Universidade de Vigo
Departamento de Economía Aplicada

Documento de Traballo
0604

Advances in the estimation of households' duration of residence

Jacobo de Uña Álvarez
Raquel Arévalo Tomé
Mª Soledad Otero Giráldez

Documentos de Traballo

Maio 2006

Departamento de Economía Aplicada
Universidade de Vigo
As Lagoas Marcosende S/N, 36310 –Vigo
Tfno: +34 986 812500 - Fax: +34 986 812401
http://webs.uvigo.es/x06/
E-mail: depx06@uvigo.es

ADVANCES IN THE ESTIMATION OF HOUSEHOLDS' DURATION OF RESIDENCE

Jacobo de Uña-Álvarez[1,*], Raquel Arévalo-Tomé[2], M. Soledad Otero-Giráldez[3],

[1]Department of Statistics and Operations Research, University of Vigo
[2]Department of Business and Marketing, University of Vigo
[3]Department of Applied Economics, University of Vigo

Abstract

In this paper we investigate the duration of residence for Spanish households from panel data. The inference is based on those households surveyed in 1994 (and followed since that year on) and moving into their 1994-current residence after 1979. We distinguish among owners, renters, and borrowers, because these groups of households are known to exhibit a different pattern regarding residence time. Our approach for estimation is purely nonparametric, being free of the equilibrium assumption (Anily *et al.* 1999) otherwise. This is an interesting feature, since we show that the renewal processes which represents the households' mobility have a non-constant rate over the period 1980-1994 for owners and borrowers.

Key Words: age of residence, censoring, equilibrium equation, left-truncation

JEL Codes: C41,D19

*Corresponding author. *Address*: Facultad de Ciencias Económicas y Empresariales, Universidad de Vigo, Campus Lagoas-Marcosende 36310, Vigo, Spain. Phone: (+34) 986812492, Fax: (+34) 986812401, E-mail: jacobo@uvigo.es.
First author's research was supported by the Grant MTM2005-01274 (FEDER funding included) of the Spanish Ministerio de Educación y Ciencia.

1 Introduction

The importance of investigating the residential housing from both an economic and social perspective is greatly connected to two specific characteristics of the houses: (a) they provide a basic service for households, which (b) is used for a very long time. The costs associated to this service represents a substantial portion of the total budget of the household, mainly when moving from one residence to another (because of the so-called transaction costs). A household decides to move when the expected benefits exceed the costs of moving, and this determines the total household's residence time.

This duration component of the housing has a big impact in the social and economic scenarios of any country. Certainly, the expected residence time influences important private decisions of the households (tenure status, emancipation or movement of household's members, investment in maintenance of the house, etc.), thus conditioning the household's social status. Also, duration of residence affects the demand of goods and services which are related to housing, see for example Henderson and Ioannides (1989) and Haurin *et al.* (1997). In this manner, many economic agents get involved in satisfying such demands (building companies, intermediate agents for renting or moving, financial sectors, furniture companies, decoration services, etc.), which results in an activation of the labor and economic sectors. In fact, Oswald (1997) and Geen and Hendershott (2001) argue and present evidences of correlations between housing ownership and unemployment rates across European countries and US states, respectively. On the other hand, using survey data from the British Household Panel Survey, Benito and Word (2005) find that households are two to tree times more likely to purchase certain durable good when they move to a new house.

Several authors have investigated the household mobility and the duration at the same residence from different perspectives. For example, Schachter and Kuenzi (2002) describe the stational behavior of the residential mobility and the occupancy time for several types of households in US. Deng *et al.* (2002) investigate explanatory factors of the duration of residence, by using a proportional hazards model, which allows for measuring the impact of the households' characteristics on the instantaneous probability of moving. One of the key problems when investigating the residence time is the lack of reliable methods for measuring the occupancy duration, as pointed out by Kidd (1977), Harsman and Quigley (1991) and Pickles and Davis (1991) (among others). Anily *et al.* (1999) discuss the importance of estimating the total residence time of the households, mentioning up to five different fields for which this information is indeed relevant (understanding social and economic phenomena; predicting household expenditure patterns; calculating mobility indexes; explaining general living systems; reducing the attrition bias in consumer panel data).

A recognized problem when measuring the total residence time is that usual microeconomic surveys do not provide this variable, since the household's future moving date is unknown. Typically, each household provides the date it moved into its current residence, which allows to compute the current residence time (at the inquiry date), or age. Of course, this current residence time is a right-censored value of the household's duration of residence, and direct estimation of parameters such as the average total residence time (or the distribution of household's duration of residence) is not possible. Anily *et al.* (1999) proposed to estimate the households' duration of residence from the current residence time on the basis of the equilibrium distribution, which can be motivated as a limit case in the scope of the renewal process theory (Feller 1966). However, it may happen that the equilibrium distribution is different from that of interest; this will occur in situations with a non-constant incidence rate, and it certainly narrows the application of the method. We report evidences on this issue for Spanish data in Sections 3 and 5.

The main goal of this paper is to provide reliable estimation of the household's duration of residence, without imposing any restriction on the rate at which the households are created. To this end, we will use methods based on truncation models, which have been largely applied since the nineties mainly in the scope of biomedical data analysis, see Sections 3 and 4 for details. However, for the best of our knowledge, the application of these methods for investigating the household mobility is new, and indeed the economic literature seems to be quite unaware about the capabilities of these powerful statistical tools. Our approach will be purely nonparametric; that is, unlike in Anily *et al.* (1999), no parametric rigid form is assumed for the distribution to be estimated. Hence, the inferences are completely free of any *a priori* judgement about the variable under investigation. As a by-product, we will propose empirical estimation of the incidence rate for the processes representing the households' formation, on the basis of Spanish panel data.

Specifically, we will provide empirical evidences for the Spanish occupied housing. We mention that the total residence time for Spanish households has not been quantified so far in the literature on housing. The results will be displayed separately for owners, renters and borrowers. We believe that the results provided in this paper may be of great interest for the sectors which are related to the housing market in Spain. Importantly, the Spanish public policies on housing could be evaluated and oriented in a better way by learning more about the households' duration of residence.

The organization of the paper is as follows. In Section 2 we briefly review the Spanish housing market's main characteristics. In Section 3 we present our data basis, and we discuss methodological aspects related to the censoring and truncation issues which arise when computing durations from panel data. Section 4 introduces the statistical methods for the estimation of the residence time distribution. Main empirical results are reported in Section 5, while Section 6 presents the main conclusions of our research.

