NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

The application of hedonic price models in the estimation of effective and imputed rents of Housing: A case study on the 2021 housing census data in Portugal.

Rogério Domingos Paulo

Dissertation

presented as partial requirement for obtaining the Master's degree program in Data Science and Advanced Analytics, with specialization in Business Analytics.

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

The application of hedonic price models in the estimation of effective and imputed rents of housing: A case study on the 2021 housing census data in Portugal.

by

Rogério Domingos Paulo

Dissertation for obtaining the Master's degree in Data Science and Advanced Analytics, with Specialization in Business Analytics.

Supervisor: Bruno Damásio

Co-Supervisor: Ângelo Teixeira

July 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the rules of conduct and code of honor from the NOVA Information Management School.

Rogério Domingos Paulo

Lisbon, July 2023

DEDICATION

Firstly, I dedicate this work, as well as this journey of training in this masters course in Data Science and Advanced Analytics to my beautiful and lovely daughter Adrielle de Rogério Paulo that for reasons of seeking to provide a better future for her, I had to make a great sacrifice of separating me from her with my destiny to Portugal when she was only one year and five months old, I confess it was not easy to make this decision, but because the reason was greater than emotion, I had to pack my bag and follow the battle.

Secondly, I dedicate this work and my entire journey to my dear mother Rosalina Alde for her special dedication during my entire childhood before leaving Nampula to Maputo to start my academic journey to degree level and it was through this journey that the will to continue investing more in training was born because knowledge is the best tool that any man should be thirsty to seek.

To my brothers in memory, Romano Napita and Hilário Napita, I dedicate also this journey to you because I always have you in my heart.

Finally, this journey is also dedicated to all those who have prayed for me and who have always believed in me, I will not leave them aside because they are part of my battles and achievements.

ACKNOWLEDGEMENTS

In first place I thank the almighty God who has been my best refuge at all times of my life and in particular during my academic journey, the right words to thank him for everything he has done and what he will do has been missing because he is wonderful in a continuous way and that so blessed me also with a beautiful daughter who is called Adrielle de Rogério Paulo and she is the reason for my decision to continue with my studies to provide a better future.

Secondly, I want to thank my supervisor Bruno Damásio for accepting me as his supervisee and also for his dedication in the supervisory process for this work to reach the desired scientific quality to serve as a work for obtaining this master degree in Data Science and Advanced Analytics.

Thirdly, my other major thanks go to my co-supervisor, master Ângelo Teixeira who, by his intelligence, expertise, mastery and vast experience in this research area, played a major role in the quality of this work.

I will never forget those who are the reason for my existence and did everything to make my birth possible, and those wonderful people are my parents, Paulo Napita and Rosalina Alde, to them I wish that the Lord our God bless them and keep them because there is no invoice that pays all their dedications to me.

To my brothers, Isildo Napita, Sérgio Napita, Lucina Napita and Wiliam Napita, I thank you for always believing in me and for the strength and encouragement you gave me at all times of my life and always put confidence in my dreams, many thanks.To my beloved nephews Kevin Napita, Paula Rosalina, Caleb Estevão, Octavia dos Santos and Gizela da Lucina, my extensive thanks to them for being special to me.

To my best friends who rise to the position of true brothers Fernando Machavane and Nelido Atumane, words fail me for the great and unconditional love that they manifested for me and to them I wish many blessings from God, along with their wonderful family.

To conclude, to all who directly or indirectly made all my journey possible, my sincere gratitude.

ABSTRACT

The output of services of owner-occupied dwellings is valued at the estimated value of rental that a tenant would pay for the same accommodation, considering factors such as location, neighbourhood amenities, as well as the size and quality of the dwelling itself. Statistics Portugal Institute, in the context of the National Accounts, whenever the households own the dwelling they occupy, must estimate a market value for the rent of that dwelling (imputed rent) based on the rent of similar or even different dwellings that are rented (effective rent). This study uses the imputation hedonic price model with the objective of estimation of the actual housing rental price and the prices for own occupation housings. Because the dependent variable is income value in numerical intervals, the method used to estimate the hedonic function was Truncated Ordered Probit Model. The functional form used is log-linear and the model included interaction and square root variables and also includes the external variable which is coefficient of location. In terms of research methodology, this is a quantitative and applied study that uses descriptive methods based on bibliographic references and also uses a secondary source of data. The main database of the study is the general housing census conducted in Portugal in the year 2021.The study concluded that the imputation hedonic regression model recognizes that heterogeneous goods can be described by their attributes and location market. The estimation of imputed and effective rents is mainly used for harmonization purposes in determining the Gross National Income (GNI), which in turn plays a key role in the analysis and coordination of economic policies.

