**Universidade de Vigo**
**Departamento de Economía Aplicada**

Documento de Traballo
**0302**

# Using data-driven prediction methods in a hedonic regression problem

Marcos Álvarez-Díaz
Manuel González Gómez
Alberto Álvarez

Documentos de Traballo

Marzo 2003

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

# Using Data-Driven Prediction Methods in a Hedonic Regression Problem

**Marcos Álvarez Díaz**[*]
Dept. of Applied Economics
University of Vigo
Lagoas-Marcosende s/n
36200 VIGO (Spain)
Phone: +34 986 81 4072
Fax: +34 986 812401
e-mail mad@uvigo.es

**Manuel González Gómez**
Dept. of Applied Economics
University of Vigo
Lagoas-Marcosende s/n
36200 VIGO (Spain)
Phone: +34 986 81 4072
Fax: +34 986 812401
e-mail mgzlez@uvigo.es

**Alberto Álvarez**
Dept. of Electrical Engineering
University of Pisa
Via Diotisalve 2
56100 Pisa, Italy
e-mail vieaaad@uib.es

[*] *Corresponding author: Departamento de Economía Aplicada, Universidad de Vigo Lagoas-Marcosende s/n, 36200 Vigo, Spain. Fax : 986812401; e-mail* mad@uvigo.es .

# Using Data-Driven Prediction Methods in a Hedonic Regression Problem


**Abstract**

The traditional studies about hedonic prices apply simple functional forms such as linear or linearity transformable structures. Nowadays, it's known in the literature the importance of introducing non-linearity to improve the models' explanatory capacity. In this work we apply data-driven methods to carry out the hedonic regression. These methods don't impose any a priori assumption about the functional form. We use the nearest neigbors technique as non-parametric method and neural networks and genetic algorithms both as semi-parametric methods. Neural Networks have already been employed to the specific hedonic regression problem but, to the authors' knowledge, this is the first time that a genetic algorithm is employed. The empirical results that we have obtained demonstrate the usefulness of applying data driven models in the estimation of hedonic price functions. They can improve traditional parametric models in terms of out-of-sample $R^2$.

## 1-. Introduction

The theory of hedonic prices attempts to evaluate the effects that certain aspects of comfort have on goods prices. The hedonic perspective has been successfully applied to understand the prices of automobiles (Murray and Sarantis, 1999), personal computers (Berndt et al., 1995), televisions (Curry et al., 2001), wine (Combris et al., 1997) and housings (Palmquist, 1984). Among them, housing is one of the goods where strong evidences of association between price and comfort are found. The location of specific environmental amenities nearby, structural characteristics, type of construction and surrounding environment constitute relevant factors on determining the price of a house. Specifically, the price (P) of a residential location could be given in terms of a hedonic price function Ψ(.):

$$P = \Psi(S, N, Q) \qquad (1)$$

where Q is a vector of location of specific environmental amenities, S a vector of structural characteristics of housing such as size, number of rooms, age, type of construction …, and N a vector of characteristics of the neighbourhood in which the house is located such as quality of local schools, accessibility to parks, stores and workplace, crime rates, etc. Thus, the hedonic approach involves estimation of the implicit prices of the neighbourhood, structural and environmental location.

While the dependence between the house's price and comfort characteristics is qualitatively well established in hedonic theory, the adequate functional relation Ψ in (1) remains controversial (Rosen, 1974; Gençay and Yang, 1996). A linear functional form is usually preferred in hedonic theory. The main reason argued for this election is that linear functions are easy to estimate and interpret. Nevertheless, it is also recognized in the literature the importance of non-linearities in the hedonic price function to increase its explanatory capacity (Rasmussen and Zuehlke, 1990). Nonlinear relationships between house's price and comfort characteristics have been incorporated in many works employing techniques that, by means of transformations, allow flexible parametric functional forms (for example, Box-Cox transformations) (Goodman, 1978). These flexible forms impose ad hoc functional structures, introducing sometimes forced and unnecessary non-linearity (Cassel and Mendelsohn, 1985). Results can be "over-parameterized" models with poor out-of-sample performance. Data-driven methods, like nearest neighbours, neural networks and genetic algorithms, that permit to obtain a model without imposing any a priori assumption on the functional form, could be employed to solve this problem.