2 Spanish housing market

The last Spanish Census of Population and Housing (year 2001) indicates that the Spanish population consists in 40.8 million people (5.1% larger than in the preceding Census, which goes back to 1991), the number of houses is 20.8 millions (21% larger than in 1991), and that the number of households is 14.3 millions. In turn, household consumption and residential investment represent 58% and 7% of Spanish GDP, respectively. Between 1976 and 2002, average Spanish house prices have risen sixteen-fold in nominal terms and have doubled in real terms (Martínez and Maza 2003).

Much has been said about "the Spanish housing problem" in connection to the prices evolution (see, for example, Martínez and Maza 2003; and López-García 2004), the government implication (De Mesa and Valiño 2001; Eastaway and San Martin 2002; Fernández, 2004), the decceleration in new households' formation (Holdsworth *et al.* 2002), or the relationship between the labor and housing markets (Díaz-Serrano 2005). For a detailed description of the housing market in Spain see, for example, Martínez and Matea (2002) and Ball (2005).

Following Martínez and Maza (2003), from an international analysis, Spain would be among the three or four OECD countries evidencing the highest long-term real growth in house prices. This rising trend is very clear in some countries (Netherlands, Luxembourg, Ireland or the United Kingdom), while in others (as Germany, Denmark, Canada or Sweden) it is scarcely perceptible. Spain shows an annual average growth rate in real terms of 2.9% over the past 26 years. This percentage is 0.2% in Sweden, 3.1% in the United Kingdom and 3.8% in Ireland. On the other hand, Spain exhibits the largest ratio of houses per household in Europe (1.5 in 2000, while the European average is 1.1), see Trigo (2003).

The Spanish case is also outstanding because of the large percentage of ownership for housing (84.8% of the households own the house they live at, as reported by the European Community Household Panel, 2001). To this regard, the European countries closest to Spain are Greece (84.6% of ownership), Ireland (81.9%), Italy (76%) and Belgium (74%). This percentage is below 60% for Sweden, Austria, The Netherlands and Germany. In US the ownership rate is 67%, see the US Census Bureau, Statistical Abstract of the United States, 2001. All the countries show an increasing trend for ownership as opposite to rentership. For example, the ownership rate in 1994 was 78.8% in Spain (77.1 % in Greece, 80.6 % in Ireland, 70.4% in Italy, and 66.8% in Belgium), and it was 64% for US (in 1993).

With respect to the residential age of the households, Spain exhibits larger figures for owners. This phenomenon can also be appreciated in the EU countries (European Community Household Panel, 2001) and in US (US Bureau of the Census, 2000). Certainly, one-year-old households represent 7% of owners and 21% of renters in Spain, the figures corresponding to US being 11% and 39%

respectively. On the other hand, the 2001-median age of Spanish households is 19 years for owners and it reduces to 8 years for renters. Other European countries show smaller median ages, such as Finland (12 years for owners, 2 years for renters), Denmark (11 and 4 years, respectively), or Sweden (13 and 4 years); while among the European countries with largest ages we find Italy and Austria, both of them showing median ages of 22 years for owners and 13 years for renters. The 1993 figures for US are 8.2 years for owners and 2.1 years for renters.

3 Data description

3.1 The data

Our data source is the European Community Household Panel (ECHP), a yearly panel of the EU-15 countries that started in 1994 and closed in 2001. This survey is carried out by Eurostat in cooperation with the national agencies of statistics of each of the EU-15 countries. Specifically, we analyze the waves covering the period 1994-2001 for Spain.

The statistical unit is defined as the house a household occupies for living. The ECHP survey reports information on several indicators of the household's status, as wage level, labor status, house characteristics, or health. Our data basis corresponds to the 7,182 Spanish households which report information in 1994. The information provided by this (longitudinal) data basis allows for a eight-years following-up of the households' characteristics as long as of each individual inside the household. The households surveyed in the 1994 wave are maintained in the following waves, even when new members have entered the household, or when some members of the household or the whole household have moved.

As relevant variables for our investigation we have considered the following: tenure status (owners, renters, and borrowers) in 1994, which separates the households in groups with a different pattern regarding the duration of residence; and year the household moved into its current residence (which allows us to compute the households' current residence time or age in 1994). Table 1 describes these two variables. This Table 1 indicates that most of the households (about 80%) own the house they live in, while the renters represent about the 14%. Regarding the age of current residence, Table 1 shows that 54.5% of the households moved for the last time after 1979, and that only about 14% correspond to the newest households (less than 5 years old). On the other hand, it is seen that the 1994-current residence time is greater for the owners (9.5% of less than 5 years old households) than for the borrowers (22%), and that the time corresponding to the borrowers is greater than that of the renters (about 35% of 'new' households).

More than a half of the observed households moved to their residence before 1980 (see Table 1). These households (unlike the remaining ones) do not provide further information about the year they moved. This issue complicates the investigation of the total duration of residence. We decided to focus on the population of households for which the entry date to the house fails into the range 1980-1994. This reduction has two advantages; first, we concentrate on those households for which the exact year of entry is available; second, we reduce the sample heterogeneity by considering only the newest households. Then, the final sample size is 3,268.

3.2 Methodological aspects

In principle, the distribution of households' residence time can be inferred from the age of current residence reported in the 1994 wave of the ECHP. Anily *et al.* (1999) suggested some estimation method relying on the so-called equilibrium equation of the underlying renewal process. This method works for processes that exhibit a constant incidence rate on the study period (1980-1994, in our case). However, when analyzing our data basis we have found that the density function of the owners and borrowers' age of current residence is not monotone decreasing, thus the equilibrium equation can not hold for these two groups. These densities, together with that corresponding to renters, are reported in Figure 1. The smooth estimates were obtained similarly as in Anily *et al.* (1999), by fitting a cubic function to the empirical log-survival function of the 1994-current residence time. The invalidity of the equilibrium assumption motivates the introduction of more flexible methods, as those discussed below.

By using the panel 1994-2001, we are allowed to observe the household mobility in that period. Through these years, some households were observed to leave the house they occupied back in 1994. The following-up of these units allows to get complete information about their total residence time. Unfortunately, these are just 497 of the 3,268 households (15.2% of the observations); the remaining 84.8% provide right-censored durations of residence, in the sense that we just know that the total residence time exceeds some observed time of stay. Right-censored observations correspond to two groups of households: those which are not observed to move in the period 1994-2001 (1,238 households, or 44.7% of the censored observations), and those for which the information is missing during the referred period (the remaining 55.3% of the censored times). The "potential censoring time" for each household, determined by the year in the household is missed (lost to follow-up), is given in Table 2. In this Table 2 we see (as expected) that the censoring proportion reduces as the following-up period increases.

