

ON THE IMPUTATION OF RENTAL PRICES TO OWNER-OCCUPIED HOUSING

Raquel Arévalo
Universidad de Vigo

Javier Ruiz-Castillo
Universidad Carlos III de Madrid

Abstract

This paper challenges the usual objections to the possibility of applying the rental equivalent approach to determine the weight that nonrental housing services should have in the Consumer Price Index (CPI). Using data from two Spanish household budget surveys, it is shown that market rents can be well represented in terms of an index of housing quality, two geographical variables, and the year of occupancy. This parsimonious empirical model is used to impute a rental value to nonrental housing units, taking into account the possible selection bias induced by systematic differences in housing characteristics between the market rental sector and the nonrental stock. On average, the estimated hedonic values are relatively close to the self-imputations provided in the household surveys by the occupants of such dwellings. Therefore, using either of the two alternatives to assess the importance of nonrental housing services in the CPI have small consequences for inflation. Instead, dropping these services from the CPI creates a downward bias in the measurement of inflation of 0.33 percentage points per year during 1985–1992, and an upward bias of 0.38 percentage points per year during 1993 to 2000. (JEL: C43, D12, R21, C21, E31)

1. Introduction

In most countries, the Consumer Price Index (CPI) is used every year to index many public expenditure programs, as well as tax brackets, personal exemptions, and other features of the income tax. Moreover, it is often used as a basic reference in wage negotiations in the private and public sectors. Finally, it serves to express, in real terms, money measures of aggregate consumption, labor earnings, and the poverty line. Therefore, even a small error in the CPI during extended periods of time may cause havoc in any economy.

Acknowledgments: This paper contains material from the first author's Ph.D. dissertation defended at the Universidad Complutense of Madrid in June 2001 (Arévalo 2001). We are grateful for comments from Raquel Carrasco, Sergi Jimenez, Eduardo Ley, and Gregorio Serrano, as well as from the members of the dissertation committee—Olimpia Bover, Omar Licandro, and Paloma Taltavull. Two referee reports and the editor Alan Krueger helped to improve the original version of the paper. Financial support from project BEC2000-0173 from the Spanish DGI is gratefully acknowledged.

E-mail addresses: Arévalo: arevalo@uvigo.es; Ruiz-Castillo: jrc@eco.uc3m.es

Journal of the European Economic Association June 2006 4(4):830–861
© 2006 by the European Economic Association

This paper analyzes the possible bias caused by the treatment of housing in the CPI, one of the more vexing problems for theorists, official statisticians, central bankers, and other users of this fundamental statistic. The difficulty is that, in the absence of observable transactions between owners and users of housing services, it is not obvious at all (i) how to determine the weight to be given to owner-occupied housing in the CPI; and (ii) exactly which prices should be monitored over time. Lacking clear answers to these questions, many countries exclude nonrental housing from the CPI. Among European countries, for example, Austria, Belgium, France, Greece, Italy, Luxembourg, Portugal, and Spain have adopted this approach. This is also the case of the Harmonized Index of Consumer Prices (HICP), the official indicator of inflation in the EU produced by Eurostat. Given the large differences in the proportion of nonrental housing across countries, dropping its services from the HICP makes inflation rates essentially noncomparable across the EU, a situation that makes it harder to conduct and evaluate European monetary policy.

This paper presents new evidence from Spain about the possibility of using the rental equivalent approach to solve the first of the described two problems.¹ Two radically different ways of imputing a rental value to each nonrental housing unit lead to practically indistinguishable inflation rates for 1985–2000. In comparison, dropping those services from the CPI leads to an important bias in the measurement of inflation during the 1985–2000 period.

To introduce the issues involved, accept that the economic approach to index number theory provides the conceptual framework for the country's CPI.² This approach suggests that the relevant commodity to be included in the CPI is the flow of housing services provided by the nonrental housing stock. As Triplett (2001, p. F327) indicates with respect to owner-occupied housing: "The concept of consumption implies that the standard of living depends on the consumption of housing services, and not on the purchase of houses." In turn, according to the rental equivalent approach, to price this flow there are two alternatives.³

The first is to simply ask occupants (or experts) the rent they think that the dwellings in question could carry in the rental market. Self-imputations by occupants and expert judgments are currently used in the US and the Netherlands,

1. As pointed out in Diewert (2003), the rental equivalent approach can be traced back to Marshall (1898, p. 594) at least, and it is the approach taken for owner-occupied housing by most countries in the system of National Accounts; see Eurostat, World Bank, International Monetary Fund, OECD, and United Nations (1993).

2. This is actually the case in the US, the Netherlands, and Sweden (United States Department of Labor 1997; Balk 1994; Dalén 1999). For in-depth discussions of the superiority of this position versus the alternatives advocated by Hill (1999) and Turvey (1999), see Triplett (2001) and Diewert (2000). For alternatives to the economic approach in the housing sector, such as the net acquisition and payment approaches, see Turvey (1989, 2000).

3. Within the economic approach, an alternative way of estimating the opportunity costs of nonrental housing is the user cost approach. See, for instance, Smith, Rosen, and Fallis (1988), Diewert (2003) and Triplett (2001), as also note 22.

respectively, to determine the official weight assigned to owner-occupied housing services in the CPI. In Spain, self-imputations are collected in the household budget surveys known as Encuestas de Presupuestos Familiares (EPF), gathered by the Instituto Nacional de Estadística (INE) with the main purpose of estimating the official weights of the Spanish CPI. These rental values were used in the CPI systems based in 1976 and 1983. However, possibly because the prices of the housing stock had been through an upward cycle since Spain joined the EU in 1986, the self-imputed values collected in the 1990–1991 EPF were thought to be too high, and nonrental housing services were eliminated from the CPI system based on 1992 (see Castro (1992), the spokesman for the INE at the time).

The second alternative is to use hedonistic regression methods to estimate rental values for nonrental housing from information collected on housing characteristics for the whole stock and on observed rents in the market rental sector. The usual objection to this procedure is that rent controls or public subsidized rents interfere with the workings of the rental housing sector, or that the characteristics of the dwellings of a very thin market rental sector are not representative of those of a very large nonrental housing stock. Spain is a case in point. The market rental sector only represents 8.9% and 5.3% of permanently occupied residential housing in 1980–1981 and 1990–1991, respectively. This small sector, which is still partly regulated, operates under the influence of a rent controlled and a publicly subsidized rental sector. In the other side of the spectrum, a dominant nonrental sector represents 71.1% and 85.0% of the housing stock in 1980–1981 and 1990–1991, respectively.

Under these conditions it would appear that, whatever its conceptual merits, the rental equivalence approach to the determination of a CPI weight to nonrental housing services is bound to fail in practice. Therefore, the scene is set for the experiment conducted in this paper. In the first place, the hedonic approach to the task of imputing rental values to nonrental housing services is applied to the information in the 1980–1981 and 1990–1991 EPFs. Second, the hedonic rental values (“objectively”) imputed to nonrental housing are compared with the self-imputations (“subjectively”) suggested by the occupants.⁴ Finally, the official inflation from August 1985 to December 2000 is compared with the inflation that would be obtained if the CPI weights for nonrental housing had been constructed using our hedonic estimates.

From a methodological perspective, the hedonic imputation method advocated in this paper introduces two novelties upon the traditional procedure. First, multiple correspondence analysis (MCA) is used to select the linear combination of physical attributes that accounts for the maximum variance in this

4. This is what is done in Francois (1989), who used the traditional hedonic approach with a sample of new tenants only. Alternatively, Kurz and Hoffmann (2004) compared the results of applying the hedonic approach to rents in the market rental sector and to self-imputations provided by the occupants in the owner-occupied sector. For their results, see footnote 17.

multiple-dimensional space for the entire housing stock. It is shown that this scalar index can be interpreted as a housing quality index, and that it provides as good an explanation of market rents in 1980–1981 as the usual alternative where all physical attributes are included as separate regressors. This justifies our choice of the more parsimonious model to also analyze the 1990–1991 data. Second, this paper recognizes the possible interdependence between the choices of tenure mode and housing characteristics. To correct for the possible selection bias caused by systematic differences in housing characteristics between the market rental sector and the nonrental housing stock, a Heckman procedure is used for the first time in this context.

From an empirical point of view, the most remarkable result of the paper is the broad agreement between the hedonic rental values and the self-imputations for nonrental housing services. On average, hedonic values in 1980–1981 are 12% higher, and in 1990–1991 they are 13% lower, than self-imputed values. Therefore, using either of the two alternatives to assess the importance of nonrental housing services in the CPI system has very small consequences for inflation. Instead, if the rental equivalence approach is abandoned and nonrental housing services are dropped from the CPI, the weight of the “nonrental housing” commodity goes down by about 10 percentage points. Given that housing prices rose by less than the prices of the remaining commodities in the CPI from August 1985 to December 1992 and by more than the remaining commodities from January 1993 to December 2000, inflation without nonrental housing services would have been 0.33 percentage points per year higher during the first period and 0.38 percentage points per year lower during the second period.

The rest of the paper consists of four sections and an Appendix devoted to the construction of the housing quality index. Section 2 presents the minimum information on the institutional background and data sources that are necessary to understand the task at hand. Section 3 serves two purposes: (i) It compares the traditional hedonic model for the explanation of market rents with an alternative in which a vector of physical attributes is substituted for a housing quality index; and (ii) It presents the regression results in a model that takes into account the possible interdependence between the choices of tenure mode and physical attributes. Section 4 compares the hedonic estimates with the self-imputed values provided by their occupants, and studies the consequences for inflation of using the hedonic estimates to determine the weight of nonrental housing services in the CPI. Section 5 concludes.

