>De las 2.061 viviendas en arrendamiento con CL conocida en la EPF 1990-91, se observa falta de información en los siguientes 94 casos: en 59 viviendas no se proporciona el alquiler y en 35 no se dispone de alguna de las variables que permiten el cálculo del ICA90 (33 hogares no conocen el año de construcción del edificio y en otros 2 no se facilitan los metros cuadrados de la vivienda).

ciento en el primer periodo y de un 62,89 por ciento en 1990-91.

### (2) Resultados de la estimación

Los resultados de la estimación de los modelos para 1980-81 y 1990-91, se muestran en el Cuadro I de forma que, en las columnas primera y segunda de cada modelo se presentan, respectivamente, el coeficiente estimado de cada variable y el estadístico Wald de significatividad. A esta información, que procede de la estimación del modelo, se añade la proporción real -no estimada- de viviendas libres de cada características (tercera columna) con el fin de comprobar el nivel de concordancia entre el signo estimado en los coeficientes y los datos observados.

A la vista de los resultados obtenidos, el nivel de asociación entre la CL predicha y observada es de un 75 por ciento en 1980-81, y de un 78,5 por ciento en 1990-91. En ambos modelos se obtiene que la probabilidad estimada de que una vivienda sea de renta libre aumenta cuando la vivienda se localiza en municipio con menos de 10,000 habitantes, en la comunidad autónoma, por ejemplo, de Baleares; cuando la vivienda fue ocupada con anterioridad a 1950, o cuanto mayor es el alquiler mensual pagado por el arrendamiento. Por el contrario, esta probabilidad disminuye a medida que aumenta la calidad de la vivienda.

### (3) Estimación de la CL

En las curvas ROC del modelo 1980-81 y 1990-91 (Gráficos 1 y 2 respectivamente) se identifica que la probabilidad umbral  $P^*$  que minimiza el error tipo I, asignar la CL de libre cuando es de PO, es de  $P^* = 0,62$  en ambos modelos. A partir de este dato, la aplicación de la regla de clasificación discriminante,  $RCD(P^*)$ , que estima una CL,  $\hat{L}$  o  $\widehat{PO}$ , a toda vivienda con información sobre las variables explicativas del modelo, arroja los resultados que se presentan en el siguiente cuadro. En la columna 1 aparece el por-

centaje de las viviendas con CL conocida sobre las que la  $RCD(P^*)$  asigna correctamente su CL. En la columna 2 se muestra el porcentaje de error tipo I asociado a la  $RCD(P^*)$  en cada uno de los modelos considerados.

Estimación de la capacidad predictiva de la $RCD(P^*)$		
Modelo	Asignación correcta	Error tipo I
1980-81	67,71 %	10,39 %
1990-91	71,18 %	9,96 %

La aplicación de las  $RCD(P^*)$  a las viviendas en arrendamiento sin CL de la EPF requiere la estimación previa de la probabilidad de pertenecer al grupo de calificación libre,  $\hat{\pi}(x)$ , en relación a los coeficientes  $\beta$  estimados y las variables  $x$  de este tipo de viviendas. En este sentido, el número de viviendas susceptible de clasificación en la EPF de 1980-81 es de 522<sup>34</sup> y de 879 en la EPF de 1990-91<sup>35</sup>. La estimación de la CL según la  $RCD(P^*)$  es de: 339 viviendas en  $\hat{L}$  y 183 en  $\widehat{PO}$ , para la muestra de 1980-81, y 482 en  $\hat{L}$  y 397 en  $\widehat{PO}$  para la muestra de 1990-91. En el siguiente cuadro se presenta la distribución poblacional de la CL para el conjunto de viviendas en régimen de arrendamiento, antes y después de considerar estas estimaciones.

<sup>34</sup>En 13 de las 535 viviendas en arrendamiento sin CL no se proporciona información sobre alguna de las variables  $x$  del modelo: en 8 no se dispone de año de construcción, en 6 falta el año de ocupación y en 1 caso los dos.

<sup>35</sup>En 35 de las 914 viviendas en arrendamiento sin CL falta la siguiente información: el alquiler pagado, en 11 casos, el año de construcción, en 21 caso, ambas variables, en 1 caso, y los metros cuadrados contruidos en otros 2 casos.