Dealing with right-censored observations is a well-known problem in economic duration analysis. When some observations are censored from the right, the natural substitute for the empirical distribution function of the sample is the Kaplan-Meier estimator, see for example Lancaster (1990), pages 278-280. Assume that we want to describe the distribution of a time variable or lifetime, say Y, given a sample which includes censored times. The Kaplan-Meier method can be used to construct a consistent estimator for the survival function of Y, $S(y) = \mathbb{P}(Y > y)$, under the assumption that the units censored at time y are representative of those units with lifetime greater than y. However, we must face an extra problem when investigating the total time of residence from our data set. This is because the sampled households are length-biased regarding their potential durations of residence, as we will explain in detail at this point.

Each one of the 3,268 sampled households moved to their 1994-current house in some year a, ranging from 1980 to 1994. Put y for the household's total duration of residence. For this duration y to be observed (sampled), it was necessary that this y exceeded the time between a and the sampling point (1994). As a result, the probability of being sampled is associated to the total residence length in the sense that, the longer the length of stay, the greater the probability of being observed. This issue is a consequence of sampling the data by cross-section at 1994, and it is typically referred as a left-truncation problem, where the truncation variable is defined as time from "onset" (the year the household moved to its current house) to the cross-section date (the sampling date in 1994). See Wang (1991) for more discussion on this. Thus, direct application of the Kaplan-Meier method would lead to an overestimation of the survival function and location parameters (such as the mean or the median).

In situations like this, some correction of the Kaplan-Meier method is required in order to compensate the initial overrepresentation of large durations in the sample. Following Wang (1991) (see also Asgharian *et al.* 2002), two general procedures are possible. (1) The "unconditional approach" implies that a specific shape for the truncation distribution is assumed. A typical choice is the uniform truncation model, because it is associated to the so-called stationarity assumption, representing the hypothesis that the incidence of the process under investigation is constant over some time interval. The drawback of this approach is that it relies on a given shape of the truncation distribution, which could be eventually misspecified. (2) As an interesting alternative, the "conditional approach" is purely empirical, in the sense that it proceeds conditionally on the observed truncation times, which are not assumed to belong to any specific distribution. We will follow this latter approach for estimating the households' duration of residence from our panel data.

4 Statistical methods

In this Section we present the statistical methods for the estimation of the households' duration of residence, given the special nature of our panel data. We will use the following notation: T stands for the left-truncation time, in our application defined as time from the date the household moved (into its 1994-current house) to the sampling year (1994); Y denotes the lifetime of ultimate interest (total household's duration of residence), and (due to the left-truncation issue) it is observable if and only if $T \leq Y$; C is used to denote the potential censoring time, assumed to satisfy $\mathbb{P}(C \geq T) = 1$; and, finally, $X = \min(Y, C)$ and $\delta = \mathbb{I}(Y \leq C)$ represent the observed duration of residence and the censoring indicator, respectively. The sampling information is $(T_1, X_1, \delta_1), ..., (T_n, X_n, \delta_n)$, independent and identically distributed as (T, X, δ) given $T \leq X$.

It is assumed that (i) T and Y are independent; and (ii) $C - T$ is independent of $(T, Y - T)$ conditionally on $T \leq X$. Assumption (i) claims that the population under study is homogeneous regarding the Y, not being influenced by the special T value; for the application we have in mind, this means that the Spanish households entering their houses between 1980 and 1994 are expected to exhibit an homogeneous behavior with respect to their lengths of stay. Assumption (ii) claims that the residual censoring time (after cross-section in 1994) $C - T$ is independent of everything else; in our case, the interpretation of this assumption is that the sampled households which are censored a years after 1994 are representative of those households which are observed to stay at their houses for a longer time (in other words, the event $\{C - T = a, \delta = 0\}$ gives no information on the total residence time other than $Y - T > a$).

This approach can be regarded as a natural way of modeling cross-sectional data. It differs from the basic model considered by Wang (1991) in the way in which the censoring effects are incorporated. The methods proposed by Wang (1991) start with the hypothesis of independence between the pair (T, C) and the lifetime Y for the *untruncated* population. However, in our application it does not make much sense to consider the existence of a potential censoring time C for the durations which are left out due to truncation. Besides, the motivation of Wang (1991)'s estimator as a conditional nonparametric maximum likelihood estimator (NPMLE) requires the (often artificial) assumption that all the potential censoring times are known, even for those units whose durations have being completely observed. Our model assumptions (i) and (ii), inspired in the ideas contained in Asgharian *et al.* (2002) (but not relying on the stationarity assumption in that paper), overcome these difficulties.

Put F and L and for the distribution functions (df's) attached to Y and T, respectively. Besides, we denote by R the conditional df of $C - T$ given $T \leq X$. Here we derive the NPMLE's of F, L and R. In order to do that, we first derive the likelihood function of the (T_i, X_i, δ_i)'s. The contribution of the observation

(T_i, X_i, δ_i) to the full likelihood, defined as

$$\mathcal{L}_i = P(T = T_i, X = X_i, \delta = \delta_i \mid T \leq X),$$

can be decomposed as a product of the conditional likelihood of (X_i, δ_i) given the T_i, and the marginal likelihood of the T_i:

$$\mathcal{L}_i = P(X = X_i, \delta = \delta_i \mid T = T_i, T \leq X) \times P(T = T_i \mid T \leq X) \equiv \mathcal{L}_{c,i} \times \mathcal{L}_{m,i}$$

Hence, the full likelihood $\mathcal{L} = \prod_i \mathcal{L}_i$ is also decomposed as a product, $\mathcal{L} = \mathcal{L}_c \times \mathcal{L}_m$, where $\mathcal{L}_c = \prod_i \mathcal{L}_{c,i}$ is the conditional likelihood of $(X_1, \delta_1), ..., (X_n, \delta_n)$ given $T_1, ..., T_n$, and $\mathcal{L}_m = \prod_i \mathcal{L}_{m,i}$ is the marginal likelihood of $T_1, ..., T_n$. Under the model (i) and (ii), the explicit form of these likelihood functions can be derived. Certainly, it is easily seen that, under (i) and (ii), we have $\mathcal{L}_c = \mathcal{L}_c^1 \times \mathcal{L}_c^2$ where

$$\mathcal{L}_c^1 = \mathcal{L}_c^1(F) = \prod_{i=1}^n \frac{dF(X_i)^{\delta_i}(1 - F(X_i))^{1-\delta_i}}{1 - F(T_i^-)},$$

and

$$\mathcal{L}_c^2 = \mathcal{L}_c^2(R) = \prod_{i=1}^n (1 - R((X_i - T_i)^-))^{\delta_i} dR(X_i - T_i)^{1-\delta_i};$$

and that

$$\mathcal{L}_m = \mathcal{L}_m(F, L) = \prod_{i=1}^n \frac{(1 - F(T_i^-))dL(T_i)}{\int_0^\infty (1 - F(u^-))dL(u)}.$$