Keywords

Hedonic price model; Hedonic function; Imputation method; Effective rental price; Imputed rental price.

Sustainable Development Goals (SGD): https://sdgs.un.org/goals,

INDEX

1. INTRODUCTION..1

2. LITERATURE REVIEW..4

2.1. What are a hedonic price models?..4

2.2. The fundamental of the hedonic equation 4

2.2.1. First stage (Single equation) 5

2.2.2. Second stage of the hedonic model: Recovering structural parameters. 5

2.3. The hedonic prices, and related pricing models: An abbreviated history 6

2.4. Theoretical underpinnings of hedonic models...7

2.4.1. Two identification problems...7

2.4.2. Equilibrium or disequilibrium models? 7

2.4.3. Single-equation models, or structural models? 8

2.5. Specification issues..8

2.5.1. What theory tells us about specification?..8

2.5.2. Choice of the dependent variable 9

2.5.3. Selection of independent variables..9

2.5.4. Functional form in general 10

2.5.5. The log-linear form has five things to recommend it. 10

2.5.6. Functional Form and Independent Variables 11

2.6. Market and submarket, definition 11

2.7. Hedonic specification: Art or science? 11

2.8. The hedonic imputation approach 12

2.8.1. Motivation 12

2.8.2. Imputation model..12

2.9. Limited dependent variables..12

2.9.1. Truncation....12
2.9.2. Truncated regression model (Tobit model)....13
2.9.3. Maximum Likelihood Estimation (MLE)....13
3. DATABASE CHARACTERIZATION....15
3.1. Data Universe....15
3.1.1. Owners....15
3.1.2. Renters....15
3.2. Characterization of the universes of owners and renters....16
4. METHODOLOGY....21
4.1. Hedonic method....21
4.1.1. The imputation hedonic method....23
4.1.2. Maximum Likelihood Estimator (MLE)....23
5. RESULTS AND DISCUSSION....25
5.1. Final hedonic rental price model....25
5.2. Descriptive Statistics....26
5.3. Coefficient estimates and statistical significance....28
5.4. Effective rental price estimation....29
5.5. Imputed rental price estimation....33
6. CONCLUSION, LIMITATIONS AND FURTHER RESEARCHS....37
7. REFERENCES....40
APPENDIX I....49

LIST OF FIGURES

Figure 1 – Universe of owners and renters in total and for estimating the model. 16

Figure 2 – Number of rooms in dwelling.. 17

Figure 3 – Size of dwelling by floor area 17

Figure 4 – Access and comfort features of the accommodation 17

Figure 5 – Buildings: type, construction period, state of conservation 18

Figure 6 – Derived indicators 19

Figure 7 – Lease agreement: season, owner entity and rent scale 19

Figure 8 – The mechanics of the imputation's hedonic method 23

Figure 9 – Histogram of rent scales.. 26

Figure 10 – Period of the lease contract 26

Figure 11 – Rental price contribution by region 32

Figure 12 – Imputed price contribution by region 36

Figure A.1 – The basic PROC QLIM algorithm of (Truncated Ordered Probit Model or Tobit Model) in SAS. 58

Figure A.2 – Convergence algorithm of Truncated Ordered Probit Model. 59

LIST OF TABLES

Table 1: Descriptive statistics of the variables associated with quantitative dwelling characteristics included in the model 27

Table 2: Descriptive statistics of the variables associated with qualitative dwelling characteristics included in the model 27

Table 3: Coefficient estimates of the variables associated with qualitative and quantitative dwelling characteristics included in the model..28

Table 4: General results by region of effective rental price 30

Table 5: Summary outputs of effective rental price 32

Table 6: General results by region of imputed rental price 33

Table 7: Summary outputs of imputed rental price...36

Table A.1: Description of Dummy variables associated with qualitative and quantitative dwellings characteristics 48

Table A.2: Descriptions of 308 Dummy variables associated with dwellings location...49

Table A.3: Census 2021 variables type and description 52

Table A.4: Coefficient estimates of dwelling locations..54

LIST OF ABBREVIATIONS AND ACRONYMS

MLE – Maximum Likelihood Estimation.

EM – Expectation-Maximization.

OLS – Ordinary Least Square.

RPPIs – Residential Property Price Indexes.

SECO – The State Secretariat for Economic Affairs.

DRE – Diário da República Eletrónico.

GRI – Global Reporting Initiative.

IMI – Imposto Municipal Sobre Imóveis.

EURATOM – European Atomic Energy Community.

EC – European Commission.