## Distribución de la calificación legal (CL) en el arrendamiento

EPF	CL	CL conocida		CL estimada <sup>(1)</sup>	
		Nº viviendas	%	Nº viviendas	% libre
1980-81					
	Libre	1.300.146	62,71	1.446.726	63,10
	PO	773.164	37,29	846.047	36,90
	Total	2.073.310	100	2.292.772	100
1990-91					
	Libre	731.938	63,16	1.022.904	61,33
	PO	426.875	36,84	645.030	38,67
	Total	1.158.813	100	1.667.934	100

(1) Viviendas con CL conocida más la estimación de la CL en caso de no disponer de ella.

Se puede observar que, en términos poblacionales, la  $RCD(P^*)$  permite estimar la CL de 219.462 viviendas en arrendamiento que en la EPF de 1980-81 no proporciona esta información, de tal forma que la distribución de la CL apenas se ve afectada. En 1990-91 se estima la CL de 509.121 viviendas y su distribución no llega a reducir un 2 por ciento la representación del arrendamiento libre respecto a la PO. En relación a esta información, se puede decir que no hay evidencia de una relación entre la falta de CL en el arrendamiento y alguna de las modalidades de dicha variable.

## II. EFECTO DE LA CL ESTIMADA EN EL MODELO DE ALQUILER.

En este apartado se analizan los efectos que ocasionaría considerar las viviendas en arrendamiento sin CL en los modelos del alquiler (4.2) estimados en este capítulo para 1980-81 y 1990-91, respectivamente. Para ello se siguen los siguientes pasos: (1) identificar las viviendas en arrendamiento con CL estimada que deben ser incluidas en el modelo, y (2) medición de su efecto en el modelo en términos de eficiencia y sesgo.

(1) Identificación de las viviendas a incluir en el modelo.

Como se ha expuesto en este capítulo, el modelo (4.1) tiene en cuenta que el 'alquiler libre' que se pretende explicar sólo se observa en las viviendas en régimen de arrendamiento libre en sentido estricto ( $AL$ ), ocupación posterior a la entrada en vigor de la Ley de Arrendamientos Urbanos de 1964. Al resto de las viviendas del parque residencial, que también serán consideradas en el modelo con objeto de corregir el efecto de la selección muestral, se las engloba en el régimen 'Otros'. Esta clasificación del régimen de tenencia ( $AL$ ,  $Otros$ ), obliga a reclasificar las viviendas en arrendamiento con CL estimada del apartado anterior, como paso previo a su consideración en el modelo (4.2). En concreto, se precisa que las viviendas estimadas en la calificación  $\widehat{L}$  sean clasificadas en el régimen estimado  $\widehat{AL}$ , si la ocupación fue anterior a 1964 (pre65), o en  $\widehat{Otros}$  si fue posterior (pos64), la denominada renta antigua. Por su parte, las viviendas clasificadas en  $\widehat{PO}$  pertenecen al régimen  $\widehat{Otros}$ , el alquiler pagado por este tipo de viviendas no corresponde a un precio de mercado. En la siguiente tabla se muestra la relación entre el régimen de tenencia y la CL estimada en las viviendas en arrendamiento de la EPF de 1980-81 y 1990-91, respectivamente.

Clasificación del arrendamiento con CL estimada					
EPF	Régimen	CL estimada			Total
		$\widehat{L}(pre65)$	$\widehat{L}(pos64)$	$\widehat{PO}$	
1980-81	$\widehat{AL}$	217			217
	$\widehat{Otros}$		122	183	305
	Total	217	122	183	522
1990-91	$\widehat{AL}$	418			418
	$\widehat{Otros}$		64	397	461
	Total	418	64	397	879