As in Wang (1991), these expressions show that: (a) for a given F, the function $L \mapsto \mathcal{L}_m = \mathcal{L}_m(F, L)$ is maximized by

$$\widehat{L}_F(t) = \frac{\int_0^t (1 - F(u^-))^{-1} dL_n^*(u)}{\int_0^\infty (1 - F(u^-))^{-1} dL_n^*(u)},$$

where L_n^* stands for the ordinary empirical df of the T_i's; (b) the maximum value $\mathcal{L}_m(F, \widehat{L}_F)$ does not depend on the particular F; (c) as a consequence of (a) and (b), the function $(F, L) \mapsto \mathcal{L}_c^1(F) \times \mathcal{L}_m(F, L)$ is maximized by (F_n, L_n), where F_n and L_n are given by

$$F_n(y) = 1 - \prod_{y_j \leq y} \left[1 - \frac{d_j}{n_j}\right] \tag{1}$$

and

$$L_n(y) = \alpha_n \int_0^y (1 - F_n(u^-))^{-1} dL_n^*(u), \tag{2}$$

respectively; where

$$\alpha_n = \left(\int_0^\infty (1 - F_n(u^-))^{-1} dL_n^*(u)\right)^{-1};$$

$y_1 < ... < y_k$ are the k distinct values among the X_i with $\delta_i = 1$, d_j is the number of individuals for which $X_i = y_j$ and $\delta_i = 1$, and $n_j = \sum_{i=1}^{n} \mathbb{I}(T_i \leq y_j \leq X_i)$. Besides, assume that censoring is noninformative, in the sense that the function R contains no information on F. Then, (d) the full likelihood $\mathcal{L}$ is maximized by (R_n, F_n, L_n), where R_n is the maximizer of $\mathcal{L}_c^2$.

The estimators (1) and (2) coincide to those proposed in Wang (1991). However, the derivation of statistical properties on the basis of the new model (i)-(ii) requires specific proofs. In de Uña-Álvarez (2005) a rigorous derivation of the asymptotic sampling distribution of (1) is provided. This asymptotic result will be used in the following Section, in order to construct pointwise confidence bands for the survival function of the households' residence time. On the other hand, the maximizer R_n of $\mathcal{L}_c^2$ is given by the Kaplan-Meier product-limit estimator of the "residual censoring" distribution, based on the pairs $(X_i - T_i, 1 - \delta_i)$. Since, under (ii), the variables $C - T$ and $Y - T$ are conditionally independent, usual properties of Kaplan-Meier estimation hold in this case. Finally, we note that the pair (F_n, R_n) maximizes the conditional likelihood $\mathcal{L}_c$, and thus the property of being a conditional NPMLE is completely justified for (1).

5 Main empirical results

We have evaluated the conditional NPMLE (1) for our 3,268 data on households' duration of residence. Specifically, we have computed the survival function $1 - F_n(y)$ separately for each of the three groups of households, regarding their 1994-tenure status: owners (2,381 households), renters (625), and borrowers (262). Besides, we have evaluated 95%-pointwise confidence limits for these curves, on the basis of the asymptotic normal distribution of the conditional NPMLE. These limits are defined as

$$1 - F_n(y) \pm 1.96 \frac{\sigma_n(y)}{\sqrt{n}},$$

where

$$\sigma_n(y) = [1 - F_n(y)] \left[\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{I}(X_i \leq y, \delta_i = 1)}{M_n(X_i)^2} \right]^{1/2}.$$

and $M_n(y) = n^{-1} \sum_i \mathbb{I}(T_i \leq y \leq X_i)$ (see de Uña-Álvarez 2005). The results are displayed in Table 3 and Figure 2.

Figure 2 reveals several interesting features regarding the households' relative mobility. First of all, it is clearly seen that the owners' residence time is superior to that of the borrowers; and that the latter stay at their houses longer than the renters. The distance between the survival functions corresponding to owners and borrowers is statistically significant from 3.42 years on; while from 6.83 years on statistically significant departures are found when comparing the borrowers' residence time to the renters'.

The renters present a median duration of residence of 3 years and 8 months. This group of households exhibits a relatively high mobility, and indeed the inference shows that 25% of the renters leave their houses in the first two years of residence. Only about 20% of the renters stay living at the same house more than ten years. The median for the borrowers is greater, 7 years and 4 months, and about 44% of these households remain at their houses after ten years of residence. Finally, the mobility of the owners is extremely low; 81% of these do not move in the first ten years, while about 68% remain at the house more than 22 years.

The figures corresponding to owners and borrowers are greater when compared to their 1994-age of current residence, see Table 3. For owners, the median age is 9 years, while for borrowers this parameter is estimated in 6 years. However, the median age for renters (5 years) is greater than their median total residence time, which demonstrates that no general conclusion can be stated regarding the relative location of both populations (age of residence and total residence time).

We have also computed the estimator $L_n(y)$ in (2) for the three groups. These curves are interesting since they provide information on the "incidence rate" for ownership, rentership, and borrowership, in the period 1980-1994. Results displayed in Figure 3 (top) suggest that a constant incidence rate (that is, uniformly distributed truncation) could be present for movements to rentership, but also that non-constant rates could probably rule the renewal processes associated to ownership and borrowership. Indeed, the initial relatively small slope of the curve corresponding to owners indicates that the incidence of ownership, being stationary in the period 1980-1990, has decreased at the beginning of the nineties. On the other hand, the incidence rate for borrowership (that is, the derivative of the corresponding L function) shows an increasing-decreasing shape through the period 1980-1994. Departures of these empirical curves when compared to the uniform model are depicted in Figure 3, bottom. In sum, this Figure 3 supports the evidences reported in Figure 1, since non monotone densities for age of residence can not correspond to uniform truncation. As mentioned, from a methodological point of view, the presence of non-constant incidence rates makes unfeasible the application of the equilibrium equation (Anily *et al.* 1999) for recovering the distribution F from our 1994-current residence times. This issue enhances the importance of the estimation methods applied in this paper, which are completely free of any assumption on the truncation distribution.

For renters, it is possible to compare the estimator depicted in Figure 2 to that obtained by methods discussed in Anily et al. (1999). On the basis of the cubic fit to the density function of the 1994-current residence time (see Figure 1), and by using the equilibrium equation, we have computed the survival function of the total residence time, which corresponds to the formula

$$1 - \widehat{F}(y) = -\frac{1}{\widehat{\beta}_1}(\widehat{\beta}_1 + 2\widehat{\beta}_2 y + 3\widehat{\beta}_3 y^2)\exp(\widehat{\beta}_1 y + \widehat{\beta}_2 y^2 + \widehat{\beta}_3 y^3).$$

Specifically, the estimated parameters are $\widehat{\beta}_1 = -0.2317$, $\widehat{\beta}_2 = 0.0200$, and $\widehat{\beta}_3 = -0.0015$. The estimators $1 - F_n(y)$ and $1 - \widehat{F}(y)$ are jointly reported in Figure 4. Note that $\widehat{F}(y)$ is reliable only on the support of the age of current residence, and hence the comparison can not be performed from 15 years on. The closeness of both curves support the assumption of a constant incidence rate for the renewal process of rentership.