2. Institutional Background and Data Sources

As indicated in the Introduction, the data sources for this paper are the 1980–1981 and 1990–1991 EPFs collected from April 1980 to March 1981, and from April 1990 to March 1991, respectively. These two household budget surveys consist

TABLE 1. Permanent residential housing by tenure mode and legal status in the 1980–1981 and 1990–1991 EPFs. Sample and population statistics.

| Tenure mode | 1980–1981 EPF | | | 1990–1991 EPF | | |
|-------------------------|---------------|------------|-------|---------------|------------|-------|
| | Sample | Population | % | Sample | Population | % |
| I. Rental housing | 5,484 | 2,297,105 | 22.9 | 2,975 | 1,694,184 | 15.0 |
| 1. Market sector | 2,181 | 888,945 | 8.9 | 1,061 | 601,970 | 5.3 |
| 2. Rent controlled | 968 | 405,290 | 4.0 | 229 | 129,968 | 1.1 |
| 3. Public housing | 1,787 | 773,164 | 7.7 | 771 | 426,875 | 3.9 |
| 4. Unknown legal status | 535 | 223,795 | 2.2 | 914 | 535,371 | 4.7 |
| II. Owner-occupied | 16,427 | 6,928,150 | 69.1 | 16,623 | 8,789,287 | 77.8 |
| 5. Market sector | 9,307 | 4,104,814 | 41.0 | 9,132 | 4,883,659 | 43.2 |
| 6. Public housing | 5,316 | 2,048,206 | 20.4 | 5,222 | 2,676,098 | 23.7 |
| 7. Unknown legal status | 1,804 | 775,130 | 7.7 | 2,269 | 1,229,530 | 10.9 |
| III. Other | 2,060 | 798,911 | 8.0 | 1,557 | 815,038 | 7.2 |
| All | 23,971 | 10,024,166 | 100.0 | 21,155 | 11,298,509 | 100.0 |

Note: This sector I includes 13 sample observations, representative of 5,911 housing units at the population level, that cannot be classified in the market or the rent controlled sector because they lack information on the year of first occupancy.

of 23,971 and 21,155 observations, respectively, representative of the household population occupying non institutional, or residential housing in all of Spain. This population is around 10 million and 11.3 million households in 1980–1981 and 1990–1991, respectively. The INE provides a set of blowing up factors to convert sample statistics into population statistics. Both EPFs heavily underestimate the amount of secondary housing recorded in the 1981 and 1991 Housing Censuses, but cover approximately 96% of permanently occupied residential housing. Therefore, only the latter will be studied herein.⁵

To understand the problem addressed in this paper, it is essential to classify the housing stock according to both tenure mode and the legal status determined by public policies in the housing sector. As can be seen in Table 1, three tenure modes are distinguished. Rented dwellings and dwellings occupied by their owners constitute the rental and the owner-occupied sector, respectively. In the third mode, referred to as the “Other” note, occupants do not own or rent the dwellings, but use them either as wages in kind or as a transfer from a public organism, a private institution or an individual person.

As far as the legal status is concerned, in Spain, as in other countries, government intervention in the housing sector takes many forms. For our purposes, it will suffice to describe the stylized features of two major policies. In the first place, the public sector has entirely financed some housing construction but, starting from the 1950s, most government intervention has operated through a variety of programs that provide incentives to housing construction by means of moderate

5. All of the information used in this paper has been made accessible at <http://www.eco.uc3m.es/epf80-81.html> and <http://www.eco.uc3m.es/epf90-91.html>. For sampling methods and a full description of the data, see Instituto Nacional de Estadística (1983, 1992).

direct subsidies, low interest rates, or fiscal credits to the producers. Complex public regulations, the details of which need not be discussed here, mandate that housing built under these policies are rented or sold below market prices. As a result of this type of government intervention, some housing units in the rental and the owner-occupied sectors will be classified as public housing (see rows 3 and 6 in Table 1, respectively).

Within the owner-occupied sector, all private housing units integrate what might be called the market sector (row 5). However, due to the public policy known as rent control, in the rental sector this is not the case. Together with the compulsory renewal clause on all leases, which dates from 1920, a 1946 law systematically froze all rentals at the level reached at the time of the first contract, following up on transitory regulations in the same vein already in effect in previous years. A 1964 law sanctioned a dramatic policy change: Although lease renewals were still compulsory, rents in new contracts were allowed to be determined by market forces; owners and renters were also allowed to include rent revision clauses subject only to annual ceilings set by the government. However, for all housing rented before 1964, the power of rent revision remained with the government and only moderate increases have been permitted. Therefore, a distinction must be made between pre-1964 or rent-controlled housing (row 2), and post-1964 or market rental housing (row 1) that resembles what has been termed second-generation rent control in Arnott (1995) or tenancy rent control in Basu and Emerson (2000). Within the period covered in this paper, which extends until 1990–1991, legislation enforced since April 1986 did away with compulsory renewal clauses and completely liberalized the rental sector.

Against this background, the problem addressed in the next section is how to use the information on housing characteristics and rents in the market rental sector in row I.1 of Table 1 (i.e., 2,181 and 1,061 observations in 1980–1981 and 1990–1991, respectively), to impute a rental value to all dwellings in the nonrental housing sector for which the EPFs provide self-imputations made by their occupants (i.e., rows II and III, or 18,487 and 18,180 observations in 1980–1981 and 1990–1991, respectively).

3. Imputing Hedonic Rental Values to Nonrental Dwellings

In the hedonic approach to the study of heterogeneous, differentiated commodities, unobservable quality differences between two product varieties are assumed to be well approximated by a set of observable physical attributes. This provides the original rationale to explain product prices in terms of product characteristics. Rosen (1974) established the microeconomic foundations of the approach by means of a perfect competitive model in which the price of an indivisible, differentiated product is jointly determined by the interaction of supply and

demand for the product's attributes. Although the underlying demand and supply behavioral relations cannot be identified from the knowledge of product prices and product characteristics, partial derivatives with respect to each characteristic in a hedonic regression can be interpreted as the implicit marginal equilibrium price of the attribute in question.⁶

As time goes by, some housing units become vacant and others remain occupied. Rent inflation in this sector manifests itself through new contracts on those units that become rented for the first time in each period. Rents on dwellings that remain occupied grow in time as a consequence of rent renewal clauses. Assuming away aging effects, on average it would be indifferent to rent a housing stock of given characteristics in the current or in some previous period if and only if renewal clauses exactly capture the impact of inflation. Only in this case should the year of occupancy not be one of the explanatory variables on a hedonic regression explaining current rents in the survey period.

However, there are good reasons to expect a negative mean impact of the year of occupancy on present rents, meaning that owners would be giving up a discount to sitting tenants renewing their rents, relative to the rents charged in new contracts for housing of the same characteristics. In the first place, there are theoretical models yielding what is known as *tenure discounts* (see, inter alia, Guasch and Marshall 1987; Miron 1990; and Hubert 1995). Landlords know that tenants will incur moving costs if they leave the unit. However, several factors work in the opposite direction. (i) Landlords also incur costs when a tenant moves out of a unit, including the cost of reconditioning, the cost of marketing a vacancy, and the rental income forgone during the vacancy. (ii) Landlords may want to retain current tenants that have shown during some period to be "good" tenants, minimizing wear and tear, avoiding trouble with neighbors, and so on. (iii) There could be a tenant's decreasing willingness to pay with the passage of time. Thus, tenure discounts may appear as an equilibrium phenomenon in the game played by tenants and landlords. Second, it should be remembered that, in the Spanish case, new contracts signed between 1964 and 1986 have an automatic clause compelling landlords not to evict tenants for up to two generations except in a very restricted set of circumstances. This shifts considerable bargaining power towards tenants in the rent renewal process, leading presumably to large tenant discounts (Börsch-Supan 1986; Nagy 1997; and Basu and Emerson 2000). Moreover, rent increases for sitting tenants had to comply with an annual governmental ceiling, typically linked to the previous year's housing inflation rate.

6. See Quigley (1979) for a survey of the early literature on housing, and Sheppard (1999) for a recent one. Triplett (1990) contains a guide to the approach, as well as an illuminating account of why statistical offices have resisted its use for 30 years.

Therefore, as Hoffman and Kurz (2002) conclude, both the peculiarities of housing markets and government regulations may result in tenancy discounts.⁷ To capture that phenomenon, the year of first occupancy should appear as an explanatory variable of current market rents. Assume that in the current year T there is a sample consisting of n_t observations of housing units first rented at time $t = 1965, 1966, \dots, T$, so that the sample size is $n = \sum_t n_t$. Denote by $r = \cup_t \{r_t\}$ the set of rents actually paid in year T , $X_1 = \cup_t \{X_{1t}\}$ the set of physical housing attributes, $X_2 = \cup_t \{X_{2t}\}$ the set of geographic characteristics, and $X_3 = \{Occup65, Occup66, \dots, OccupT\}$ an index set of years of occupancy where, for each t , $Occup_t = 1$ if the housing unit was first rented in year t and $Occup_t = 0$ otherwise. Each housing observation, indexed by $i = 1, \dots, n$, can be described by

$$(r_i, X_{1i}, X_{2i}, X_{3i}) \in r \times X_1 \times X_2 \times X_3.$$

Assume for the time being that the impact of physical and geographic characteristics on rents is independent of the year the housing unit was first rented. Under this simplifying assumption, rather than working with several separate models $r_t = f(t, X_{1t}, X_{2t}, \varepsilon_t)$, $t = 1965, \dots, T$, it is possible to work with a single model:

$$r = g(X_1, X_2, X_3, \varepsilon), \quad (1)$$

where ε is a normally distributed random term with $E[\varepsilon] = 0$ and $\text{Var}[\varepsilon] = \sigma_\varepsilon^2$. Two empirical specifications of this model are contrasted using data from the 1980–1981 EPF. In the first place, the inspection of the sample rent distribution, as well as the residuals of some preliminary linear specification of model (1), led to the following semi-logarithmic functional form with parameters $(\alpha, \beta_1, \beta_2$ and $\gamma)$:⁸

$$\ln r = \alpha + X_1\beta_1 + X_2\beta_2 + \sum_t \gamma_t \text{Occup}_t + \varepsilon. \quad (2)$$

Second, as explained in the Appendix, a (ordinal) housing quality index (*Qindex*) that summarizes the physical attributes has been constructed (see Table 2 for some descriptive statistics of this variable). Thus, the empirical specification in equation (2) will be contrasted with the following one where this index substitutes for the entire vector X_1 of physical characteristics:

$$\ln r = \alpha + \beta \text{Qindex} + X_2\beta_2 + \sum_t \gamma_t \text{Occup}_t + \nu. \quad (3)$$

7. There is ample empirical evidence on tenure discounts in several countries under different legal arrangements. For the US, see, for instance, Lowry (1981), Goodman and Kawai (1985) and Clark and Heskin (1982). For Germany, see Börsch-Supan (1986), Schlitch (1983), and Hoffman and Kurz (2002). For Spain, see Peña and Ruiz-Castillo (1984).

8. There is a large amount of literature on the appropriate choice of functional form for the hedonic price function (see the discussion in Sheppard (1999), for example), but the simple log-linear form generally performs well.