Con objeto de ser exhaustivos a la hora de analizar el efecto de considerar las viviendas  $\widehat{AL}$  y  $\widehat{Otros}$  en el modelo (4.2), se plantea la posibilidad de que en el grupo de viviendas  $\widehat{AL}$  existan observaciones que, al igual que se detectó

en el Capítulo III para el grupo de viviendas  $AL$ , tengan un comportamiento atípico en relación al modelo de alquiler lineal (4.1). Por este motivo, se vuelve a aplicar el método de Peña y Yohai (1995) a los respectivos modelos (4.1) estimados para 1980-81 y 1990-91, pero ahora incluyendo las viviendas en arrendamiento con CL estimada en régimen  $\widehat{AL}$ : 217 viviendas en 1980-81 y 418 en 1990-91. Los resultados obtenidos son: 1) se confirma el carácter atípico de las mismas observaciones detectadas como tales en el Capítulo III entre las viviendas en  $AL$  (90 en 1980-81 y 53 en 1990-91), y 2) se detectan, además, 14 atípicos entre las viviendas  $\widehat{AL}$  de 1980-81 y 21 en 1990-91.

Con la eliminación de estas observaciones quedan identificadas las viviendas en arrendamiento con CL estimada que pueden ser incluidas en los modelos (4.2) de 1980-81 y 1990-91 respectivamente. En la siguiente tabla se presenta la distribución (muestral y poblacional) del régimen de tenencia, tanto para el total de viviendas con información sobre la CL (columnas (1)), como para las viviendas en arrendamiento con CL estimada (columnas (2)) que pueden ser incluidas en el modelo.

Distribución del régimen de tenencia (RT)

EPF	RT	Muestra		Población	
		(1)	(2)	(1)	(2)
1980-81	AL	2.052	203	834.046	88.942
	Otros	19.470	305	8.211.540	125.019
	Total	21.522	508	9.045.586	213.961
1990-91	AL	982	397	562.013	234.765
	Otros	17.580	461	9.325.481	266.159
	Total	18.562	858	9.887.494	500.924

(1) Parque residencial con CL conocida. (2) Arrendamiento con CL estimada.

## (2) Efectos en el modelo

El efecto de incluir en el modelo (4.2) de 1980-81 (1990-91) las 508 (858) viviendas en arrendamiento con CL estimada será medido en términos de

eficiencia y sesgo en los parámetros estimados en el modelo (coeficiente  $\beta_k$  y desviación típica  $\sigma_{\beta_k}$  de cada variable  $k = 1, 2, \dots, K$ , siendo  $K$  el número total de variables explicativas,  $x$  y  $z$ , del modelo<sup>36</sup>). Con objeto de evaluar ambos efectos: eficiencia y sesgo, de forma independiente, se recurre a utilizar la metodología *bootstrap*<sup>37</sup>. Seguidamente se exponen: i) las aplicaciones que se realizan de esta técnica, y ii) las medidas que se proponen para definir eficacia y sesgo en el modelo, así como los resultados obtenidos en 1980-81 y 1990-91, respectivamente.

#### Aplicación Bootstrap

En la aplicación que nos ocupa, la técnica *bootstrap* permite seleccionar muestras aleatorias con reemplazamiento y reestimar el modelo (4.2) en tres procesos distintos (*Bootstrap I*, *Bootstrap II* y *Bootstrap III*) que, como se señala a continuación, hacen referencia al distinto tipo y número de observaciones consideradas según los objetivos buscados. Efectivamente, en cada aplicación *bootstrap* se efectúan 1.000 réplicas del modelo (4.2) de forma que,

- En el *Bootstrap I* se identifica la distribución de los 'estimadores *bootstrap*' (coeficiente  $\beta_k^I$  y desviación típica  $\sigma_{\beta_k}^I$  de cada variable  $k$  del modelo) que servirán de referencia para evaluar la eficiencia y sesgo en los parámetros del modelo (4.2) cuando se introduzcan las viviendas con CL estimada. En esta aplicación, las muestras aleatorias que permiten replicar el modelo se extraen del conjunto de viviendas con CL conocida: del grupo de viviendas *AL* y *Otros*, respectivamente, y el tamaño muestral coincide con el tamaño de los respectivos grupos de viviendas.

<sup>36</sup>Se recuerda que las variables  $x$  determinan el alquiler pagado por una vivienda en *AL* mientras que, las variables  $z$  determinan la probabilidad de pertenecer al régimen *AL* frente a '*Otros*' regímenes.

<sup>37</sup>Algunas referencias básicas sobre la aplicación del *bootstrap* son, por ejemplo, Efron and Tibshirani (1993) y Davison (1997).