6 Main conclusions

In this paper we investigate the duration of residence for Spanish households from panel data. The inference is based on those households surveyed in 1994 (and followed since that year on) and moving into their 1994-current residence after 1979. We distinguish among owners, renters, and borrowers, since these groups of households are known to exhibit a different pattern regarding residence time. Our main empirical findings are the following: (a) The median residence time for owners (>22 years) is more than six times that of the renters (3.67 years); while the borrowers' estimated median duration of residence is 7.33 years. (b) The figures corresponding to owners and borrowers are greater when compared to the 1994-age of current residence (the median age is 9 years for owners, and 6 years for borrowers); however, the median age for renters (5 years) is greater than their median total residence time. (c) The renewal processes which represents the households' mobility has a non-constant rate over the period 1980-1994 for owners and borrowers. Finding (c) is important since it invalidates the inference methods based on the equilibrium equation discussed in Anily *et al.* (1999). In fact, this paper introduces, in the scope of household's duration analysis, alternative inference methods based on truncation models, which overcome restrictions involved by the equilibrium assumption.

7 References

Anily, S., Hornik, J. and Israeli, M. (1999). Inferring the distribution of households' duration of residence from data on current residence time. *Journal of Business & Economic Statistics* 17, 373-381.

Asgharian, M., M'Lan, C.E. and Wolfson, D.B. (2002). "Length-biased sampling with right censoring: an unconditional approach". *Journal of the American Statistical Association* 97, 201-209.

Ball, M. (2005). RICS European Housing Review 2005. *Royal Institution of Chartered Surveyors Research Report*, London: RICS, 2005, 155 pp.

Benito, A. and Word, R. (2005). How important is housing market activity for durables spending?. *Bank of England Quarterly Bulletin*, Summer, 153-159.

De Mesa. L. and Valiño, A. (2001). Política de vivienda: aspectos teóricos y características en Europa y España. Documento de Trabajo n^{o}27, Facultad de CC. Económicas y Empresariales, Universidad Complutense de Madrid.

Deng, Y., Gabriel, S. A. and Nothaft, F. E. (2002). Duration of residence in the rental housing market. *The Journal of Real Estate Finance and Economics* 26, 267-285.

Díaz-Serrano, L. (2005). Labor income uncentainty, skewness and homeownership: a panel data study for Germany and Spain. *Journal of Urban Economics* 58, 156-176.

Eastaway, M. P. and San Martin, I. (2002). The tenure imbalance in Spain: The need for social housing policy. *Urban Studies* 39, 283-295.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications* (Vol. 2, 2nd ed.), New York: Wiley.

Fernández, A. (2004). Veinticinco años de política de vivienda en España (1976-2001): una visión panorámica. ICE, Tribuna de Economía, 816, 145-161.

Green, R. K. and Hendershott, P. H. (2001). Homeownership and unemployment in the US. *Urban Studies* 38, 1509-1520.

Harsman, B. and Quigley, J.M. (1991). *Housing Markets and Housing Institutions: An International Comparison*, Boston: Kluwer.

Haurin, D. R., Hendershott, P. H. and Wachter, S. M. (1997). Borrowing constraints and the tenure choice of young households. *Journal of Housing Research* 8, 137-154.

Henderson, J. V. and Ioannides, Y. M. (1989). Dynamic aspects of consumer decisions in housing markets. *Journal of Housing Economics* 26, 212-230.

Holdsworth, C., Voas, D. and Tranmer, M. (2002). Leaving home in Spain: when, where and why? *Regional Studies* 36, 989-1004.

Kidd, P. E. (1977). Housing statistics: some sources and some limitations. *American Real State and Urban Economic Journal* 5, 335-354.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press: Cambridge.

López-García M.A. (2004). Housing, prices and tax policy in Spain. *Spanish Economic Review* 6, 29-52.

Martínez, J. and Matea, M. LL. (2002) . The housing market in Spain. Banco de España, Economic Bolletin, October, 71-80.

Martínez, J. and Maza, L. A. (2003). Analysis of house prices in Spain. Documento de Trabajo n^{o}0307, Banco de España-Servicio de Estudios.

Oswald, A. (1997). Thoughts on NAIRU. Correspondence to *Journal of Economic Perspectives* 11, 227-228.

Pickles, A.R. and Davis, R.B. (1991). An empirical analysis of housing careers: a review and a general statistical modeling framework. *Environment and Planning* 23, 465-484.

Schachter, J. P. and Kuenzi J. J. (2002). Seasonality of moves and the duration and tenure of residence: 1996. Population Division, Working Paper n^{o}69.

Trigo, J. (2003). Perspectiva del Mercado de la vivienda en España. Boletín Económico de ICE, n^{o}2774, pp.13-22.

de Uña-Álvarez, J. (2005). On the conditional NPMLE for cross-sectional survival data. Unpublished manuscript.

Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* 86, 130-143.

year of occupation	all houses	owners	renters	borrowers
1994	3.2	1.7	11.9	3.3
1993	2.3	1.4	7.4	2.0
1992	2.4	1.5	6.7	4.1
1991	2.3	1.7	4.2	5.9
1990	3.6	3.2	4.6	7.2
1989	4.0	3.8	4.0	6.1
1988	3.3	3.4	2.9	3.1
1987	3.7	3.8	3.0	4.8
1986	3.3	3.1	3.8	4.4
1985	2.5	2.6	1.7	2.8
1984	3.6	3.7	3.0	3.5
1983	2.5	2.7	2.2	1.5
1982	3.5	3.7	2.5	3.3
1981	2.0	2.2	1.0	1.3
1980	3.3	3.5	2.3	3.9
before 1980	54.5	58.2	38.9	42.9
all houses	100.0	79.4	14.2	6.4

Table 1. Classification of 7,182 Spanish households, wave of 1994, according to the year they moved into their current residence and their tenure status.

follow-up years	censored	uncensored	total	% censoring
0 (1994-1994)	427	0	427	100.0
1 (1994-1995)	255	18	273	93.4
2 (1994-1996)	288	29	317	90.9
3 (1994-1997)	208	31	239	87.0
4 (1994-1998)	108	21	129	83.7
5 (1994-1999)	148	33	181	81.8
6 (1994-2000)	99	17	116	85.3
7 (1994-2001)	1,238	348	1,586	78.1
Total	2,771	497	3,268	84.8