1. INTRODUCTION

Statistics Portugal Institute, in the context of the National Accounts, whenever the households own the dwelling they occupy, must estimate a market value for the rent of that dwelling (imputed rent) based on the rent of similar dwellings that are rented (effective rent). Henceforth, following up on this subject raised, usually in the real estate market to calculate these market values of the properties, the methodology used is hedonic price models. The hedonic price model is a regression model that makes use of a hedonic function. Therefore, a hedonic function is a relation between the prices of different varieties of a product or housing, and the quantities of characteristics in them (Eurostat ,2013).

According to Eurostat (2013), the housing sector has great importance since a residential property is both a source of wealth and to the extent that owners live in their own kind of property as well as renting. And yet, the housing sector is very much associated with the economic health and wealth of any nation and high demand for housing would trigger growth in many other economic sectors. Therefore, the Council of the European Union, having regard to the Treaty establishing the European Atomic Energy Community (EURATOM), based on Council Decision 2000/597/EC of 29 September 2000 on the system of the European Communities' own resources, states that Gross National Product (GNP) at market prices shall be considered equal to Gross National Income (GNI) at market prices as provided by the Commission in application of the European System of Accounts (ESA). GNI, replaced GNP as a criterion for own resources with effect from budget year 2002. However, regarding the harmonization of the compilation of Gross National Income (GNI) at market prices, the principles for the calculation of dwelling services were laid down by Decision 95/309/EC[1]. However, whenever the census takes place, the statistical offices use the price data on dwelling rents to estimate imputed and effective rents for harmonization purposes in the determination of Gross National Income (GNI).

However, based on the explanation above, the raised research question it is concerned with: How can make use of the universe of rented dwellings identified in Census 2021 to define

---

[1] Commission Regulation (EC) No 1722/2005 of the European Community (EC) and the European Atomic Energy Community (EURATOM).

a hedonic model with their characteristics and locations, to estimate the value of the actual rents (effective rents), and also for dwellings used for own occupation (imputed rents)?

In regard to the research question above raised, the research gap it is centered on the estimation of the market value for the rental of an owner-occupied dwelling (imputed rent) based on the rental of similar dwellings that are rented (effective rent). The general research objective it is to estimate the value of housing rents, effective and imputed, based on the Census 2021 data.

Throughout the development of this research and as the concern focuses on answering the research question, it is important to notice that five specific objectives will be devoted in which are split into theoretical standpoint (the first two ones) and practical standpoint (the last three ones), namely: 1) to conceptualize the hedonic price model; 2) to conceptualize the hedonic function; 3) to estimate the "shadow prices"; 4) to estimate the effective rents; and 5) to estimate the imputed rents.

According to Triplett (2006), the "shadow prices" are the implicit rental price obtained in the universe of housings represented with a bundle of characteristics located in the market.

Before we take the subsequent steps into the development of the study, it is important to classify in terms of the methodological approach. According to Kumar & Ranjit (2011), this is a quantitative and applied research that uses descriptive methods based on bibliographic references and also uses a secondary source of data.

By researching various aspects of real estate markets, the hedonic pricing model, derived from consumer Lancaster (1966) and model theory Rosen (1974), has been widely used.

The hedonic pricing model is the linear relationship between the price of goods and the characteristics contained in those goods (Lancaster,1966). Numerous studies have used this technique to examine the relationship between attribute preference and property price (Li & Brown, 1980; Gillard, 1981; Walden, 1990; Sirpal, 1994). This is because the market price of a housing unit can be determined by buyers' evaluations of the set of attributes inherent in the housing unit, such as location, structural or neighborhood attributes (Freeman M., 1979).

In hedonic models, the set of characteristics that influence prices is different in different markets, and the variables included in the model are also constrained by data availability.

In the market, housing prices are heterogeneous because they can be differentiated in terms of local, structural or neighbourhood attributes, or based also on some other criteria such as housing type (bungalow, terrace house, high-rise flat, or condominium)(Kwong, 2003).

Taking into consideration all the situations that can be found in the market, in applying the hedonic pricing model to the real estate market, it is important to take into account the assumptions such as the homogeneity of the real estate product, the market operates under perfect competition, and there are numerous buyers and sellers, and many buyers looking for housing in the market, buyers and sellers have perfect information about the real estate products and the price is quite reasonable, and there are also many real estate developers providing the housing, and buyers and developers are considered to have the freedom to enter and exit the market (Kwong, 2003).

Finally, the hedonic pricing model only works under the assumption of market equilibrium, and there are no interrelationships between the implicit prices of attributes (Dusse & Jones, 1998). Market equilibrium is not plausible because there are imperfections in the real-world housing market. It is idealistic to assume that the price vector will adjust instantaneously to changes in either demand or supply at any point in time.