## Extensiones

Las aportaciones de esta tesis se pueden concretar en dos puntos que hacen referencia a la heterogeneidad observada en la vivienda ocupada en España con fines residenciales: 1) se proporcionan tres indicadores del nivel de calidad que permite comparar la disponibilidad de servicios entre el conjunto de viviendas que compone la EPF de 1980-81, la EPF de 1990-91 y ambas EPF respectivamente, y 2) se proporciona un alquiler de mercado a toda vivienda de la EPF de 1980-81 y 1990-91 que refleja la estimación del valor del flujo de servicios generados por la vivienda, así como el efecto de su localización, a precios de 1980 y 1990 respectivamente.

El valor añadido que supone trabajar con dos encuestas de representatividad nacional y con información tan completa sobre el hogar como son las EPF de 1980-81 y 1990-91, es la posibilidad de realizar el mismo estudio desde dos perspectivas temporales, lo que contribuye a completar y enriquecer los resultados de cualquier tipo de análisis que trate de explicar la realidad observada en este tipo de encuestas, así como los cambios producidos en estos diez años.

En este apartado se exponen tres posibles líneas de investigación para las que podría ser de utilidad las aportaciones de esta tesis: 1) una evaluación de los efectos de la política gubernamental en materia de vivienda respecto al tipo de hogar afectado, 2) un análisis sobre desigualdad y bienestar de los

hogares españoles en 1980-81 y 1990-91 respecto al gasto y nivel de calidad en el bien vivienda, y 3) una propuesta sobre la consideración de la vivienda en la medición del índice de precios de consumo (IPC) que elabora el Instituto Nacional de Estadística (INE) en base a la información proporcionada en las EPF.

En referencia al primer punto, es sabido que la intervención por parte del Estado en el mercado de la vivienda en propiedad y/o alquiler es una práctica habitual en Europa desde la Primera Guerra Mundial y en Estados Unidos desde la Segunda.

El sector de la vivienda en España ha estado fuertemente intervenido debido principalmente a dos tipos de medidas: i) los diversos sistemas de protección oficial (PO) desarrollados a partir de los años 50 tanto en el régimen de propiedad como de arrendamiento, y ii) la política de control de alquileres donde destaca la congelación el precio del arrendamiento para las viviendas arrendadas con anterioridad a la entrada en vigor de la Ley de Arrendamientos Urbanos (LAU) de 1964. Esta ley permite la libre fijación de precios y su actualización para los nuevos contratos, aunque se mantiene la prorroga forzosa hasta la entra en vigor de las Medidas de Política Económicas de 1985. De esta forma, junto a la vivienda en régimen de cesión, el sector de la vivienda queda dividido en cinco subgrupos: propiedad libre, propiedad de PO, arrendamiento de PO, arrendamiento de renta antigua (ocupación anterior a la LAU de 1964) y arrendamiento libre en sentido estricto, *AL*, (ocupación posterior a 1964).

Los objetivos de la intervención del Estado en el libre funcionamiento del mercado, con independencia de su impacto en el sector de la construcción, se podrían resumir en uno: facilitar el acceso a viviendas, en propiedad o arrendamiento, a la población más necesitada a precios, o alquileres, inferiores a

los del mercado libre. Es obvio que un análisis de la eficacia distributiva de la política gubernamental precisaría disponer de información sobre las condiciones socio económicas de los potenciales beneficiados (y los perjudicados) en el momento de poner en marcha dichas medidas. De esta forma se podría evaluar tanto el nivel de discriminación a la hora de obtener dichas ventajas como el nivel de equidad distributiva entre los directamente beneficiados. Las aportaciones de esta tesis hacen posible analizar las implicaciones últimas de la intervención gubernamental, al menos, desde dos puntos de vista: 1) efectos en la asignación de recursos y 2) efectos distributivos. Efectivamente, disponer de un indicador sobre el nivel de calidad y el coste corriente de mercado del flujo de servicios del parque de vivienda ocupadas en 1980-81 y 1990-91 permite profundizar en ciertas cuestiones, algunas de las cuales ya fueron detectadas en la tesis: reducción sistemática del peso de los distintos tipos de arrendamiento acompañado de un deterioro de la calidad en la vivienda de renta antigua, alteración de la estructura de gasto en vivienda, grado de congelación de los alquileres según el año de ocupación, incentivos a la inmovilización de la población beneficiada,... Por otra parte, si se define el beneficio que comporta la intervención en el régimen de arrendamiento como la diferencia entre la estimación hallada del alquiler corriente de mercado y el alquiler realmente pagado, es posible analizar su distribución entre los distintos tipos de arrendamiento para multitud de factores disponibles en la EPF: según el año de ocupación, tamaño de municipio donde está ubicada la vivienda, provincia, comunidad autónoma, etc. De forma equivalente, si se relacionan dichos beneficios con las características personales de quienes los reciben, se pueden analizar las consecuencias de dicha intervención desde un punto de vista de la justicia distributiva.