TABLE 2. Frequency distribution and mean monthly rent (in euros) for geographic characteristics, years of occupancy, and quality index in the market rental sector, 1980–1981 and 1990–1991.

| | 1980–1981 | | 1990–1991 | |
|---|------------------------|-----------|------------------------|-----------|
| | Frequency distribution | Mean rent | Frequency distribution | Mean rent |
| All | 888,945 = 100.0% | 33.8 | 601,970 = 100.0% | 95.5 |
| 1. Municipal size dummies | | | | |
| Less than 2,000 Inhabitants | | | | |
| 2,001–10,000 | 7.6 | 19.9 | 3.0 | 53.0 |
| 10,001–50,000 | 14.7 | 20.6 | 13.3 | 74.4 |
| 50,001–500,000 | 20.7 | 25.0 | 23.6 | 88.5 |
| More than 500,000 | 34.8 | 36.6 | 35.8 | 107.4 |
| | 22.2 | 51.3 | 24.2 | 132.6 |
| 2. Provinces with mean housing price ¹ | | | | |
| ² Below 70 euros/m ² | 5.5 | 26.8 | 5.2 | 62.8 |
| ³ 71–100 euros/m ² | 40.8 | 26.5 | 36.1 | 99.6 |
| ⁴ 101–130 euros/m ² | 19.0 | 33.1 | 20.2 | 101.0 |
| ⁵ Above 130 euros/m ² | 34.7 | 44.0 | 38.5 | 112.7 |
| 3. Years of occupation | | | | |
| During the 1964–1973 period | 43.3 | 20.5 | 29.2 | 40.8 |
| During the 1974–1980 period | 56.7 | 44.0 | 70.8 | 128.7 |
| Dummy variable for the 1965–1973 period | 43.3 | 20.5 | 29.2 | 40.8 |
| Dummy variable for the 1974–1980 period | 56.7 | 44.0 | 24.8 | 104.6 |
| 4. <i>Q</i> index, quartiles | Mean | Mean rent | mean | mean rent |
| I | <i>Q</i> index | 17.3 | <i>Q</i> index | 56.6 |
| II | -0.72 | 27.1 | -5.14 | 84.6 |
| III | -0.12 | 40.7 | -1.80 | 118.7 |
| IV | 0.22 | 71.3 | 0.76 | 176.8 |
| | 0.65 | | 6.24 | |
| 5. <i>Q</i> index if occupation took place: | | | | |
| During 1965–1973 | -0.15 | 20.5 | -1.70 | 40.8 |
| During 1974–1980 | -0.09 | 44.0 | -0.27 | 104.6 |
| | | | 0.05 | 141.7 |

Notes: ¹ The source for the construction of the dummy variables for the following provincial groups is *Ministerio de Fomento*, <http://www.infom.estadisticas>.² Cáceres, Murcia, Badajoz, Lugo, Ceuta, Melilla.³ Cádiz, Girona, Málaga, Segovia, Granada, Palencia, León, A Coruña, Almería, Soria, Toledo, Tarragona, Ávila, Guadalajara, Huelva, Valencia, Castellón, Huesca, Córdoba, Ourense, Alicante, Ciudad Real, Lleida, Cuenca, Albacete, Zamora, Teruel, Jaén, Pontevedra, Santa Cruz de Tenerife.⁴ Burgos, Palma de Mallorca, Santander, Salamanca, Logroño, Las Palmas de Gran Canaria, Valladolid, Zaragoza, Oviedo, Sevilla.⁵ Madrid, Barcelona, San Sebastián, Bilbao, Vitoria, Pamplona.

In this case, consistently with the hedonic approach, the β coefficient can be interpreted as the implicit marginal equilibrium price of housing quality.

There are several reasons why the specification in equation (3) might be preferred to the traditional one. (i) It permits the multicollinearity problem, which may disturb the precise estimation of the physical attributes in vector X_1 , to be overcome. (ii) By construction, the quality index is orthogonal to the remaining indicators in the MCA. Therefore, using only $Qindex$ in equation (3) filters possibly irrelevant information as far as explaining market rents is concerned. (iii) It permits estimating the impact of housing characteristics that, due to their infrequency, would have no explanatory power in the traditional approach. For example, in 1980–1981 only 14 out of 2,181 observations in the market rental sector lacked electricity. In the second approach, this attribute may influence market rents through its effect on the housing quality index. (iv) The study of interactions between the variables synthesized in the scalar index and other explanatory variables become considerably easier. In this context, this is the case with respect to the interactions with the year of occupancy.

From an empirical point of view, the comparison of equations (2) and (3) shows that the coefficients of the geographical variables and the year of occupancy follow essentially the same pattern. Interestingly enough, Peña and Yohai's (1995) procedure yields exactly the same 90 outliers in both cases. More important, in spite of the reduction in the number of parameters, the goodness of fit in the second model is essentially preserved (the R^2 is equal to 0.586 and 0.578 in the two models). The conclusion is that the two models perform equally well in explaining 1980–1981 market rents.⁹ This reinforces the idea that the index constructed in the Appendix can be appropriately interpreted as a housing quality index, a fact that might prove important in other contexts. For instance, together with household income or total expenditures, a meaningful housing quality index could be used in the evaluation of household welfare or as a conditioning variable in econometric studies of household demand. These considerations justify our choice of the more parsimonious model that uses only the variable $Qindex$ in the analysis of both the 1980–1981 and the 1990–1991 market rental sectors.

Given the possibility of explaining a large part of the variance in market rents by means of a satisfactory statistical model, it would appear that the next task is to impute a monetary value to the flow of housing services provided by nonrental dwellings. However, if the choices of tenure mode and housing characteristics are not independent, the OLS estimation in the market rental sector might be inconsistent, as well as unbiased.

In order to deal with this problem, we begin by recognizing that the total sample of housing units, indexed by $i = 1, \dots, M$, consists of two sets of

9. For a full discussion, including descriptive statistics and detailed regression results, see Arévalo and Ruiz-Castillo (2004, Section III.2).

observations: Uncensored observations in the set $\{i = 1, \dots, N \mid i \in \Lambda\}$, where Λ denotes the set of market rental units, and censored observations in $\{i = N + 1, \dots, M \mid i \notin \Lambda\}$ for which market rents r_i are obviously not observed. Let z_i^* be the latent variable that determines the sample selection, so that r_i is only observable when $z_i^* \geq 0$. The implication is that the regression model (3) has to be reformulated introducing a selection mechanism for z_i^* .

For that purpose, let Y be the vector that includes the variables that have been shown in Arévalo and Ruiz-Castillo (2004) to have a significant role in explaining market rents; namely a constant, the geographical variables, the year of occupancy variables, the housing quality index and its interactions with the year of occupancy (see Table 2 for some descriptive statistics, including mean market rents). Then we have

$$\ln r_i = Y_i \beta + v_i \quad \text{if } i \in \Lambda \quad \text{and} \quad z_i^* = Z_i \delta + u_i, \quad (4)$$

where v is normally distributed with $E[v_i] = 0$ and $\text{Var}[v_i] = \sigma_v$, Z_i is the set of variables that permit classifying each housing unit according to its tenure mode, and u is a normally distributed variable with $E[u_i] = 0$ and $\text{Var}[u_i] = 1$. Therefore,

$$\Pr(i \in \Lambda) = \Phi(Z_i \delta) \quad \text{and} \quad \Pr(i \notin \Lambda) = 1 - \Phi(Z_i \delta),$$

where Φ is the distribution function of a standard normal.

In order to get consistent and efficient estimates, the model is estimated by maximum likelihood methods, taking as initial values the results of the estimation of Heckman (1979) selection model in two stages. In this context, the log likelihood for each observation i , l_i , is equal to

$$w_i \ln \Phi([Z_i \delta + (\ln r_i - Y_i \beta) \rho / \sigma] / (1 - \rho^2)^{1/2}) - (w_i / 2) ([\ln r_i - Y_i \beta] / \sigma)^2 - w_i \ln(2\pi)^{1/2} \sigma_v$$

if $i \in \Lambda$ and to $w_i \ln \Phi(-Z_i \delta)$ if $i \notin \Lambda$, where ρ is the correlation coefficient between the random terms v and u , and w_i is the blowing up factor for each unit that permits to go from sample to population statistics. Given that the M observations are assumed to be independent, the function to be maximized is

$$L(\beta, \delta, \sigma, \rho \mid r, Y, Z) = \sum_i l_i.$$

Provided that the selection of the Y and the Z variables ensures the model's identification, the consistent estimation of the β parameters and the availability of the Y variables for all dwellings will make possible the imputation of a monthly rent to all units in the nonrental sector.

Together with some of the physical attributes, the geographical variables and the year of occupancy, the set Z includes the migrant condition of the household,

measured in the EPFs by means of a question asking whether the household had moved into the current municipality before or after 1975 or 1985 in the 1980–1981 and the 1990–1991 cases, respectively. According to Table 1, there are 18,487 nonrental units in 1980–1981 and 18,180 in 1990–1991. However, 1,745 and 1,570 in these two periods lack information on building age, square meters or, above all, year of occupation.¹⁰ Therefore, only 16,742 and 16,610 nonrental dwellings, representative of 7,044,514 and 8,782,425 population units have been considered in 1980–1981 and 1990–1991, respectively. Descriptive statistics about all Z variables in the market rental and the nonrental sectors are presented in Table 3.

In 1980–1981, nonrental units seem to be larger and have more hygienic services, water, garage, and telephone facilities than market rental dwellings; moreover, nonrental units are older than rental market dwellings in spite of the fact that all of the latter have been occupied after 1964. Relative to owner-occupied and transferred units, rented dwellings seem to be more prevalent in more populated municipalities and in the group of provinces with higher stock prices.¹¹ The differences in all these dimensions, however, are not very large. In 1990–1991 the pattern is exactly the same, including a greater proportion of detached units with garden facilities in the nonrental sector.

Not surprisingly, the proportion of migrant households living in rented accommodations is more than four times that of those living in nonrental housing. Given that the migration status of the household does not appear as a determinant of housing rents, this is the variable that permits identification of equation (4). The results of the estimation in both years are presented in Table 4.

In 1980–1981, but not in 1990–1991, there is evidence that the sample selection bias must be corrected.¹² In the hedonic model in the upper panel of Table 4, the observed pattern for the geographical variables is practically the same in both models: First, the greater the population of the municipality where the unit is located, the higher the estimated rent is; second, in 1980–1981 rents are higher in the provinces where the housing stock has a higher mean price, whereas in 1990–1991 rents are lower in those provinces where the housing stock has a lower mean price. In any case, in both years greater quality implies larger market rents, but the year of occupancy, as well as the interaction between this variable and housing quality, displays opposite patterns in the two samples, a feature to which we will return in the following sections.

10. Among the observations lacking some information, as many as 1,648 and 1,557 are occupied as wages in kind or as a result of a private or a public transfer.

11. Provinces have been classified into four groups taking into account the average for 1980 and 1990 of mean provincial housing prices available at the Ministerio de Fomento, <http://www.mfom.estadisticas>.

12. As pointed out in Table 4, in 1980–1981 the hypothesis of independence ($H_0: \rho = 0$) should be rejected at a 5% significance level.