2. LITERATURE REVIEW

In this session, the main objective is to provide the theoretical and scientific basis that sustains the theme under study in greater depth. To this end, the following topics will be addressed in this session: The concept of the hedonic price models, the fundamental of the hedonic equation, the hedonic prices and related price models: An abbreviated history, the theoretical underpinnings of hedonic models, the specification issues in the hedonic models, the definition of market and submarket, the hedonic specification: Art or science, the imputation approach and the approach of limited dependent variables.

2.1. WHAT ARE A HEDONIC PRICE MODELS?

Hedonic models are those that make use of all available information and allow to control of quality and quantity differences by formulating the price of a dwelling as a function of its characteristics and location (Hill, 2004). According to Malpezzi (2003), hedonic equations can be described as one-way expenditures on housing that can be decomposed into measurable prices and quantities so that rents for different or identical housings in different locations can be compared and predicted. Sirmans *et al*. (2005), define hedonic model as a regression that allows to estimate the value of individual attributes or characteristics of the housing. The independent variables assume the individual attributes or characteristics of the housing, and the regression coefficients may be transferred into estimates of the implicit rental prices "shadow prices "of these attributes or characteristics.

The use of hedonic models in rent pricing can also provide a framework in which depreciation and renovations can be accommodated and, moreover, avoid the need for retrospective reviews. It is important to emphasise that there are some drawbacks associated with the use of hedonic models, which include incorrect specification, omitted variable bias, variable selection bias and the treatment of sparse data. Furthermore, if the number of transactions is low and the number of attributes or characteristics in the hedonic price functions is large, there may be a problem in estimating the model (Malpezzi, 2003).

2.2. THE FUNDAMENTAL OF THE HEDONIC EQUATION

Before moving forward, it is important in a short overview to distinguish between the first-stage hedonic model (single equation) and the two-stage hedonic model. The below of this section will be discussed these topics and therefore, both stages in this research will be treated as a hedonic model at all.

2.2.1. First stage (Single equation)

According to Malpezzi (2003), the first stage of the hedonic model, simply estimates the effect of characteristics on prices and does not examine the structural parameters of the individual attributes or characteristics. Taylor (2008), says that in the first stage the hedonic model is estimated using information about prices and housing attributes or characteristics. In addition, Sirmans *et al*. (2005), sustain that the hedonic model estimated as a single equation is also estimated in different forms, regarding the response variable (rental price). However, the most and typical model structural form has been the semi-log, with the rental price specified in natural logarithms and regressed against unlogged independent variables. This technique allows for variation in attribute or characteristics across different rental price range within the universe.

2.2.2. Second stage of the hedonic model: Recovering structural parameters.

Rosen (1974) cited by Malpezzi (2003), explains that the equation presented above is a reduced form, but he also admits that under certain hypotheses, hedonic equations can be interpreted structurally. As a specific case, we know that rent prices vary within the universe, but if the level of characteristics also varies, we can use these variations to estimate the elasticities of rent prices for the individual coefficients. However, it is common in studies to use a second stage linear demand model after having used a logarithmic hedonic model in the first stage. Taylor (2008), succinctly describes the second stage of hedonic model as the procedure in which the implicit prices for housing attributes or characteristics recovered in the first stage could be combined with observed quantities of attributes or characteristics chosen to estimate underlying demands for attributes or characteristics.

Witte *et al*. (1979); Follain & Jimenez (1985), present studies on estimating the demand for attributes or characteristics of housing using hedonic price models. They assume that supply is elastic, so only those factors changing demand are included, such as household income or even family size, which are readily available. However, the prototypical data set for the first stage includes household-level data on a response variable, such as rent price, and also includes the attributes or characteristics of the dwelling. The prototypical dataset for the second stage adds information on household/unit level demand and supply shifters. In other circumstances, also neighbourhood-level data can be appended to household level data[2].

2.3. THE HEDONIC PRICES, AND RELATED PRICING MODELS: AN ABBREVIATED HISTORY

Although Maclennan (1977), some thirty years ago or more made a crucial contribution to the hedonic models that emerged during explosive growth, in the year two thousand and twenty-three Khoshnoud *et al*. (2023), provides an important review of the hedonic price model literature focusing on the studies of top three traditional real state journals published between years thousand and five and thousand and twenty-one.

However, both provided several important articles and relevant discussion about the hedonic literature. For instance, two oft-cited predecessors of MacLennan's study were classic papers of Lancaster (1966) and Rosen (1974)[3]. Another relevant article discussed by Maclennan point out the studies of Haas (1922); Wallace (1926); Court (1939); Hicks (1939, 1960); Griliches (1971); Shell & Fisher (1971); Witte *et al*. (1979); Awan *et al*. (1982); A. Goodman (1998) and P. Colwell (1999). A full history discussing the results obtained in each study would be centred on developing another study, so the reader is advised to consult these studies for further details.