En relación a la segunda línea de investigación, se puede considerar la

posibilidad de incorporar la estimación del valor del flujo de servicios de la vivienda como una dimensión más en los análisis de desigualdad, pobreza y polarización basado en los gastos (ingresos) de los hogares que componen las EPF de 1980-81 y/o 1990-91. Esto permitirá contrastar los resultados obtenidos en diversos trabajos como, por ejemplo: en INE (1996) se obtiene que el porcentaje de pobres desciende de un 20,6 por ciento en 1980-81 a un 19 por ciento en 1990-91; en Del Río y Ruíz-Castillo (1996) se obtiene una disminución de la desigualdad relativa y aumento de la absoluta para todos los índices consistentes con Lorenz y para diferentes economías de escala y variables de gasto utilizadas; en Del Río y Ruíz-Castillo (2001) se obtiene una disminución de la pobreza absoluta y relativa según una amplia clase de índices y para diferentes supuestos sobre las economías de escala y líneas de pobreza; y en Gradín (2000) se comprueba un aumento de la polarización geográfica del gasto en la década de los ochenta y se identifica el nivel educativo como la característica más determinantes en la bipolarización.

Como tercera línea de investigación abierta, se señala la posibilidad de considerar la estimación del alquiler de mercado imputable a las viviendas en propiedad y cesión en el sistema del índice de precios al consumo (IPC). Como es conocido, el IPC que elabora el Instituto Nacional de Estadística (INE), es un indicador de la evolución del nivel de inflación de referencia obligada en multitud de actuaciones con repercusión económica y social. El INE utiliza los alquileres observados en las Encuestas de Presupuestos Familiares (EPF) para establecer las ponderaciones del bien vivienda en el sistema del IPC de nuestro país. El alquiler que proporciona la EPF, real en caso de arrendamiento y autoestimación en caso de propiedad o cesión, fue utilizado por el INE para elaborar el sistema del IPC con base en 1976 y 1983; sin embargo, el INE decidió eliminar del IPC con base 1992 la vivienda en régimen

distinto del arrendamiento. Sobre esta medida, el subdirector general de esta institución, responsable del IPC comenta que: "*los alquileres autoimputados por los ocupantes de su propia viviendas (en la EPF de 1990-91) muestran un sesgo sistemático hacia una hipervaloración, con lo que su ponderación (en la elaboración del IPC con base 1992) estaría sobreestimada*" (Castro, 1992). En esta tesis se muestra evidencia empírica sobre esta afirmación, sin embargo, y en cualquier caso, esto no es suficiente, como se discute en Ruíz-Castillo et al., (1999), para justificar el cambio de procedimiento seguido por el INE en 1992 respecto a los sistemas del IPC anteriores. Estos autores analizan los efectos de esta decisión y estiman que: "*...omitir los alquileres imputados (autoestimación de la EPF de 1992),...supone incurrir en un sesgo (en la medición de la inflación) de -0,241 por ciento al año*" (p. 118).

Como aportación de interés, en esta tesis se estima el valor del flujo de servicios de las viviendas ocupadas en propiedad y cesión según las EPF de 1980-81 y 1990-91 respectivamente, lo que permitirá estimar los efectos en la inflación oficial del INE, una vez se considera dicha estimación del gasto en vivienda en la elaboración del IPC con base 1983 y 1992 respectivamente.