TABLE 3. The market rental sector versus the nonrental housing stock: Differences in physical attributes, geographic characteristics, and other characteristics, 1980–1981 and 1990–1991.

| Physical attributes | 1980–1981 | | 1990–1991 | |
|---|---------------|-----------------|---------------|-----------------|
| | Market rental | Nonrental stock | Market rental | Nonrental stock |
| All | 100.0 | 100.0 | 100.0 | 100.0 |
| 1. Hygienic services ¹ | | | | |
| No bathroom inside the building | 4.9 | 7.3 | 7.4 | 4.1 |
| Less than one bathroom inside | 13.9 | 8.0 | 74.1 | 68.9 |
| One or more bathrooms | 81.2 | 84.7 | 18.5 | 27.0 |
| 2. Water facilities | | | | |
| Hot water | 72.2 | 78.8 | 91.5 | 95.3 |
| 3. Garage facilities | | | | |
| Garage | 11.3 | 18.6 | 19.7 | 30.4 |
| 4. Telephone facilities | | | | |
| Telephone | 37.1 | 53.6 | 65.5 | 79.8 |
| 5. Housing size | | | | |
| Less than 60 m ² | 18.6 | 9.5 | 17.8 | 7.5 |
| 61–90 m ² | 46.5 | 40.0 | 39.4 | 36.2 |
| 91–130 m ² | 28.3 | 36.3 | 30.3 | 38.9 |
| More than 130 m ² | 6.6 | 14.2 | 12.5 | 17.4 |
| 6. Building age | | | | |
| More than 30 years | 46.6 | 34.8 | 42.3 | 26.8 |
| 11–30 years | 41.0 | 41.7 | 53.1 | 58.6 |
| Less than 11 years | 12.4 | 23.5 | 4.6 | 14.6 |
| 7. Building type | | | | |
| Detached house | | | 18.1 | 38.9 |
| 8. Garden facilities | | | | |
| Garden | | | 8.5 | 16.5 |
| Geographical characteristics | | | | |
| 1. Municipal size (inhabitants) | | | | |
| Less than 2,000 | 7.6 | 12.9 | 2.9 | 8.3 |
| 2,001–10,000 | 14.6 | 21.0 | 13.1 | 20.0 |
| 10,001–50,000 | 20.5 | 21.0 | 23.9 | 22.6 |
| More than 50,000 | 57.3 | 45.1 | 60.1 | 49.1 |
| 2. Provinces with mean housing price ² | | | | |
| Below 860 euros/m ² | 5.5 | 7.2 | 4.7 | 6.7 |
| 861–1,230 euros/m ² | 40.6 | 45.2 | 37.0 | 45.0 |
| 1,231–1,700 euros/m ² | 18.6 | 18.4 | 19.8 | 17.8 |
| Above 1,700 euros/m ² | 35.3 | 29.2 | 38.5 | 30.5 |
| Other characteristics | | | | |
| 1. Years of occupation | | | | |
| More than 30 | 0.0 | 14.5 | 0.0 | 13.8 |
| 11–30 | 24.3 | 46.5 | 46.4 | 58.0 |
| Less than 11 | 75.7 | 39.0 | 53.6 | 28.2 |
| 2. Migrant household head | | | | |
| Arrived during 1975–1981 | 13.3 | 3.2 | | |
| Arrived during 1985–1991 | | | 12.9 | 3.4 |

Note: ¹ In 1990–1991: less than a full bathroom; one bathroom; more than one bathroom.

² See Table 2 for sources and provincial groupings.

TABLE 4. Hedonic regression models with a selection mechanism.

| 1980-1981 | | 1990-1991 | |
|--|-------------|-----------|----------|
| Dependent variable: log rent | | | |
| | Coefficient | t | t |
| Constant | 2.0735 | 16.6 | 3.8773 |
| Less than 2,000 inhabitants | -0.4250 | -5.2 | -0.4106 |
| 2,001-10,000 inhabitants | -0.2949 | -6.4 | -0.2793 |
| 10,001-50,000 inhabitants | -0.2480 | -6.6 | -0.2033 |
| Provinces with high housing price ⁴ | 0.2520 | 6.8 | -0.3712 |
| Years of occupation during 1965-1973 | -0.0471 | -4.4 | -0.0668 |
| Years of occupation during 1974-1980 | -0.1380 | -13.4 | -0.0325 |
| Dummy for 1965-1973 occupation period | -0.4120 | -2.0 | -0.9067 |
| Qindex during 1965-1973 | 0.6496 | 11.2 | -0.7750 |
| Qindex during 1974-1980 | 0.7886 | 14.8 | 0.1163 |
| | | | 0.0926 |
| | | | 0.0642 |
| Selection mechanism | | | |
| | Coefficient | t | t |
| Constant | -153.63 | -16.6 | -38.1483 |
| No bathroom inside the building | -0.4688 | -4.9 | 0.2149 |
| One or more bathrooms | -0.1085 | -1.5 | -0.1698 |
| Hot water | -0.2637 | -4.4 | -0.1100 |
| Garage | -0.2271 | -3.1 | -0.4808 |
| | | | 8.6 |

(Continued)

TABLE 4. CONTINUED

| | 1980–1981 | 1990–1991 |
|---|---------------------|---|
| | Selection mechanism | |
| Phone | -0.4408 | -9.5 |
| Housing size in m ² | -0.0038 | -6.1 |
| More than 30 years of building age | 1.4250 | 24.0 |
| Between 11 and 30 years | 0.7870 | 14.4 |
| Years of occupation | 0.0778 | 16.6 |
| Less than 2,000 inhabitants | -0.5153 | -6.0 |
| 2,001–10,000 inhabitants | -0.4473 | -6.9 |
| 10,001–50,000 inhabitants | -0.2796 | -5.6 |
| Provinces with low housing price ¹ | -0.2570 | -2.8 |
| Provinces with medium-low housing price ² | -0.1272 | -2.4 |
| Provinces with medium-high housing price ³ | -0.1154 | -2.1 |
| Migrant head arrived during 1975–1980 | 0.5631 | 7.3 |
| ρ | 0.214 | 2.8 |
| σ | 0.591 | 40.4 |
| $n = 18,794$ | | $n = 17,804$ |
| Wald test ($\rho = 0$) : $\chi_2(1) = 7.15$; | | Wald test ($\rho = 0$) : $\chi_2(1) = 1.85$ |
| Prob > $\chi_2 = 0.0075$ | | Prob > $\chi_2 = 0.1740$ |
| | | 0.2787 |
| | | -0.0106 |
| | | -0.4432 |
| | | -0.0958 |
| | | 0.0294 |
| | | -0.3835 |
| | | -0.2255 |
| | | -0.0913 |
| | | -0.0370 |
| | | 0.4130 |
| | | 0.149 |
| | | 0.637 |
| | | 1.4 |
| | | 25.7 |
| | | 3.6 |
| | | -10.7 |
| | | -6.6 |
| | | -2.0 |
| | | 7.8 |
| | | -3.9 |
| | | -3.5 |
| | | -1.1 |
| | | -0.8 |
| | | 5.3 |

Notes: ¹ Cáceres, Murcia, Badajoz, Lugo, Ceuta, Melilla.

² Cádiz, Girona, Málaga, Segovia, Granada, Palencia, León, A Coruña, Almería, Soria, Toledo, Tarragona, Ávila, Guadalajara, Huelva, Valencia, Castellón, Huesca, Córdoba, Ourense, Alicante, Ciudad Real, Lleida, Cuenca, Albacete, Zamora, Teruel, Jaén, Pontevedra, Santa Cruz de Tenerife.

³ Burgos, Palma de Mallorca, Santander, Salamanca, Logroño, Las Palmas de Gran Canaria, Valladolid, Zaragoza, Oviedo, Sevilla.

⁴ Madrid, Barcelona, San Sebastián, Bilbao, Vitoria, Pamplona.

As has been noted before, there are good reasons for the rent actualization process that takes place via rent renewal clauses for sitting tenants to proceed more slowly than rent increases in new contracts due to inflation. According to the 1980–1981 hedonic model in the upper panel of Table 4, the annual discount generated by the difference between the sector's inflation and the actualization clauses is 4.7% during 1965–1973 and 13.8% during 1974–1980. The accumulated discounts on rents of a dwelling of average quality in a given location, relative to a comparable unit first rented in 1980, are very large indeed since they range from 12.9% in a single year for units rented in 1979 to 74.3% for units first occupied in 1965 (for further details, see Arévalo and Ruiz-Castillo 2004).

According to the 1990–1991 hedonic model in the same panel in Table 4, the direct effect of the year of occupancy has two structural changes in 1975 and 1982, but now the annual discounts decrease with time: they are 6.7% in 1965–1975, and 3.2% in 1976–1990. Likewise, the coefficients of the quality index tend to decrease with time in the corresponding sub-periods. It would appear that, with the passage of time, an increasing number of new contracts do include rent renewal clauses. Nevertheless, the evidence of tenure discounts remains impressive, ranging from 3.2% in a single year for units rented in 1989 to 83.5% for units first occupied in 1965.¹³

To put these results in perspective, it is necessary to pay attention to the actual inflation that took place during this period in the market rental sector. Unfortunately, the Spanish INE measures the inflation rate for the entire rental sector that also includes the rent controlled and the public rental sectors. Nevertheless, Arévalo and Ruiz-Castillo (2004) concluded that the large accumulated discounts that have been estimated indicate that, relative to the official inflation for the entire rental sector, the rent increases achieved through the renewal clauses legally allowed after 1965 have been negligible—a situation that possibly requires further research beyond this paper's scope.

4. Hedonic Estimates versus Self-imputed Values

As explained in the Introduction, for all nonrental units the EPFs record the monthly rent that an informer for the household occupying the unit thinks that their dwelling would command if it were rented in the market at the time of the survey. Answers to this question will be referred to as self-imputed rental values. Because there are 13 and 4 nonrental units missing self-imputed values in 1980–1981 and 1990–1991, respectively, we are left with 18,474 and 18,176 units with complete

13. Interestingly enough, the discount estimated in Peña and Ruiz-Castillo (1984) for a sample of housing units rented in the Madrid Metropolitan Area in 1975 is 8% per year. The accumulated discount for the 1965–1975 period is 56.56%, a figure of the same order of magnitude as those shown in Arévalo and Ruiz-Castillo (2004) for the same period using the 1980–1981 and 1990–1991 EPFs.

information, representing 7,772,078 and 9,602,498 population dwellings in 1980–1981 and 1990–1991, respectively.