Table 2. Following-up years and censoring proportions for the 3,268 Spanish households which moved to their current residence, wave of 1994, after 1979.

residence time (years)	owners	borrowers	renters
1.5	1.000 (1.000-1.000)	.9167 (.7669-1.000)	.9286 (.8741-.9831)
2	.9937 (.9815-1.000)	.9167 (.7669-1.000)	.7500 (.6604-.8396)
3	.9621 (.9321-.9920)	.8202 (.6383-1.000)	.5952 (.5049-.6856)
4	.9436 (.9093-.9780)	.6754 (.4938-.8571)	.4604 (.3786-.5423)
5	.9187 (.8806-.9568)	.6323 (.4560-.8086)	.3914 (.3156-.4672)
10	.8069 (.7662-.8477)	.4427 (.3066-.5788)	.1980 (.1495-.2465)
15	.7289 (.6887-.7691)	.3715 (.2518-.4912)	.1321 (.0957-.1685)
20	.6875 (.6465-.7285)	.2912 (.1821-.4002)	.0944 (.0639-.1249)
22	.6760 (.6324-.7195)	.2912 (.1821-.4002)	.0866 (.0552-.1179)
median (years)	>22	7.33	3.67
median age (years)	9	6	5

Table 3. Estimated survival function and median for the residence time, together with 95% confidence intervals, for owners, renters and borrowers. For comparison, the median 1994-age of residence is also reported.

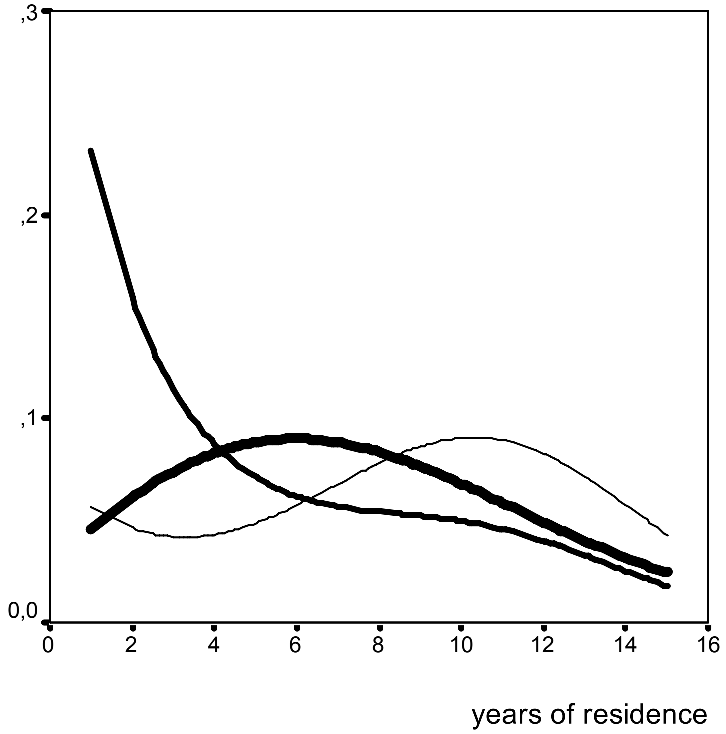

Figure 1. Density function of 1994-current residence time for 3,268 Spanish households: owners (thin line), renters (medium), and borrowers (thick line).

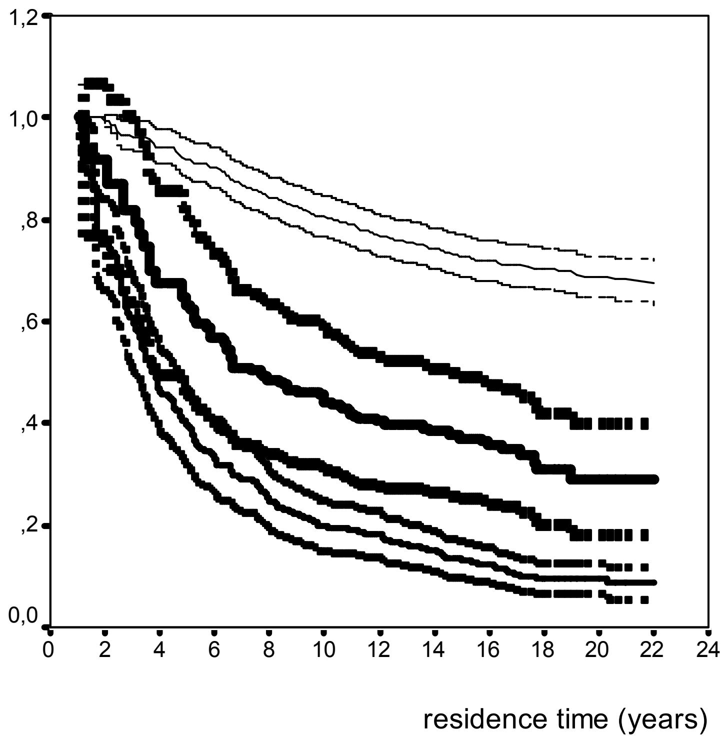

Figure 2. Survival function for residence time (solid lines), with 95% pointwise confidence bands (dotted lines), for 3,268 Spanish households: owners (thin lines), renters (medium), and borrowers (thick lines).