In this paper we investigate if nonparametric and semi-parametric data-driven models can improve the accuracy on determining the house's price with respect to parametric models. Our empirical study is centered in the real state market in the city of Vigo (Spain). To carry out our research, we collected information about renting prices and houses' information such as their structural attributes and neighborhood conditions. We structure the work as follow. In section 2, we briefly present the different functional estimation methods that we have employed (linear regressions as parametric method, nearest neighbors as non-parametric method and neural networks and genetic algorithms as semi-parametric methods). In section 3, we define the data and show our empirical results on out-of-sample predictions. Finally, section 4 concludes the work.

## 2-. Methodology

### 2.1-. Parametric Techniques

The simplest approach to the hedonic regression problem is to postulate that the functional form $\Psi(.)$ in (1) is linear. To achieve more flexibility, it's very usual to do some non-linear transformation of the data. Most studies employ simple functional relations such as linear (Cameron and Collins, 1997), semi-logarithmic (Chau et al., 2001), double-log (Berndt et al., 1995) or quadratic semi-log (Rasmussen and Zuehlke, 1990), among others. The justification for such functional forms is based on using traditional estimation methods like ordinary least squares, the success of the technique in previous studies or to facilitate statistical inference and hypothesis testing.

The procedure employed in this work consists on selecting the linear functional with highest accuracy and, at the same time, with all variables statistically significant. The selection of variables is carried out by the backward method: a model including all the initial available variables and linear in parameters (linear, semi-log, double-log and quadratic semi-log) is estimated by ordinary least square. Then, the less significant variable is discarded from the model and a new linear function without the deleted variable is computed. The process is repeated until all the survivor variables are statistically significant. The significant threshold employed is 5%.

### 2.2-. Nonparametric Approach

The method of local linear regression is based on the idea that similar situations should show similar effects. Given a sample of observations of input-output pairs of a system, estimation of the unknown output from a new input vector can be obtained by performing a linear regression of the outputs of its K-nearest neighbours. Of great interest is the choice of the nearest neighbours number (K). The rule $K = T^{\alpha}$, with T the sample size and $0 < \alpha \leq 1$ (it's usual to assume $\alpha = 0.5$), is suggested in the literature. In this study, the empirical perspective of proving many values of K and selecting the one with highest out-of-sample performance was adopted.

### 2.3-. Semi-parametric Approach

#### a) Neural Networks

Neural Networks (NN) are a class of semi-parametric models inspired on brain and nerve system working procedure (Bishop, 1995). They are composed of interconnected elements, called neurons, linked between them through weights and grouped in layers. The first layer is called the input layer and the last is the output layer. The middle layers are denominated hidden layers. Each neuron in the input layer brings into the network the value of one independent variable and propagates it towards the neurons of the next layer. In its turn, each neuron of the next layer makes a weighted linear combination of each received input signal, processes this weighted information through a transfer function and sends an output signal. The signals from all neurons are propagated across the NN in the same way as far as the final layer, where the NN's output is offered. The difference between the NN's output and the known value of the dependent variable is calculated. The NN try to minimize this error modifying the weights between links. This process continues iteratively to find the optimal weight's values, finishing when a determined error level is achieved or after a determined number of iterations.

The construction of a good NN for a particular application is not trivial. An appropriate architecture must be chosen to avoid lackness of generalization (for example, number of hidden layers, number of units in each layer, connections between units and transfer functions). Usually, a common practice to build a NN is to select the architecture by a process of "trial and error" searching the highest performance (Tenti, 1996).