Regarding Khoshnoud *et al*. (2023) reviews of the hedonic price model studies developed between the interval above mentioned, his research shows that the real estate research is global in both authorship and data usage and most of the studies use U.S. data and are performed by U.S. researchers and the real estate modelling is quite expansive and includes some geospatial autoregressive models and other advanced techniques.

[2] Some hedonic studies use aggregate data, e.g., average levels of variables over, say, Census tracts; but these have gone out of favour, partly because of aggregation bias (as discussed in Ball (1973)) and partly because household/property level data has become more readily available.

[3] Interestingly, a casual perusal through stacks of papers suggest that UK scholars more often cite Lancaster as their fundamental reference, and U.S. authors more readily cite Rosen. Of course, both papers are often cited by writers from any country. See also Lancaster's later elaboration of his ideas in Lancaster (1971).

2.4. THEORETICAL UNDERPINNINGS OF HEDONIC MODELS

In this study will be covered a briefly discussion about the theoretical underpinnings of hedonic models, however the reader can find most cited surveys such as: Follain & Jimenez (1985); Sheppard (1999); Sirmans *et al*. (2005); Taylor (2008); Hill & Rambaldi (2021) and Khoshnoud *et al*. (2023).

The underpinning theoretical point of the hedonic models discussed by Malpezzi (2003), arises because of existence of heterogeneous housing stock and heterogeneous consumers in the market. Although each housing can contain different attributes or characteristics, this attributes or characteristics may be valued differently by different consumers.

Another issue faced by the hedonic model researchers is related with distinguishing between supply and demand, and it is called identification problem that we will discuss in the below (Shell & Fisher, 1971; Triplett, 2006).

2.4.1. Two identification problems

According to Malpezzi (2003), in hedonic models we can face two problems namely: The first problem is related with separation between supply and demand, where the price of the goods is exogenous and the consumer is a price-taker, he decides how much to consume based on the price. And the second problem is related with non-linearity where the price varies with the quantity, the consumer chooses a price and quantity.

These problems have been addressed by Blomquist & Worley (1982) and Diamond & Smith (1985), and in the Follain & Jimenez (1985) and Sheppard (1999) research. The reader is referred to their more discursive treatment.

2.4.2. Equilibrium or disequilibrium models?

One of the characteristics of the housing market is the issue of market equilibrium or disequilibrium. However, for hedonic models, the aim is for markets to be balanced. In this study, we will discuss how to identify and correct the issue of market imbalance. However, for more details we recommend that the reader consult the work of R. C. Fair & Jaffee (1972); Bowden (1978); Grigsby *et al*. (1985); Anas & Eum (1984); Abraham & Hendershott (1996); Dreiman & Follain (2000); Malpezzi (2000); Boyle & Kiel (2001) and Taylor (2008).

Malpezzi (2003) points to three fundamental steps. The first is to specify the nature of the process that distinguishes equilibrium from disequilibrium observations. The second is to focus on the effects of disequilibrium in analysing the relationship between the measured process, the amount of disequilibrium in each market or period and also the process of adjustment back to equilibrium. The third and final step is to analyse the determinants of imbalance and the temporal nature of the return to equilibrium.

2.4.3. Single-equation models, or structural models?

According to the studies carried out by Follain & Jimenez (1985); Gross (1986) and Mayo & Gross (1987), the World Bank's research projects on housing demand concluded that, after investing significant resources in an attempt to develop demand models for characteristics that would improve the concentration of low-cost housing projects, the problems of identification, specification and the lack of robustness of the coefficient estimates led to the question of the reliability of the demand for characteristics in hedonic models applying only the second stage, which is why Triplett (2006) suggests that the best technique is to use the first stage and then the second stage to overcome these problems. See also researches like King (1975); Mayo (1981); Olsen (1987) and Whitehead (1999), to add more knowledge.

2.5. SPECIFICATION ISSUES

In this section, we discuss several practical issues such as which variables or functional form to use, or how to define a market. But first, we give a brief overarching comment explaining why these efforts so often seem ad hoc.

2.5.1. What theory tells us about specification?

According to Sirmans *et al*. (2005), the question of specification in hedonic models is raised because of the difficulty of applying them in practice, i.e., a priori the choice of functional form does not depend on the domain of science itself, but is also associated with art. The articles by Lancaster (1966) and Rosen (1974) elegantly present models of housing characteristics without having much to say about what these characteristics are or how exactly they are related to rental prices. However, in essence, in hedonic models, the dependent variable in the model is usually a sales price, which serves as an indicator of the value of the home and the use of this practice is due to the fact that the observed price minimises bias better than other measures, such as the owner's self-assessment.