A expensas de este resultado, disponer de un alquiler imputado de mercado hallado con criterios objetivos, similar al proporcionado en esta tesis, puede ser una propuesta válida para estimar el cambio en el coste de los servicios de la vivienda en el régimen distinto al arrendamiento y, por consiguiente, una herramienta apropiada para que la totalidad de los hogares estén representados a la hora de establecer las ponderaciones del bien vivienda en el sistema del IPC.



## Referencias bibliográficas

Albácar, J.L. (1996). '*Legislación de Arrendamientos Urbanos: Doctrina y Jurisprudencia*', Trivium, S.A. Madrid.

Alonso-Colmenares, M.D., Lara, A., Arévalo, R., y Ruiz-Castillo, J. (1994). "*La Encuesta de Presupuestos Familiares de 1980-81*", Documento de Trabajo 94-12, Serie de Economía 05, Universidad Carlos III de Madrid. Publicación electrónica accesible en la página web <<http://www.eco.uc3m.es/epf80-81.html>>.

Arévalo, R., Cardelus, M.T y Ruiz-Castillo, J. (1995). "*La Encuesta de Presupuestos Familiares de 1990-91*", Documento de Trabajo 95-07, Serie de Economía 05, Universidad Carlos III de Madrid. Publicación electrónica accesible en la página web <<http://www.eco.uc3m.es/epf90-91.html>>.

Ball, M.J. (1973), "Recent Empirical Work on the Determinantes of Relative House Prices", *Urban Studies*, 10, 213-231.

Benzécri, J-P.(1979). 'Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum [BIN.MULT.]'. *Cahiers de l'Analyse des Données* 4, 377-378.

Bösch-Supan, A. (1986). "On the West German Tenants Protection

Legislation". *Journal of Institutional and Theoretical Economics*, 142, Nº2, 380-404.

Bover, O. (1993). "Un modelo empírico de la evolución de los precios de la vivienda en España (1976-1991)", *Investigaciones Económicas*, Vol. XVII, 65-86.

Court, A.T. (1939). "Hedonic Price Indexes with Automotive Examples", *The Dynamics of Automobile Demand*, New York: The General Motors Corporation, 99-117.

García, J y Mas, M. (2000). "La Vivienda y el Sector de la Construcción en España", *Instituto Valenciano de Investigaciones Económicas*, Caja de Ahorros del Mediterraneo, CAM. Depósito Legal: V-226-2000.

González-Páramo, J.M y J. Onrrubia (1992). "El Gasto en Vivienda en España", *Hacienda Pública Española*, 120/121, 189-231.

Goodman, A. C. (1998). "Andrew Court and the Invention of Hedonic Price Analysis", *Journal of Urban Economics*, Nº 44, pp. 291-298.

Greenacre, M.J (1984). 'Theory and applications of correspondence analysis'. London: Academic Press.

Greenacre, M.J (1990). 'Some Limitations of Multiple Correspondence Analysis'. *Computational Statistics Quarterly*, 3, 249-256.

Greenacre, M.J (1991). 'Interpreting Multiple Correspondence Analysis'. *Applied Stochastic Models and Data Analysis*, 7, 195-210.

Griliches, Z. (1961). "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change", in *The Price Statistics of the*

*Federal Government*, General Series N° 73. New York: Columbia University Press for the National Bureau of Economic Research, pp. 137-196. Reprinted in Zvi Griliches, ed., *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass.: Harvard University Press, 1971, pp. 55-87.

Griliches, Z. (1971). "*Prices Indexes and Quality Change*", Harvard University Press.

Heckman, J. (1979). "Sample selection bias as a specification error". *Econometrica* Vol. 47, 1, 153-161.

Hubert, F. (1995). "Contracting with costly tenants". *Regional Science and Urban Economics* 25, 631-654.

Instituto Nacional de Estadística (1983a) 'Encuesta de Presupuestos Familiares. 1980-81. Tomo I: El gasto y el ingreso de los hogares. Conjunto Nacional'.

Instituto Nacional de Estadística (1983b) 'Encuesta de Presupuestos Familiares 1980-81. Tomo II: Equipamiento y condiciones de las viviendas familiares. Conjunto Nacional', Madrid.

Instituto Nacional de Estadística (1993a) 'Encuesta de Presupuestos Familiares. 1990-91. Metodología', Madrid.