A feed-forward back-propagation NN is employed in this work. The input-output relation in this NN is given by:

$$Y_i = \Phi\left(\beta_0 + \sum_{i=1}^{q} \beta_i \cdot \Psi\left(\alpha_{i0} + \sum_{j=1}^{n} \alpha_{ij} \cdot X_{ij}\right)\right) + \varepsilon_i \qquad (2)$$

where $Y_i$ is the dependent variable, $X_i$ the input vector, the parameters $\alpha$ and $\beta$ are the weighs to be adjusted, n is the number of inputs and q the number of hidden units, $\Phi$ and $\Psi$ are the transfer functions and $\varepsilon_i$ a disturb term. It's known and accepted that a three layers feed-forward NN with a linear transfer function in the output unit $(\Phi)$ and a logistic transfer function in the hidden layer neurons $(\Psi)$ is able to approximate any non-linear function to an arbitrary degree of accuracy (Qi, 1999). An important question is how to select the NN inputs. In other words, we have to determine the independent input variables to the NN. Medeiros and Teräsvirta (2001) suggest to carrying out the selection by linearizing the model (2) and applying some linear variable selection method. In our case, the backward method was used to select the relevant variables. More details about NN can be found in Smith (1995) and economic applications in Deboeck (1994).

NN have been employed to solve a huge range of economic problems such as financial time series forecasting (Gately, 1996) and bankruptcy prediction (Shah and Murtaza, 2000). Some works have successfully employed NN to the specific problem of hedonic regression (Curry et al., 2001).

**b) Genetic Algorithms**

Genetic Algorithms (GA) are a functional search procedure based on the Darwinian theories of natural selection and survival. This procedure have been developed by Holland (1975) and divulged by Goldberg (1989) and Koza (1992). In general, its application to economic problems is very scarce and, to the authors' knowledge, it hasn't been used to the hedonic regression problem.

GAs show advantages respect to the neural networks and nearest neighbors methods. First, this procedure permits to obtain explicitly a mathematic equation that can be analyzed. Unlike neural networks, GAs are very flexible because they don't require the specification of a previous architecture.

In this work a GA called DARWIN (Álvarez et al., 2001) was employed. DARWIN performs an optimization process that finds an optimal functional form from a developing initial population of equations. The algorithm simulates in a computer the process of selection and survival observed in the Nature. Briefly, we can explain how DARWIN works in the following way. First, a set of candidate equations (the initial population) for representing the relation between variables is randomly generated. These equations are initially of the form $S = (A \otimes B) \otimes (C \otimes D)$ where the arguments A, B, C and D are the explicatory variables or real-number constants (the coefficients in the equations), and the symbol $\otimes$ stands for one of the four basic arithmetic operators $(+, -, \cdot \text{ and } \div)$. Other mathematical operators are conceivable but increasing the number of available operators complicates the functional optimization process. Each equation of the initial population is evaluated and classified according to its fitness or explained variance $R^2$. The equations with highest values of $R^2$ are selected to exchange parts of the character between them (reproduction and crossover) while the individuals less fitted to the data are discarded. As a result of this crossover, offspring more complicated than the parents are generated. The total number of characters in the equations is upper bounded to avoid the generation of offspring with excessive length. Finally, a small percentage of the equations' most basic elements, single operators and variables, are mutated at random. The process is repeated a large number of times to improve the fitness of the evolving population. At the end of the evolutionary process, DARWIN offers as result an equation that it considers optimal to represent the true functional relation between variables.

**3-. Data and Results**

The sample used consists of 110 observations obtained through interviewing estate agencies in the city of Vigo (Spain), from March to May 1998. Information about renting price and housing characteristics was collected for each house. The housing characteristics are presented in table 1. Nonparametric, semi-parametric and parametric models were estimated from the data. Specifically, models were estimated from the first to the 85$^{th}$ observation (training set). Remaining observations were used to test the model and to obtain the out-of-sample predictions (validation set). As it was already mentioned, variable selection was carried out with the backward method. Finally, the measure to compare the forecasting accuracy of the models is the R-square out-of-sample, $R^2$.