On the other hand, given the parameter vector β estimated in the hedonic model (4), knowledge of the Y_i variables for a nonrental unit i will suffice to produce an imputed rental value dependent on the year of occupancy. To obtain a comparable imputation to the self-imputed value provided by the occupant it is necessary to compute the rental value that the unit would command at the survey time, that is, making the variable year of occupancy equal to 1980 and 1990 in the two samples, 1980–1981 and 1990–1991, respectively.¹⁴ The resulting estimates will be referred to as hedonic rental values.¹⁵

The frequency distribution of nonrental units by tenure mode and the quartiles of the distribution of hedonic rental values (ordered from smaller to larger values) are presented in the upper panel of Tables 5 and 6. In 1980–1981 more than 50% of nonrental units are part of the owner-occupied market sector, and slightly more than 25% are public housing. The remaining 20% is equally divided between owner-occupied housing of unknown legal status and other units occupied as wages in kind or as a result of a public or a private transfer. In 1990–1991 the distribution by tenure mode is very similar, with a slight increase in the percentage of public housing and owner-occupied of unknown legal status and a corresponding decrease in the market sector and the last tenure mode. In both years, public housing is overrepresented in the third and the fourth quartiles, whereas the other three tenure modes are overrepresented in the second and, above all, in the first quartile.

The comparison between hedonic and self-imputed values takes two forms. In the first place, the ranking of nonrental units according to the two sets of values has been compared in Arévalo and Ruiz-Castillo (2004). The main findings can be summarized as follows. In both years, the percentage of units classified in the same quartile of the two distributions is greater in quartiles I and IV. By tenure mode, the minimum agreement between the two rankings takes place in the public sector, which is not surprising because the large variety of public housing policies developed in Spain since 1950 makes harder any kind of imputation in this sector. In general, the agreement between the two classification criteria is somewhat better in 1980–1981 than in 1990–1991. Thus, for example, in the first survey

14. In terms of the variables appearing in the regression models in the upper panel of Table 4, making the year of occupancy equal to 1980 means making the variables Years of Occupation During 1965–1973 and 1974–1980, Dummy for the 1965–1973 Occupation Period, and *Qindex* During 1965–1973 equal to zero; whereas making the year of occupancy equal to 1990 means making the variables Years of Occupation During 1965–1975 and 1976–1990, Dummy for the 1965–1975 and 1976–1982 Occupation Periods, and *Qindex* During 1965–1975 and 1976–1982 equal to zero.

15. There are 20 nonrental dwellings in 1980–1981 and 233 in 1990–1991 without information on building age, as well as 62 in 1990–1991 without information on square meters. These units have been assigned the mean value of those variables in the tenure mode to which they belong. In this way, a housing quality index and a hedonic value for them have been computed.

TABLE 5. A comparison of hedonic and self-imputed rental values for the nonrental housing stock, 1980–1981.

| Tenure mode | All | Frequency distribution of the nonrental housing stock by tenure mode and quartiles of the distribution of hedonic values | | | |
|----------------------|-------|--|-------|-------|-------|
| | | I | II | III | IV |
| Owner-Occupied | | | | | |
| Market sector | 53.2 | 66.3 | 55.9 | 44.8 | 46.5 |
| Public housing | 26.5 | 3.6 | 19.7 | 37.9 | 42.7 |
| Unknown legal status | 10.0 | 15.7 | 10.9 | 8.1 | 6.0 |
| Other | 10.3 | 14.4 | 13.6 | 9.2 | 4.8 |
| All | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

| Tenure mode | Mean Monthly Rental Values (in euros) | | Difference between hedonic and self-imputed mean values (%) | | | | |
|----------------------|---------------------------------------|----------------|---|---------------------------------|------|------|------|
| | Hedonic | Self-imputed | All | Quartiles of the Hedonic Values | | | |
| | | | | I | II | III | IV |
| Owner-occupied | | | | | | | |
| Market sector | 54.2 (31.9) | 51.9 (45.2) | 4.5 | 2.5 | -1.0 | 3.5 | 8.2 |
| Public housing | 73.3 (26.4) | 57.8 (35.4) | 26.9 | -2.1 | 11.0 | 25.9 | 31.6 |
| Unknown legal status | 46.9 (26.8) | 44.4 (25.5) | 5.8 | -5.5 | 0.0 | 4.9 | 19.2 |
| Other | 46.1 (25.1) | 41.5 (34.3) | 11.1 | 6.3 | 4.7 | 17.3 | 15.5 |
| All | 57.7 (31.0) | 51.6 (41.1) | 11.8 | 6.3 | 2.8 | 11.2 | 16.4 |

Note: Standard deviations are in parentheses.

the Spearman coefficient ranges from a minimum of 0.49 for the public sector to a maximum of 0.72 for the market sector, being equal to 0.66 for the nonrental stock as a whole; instead, in 1990–1991 these values are equal to 0.46, 0.61, and 0.57, respectively.

Beyond this ordinal analysis, in the bottom panel of Tables 5 and 6 the differences in percentage terms between mean hedonic and self-imputed values by tenure mode are presented for the entire nonrental sector and the quartiles of the distribution of hedonic rental values. In 1980–1981, mean hedonic values are 11.8% greater than mean self-imputed values, although standard deviations are generally lower for hedonic values.¹⁶ The proximity of mean values in the first two quartiles is remarkable, but the discrepancies are somewhat larger in quartile IV. Clearly, the main difficulty lies in assessing rental values for public housing dwellings of greater than average quality. In 1990–1991, mean hedonic values are 13.2% lower than self-imputed ones, and the latter present much larger

16. Mean hedonic values using the whole vector of physical housing characteristics instead of the housing quality index are only 4.5% greater on average than self-imputed values. The standard deviation is 29.4 versus 31.0 for our preferred choice of hedonic values in Table 5.

TABLE 6. A comparison of hedonic and self-imputed rental values for the nonrental housing stock, 1990–1991.

| Tenure mode | Frequency distribution of the nonrental housing stock by tenure mode and quartiles of the distribution of hedonic values | | | | |
|----------------------|--|-------|-------|-------|-------|
| | All | I | II | III | IV |
| Owner-occupied | | | | | |
| Market sector | 50.9 | 67.4 | 59.0 | 41.7 | 38.6 |
| Public housing | 27.9 | 5.8 | 16.8 | 38.0 | 46.3 |
| Unknown legal status | 12.8 | 16.1 | 13.8 | 12.2 | 9.8 |
| Other | 8.5 | 10.7 | 10.4 | 8.1 | 5.3 |
| All | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

| Tenure mode | Mean Monthly Rental Values (in euros) | | Difference between hedonic and self-imputed mean values (%) | | | | |
|----------------------|---------------------------------------|------------------|---|---------------------------------|-------|-------|-------|
| | Hedonic | Self-imputed | All | Quartiles of the Hedonic Values | | | |
| | | | | I | II | III | IV |
| Owner-occupied | | | | | | | |
| Market sector | 147.5 (70.1) | 184.8 (166.0) | -20.1 | -10.0 | -19.9 | -23.0 | -22.7 |
| Public housing | 191.8 (63.5) | 195.9 (122.5) | -2.1 | -18.3 | -8.2 | -5.0 | 1.4 |
| Unknown legal status | 146.6 (63.9) | 173.9 (127.2) | -15.7 | -11.0 | -19.2 | -17.6 | -13.9 |
| Other | 140.7 (59.3) | 149.8 (128.8) | -6.1 | -5.6 | -7.7 | -9.5 | -1.2 |
| All | 159.2 (69.7) | 183.5 (147.7) | -13.3 | -10.3 | -16.8 | -15.1 | -11.3 |

Note: Standard deviations in parentheses.

standard deviations. This time the largest differences in mean values are in quartiles II and III. This seems to be due to the discrepancies found with respect to owner-occupied units of intermediate quality and private or of unknown legal status. As a matter of fact, differences in mean values in these sectors reach 20.1% and 15.7%, respectively.¹⁷

Unfortunately, there exists no reliable statistical sources about housing prices at a national level prior to 1985.¹⁸ However, there is no doubt about the existence of a continuous increase in housing prices starting around 1985, a year before Spain became a full member of the EU, which lasts until 1991, shortly before the

17. Francois (1989) presented preliminary research within the Bureau of Labor Statistics (BLS) in the US suggesting that owner estimates of implicit rents may be biased on average by as much as 10% or 30% above actual values. However, he criticized that research for underestimating spot rents, and pointed out that these results have not been confirmed by comparisons using data collector estimates. More importantly, using a sample of new tenants consisting of 3,706 observations in 27 different major metropolitan areas, this author showed that on average owner estimates are only 1.4% lower than hedonic estimates. Kurz and Hoffman (2004) estimated hedonic functions in West Germany for rents and self-imputed values in the rental and the owner occupied sectors, respectively, and found that in most cases the sign and the size of the estimated coefficients do not differ significantly.

18. Bover (1993) contains a discussion of available statistical sources.

beginning of a downswing in the general business cycle for the Spanish economy. Evidence about housing prices for new transactions in Madrid presented in Bover (1993) speaks of a previous boom around 1979. Based on such tentative evidence, it would appear that the sign of the difference between hedonic and self-imputed values might be influenced by the phase of the housing cycle in which the self-imputations take place: a downswing in the case of the EPF collected from April 1980 to March 1981, and at the very end of an upswing for the EPF collected from April 1990 to March 1991.

Be that as it may, the important question is whether such differences are large or small for the purpose at hand. As explained in the Introduction, there are two ways to apply the rental equivalence approach to the determination of the weight that nonrental housing should receive in the CPI: Using the evaluation by experts or occupants, or using the imputation obtained through hedonic methods. Therefore, in the context of this paper, the answer to the above question hinges on the implications for inflation of following either of the two versions of that approach.

Among other classifications, the CPI system based in 1983 uses a 57-dimensional commodity breakdown, where commodity 35 includes two main items: A “nonrental housing services” component, whose official weight is estimated using the self-imputed rental values declared in the 1980–1981 EPF by occupants of these dwellings; and a second component mainly consisting of “housing repair and maintenance.” Secondary housing must be excluded here because, as explained in Section 2, no quality index has been computed for it. Based on self-imputed values, our own estimates of the weights given to nonrental housing services and commodity 35 as a whole are 9.4% and 13.5%, respectively.¹⁹ The estimates of these two magnitudes using hedonic values are 10.4% and 14.5%, respectively. Based on these two alternatives, interannual inflation rates and the corresponding mean annual inflation for the entire period August 1985 to December 1992, are presented in columns 1 and 2 in the upper panel of Table 7. The differences between these two series are small (of the order of 0.03 percentage points per year). Therefore, whether we use self-imputations or hedonic values in assessing the importance of nonrental housing services in the CPI has a small impact on the measurement of inflation from 1985 to 1992.