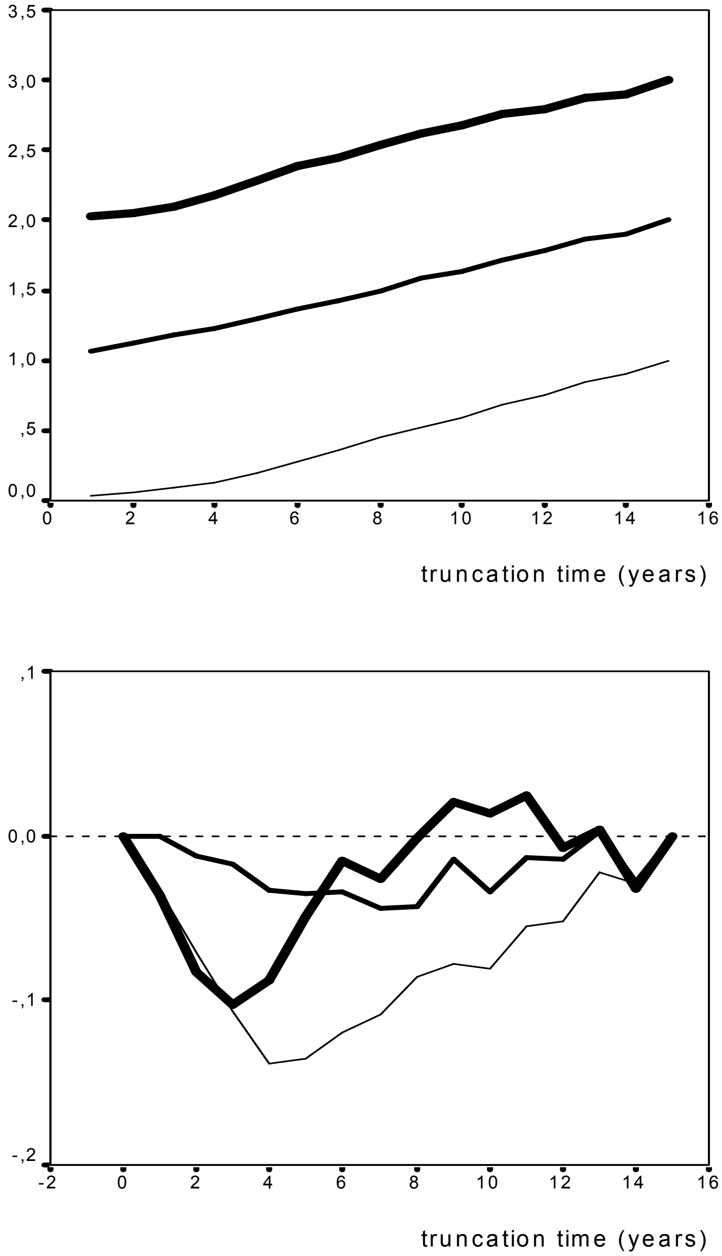

Figure 3. Top: Truncation distribution function for ownership (thin line), rentership (medium) and borrowership (thick line); curves corresponding to renters and borrowers have been moved up for clear visualization. Bottom: Differences between the depicted empirical truncation distribution functions and the uniform distribution (the dotted line corresponds to a perfect fit).

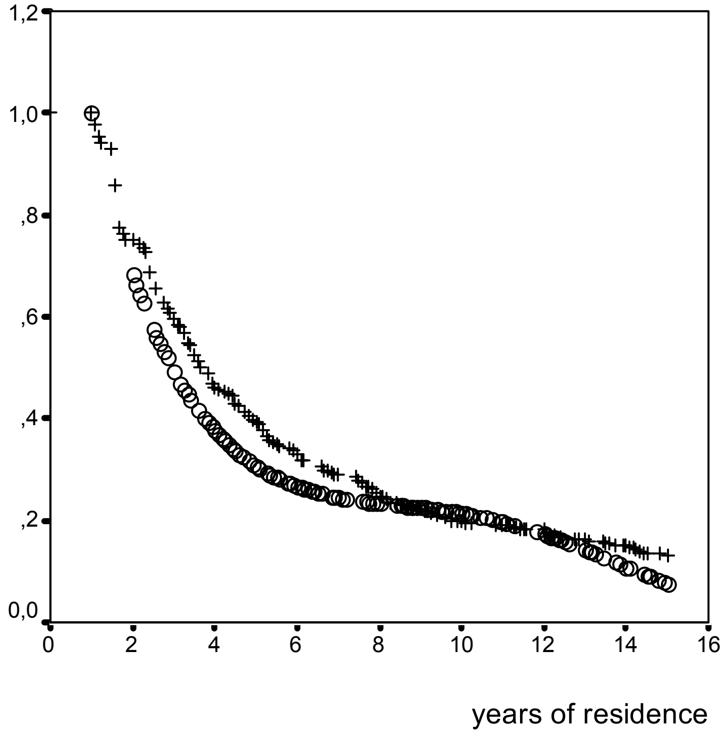

Figure 4. Survival function of total residence time for renters: equilibrium survival (circles) and flexible estimator (crosses).

Relación de Documentos de Traballo publicados

9901 Philippe Polomé: Experimental E vidence on Voting Manipulation in Referendum Contingent Valuation with Induced Value
9902 Xosé M. González e Daniel Miles: Análisis Envolvente de Datos: Un Estudio de Sensibilidad
9903 Philippe Polomé: Combining contingent valuation and revealed preferences by simulated maximum likelihood
9904 Eva Rodríguez: Social value of health programs: is the age a relevant factor?
9905 Carlos Gradín e Mª Soledad Giráldez: Incorporación laboral de la mujer en España: efecto sobre la desigualdad en la renta familiar
9906 Carlos Gradín: Polarization by sub-populations in Spain, 1973-91
9907 Carlos Gradín: Polarization and inequality in Spain: 1973-91
0001 Olga Alonso e José María Chamorro: How do producer services affect the location of manufacturing firms?. The role of información accesibility
0002 Coral del Río Otero: Desigualdad *Intermedia* Paretiana
0003 Miguel Rodríguez Méndez: Margins, Unions and the Business Cycle in High nd Low Concentrated Industries
0004 Olga Alonso Villar: Large metropolies in the Third World: an explanation
0005 Xulia González e Daniel Miles**:** Wage Inequality in a Developing Country: Decrease of Minimum Wage or Increase of Education Returns
0006 Daniel Miles: Infecuencia de las Compras y Errores de Medida
0007 Lucy Amigo: Integración de los Mercados de Cambio: Análisis rentabilidad-riesgo de la cotización Peseta/Dólar
0008 Eduardo L. Giménez e Manuel González-Gómez: Efficient Allocation of Land Between Productive Use and Recreational Use.
0009 Manuel González-Gómez, P.Polomé e A. Prada Blanco: Sesgo sobre la Información Obtenida y la Estimación de Beneficios en Entrevistas a Visitantes de un Espacio Natural
0010 M. Xosé Vázquez Rodríguez e Carmelo León: Preferencias Imprecisas y Contexto en la Valoración de Cambios en la Salud.
0011 Begoña Alvarez: Can we Identify Fraudulent Behaviour?. An Aplication to Sickness Absence in Spain
0012 Xulia González, Xosé M. González e Daniel Miles: La Transición de la Universidad al Trabajo: una Aproximación Empírica.
0013 Olga Cantó: Climbing out of poverty, Falling back in: Low Incomes' Stability in Spain
0101 Arancha Murillas: Investment and Development of Fishing Resources: A Real Options Approach
0102 Arancha Murillas: Sole Ownership and Common Property Under Management Flexibility: Valuation, Optimal Exploitation and Regulation
0103 Olga Alonso Villar; José-María Chamorro Rivas e Xulia González Cerdeira: An análisis of the Geographic Concentratin of Industry in Spain
0104 Antonio Molina Abraldes e Juan Pinto-Clapés: A Complete Characterization of Pareto Optimality for General OLG Economies
0105 José María Chamorro Rivas: Communications technology and the incentives of firms to suburbanize
0106 Luci Amigo Dobaño e Francisco Rodríguez de Prado: Incidencia del efecto día en los valores tecnológicos en España