Results obtained from the different models are shown in Table 2. In the case of parametric models, results show that simple models perform well in terms of explanatory power and predictive ability. The best results are obtained with the quadratic semi-log and semi-log models with an out-of-sample $R^2$ of 0.859 and 0.83, respectively. However, notice that the sign showed by the variable *IndCond* in the case of quadratic semi-log is unexpected. A positive contribution of this term would seem more appropriate. The poorest results are obtained with the linear and double-log models (0.72 and 0.656, respectively).

For the non-parametric case, Figure 1 shows the sensibility of $R^2$ versus the number of nearest neighbours considered. The highest accuracy is achieved for K=30 obtaining a value $R^2$=0.8575. The accuracy degrades when a bigger number of nearest neighbours is considered, being indicative of the non-linear nature of the hedonic regression function.

Concerning the semi-parametric methods, the NN model is first analysed. A total of three hidden units was selected. As it can be observed, the NN model produces the best result, $R^2$=0.8621, with a slight improvement respect to the quadratic semi-log, local regression and semi-log models. Conversely, GA presents an accuracy worse than parametric models (except linear and double-log), non-parametric and NN. However, GA permit to obtain a flexible and explicit non-linear expression that approximates the relation between variables. In this way, it can be emphasized two important aspects. First, the expression is conformed by a non-linear component (it affects the variables *M2* and *Cond*) and a linear component (variables *Actecon* and *Ind*). Second, the variables effects on the renting price are the expected a priori corroborating the GA's robustness .

In summary, we have been able to prove how data-driven methods such as Neural Network, Genetic Algorithm (semi-parametric methods) and nearest neighbour (non-parametric method) work acceptably in a hedonic problem. These methods permit to capture a non-linear structure which cannot be fully captured by linear and double-log models. However, linear transformation models such as quadratic semi-log and semi-log can get as goods results as data-driven methods in terms of out-of-sample performance.

## IV-. Conclusions

Hedonic price theory provides a coherent basis for explaining the prices of houses in an urban market as a function of the levels of characteristics embedded in each house. There is an emerging literature in the field of hedonic regressions which successfully applies flexible functional forms. This is due to a recognition of the importance of non-linearities in terms of increasing the explanatory power of the hedonic price function. The empirical results that we presented in this paper demonstrate the usefulness of applying nonparametric and semi-parametric data driven models in the estimation of hedonic price functions. The data-driven models outperform the linear and double-log regression model, and show similar results than others simple functional forms such as quadratic semi-log and semi-log models, in terms of out-of-sample $R^2$. In comparison with the data-driven methods, we can affirm that these last models are capable of capturing the non-linearity existing in the data.

## References

Álvarez A., A. Orfila y J. Tintore (2001) DARWIN- an Evolutionary Program for Nonlinear Modeling of Chaotic Time Series, Computer Physics Communications, 136, 334-349.

Berndt E. R,, Grillches Z. and Rappaport N. J. (1995) "Econometric estimates of price indexes for personal computers in the 1990s", Journal of Econometrics, 68, 243-268.

Bishop C. M.(1995) Neural Networks for pattern Recognition, Oxford University Press.

Cameron J. and Collins A. (1997) "Estimates Hedonic ageing equation for partner search", Kyklos, 50(3), 409-418.

Cassel E. and Mendelsohn R. (1985) "The choice of functional forms for hedonic price equations: comment" Journal of Urban Economics, 18, 135-142.

Chau K. W., Vincent S. M. and Ho D.C. (2001) "The Pricing of `Luckiness´ in the apartment market" Journal of Real State Literature, 9, 1, 31-40 .

Combris P., Lecocq S. and Visser M. (1997) "Estimates of a Hedonic price equation for Bordeaux wine: does quality matter?, Economic Journal, 107, 390-402.