As an alternative, consider the possibility of abandoning the two versions of the rental equivalence approach. This means dropping the nonrental housing services component from the CPI altogether—as the Spanish INE did officially in the 1992 base. In this case, our estimate of commodity 35’s weight becomes 4.6%. The mean inflation rate of the nonrental housing services component for the 1985–1992 period is almost 3 percentage points below the mean inflation rate for the economy as a whole (column 5 in Table 7). Therefore, overall inflation

19. The official values are 10.33% and 13.61%, respectively.

TABLE 7. Inter-annual inflation rates and mean annual inflation rate for 1985–1992 and 1993–2000 (percentage points per year).

| Base = 1983 Subperiods | Treatment of non-rental housing services in the CPI | | | (4) = (1) – (3) | Nonrental housing prices (5) |
|--|--|---------------|-----------------|-----------------|---------------------------------------|
| | Hedonic | Self-imputed | Excluded | | |
| | imputations (1) | values (2) | from CPI (3) | | |
| August 1985–December 1985 | 2.82 | 2.84 | 3.05 | –0.24 | 0.62 |
| 1986 | 8.26 | 8.30 | 8.66 | –0.40 | 4.47 |
| 1987 | 4.63 | 4.66 | 4.89 | –0.26 | 2.09 |
| 1988 | 5.91 | 5.89 | 5.77 | 0.13 | 7.26 |
| 1989 | 6.92 | 6.91 | 6.84 | 0.08 | 7.76 |
| 1990 | 6.61 | 6.62 | 6.78 | –0.17 | 4.95 |
| 1991 | 5.61 | 5.60 | 5.55 | 0.06 | 6.23 |
| 1992 | 5.41 | 5.44 | 5.69 | –0.28 | 2.63 |
| Mean annual rate, August 1985–December 1992 | 9.26 | 9.29 | 9.58 | –0.33 | 6.45 |

| Base = 1992 Subperiods | Treatment of non-rental housing services in the CPI | | | (4) = (1) – (3) | Nonrental housing prices (5) |
|---|--|---------------|-----------------|-----------------|---------------------------------------|
| | Hedonic | Self-imputed | Excluded | | |
| | imputations (1) | values (2) | from CPI (3) | | |
| January 1993–December 1993 | 4.41 | 4.47 | 3.92 | 0.49 | 8.23 |
| 1994 | 4.40 | 4.41 | 4.33 | 0.07 | 4.92 |
| 1995 | 4.52 | 4.56 | 4.26 | 0.26 | 6.48 |
| 1996 | 3.78 | 3.85 | 3.23 | 0.55 | 7.85 |
| 1997 | 2.56 | 2.62 | 2.05 | 0.50 | 6.09 |
| 1998 | 1.76 | 1.80 | 1.39 | 0.36 | 4.20 |
| 1999 | 2.97 | 2.98 | 2.88 | 0.09 | 3.56 |
| 2000 | 3.99 | 3.99 | 4.00 | –0.01 | 3.89 |
| Mean annual rate, January 1993–December 2000 | 4.19 | 4.24 | 3.81 | 0.38 | 7.19 |

rates are now significantly lower than before (column 3 in Table 7). In particular, abandoning the rental equivalence approach represented by the hedonic alternative would have led to a downward bias in inflation of 0.33 percentage points per year during the 1985–1992 period (column 4 in Table 7). For comparison purposes, when all sources of bias considered by the Boskin Commission in the US (Boskin et al. 1996) are taken into account,²⁰ it has been estimated that the official CPI in Spain suffers from an upward bias of 0.6 percentage points per year (Ruiz-Castillo, Izquierdo, and Ley 1999).

Apart from other minor changes relative to the 1983 base system, in the CPI system based in 1992 commodity 35 includes only housing repair and maintenance, local housing taxes, and other items pertaining to nonrental housing,

20. The more important biases are due to the lack of substitution of a fixed weights CPI of the Laspeyres type, and to insufficient price corrections in the presence of quality change.

but excludes all imputations for nonrental housing services. In order to estimate the inflation rates that would be obtained if the rental equivalence approach were applied in the 1993–2000 period, there is no difficulty in creating an additional commodity 58 whose weight can be estimated using the hedonic or the self-imputed rental values for nonrental housing. The sum of the weights for commodities 35 and 58 in the hedonic and the self-imputed cases is 15.6% and 17.0%, respectively, whereas the weight of commodity 35 in our version of the official system is 4.8%.²¹ In the absence of an official price index for commodity 58, the index for commodity 33, “rents paid in rental housing,” has been used. The estimates of the three series of inflation rates, as well as the information about the evolution of commodity 33’s prices, appear in the lower part of Table 7. Again, the differences between the two versions of the rental equivalence approach in columns 1 and 2 are small (about 0.05 percentage points per year). However, abandoning the rental equivalence approach as was done in the official system (column 3) leads to an upward bias of 0.38 percentage points per year (column 4)—a considerable amount.

The conclusion is inescapable. The two versions of the rental equivalence approach lead to comparable inflation rates in both periods. Relative to this option, dropping nonrental housing services from the CPI considerably reduces the weight this sector receives in the price index and leads to a sizable bias in the measurement of inflation, whose sign depends on the evolution of the price of housing services relative to the CPI as a whole. During the 1985–2000 period, in Spain this bias has been approximately 0.35 percentage points per year in absolute terms.

5. Conclusions

The treatment of nonrental housing in the CPI is an old topic in index number theory and practice. The issue requires a solution to two problems: (i) How to determine the weight that nonrental housing should have in the CPI; and (ii) how nonrental housing prices should be monitored over time. This paper has presented an empirical contribution to the first question using the 1980–1981 and 1990–1991 household budget surveys for Spain that are used to determine the official CPI weights.

The starting point is the economic approach to price index numbers for consumption goods and services. As far as housing is concerned, the standard of living depends on the consumption of the housing services that a dwelling provides regardless of the tenure mode in which the dwelling is held. The empirical

21. The official weight itself is 5.30%. For a discussion of the reasons why our estimate differs from the official one, see Ruiz-Castillo, Izquierdo, and Ley (1999, chapters 2 and 3).

problem addressed in this paper is how to impute a rental value to the flow of housing services provided by the nonrental housing stock in the base period. The CPI weight for such services will then be equal to the ratio of the imputed value to the total household expenditures in all commodities (including the nonrental housing services in question).

In a novel version of the hedonic approach, it has been shown that the use of a housing quality index leads to a satisfactory and parsimonious model explaining market rents. To correct for the possible selection bias caused by systematic differences in housing characteristics between the market rental sector and the nonrental housing stock, a Heckman procedure has been used for the first time in this context. In this way, the estimated coefficients of the variables explaining market rents can be used to impute a rental value to nonrental housing units.

The more remarkable result of the paper is that, in spite of a thin market rental sector whose working is likely to be affected in complex ways by a number of public policies, occupants of nonrental housing in Spain and hedonic procedures arrive, on average, at comparable imputed values. The differences between the two procedures do not lead to important departures in average annual inflation rates for the 1985–2000 period.

What is the explanation? On the statistical side, it should be remembered that hedonic methods to explain rents and prices in rental and stock housing markets have been working reasonably well for 30 years in a variety of institutional scenarios. On the other hand, it seems that people are keenly aware of how housing markets work, so that self-imputed rental values, although more dispersed than hedonic values, do contain useful information. Finally, it should be emphasized that the aim of the exercise is simply to assess the weight that a given commodity should be assigned in a large commodity space. Tolerable differences in the numerator of the expression for determining that weight need not lead to significant differences in the magnitude of the weight itself. This means that even in a priori unfavorable circumstances, as long as good data are available it should be quite possible to find an acceptable solution to the problem at hand.

There is no room in this paper for a proper discussion of how the prices of nonrental housing services should be monitored over time. However, the treatment of quality change, as well as depreciation and aging bias in both rental and nonrental housing sectors is bound to rely on hedonic methods (see, *inter alia*, Randolph 1988; Silver 1999; Hoffman and Kurz 2002; and Kurz and Hoffmann 2004).²² The results of this paper that add to the successful literature show that these methods are worth pursuing.

This paper contains a warning for the specialists in the housing sector, the officials in charge of these matters, and public opinion in general: In a country with

22. This is also the case in the more promising version of the user cost approach as exemplified in Linneman and Voith (1991) and Crone, Nakamura, and Voith (2000).

a large nonrental housing stock, eliminating its services from the CPI has been estimated to give rise to a large bias in the measurement of inflation. Given the alternatives, this is an unnecessarily crude form of dealing with a difficult problem. It appears that there is room for national and international statistical offices to keep experimenting within the sound economic approach assumed in this paper.

Appendix

This Appendix has two sections. Section A.1 is devoted to a summary of the essentials of MCA as a descriptive technique for representing contingency tables, that is, tables consisting of the frequencies with which a set of values of two or more qualitative variables appear in a data set (see Tenenhouse and Young (1985), and Greenacre (1984) for a detailed treatment of MCA). Section A.2 contains an application of MCA to the data on housing attributes coming from the 1980–1981 and 1990–1991 EPFs.

A.1. Description of MCA

Assume that there are N observations on housing units, each one characterized by Q physical attributes, indexed by $q = 1, \dots, Q$, and possibly correlated. Let J_q be the number of categories or modalities of variable q , and let $J = \sum_q J_q$ be the total number of categories with $N \gg J$. The data can be represented by a $(N \times J)$ matrix Z , whose element z_{ij} takes the value 1 when housing unit i has modality j and zero otherwise, $i = 1, \dots, N$ and $j = 1, \dots, J$. Note that because each housing unit can only have one of the J_q categories of variable q , the row sums in Z must all be equal to Q , i.e., $\sum_j z_{ij} = Q$ for each $i = 1, \dots, N$. Denote by N_j the absolute frequency of category j , or the sum of elements in column j , $\sum_i z_{ij} = N_j$, so that for each variable q , $\sum_{j \in I_q} \sum_i z_{ij} = N$. Finally, $\sum_i \sum_j z_{ij} = NQ$.

The objective of MCA is to obtain a set of uncorrelated variables w_k , indexed by $k = 1, \dots, K$, with $K < Q$, where each w_k is a linear combination of the J categories. In other words, the objective of MCA is to find a $(J \times K)$ matrix M where m_{jk} is the contribution of the j th category to the new variable w_k , so that the information in the original data can be expressed through the lower-dimensionality $(N \times K)$ matrix W defined by

$$W = ZM,$$

whose columns are the w_k variables. As will be seen in A.2, the first variable w_1 , which will be interpreted as a housing quality index, explains a large part of the variance in the data.

The matrix M is constructed as follows. Denote by F the relative frequency matrix, that is, $F = (1/NQ)Z$. The average column profile is the $(N \times 1)$ vector

$r = F1_J$ with $r_i = 1/N$ for all $i = 1, \dots, N$, and the average row profile is the $(J \times 1)$ vector $c = F^T 1_N$ with $c_j = (N_j/NQ)$ for each $j = 1, \dots, J$. The corresponding $(N \times N)$ and $(J \times J)$ diagonal matrices are denoted D_r and D_c , respectively. Define the $(N \times J)$ matrix E by

$$E = D_r^{-1/2}(F - rc^T)D_c^{-1/2}.$$

The element $e_j^T e_j$ in the diagonal of the $(J \times J)$ matrix $E^T E$ is the χ^2 distance between the j th column profile in matrix F and the average column profile r , weighted by its relative frequency c_j .²³ The sum is called the total inertia, TI , and it can be shown to be equal to $(J/Q) - 1$, namely $TI = \sum_j e_j^T e_j = (J/Q) - 1$.