0107 Eva Rodríguez-Míguez; C. Herrero e J. L. Pinto-Prades: Using a point system in the management of waiting lists: the case of cataracts
0108 Xosé M. González e D. Miles: Análisis de los incentivos en el empleo público
0109 Begoña Álvarez e D. Miles: Gender effect on housework allocation: evidence from spanish two-earned couples
0110 Pilar Abad: Transmisión de volatilidad a lo largo de la estructura temporal de swaps: evidencia internacional
0111 Pilar Abad: Inestabilidad en la relación entre los tipos forward y los tipos de contado futuros en la estructura temporal del mercado de swaps de tipos de interés
0112 Xulia González, Consuelo Pazó e Jordi Jaumandreu: Barriers to innovation and subsidies effectiveness
0201 Olga Cantó, Coral del Río e Carlos Gradín: What helps households with children in leaving poverty?: Evidence from Spain in contrast with other EU countries
0202 Olga Alonso-Villar, José María Chamorro-Rivas e Xulia González: Agglomeration eocnomies in manufacturing industries: the case of Spain
0203 Lucy Amigo Dobaño, Marcos Alvarez Díaz e Francisco Rodríguez de Prado: Efficiency in the spanish stock market. A test of the weak hypothesis based on cluster prediction technique
0204 Jaime Alonso-Carrera e María Jesús Freire-Serén: Multiple equilibria, fiscal policy, and human capital accumulation
0205 Marcos Alvarez Díaz e Alberto Alvarez: Predicción no-lineal de tipos de cambio. Aplicación de un algoritmo genético
0206 María J. Moral: Optimal multiproduct prices in differenciated product market
0207 Jaime Alonso-Carrera y Baltasar Manzano: Análisis dinámico del coste de bienestar del sistema impositivo español. Una explotación cuantitativa
0208 Xulia González e Consuelo Pazó: Firms' R&D dilemma: to undertake or not to undertake R&D
0209 Begoña Álvarez: The use of medicines in a comparative study across European interview-based surveys
0210 Begoña Álvarez: Family illness, work absence and gender
0301 Marcos Álvarez-Díaz e Alberto Álvarez: Predicción no-lineal de tipos de cambio: algoritmos genéticos, redes neuronales y fusión de datos
0302 Marcos Álvarez-Díaz, Manuel González Gómez e Alberto Álvarez: Using data-driven prediction methods in a hedonic regression problem
0303 Marcos Álvarez-Díaz e Lucy Amigo Dobaño: Predicción no lineal en el mercado de valores tecnológicos español. Una verificación de la hipótesis débil de eficiencia
0304 Arantza Murillas Maza: Option value and optimal rotation policies for aquaculture explotations
0305 Arantza Murillas Maza: Interdependence between pollution and fish resource harvest policies
0306 Abad, Pilar: Un contraste alternativo de la hipótesis de las expectativas en Swaps de tipos de interés
0307 Xulio Pardellas de Blas e Carmen Padín Fabeiro: A tourist destination planning and design model: application to the area around the Miño river in the south of Galicia and the north of Portugal
0308 Lucy Amigo Dobaño e Francisco Rodríguez de Prado: Alteraciones en el comportamiento bursátil de las acciones de empresas tecnológicas inducidas por el vencimiento de derivados

0309 Raquel Arévalo Tomé e José María Chamorro Rivas: A Quality Index for Spanish Housing

0310 Xulia González e Ruben Tansini: Eficiencia técnica en la industria española: tamaño, I+D y localización

0311 Jaime Alonso Carrera e José-María Chamorro Rivas: Environmental fiscal competition under product differenciation and endogeous firm location

0312 José Carlos Álvarez Villamarín, Mª José Caride Estévez e Xosé Manuel González Martínez: Demanda de transporte. Efectos del cambio en la oferta ferroviaria del corredor Galicia-Madrid

0313 José Carlos Álvarez Villamarín, Mª José Caride Estévez e Xosé Manuel González Martínez: Análisis coste-beneficio de la conexión Galicia-Madrid con un servicio de Alta Velocidad.

0401 María José Caride e Eduardo L. Giménez: Thaler's "all-you-can-eat" puzzle: two alternative explanations.

0402 Begoña Álvarez e Daniel Miles: Husbands' Housework Time: Does Wives' Paid Employment Make a Difference?

0403 María José Caride e Eduardo L. Giménez: Leisure and Travel Choice.

0404 Raquel Arévalo Tomé e José María Chamorro-Rivas: Credible collusion in a model of spatial competition.

0405 Coral del Río Otero, Carlos Gradín Lago e Olga Cantó Sánchez: El enfoque distributivo en el análisis de la discriminación salarial por razón de género.

0406 Olga Alonso Villar: Ciudades y globalización en la Nueva Geografía Económica.

0407 Olga Alonso Villar: The effects of transport costs revisited

0408 Xavier Labandeira e Miguel Rodríguez: The effects of a sudden CO_2 reduction in Spain.

0409 Gema Álvarez Llorente, Mª Soledad Otero Giráldez, Alberto Rodríguez Casal e Jacobo de Uña Álvarez: La duración del desempleo de la mujer casada en Galicia.

0410 Jacobo de Uña-Álvarez, Gema Álvarez-Llorente e Mª Soledad Otero-Giráldez: Estimation of time spent in unemployment for married women: An application at regional level.

0411 Mª José Moral: Modelos empíricos de oligopolio con producto diferenciado: un panorama.

0412 Mª José Moral: An approach to the demand of durable and differentiated products.

0501 Raquel Arévalo-Tomé e José-María Chamorro-Rivas: Location as an instrument for social welfare improvement in a spatial model of Cournot competition.

0502 Olga Alonso-Villar: The effects of transport costs within the new economic geography.

0503 Raquel Arévalo Tomé, Mª Soledad Otero Giráldez e Jacobo de Uña Álvarez: Estimación de la duración residencial a partir del periodo de ocupación declarado por los hogares españoles.

0504 Olga Alonso-Villar, Coral de Río, Luis Toharia: Un análisis espacial del desempleo a nivel municipal.

0601 Xulia González, Consuelo Pazó: Do public subsidies stimulate private R&D spending?

0602 Lucy Amigo Dobaño: Anomalías de los mercados financieros. Análisis de las empresas gallegas que cotizan en el mercado de renta variable.

0603 Daniel Miles Touya: Can we teach civic attitudes?

0604 Jacobo de Uña Álvarez, Raquel Arévalo Tomé, Mª Soledad Otero Giráldez: Advances in the estimation of households' duration of residence.