Curry B., Morgan P. and Silver M. (2001) "Hedonic regressions: miss-specification and neural networks" Applied Economics, 33, 659-671.

Deboeck G. (1994) "Trading on the edge: Neural, Genetic and fuzzy systems for chaotic financial markets", eds. Guido Deboeck, John Wiley & Sons.

Gately E. (1996) "Neural networks for financial forecasting", eds. P. J. Kaufman, John Wiley & Sons.

Goldberg D. E. (1989) "Genetic Algorithms in Search, Optimization and Machine Learning". Reading, MA: Addison-Wesley.

Holland J. H. (1975) "Adaptation in Natural and Artificial Systems", Ann Arbor. The University of Michigan Press.

Koza (1992) Genetic programming: On the programming of computers by means of natural seleccion", The MIT Press, Cambridge.

Medeiros M. and Teräsvirta T. (2001) "Statistical methods for modelling neural networks" paper in process.

Qi M. (1999) "Nonlinear predictability of stock returns using financial and economic variables", Journal of Business & Economic Statistics, 17, 4, 419-429.

Rasmussen D. W. and Zuehlke T. W. (1990) "On the choice of functional form for hedonic price functions", Applied Economics, 22, 431-438.

Sha J. R. and Murtaza M. B. (2000) "A Neural Network based Clustering procedure for bankrupcy prediction", American Business Review, June 2000, 80-86.

Smith M. (1995) "Neural networks for statistical modelling" Van Nostrand Reinhold, New York.

Tenti P. (1996) "Forecasting foreign exchange rates using recurrent neural networks" Applied Artificial Intelligence, 10, 567-581.

Table 1. Description of the Independent Variables

| VARIABLES | DESCRIPTION |
|---|---|
| **Rp** | Renting Price in Pesetas. For the NN case this variable was normalized and, for the GA, was divided by 1000. |
| **M2** | Square Meters |
| **Cond** | Dicotomic Variable that takes the value 1 if the house is catalogued by the real state agent as excellent to occupy. |
| **Ind** | Index built adding 5 structural characteristics: existence of lumber-room, grocery-store, central heating, elevator and if the kitchen is furnished. |
| **Actecon** | Variable collecting the economic activity of the street where the house is located. It's calculated as the ratio between the number of business in the street and the number of houses. |
| **Npg** | Number of garage places |
| **Ncb** | Number of bathrooms |