MCA computes the singular value decomposition of E , say $UD_\alpha V^T$, where U and V are orthogonal matrices, and D_α has $J - Q$ nonzero eigenvalues. In practice, because $N \gg J$, it is more convenient to compute the singular value decomposition of the diagonal matrix $E^T E$, say $E^T E = \Gamma D_\lambda \Gamma^T$, where $D_\lambda = D_\alpha^2$. The eigenvalues of D_λ quantify the inertia projected through each of the associated eigenvectors (which form the columns of Γ). The eigenvectors represent orthogonal directions of projection of centered column profiles. The direction of the first eigenvector associated to the largest eigenvalue is the optimal projection in the sense that it is the linear orientation that collects the maximum disparity between individuals according to the Q variables. The second eigenvector is orthogonal to the first and represents the linear orientation that captures the maximum residual disparity not taken into account by the first projection axis. The remaining eigenvectors can be similarly interpreted until the total inertia is accounted for by K orthogonal axis with $K \leq J - Q$. In geometric terms, a change of axis is being performed where the original space of profiles that has dimension $J - Q$ is projected in a reduced space with dimension K . The coordinates used in the projection are contained in the $(J \times K)$ matrix M defined by

$$M = D_c^{-1/2} \Gamma D_\lambda^{1/2}.$$

For each k , the eigenvalue λ_k represents a percentage of the total inertia, but these percentages tend to be small and show a pessimistic idea about the proportion of the projected inertia by each axis. Thus, Benzécri (1979) proposed considering solely the relevant P axis, that is, the axis associated with those eigenvalues with $\lambda_p > 1/Q$, $p = 1, \dots, P$, and $P \leq K$. Analogously, he proposed to correct the eigenvectors with the transformation

$$\lambda_p^c = [Q/(Q - 1)]^2 [\lambda_p - (1/Q)]^2$$

23. There is an analogous interpretation for the elements $e_j e_j^T$ in the diagonal of the $(J \times J)$ matrix EE^T that refers to the rows.

and show the proportion of inertia explained in relation to $\sum_p \lambda_p^c$. In this way, the dimensionality of the original matrix is reduced from $J - Q$ categories to P indicator variables losing a small quantity of information. Once the importance of each indicator is evaluated in this manner, its interpretation can proceed in terms of its correlation with all initial variables and the weight that each category receives in the corresponding linear combination.

A.2. An Index of Housing Quality

In 1980–1981, there are $Q = 8$ qualitative variables and $J = 23$ categories, whereas in 1990–1991 $Q = 18$ and $J = 52$. The number of observations with complete information along these dimensions is 23,898 in 1980–1981 and 20,799 in 1990–1991,²⁴ representing a population of 9,992,051 and 11,105,215 housing units, respectively. The frequency distribution of all physical attributes in both years is presented in columns 1 and 3 in Table A.1.

The percentage of inertia accounted for by the first two factors is 86.2% and 12.1% in 1980–1981, and 73.2% and 10.5% in 1990–1991. Moreover, variables with two or more categories show a parabolic structure when depicted in the plane defined by those two factors. This is a very frequent phenomenon, known as the *Guttman effect* (Greenacre 1991), which simplifies the interpretation. This effect reveals that whereas the first factor summarizes the order structure of all modalities, the second factor shows an opposition between extreme categories (low frequency) and average categories (high frequency) of a variable. Thus, in the following the analysis will continue in terms of the first factor that contains the relevant information about the variability among housing units as far as their physical attributes is concerned. Relative to the initial number of variables and categories, this constitutes a very drastic simplification indeed.

The correlation between the first factor and each of the Q physical attributes for both years is shown in brackets in Table A.1. Among the eight common variables, hygienic services and water facilities are the most influential. Telephone, building age, and heating facilities occupy an intermediate position above garage and housing size, whereas having electricity is the most prevalent characteristic in both years and that with the smallest correlation with the index whose meaning is being discussed. Apart from hygienic services and water facilities, all of the remaining six common variables gain some importance in the determination of the index in 1990–1991; for instance, the correlation coefficient of garage increases from 0.41 in 1980–1981 to 0.58 in 1990–1991. Nevertheless, having an

24. In 1980–1981 there are 73 observations without information on building age and in 1990–1991 this number is 289, plus 67 dwellings without information on housing size in square meters.

TABLE A.1. Frequency distribution, normalized category weights, and correlations between the first factor obtained by MCA and physical attributes in 1980–1981/1990–1991 (correlation in parentheses).

| Physical attributes | 1980–1981 | | 1990–1991 | |
|---------------------------------------|--------------------|--------------------|-------------------|--------------------|
| | Frequency | Normalized weights | Frequency | Normalized weights |
| | 9,992,051 = 100.0% | 0.00 | 11,105,215 = 100% | 0.00 |
| | Sample Size | | | |
| 1. Hygienic services | | (0.76) | | (0.72) |
| None or outside the dwelling | 7.6 | –8.09 | 1.8 | –8.32 |
| One full bathroom or less | 77.3 | –0.07 | 73.3 | –0.98 |
| More than a full bathroom | 15.1 | 4.34 | 24.9 | 3.41 |
| 2. Water facilities | | (0.75) | | (0.70) |
| No water | 3.8 | –8.41 | 0.4 | –10.00 |
| Only cold water | 19.4 | –4.21 | 5.1 | –6.94 |
| Cold and hot water | 76.7 | 1.49 | 94.6 | 0.32 |
| 3. Heating facilities | | (0.41) | | (0.65) |
| No heating | 61.6 | –1.73 | 10.9 | –3.62 |
| Mobile tools, ind. or central heating | 38.4 | 2.79 | 89.1 | 0.22 |
| 4. Garage facilities | | (0.41) | | (0.58) |
| No garage | 83.6 | –0.68 | 72.3 | –1.02 |
| Garage | 16.4 | 3.47 | 27.7 | 2.41 |
| 5. Telephone facilities | | (0.62) | | (0.69) |
| No telephone | 49.3 | –2.17 | 22.7 | –3.27 |
| Telephone | 50.7 | 2.29 | 77.3 | 1.14 |
| 6. Electricity | | (0.26) | | (0.33) |
| No electricity | 0.9 | –10.0 | 0.2 | –8.72 |
| Electric facilities | 99.1 | 0.09 | 99.8 | 0.01 |
| 7. Housing size | | (0.40) | | (0.49) |
| Less than 60 m ² | 12.7 | –2.59 | 9.0 | –3.27 |
| 61–90 m ² | 41.6 | 0.36 | 37.8 | –0.33 |
| 91–130 m ² | 33.4 | 1.43 | 33.4 | 0.32 |
| More than 130 m ² | 12.3 | 1.65 | 15.8 | 0.81 |
| 8. Building age | | (0.59) | | (0.68) |
| More than 50 years | 25.5 | –3.25 | 11.3 | –3.66 |
| 31–50 years | 27.7 | –1.49 | 13.6 | –2.82 |
| 21–30 years | 11.5 | –0.22 | 25.0 | –2.01 |
| 11–20 years | 11.0 | 1.08 | 33.4 | –0.81 |
| Less than 11 years | 24.3 | 2.68 | 16.7 | 2.14 |
| 9. Elevator facilities | | | | (0.80) |
| None | | | 70.3 | –1.82 |
| Elevator | | | 29.7 | 4.69 |
| 10. Air conditioning | | | | (0.80) |
| None | | | 97.6 | –0.08 |
| Air conditioning | | | 2.4 | 3.29 |
| 11. Water heating fuel | | | | (0.71) |
| None | | | 5.4 | –7.16 |
| Wood, coal | | | 2.1 | –0.81 |
| Gas | | | 59.8 | –0.89 |
| Electric energy | | | 16.2 | 0.41 |
| Natural gas | | | 16.6 | 3.65 |
| 12. Heating system fuel | | | | (0.57) |
| None | | | 10.9 | –2.25 |
| Wood, coal | | | 19.4 | –1.35 |
| Electric energy | | | 40.6 | 0.15 |
| Natural gas | | | 16.0 | 2.65 |

(Continued)

TABLE A.1. CONTINUED.

| | 1980–1981 | | 1990–1991 | |
|------------------------------|--------------------|--------------------|-------------------|--------------------|
| | Frequency | Normalized weights | Frequency | Normalized weights |
| Physical attributes | 9,992,051 = 100.0% | 0.00 | 11,105,215 = 100% | 0.00 |
| | Sample Size | | | |
| 13. Cooking fuel | | | (0.72) | |
| Wood, carbon | | 3.4 | | –4.97 |
| Gas | | 76.8 | | –0.69 |
| Electric energy | | 6.8 | | 4.46 |
| Natural gas | | 13.0 | | 5.19 |
| 14. Garden facilities | | | (0.62) | |
| None | | 84.9 | | –0.01 |
| Garden | | 15.1 | | 0.04 |
| 15. Building type | | | (0.73) | |
| Inferior tenement | | 0.3 | | –4.91 |
| One floor | | 35.6 | | –2.93 |
| Two floors | | 4.6 | | –1.93 |
| Three or more floors | | 59.5 | | 2.21 |
| 16. Swimming pool | | | (0.40) | |
| None | | 98.8 | | –0.04 |
| Swimming pool | | 0.2 | | 4.69 |
| 17. Sports area | | | (0.78) | |
| None | | 98.9 | | –0.05 |
| Sports area | | 1.1 | | 5.49 |
| 18. Other community services | | | (0.83) | |
| None | | 52.7 | | –2.44 |
| Some community services | | 47.3 | | 3.16 |

elevator or air conditioning inside the house, or having sport facilities and other community services around it, are the characteristics with the largest correlation coefficients with the index in 1990–1991.

In each year, each modality j receives a certain weight m_j , which can be positive or negative, but the more frequent a modality j is, the closer to zero m_j will be. Columns 2 and 4 in Table A.1 present the rankings of modalities in both years in terms of normalized weights m_j' that preserve the sign and the ratios of the original weights. The normalization consists of assigning the values -10 and 10 to the modalities with extreme m_j according to the formula

$$m_j' = m_j(10/m^*),$$

where $m^* = \max\{|m_1|, \dots, |m_J|\}$. In both years, the maximum normalized weight is below 10, indicating that there is no modality that plays the exact opposite role to the one that reaches the maximum negative value -10 .