As far as the independent variables are concerned, there are an almost unlimited number of independent variables that can be included in the model and the high correlation of some of these variables with each other can create estimation problems, even if not all the variables are included in the model. For example, a location variable that is highly significant in the model may in fact be reflecting another reality, such as the quality of the hospital, and this makes it more difficult to interpret the individual coefficients obtained. To this end, it is crucial to highlight the issue of specification in hedonic models.

2.5.2. Choice of the dependent variable

For hedonic price models, the issue of choosing the dependent variable is one of the particularities related to the specification problem discussed above. According to Malpezzi (2003) cited by Sirmans *et al*. (2005), although there is confusion due to the loose terminology used by real estate professionals and sometimes by housing economists, it is first important that the choice of dependent variable is the rental price or value, one of the reasons for choosing this variable, as mentioned in the previous paragraph, is intrinsically related to minimising bias. Some literature provides more detail on the advantages of this practice, such as the research by Kain & Quigley (1972); Follain & Malpezzi (1981); Maddala (1983); J. L. Goodman & Ittner (1992); Gatzlaff & Haurin (1997) and Malpezzi (1998).

2.5.3. Selection of independent variables

We have already mentioned that the choice of independent variables to be included in the hedonic price model is also part of the specification problem, however, Sirmans *et al*. (2005) explains that there is an unlimited number of independent variables that can be included in the model, but the important thing is to observe the issue of high correlations between these variables in order to avoid estimation problems and this can even happen without all the variables being included in the model. Malpezzi (2003) makes it clear that theoretical knowledge may not be a great guide, but the art of the practices employed in the work related to these models helps a lot to avoid the inclusion of problematic variables. These studies that provide good practice in the choice of variables for hedonic models include the works of Hocking (1976); Leamer (1978); Amemiya (1980); Butler (1982b); Ozanne & Malpezzi (1985) e Malpezzi (2003).

2.5.4. Functional form in general

The choice of functional form in hedonic models plays a major role and provides numerous advantages, but Rosen (1974) had already made it clear that there is no solid theoretical basis for the correct functional form to be used in hedonic models. This idea is also supported by econometric literature, such as Wooldridge *et al*. (2016), among others. However, whenever we want to choose the functional form, it is important that we analyse the advantages of using it for the nature of our data to be studied. For this study, below we present the advantages of choosing a log-linear functional form, for example, but for more details the reader can consult the literature by Follain & Malpezzi (1980); Halverson & Pollakowski (1981); Malpezzi (2003) e Sirmans *et al*. (2005).

2.5.5. The log-linear form has five things to recommend it.

In the study conducted by Sirmans *et al*. (2005), he points out three advantages of the semi-log functional form in the hedonic models. However, Malpezzi (2003) goes further and points out five advantages of using this functional form for hedonic models. Firstly, the semi-log model allows the value of a particular characteristic to vary, so that the price of one component depends partly on the other characteristics of the house. Second, the coefficients of a semi-log model have a simple and appealing interpretation. Third, the semi-log form often mitigates the common statistical problem known as heteroskedasticity or changing variance of the error term. Fourth, semi-log models are computationally simple, and so well suited to examples. And finally, we note that it is possible to build specification flexibility into the right-hand side, using dummy (or indicator) variables, splines, and the like (of which more shortly). The reader is also recommended to see others functional form aborted in the literature such as (Goldberger, 1968); Christensen *et al*. (1973); (Box & Cox, 1964); Halverson & Pollakowski (1981); Capozza *et al*. (1996); Capozza *et al*. (1997) and Triplett (2006).

2.5.6. Functional Form and Independent Variables

When it comes to the functional form and the independent variables to be used, Malpezzi *et al*. (1980) clearly explains that if the data allows it, i.e. if the practice does not result in problematic estimates for the models, it can be useful to carefully use dummy variables or indicators as independent variables, including also continuous variables, interaction variables and even quadratic or cubic variables and higher power variables. These practices, as mentioned above, are based not only on the scientific domain, but also on the art resulting from the practice of studies carried out on hedonic models. The reader is recommended to read the works of Harold (1964); Suits *et al*. (1978); Halvorsen & Palmquist (1980); Kennedy (1981) and Suits (1984).

2.6. MARKET AND SUBMARKET, DEFINITION

For hedonic models, knowledge of the concepts of what are considered markets or sub-markets is crucial, since dwellings are located in geographically identifiable and characterizable regions.

According to Struyk (1980) and Linneman (1981), a market is an entire nation, or a large region, or even a state. While Follain & Malpezzi (1980) and Malpezzi *et al*. (1980), define a market as a metropolitan area and these metropolitan areas are considered to be the labour market or the housing market.