Graphic 1. Local Regression. K Nearest Neighbours Determination

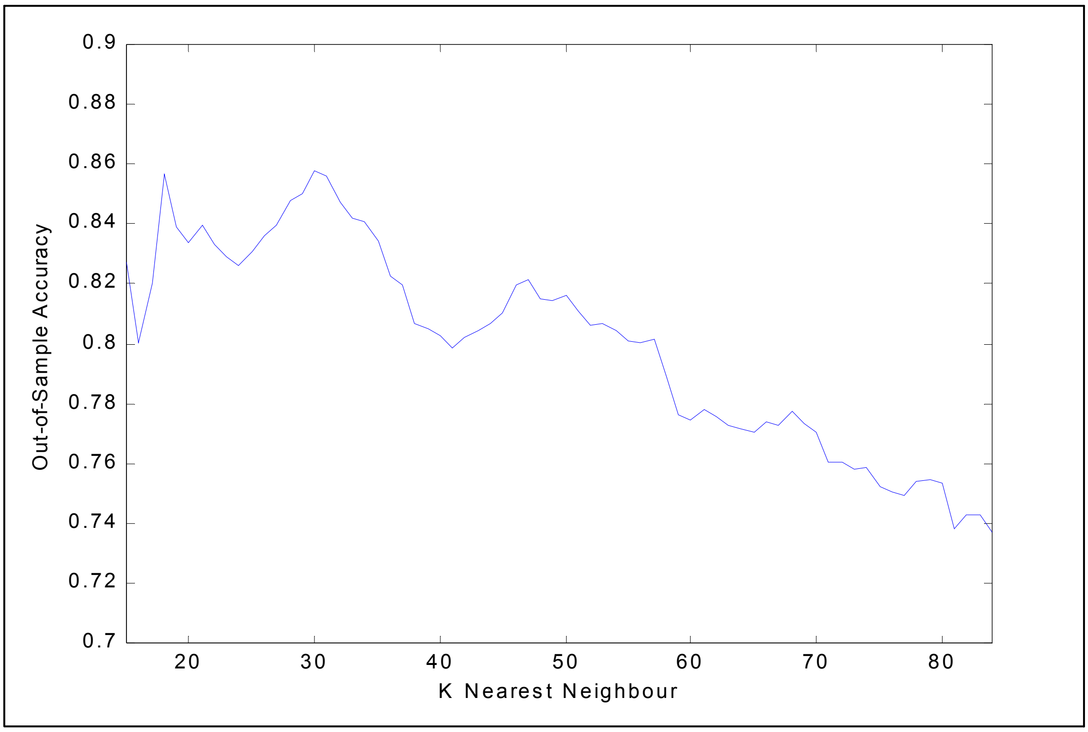

Table 2. Out-of-Sample Accuracy and Comparison between Models

| HEDONIC REGRESSION | METHOD | MODEL | $R^2$ Out-Of-Sample |
|---|---|---|---|
| PARAMETRIC METHODS | Linear Regression | $\hat{R}pi = \underset{(t=10.12)}{366.28} \cdot M2_i + \underset{(t=4.541)}{6205.5} \cdot Ind_i + \underset{(t=3.553)}{5035.1} \cdot Actecon_i + \underset{(t=4.89)}{18924} \cdot Cond_i$ | 0.7232 |
| | Semi-log | $\log \hat{R}pi = \underset{(t=107.7)}{10.074} + \underset{(t=6.123)}{0.0047391} \cdot M2_i + \underset{(t=4.475)}{0.10679} \cdot Ind_i + \underset{(t=3.208)}{0.084365} \cdot Actecon_i + \underset{(t=4.329)}{0.2749} \cdot Cond_i$ | 0.830 |
| | Double-Log | $\log \hat{R}p_i = \underset{(t=29.99)}{8.8911} + \underset{(t=5.685)}{0.38529} \cdot \log M2_i + \underset{(t=4.39)}{0.10743} \cdot \log Ind_i + \underset{(t=2.959)}{0.10166} \cdot \log Actecon_i + \underset{(t=4.342)}{0.28185} \cdot Cond_i$ | 0.656 |
| | Quadratic Semi-log | $\log \hat{R}p_i = \underset{(t=166.1)}{10.393} + \underset{(t=3.170)}{0.081046} \cdot Actecon_i + \underset{(t=3.131)}{0.70081} \cdot Cond_i + 0.5 \cdot \left[ + \underset{(t=8.640)}{0.00341} \cdot M2Ind_i - \underset{(t=-2.177)}{0.30031} \cdot IndCond_i \right]$ | 0.859 |
| NON PARAMETRIC METHOD | Local Regression | The number of nearest neighbours considered was 30 | 0.8575 |
| SEMI-PARAMETRIC METHODS | Neuronal Network | Feed-Forward Back-Propagation with 3 layers and 1 neuron in the hidden layer. | 0.8621 |
| | Genetic Algorithm | $\hat{R}p_i = \left\{ M2i \cdot \left[ 1 - \frac{1}{1.76 + Cond_i} \right] + 4.52 \cdot Actecon_i + 4.23 \cdot Ind_i \right\}$ | 0.8220 |

In bracket the t-student statistics.

**Relación de Documentos de Traballo publicados**

**9901** Philippe Polomé: Experimental Evidence 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 adn 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 Álvarez Díaz e Alberto Álvarez: 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 e 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