As far as the interpretation of the index is concerned, it is very revealing that for all variables with two categories, the worst receives a negative weight whereas the best receives a positive weight. For the remaining variables with three or more categories, weights are always naturally ordered from worst to best modality. In particular, the categories with the maximum negative influence in

both years are having no hygienic services, no water (or only cold water), or no light. Moreover, the first two are highly correlated with the index (see Table A.1), which indicates that both of them are very influential in the interpretation. Among the remaining common variables in both years, having a telephone, being in a building less than 11 years old, and having a garage or more than a full bathroom are categories associated with high and increasing positive weights. In 1990–1991, having a swimming pool, elevator, electric energy or natural gas for cooking, and a sports area are the modalities exercising the maximum positive weight in the index.²⁵

In our opinion, both the correlation between the index and the different variables and the way the weights are naturally ordered for all variables without exception, indicate that the first factor obtained from MCA can be safely interpreted as an index of housing quality. The index values assigned to each housing unit indicates the positive or negative deviation relative to the quality attributed to the “average housing unit” with the more frequent categories, whose index value is zero by construction.²⁶

The housing quality index has, of course, several shortcomings. First, the index does not take into account the heterogeneity in the way that certain categories influence quality for different housing types. For instance, heating in the Canary Islands or Andalucía should receive a smaller weight on quality than in the north of the country. Although such differences could be recognized by computing a different index for each housing type, the resulting quality indexes will not be comparable across types. Consequently, a single housing quality index including the categories “no heating” and “heating of any kind” has been computed. Second, the housing quality concept is restricted to the Q variables for which there is information in the EPFs. Judging from the existing hedonic literature on housing, there are several quality dimensions that are potentially important, such as how well preserved dwellings are, as well as basic neighborhood characteristics like pollution, safety, public transport facilities, and distance to the main centers of economic and recreational activity.

In any case, the real issue is whether, with all its shortcomings, the housing quality index here constructed behaves well in explaining market rents, an issue covered in Section 3 of the paper.

25. For the more salient features in a comparison between 1980–1981 and 1990–1991, see Arévalo and Ruiz-Castillo (2004).

26. The common categories associated with the average housing unit in both years are the following: Building age between 20 and 30 years, presence of electricity facilities, one bathroom (although perhaps not a full one), hot water, between 61 and 90 m², and without a garage. The average dwelling in 1980–1981 does not possess telephone or heating, whereas in 1990–1991 it has both and it is situated in buildings with more than three floors, without an elevator, air conditioning, garden, swimming pool or sports area; gas is the fuel used for cooking and heating water, whereas for the remaining heating uses, served by mobile tools, this fuel is combined with electric energy.

References

- Arévalo, Raquel (2001). "El mercado de la vivienda en España." Ph.D. dissertation, Universidad Complutense de Madrid.
- Arévalo, Raquel, and Javier Ruiz-Castillo (2004). "The Rental Equivalence Approach to Non-rental Housing in the Consumer Price Index. Evidence from Spain." Working Paper 04-17, Economic Series 04, Universidad Carlos III, Madrid, available at <http://docubib.uc3m.es/WORKINGPAPERS/WE/we041704.pdf>.
- Arnott, Richard (1995). "Time for Revisionism in Rent Control?" *Journal of Economic Perspectives*, 9, 99-120.
- Balk, Bert (1994). "The New Consumer Price Indices of Statistics Netherlands." *Statistical Journal of the United Nations*, 11, 119-123.
- Basu, Kaushik, and Patrick Emerson (2000). "The Economics of Rent Control." *Economic Journal*, 110, 939-962.
- Benzécri, Jean-Paul (1979). "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum." *Cahiers de l'Analyse des Données*, 4, 377-378.
- Börsch-Supan, Axel (1986). "On the West German Tenants Protection Legislation." *Journal of Institutional and Theoretical Economics*, 142, 380-404.
- Boskin, Michael, Ellen Dulberger, Robert Gordon, Zvi Griliches, and Dale Jorgenson (1996). "Toward a More Accurate Measure of the Cost of Living." Final Report to the U.S. Senate Finance Committee.
- Bover, Olympia (1993). "Un modelo empírico de la evolución de los precios de la vivienda en España (1976-1991)." *Investigaciones Económicas*, XVII, 65-86.
- Castro, Miguel Angel (1992). "El nuevo sistema de Indices de Precios de Consumo." *Situación*, BBV, 3-4, 167-179.
- Clark, William, and Alan Heskin (1982). "The Impact of Rent Control on Tenure Discounts and Residential Mobility." *Land Economics*, 58, 109-117.
- Crone, Theodore, Leonard Nakamura, and Richard Voith (2000). "Measuring Housing Services Inflation." *Journal of Economic and Social Measurement*, 26, 153-171.
- Dalén, Jörgen (1999). "A Proposal for a New System of Aggregation in the Swedish Consumer Price Index." Proceedings of the Ottawa Group, Fifth Meeting, Reykjavík, Iceland, August 25-27.
- Diewert, Erwin (2000). "The Consumer Price Index and Index Number Purpose." Discussion Paper No. 00-02, Department of Economics, University of British Columbia.
- Diewert, Erwin (2003). "The Treatment of Owner Occupied Housing and Other Durables in a Consumer Price Index." Discussion Paper No. 03-08, Department of Economics, University of British Columbia.
- Eurostat, World Bank, International Monetary Fund, OECD, and United Nations (1993). *System of National Accounts*. United Nations. New York.
- Francois, Joseph (1989). "Estimating Homeownership Costs: Owners' Estimates of Implicit Rents and the Relative Importance of Rental Equivalence in the Consumer Price Index." *American Real Estate Urban Economics Association Journal*, 17, 87-99.
- Goodman, Allen, and Masahiro Kawai (1985). "Length-of-Residence Discounts and Rental Housing Demand: Theory and Evidence." *Land Economics*, 61, 93-105.
- Greenacre, Michael (1984). *Theory and applications of correspondence analysis*. Academic Press.
- Greenacre, Michael (1991). "Interpreting Multiple Correspondence Analysis." *Applied Stochastic Models and Data Analysis*, 7, 195-210.
- Guasch, J. Luis, and Robert Marshall (1987). "A Theoretical and Empirical Analysis of the Length of Residency Discount in the Rental Housing Market." *Journal of Urban Economics*, 22, 291-311.

- Heckman, James (1979). "Sample Selection Bias As A Specification Error." *Econometrica*, 47, 153–161.
- Hill, Peter (1999). "Inflation, the Cost of Living and the Domain of a Consumer Price Index." Prepared for the Conference of European Statisticians, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, November 24–27.
- Hoffman, Johannes, and Claudia Kurz (2002). "Rent Indices for Housing in West Germany, 1985 to 1989." Discussion Paper 01/02, Economic Research Center of the Deutsche Bundesbank.
- Hubert, Franz (1995). "Contracting with Costly Tenants." *Regional Science and Urban Economics*, 25, 631–654.
- Instituto Nacional de Estadística (1983). *Encuesta de Presupuestos Familiares. Metodología y resultados*. Instituto Nacional de Estadística, Madrid.
- Instituto Nacional de Estadística (1992). *Encuesta de Presupuestos Familiares. Metodología*. Instituto Nacional de Estadística, Madrid.
- Kurz, Claudia, and Johanne Hoffmann (2004). "A Rental-equivalent Index for Owner Occupied Housing in West Germany 1985 to 1998." Discussion Paper 08/2004, Economic Research Center of the Deutsche Bundesbank.
- Linneman, Peter, and Richard Voith (1991). "Housing Price Functions and Ownership Capitalization Rates." *Journal of Urban Economics*, 30, 100–111.
- Lowry, Ira (1981). "Rental Housing in the 1980s: Searching for the Crisis." In *Rental Housing: Is There a Crisis*, edited by J. Weicher. Urban Institute Press.
- Marshall, Alfred (1898). *Principles of Economics*, 4th ed. Macmillan.
- Miron, John (1990). "Security of Tenure, Costly Tenants and Rent Regulation." *Urban Studies*, 27, 167–184.
- Nagy, John (1997). "Do Vacancy Decontrol Provisions Undo Rent Control?" *Journal of Urban Economics*, 42, 64–78.
- Peña, Daniel, and Javier Ruiz-Castillo (1984). "Robust Methods of Building Regression Models. An Application to the Housing Sector." *Journal of Business and Economic Statistics*, 2, 10–20.
- Peña, Daniel, and Víctor Yohai (1995). "The Detection of Influential Subsets in Linear Regression by Using an Influence Matrix." *Journal of the Royal Statistical Society B*, 57, 145–156.
- Quigley, John (1979). "What Have We Learned about Urban Housing Markets?" In *Current Issues in Urban Economics*, edited by P. Mieszkowski and M. Straszheim. Johns Hopkins University Press.
- Randolph, William (1988). "Housing Depreciation and Aging Bias in the Consumer Price Index." *Journal of Business and Economic Statistics*, 6, 359–371.
- Rosen, Sherwin (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82, 34–55.
- Ruiz-Castillo, Javier, Mario Izquierdo, and Eduardo Ley (1999). *La medición de la inflación en España*. La Caixa.
- Schlicht, Ekkehart (1983). "The Tenant's Decreasing Willingness to Pay and the Rent Abatement Phenomenon." *Journal of Institutional and Theoretical Economics*, 139, 155–159.
- Sheppard, Stephen (1999). "Hedonic Analysis of Housing Markets." In *Handbook of Regional and Urban Economics*, edited by E. S. Mills and P. Chesire. Elsevier Science.
- Silver, Mick (1999). "An Evaluation of the Use of Hedonic Regressions for Basic Components of Consumer Price Indices." *Review of Income and Wealth*, 45, 41–56.
- Smith, Lawrence, Kenneth Rosen, and George Fallis (1988). "Recent Developments in Economic Models of Housing Markets." *Journal of Economic Literature*, XXVI, 29–64.

- Tenenhaus, Michael, and Forrest Young (1985). "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods of Quantifying Categorical Multivariate Data." *Psychometrika*, 50, 91–119.
- Triplett, Jack (1990). "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy." In *Fifty Years of Economic Measurement*, edited by E. Berndt and J. Triplett. University of Chicago Press.
- Triplett, Jack (2001). "Should the Cost-of-living Index Provide the Conceptual Framework for a Consumer Price Index?" *Economic Journal*, 111, F311–F334.
- Turvey, Ralf (1989). *Consumer Price Indices: An ILO Manual*. Geneva: International Labour Organization.
- Turvey, Ralf (1999). "True Cost of Living Indices." Proceedings of the Ottawa Group, Fifth Meeting, Reykjavík, Iceland, August 25–27.
- Turvey, Ralf (2000). "Owner-occupiers and the Price Index." *World Economics*, 1, 153–159.
- United States Department of Labor, Bureau of Labor Statistics (1997). "Published Estimates of Consumer Price Elasticities of Demand." In *The Use of the Geometric Mean in the Elementary Indexes of the Consumer Price Index*, unpublished study quoted in Triplett (2001).