Straszheim (1975); Gabriel (1984); Grigsby *et al*. (1985); Rothenberg *et al*. (1991); Maclennan & Tu (1996); Bourassa *et al*. (1999), and Vaz *et al*. (2023), they define the sub-market as those below the metropolitan level and can be segmented by housing location (city centre or suburb), housing quality level and even by the race or income level of the residents.

2.7. HEDONIC SPECIFICATION: ART OR SCIENCE?

With regard to the question of specifying hedonic models, Malpezzi (2003) explains that there is a combination of science and art in the choice of variables, functional form, markets and sub-markets. In the previous paragraphs, we also made clear the need to consult theoretical research and practical examples of the application of techniques that will minimise errors in obtaining problematic or even biased results in the estimation of hedonic models. In the literature by of Hausman (1978) and Burgess & Harmon (1991), the reader will also find examples to be replicated in various studies.

2.8. THE HEDONIC IMPUTATION APPROACH

Some literature has focused on explaining the hedonic imputation method based on examples of how it works, but Teixeira *et al*. (2019) more clearly defines the hedonic imputation method as one that, based on the hedonic model, uses the matching algorithm to relate house prices to the effects of period and characteristics, to assign a price to a house that is missing or new. And according to Triplett (2006), in the matching algorithm the price comparison is made when there is only a possible combination between the models.

2.8.1. Motivation

Briefly, the motivating question for the hedonic price imputation method lies in the fundamental methodology of index number matching, which keeps the quality of a good and service constant with the price for the equivalent model. Triplett (2006) defines an equivalent model as one in which price comparisons can only be made when there is an equiva lence.

2.8.2. Imputation model

We already discussed above that the main purpose in the imputation model is to make a use of the hedonic function to estimate one, or both, of the "missing" prices or new. With regards the imputation model techniques, we recommend the reader to see Triplett (2006) Teixeira *et al*. (2019) and Global Reporting Initiative (2023).

2.9. LIMITED DEPENDENT VARIABLES

When the dependent variable is in numerical intervals, it is important to bring a concept of truncation. In this session, the aim is to provide a definition of truncation, the truncated regression model or Tobit model and maximum likelihood estimation. Some literature distinguishes between truncation and censuring, in this study, we refer to both as truncation.

2.9.1. Truncation

A very common problem in microeconomic data is truncation of the response variable. According to Sedeeq & Meran (2022), truncation occurs when some observations of the response variable has been restricted from the sample. Greene (2003) says that truncation is essentially a characteristic of the distribution from which the sample data are drawn. When the response variable is truncated, values in a certain range are all transformed to (or reported as) a single value.

2.9.2. Truncated regression model (Tobit model)

Conventional regression methods fail to account for the qualitative difference between limit (zero) observations and non-limit (continuous) observations (Greene, 2003). A regression model is said to be-truncated when the recorded data on the dependent variable (the response) cut-off outside a certain-range with multiple observations at the-extreme of that range. The regression is obtained by making the mean in the preceding correspond to a classical regression model. When the data are truncated, variation in the observed response variable will underestimate the impact of the regression on the actual response variable (Chay & Powell, 2001). As a result, coefficient estimates from standard ordinary least squares (OLS) regression employing truncated data are often biased toward zero, therefore, Tobin proposed the first Tobit model in 1958 (Sedeeq & Meran, 2022).

Obtaining the equation for truncated regression models requires complex mathematical demonstrations developed by Tobin that allowed him to solve the difficulties that researchers faced when faced with a dependent variable with an interval nature. However, to avoid such a complex approach, we feel it is unnecessary to present it in this section, but we guarantee that in the methodology chapter below we will clearly demonstrate how it is applied. However, we recommend that the reader consult the literature in which we searched for the knowledge that enabled us to carry out our research, namely the works of Tobin (1958); Hausman & Wise (1977); Heckman (1979); Maddala (1983); Amemiya (1984); Shaw (1988); Greene (2003); Ahmed (2018) and Flaih *et al*. (2019).

2.9.3. Maximum Likelihood Estimation (MLE)

This study is concerned with estimating an equation based on data in which the response variable is only observed to fall in a certain interval on a continuous scale, its actual value remaining unobserved. The data are also truncated in the usual sense in that both end intervals are assumed to be open-ended[4].In the previous paragraph it was made clear that when the response variable is truncated, the ordinary least squares (OLS) estimate can involve assigning the midpoint to the observations in any group (possibly after transforming the variable). However, these methods do not generally result in a consistent estimate. According to Stewart (1983), see also Damásio & Mendonça (2019,2023), to overcome this problem of inconsistent estimates, it is recommended to opt for maximum likelihood estimation (MLE). In the next chapter we will show you how to apply it in practice, but the manuals referenced in this paragraph will serve as a basis for implementation.