Instituto Nacional de Estadística (1993b) 'Encuesta de Presupuestos Familiares. 1990-91. Tomo I: Gastos e ingresos de los hogares', Madrid.

Instituto Nacional de Estadística (1993c) 'Encuesta de Presupuestos Familiares. 1990-91. Tomo III: Equipamiento de los hogares y condiciones de las viviendas familiares', Madrid.

Instituto Nacional de Estadística (1982) 'Censos de Población y Viviendas. Año 1981', Madrid.

Instituto Sindical de Estudios (1989), 'Evolución social en España 1989. La situación de la vivienda en España durante 1989 y tendencias', Madrid.

Lasheras, M.A., Salas, R., y Pérez, E. (1994). "Efectos de los incentivos fiscales en España sobre la adquisición de viviendas, en Arellano, M. (ed.), *Modelos Microeconómicos y Política Fiscal*, Instituto de Estudios Fiscales, Madrid, 147-173.

Lebart, L., Morineau, A. y Warwick, K.M. (1984). ' *Multivariate descriptive statistical analysis*'. Wiley, Chichester, UK.

Linneman, P. (1980). "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market", *Journal of Urban Economics*, 8, 47-68.

López-García, M.A. (1992a). "Algunos aspectos de la economía y la política de la vivienda", *Investigaciones Económicas*, Vol. XVI, 1, 3-41.

López-García, M.A. (1992b). "El Informe del Comité de Expertos sobre Vivienda: una valoración tentativa", *Hacienda Pública Española*, 122, 161-194.

López-García, M.A. (1997). "La Fiscalidad y la Política de la Vivienda en España", *Hacienda Pública Española*, 141/142, 287-334.

Marks, D. (1984). "The Effect of Rent Control on the Price of Rental Housing: an Hedonic Approach", *Land Economics*, 60, 81-94.

Miron, J.R. (1990), "Security of tenure, costly tenants and rent regulation", *Urban Studies*, 27, N°2, 167-184.

Ministerio de Obras Públicas y Transportes - Comité de Expertos de Vivienda (1992), *Informe para una Nueva Política de Vivienda*, V90, Centro de Publicaciones, M.O.P.T., Madrid.

Ministerio de Obras Públicas y Urbanismo (1989) "Situación de la vivienda en alquiler". Subdirección General de Estadística y Estudios Económicos. Dirección General de Programación y Coordinación Económica.

Moulton, B. (1997), "Issues in Measuring Prices Changes for Rent of Shelter", mimeo, Bureau of Labor Statistics.

Peña, D. y Ruiz-Castillo, J. (1984). "Robust Methods of Building Regression Model -An application to the Housing Sector", *Journal of Business Statistics*, Vol. 2, N°1, 10-20.

Peña, D. y Yohai, V.J. (1995). "The Detection of Influential Subsets in Linear Regression by using an Influence Matrix", *Journal of the Royal Statistical Society B*, 57, N°1, 145-156.

Quigley, J.M. (1979), "What Have We Learned about Urban Housing Markets?" *Current Issues in Urban Economics*, eds. P.Mieszkowski and M. Straszheim, The Johns Hopkins University Press.

Randolph (1988). "Housing Depreciation and Aging Bias in the Consumer Price Index", *Journal of Business and Economic Statistics*, 6, pp. 359-371.

Rosen, S. (1974). "Hedonic prices and implicit markets: product differentiation in pure competition", *Journal of Political Economy*, 82, 34-55.

Saura, P. (1995). "Demanda de características de la vivienda en Murcia, *Secretariado de Publicaciones de la Universidad*, Bibliografma.

Taltavull, P. (2000). "*Vivienda y Familia*", Colección Economía Española, Vol. XIII, Fundación Argentaria, Madrid.

Tenenhaus, M. y Young, F.W. (1985). 'An analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data', *Psychometrika*, 50, 1, 91-119.

Triplett, J. (1990). "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy", en Berndt, E y J. Triplett (eds.), *Fifty Years of Economic Measurement*, NBER, University of Chicago Press.

Waugh, F. V. (1928). "Quality as a Determinant of Vegetable Prices", *Journal of Farm Economics*, 10,2, April, 185-196.