[4] An important difference from the Tobit and related models is the absence of any completely observed data. This prevents direct extension of the methods of, for example, Heckman (1979) or C. Fair (1977) to this problem from being useful.

3. DATABASE CHARACTERIZATION

The aims of this section are to clarify the nature and provenance of the database used for the present research. However, the first step focus on the explanation of the database universe, in the second step will focus on the explanation of the criteria used for obtaining the universes of research, such as: i) Owners - database used for the imputation of the implicit rental prices, "shadow prices". ii) Renters - database used for the estimation of the implicit rental prices, "shadow prices" and estimation of effective rents. And finally, the third step focuses on the descriptive characterization of both research universes.

3.1. DATA UNIVERSE

The microdata on dwellings and buildings were obtained from Statistics Portugal (INE) and it is information referring to the last housing census carried out in the year 2021. The two microdata were aggregated to compose a total of 5,981,482 observations with 71 variables (see Table A.3, Appendix I). To obtain the below universes, we can find the restrictions, such as:

3.1.1. Owners

From the total of 5,981,482 observations among which constitute the joint database of dwellings and buildings[5] of the 2021 census, the total number of owners was identified by limiting the universe of census dwellings to those presenting the "family and classic" characteristics in the situation of "usual residence", occupied under the conditions of "ownership or co-ownership". The total number of owners is 2,900,093 entities.

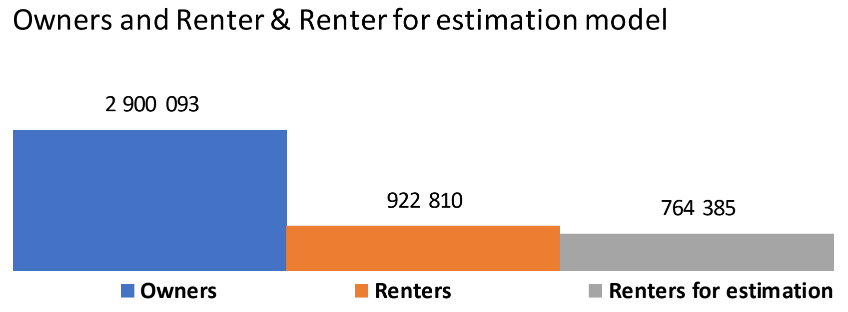
3.1.2. Renters

On the same joint database mentioned above, the universe of renters is composed of the total of census dwellings with intersection with all dwellings with the characteristics of "family and classic" in the situation of "usual residence", occupied under "renting or sub-renting" conditions. The total number of renters is 922,810 entities.

[5] Each record in the database corresponds to a census dwelling, to which the building variables were associated.

To serve as a basis for estimating the implicit rental prices, "shadow prices" and the effective rents, among the universe of renters made up of 922,810 renters, the present study consider only those whose contract has a "fixed or undetermined duration" and the owner of the accommodation is a "private or private entity or an ascendant or descendant" who is "renting the accommodation at a rate higher than 20 euros per month" and "does not have any support such as (social rent or supported rent, central government rent subsidy or municipal rent subsidy)", this total corresponds to 764,385 renters[6].

Figure 1 – Universe of owners and renters in total and for estimating the model.



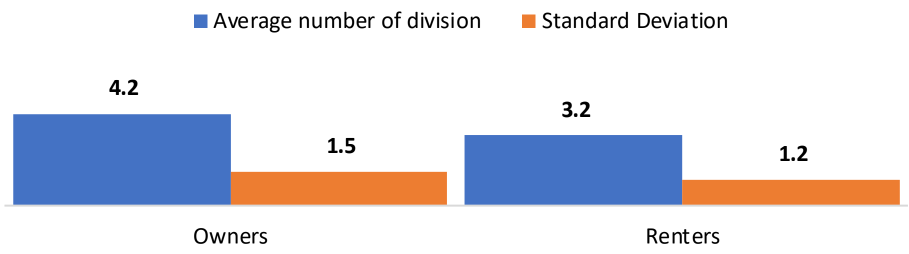
The above Figure 1 shows the universes considered in this research and in the procedures below, the aim is to characterize those universes with the purpose of analyzing the descriptive patterns presented by the data.

3.2. Characterization of the universes of owners and renters

Given that the universe of renters will be used to define the implicit rental prices, "shadow prices" and subsequently obtain the imputed rents of the universe of owners and effective rents from the same universe, it is important to characterize the respective universes and assess any differences. And the presented figures below, are intended to help us understand the characterization patterns of both universes.

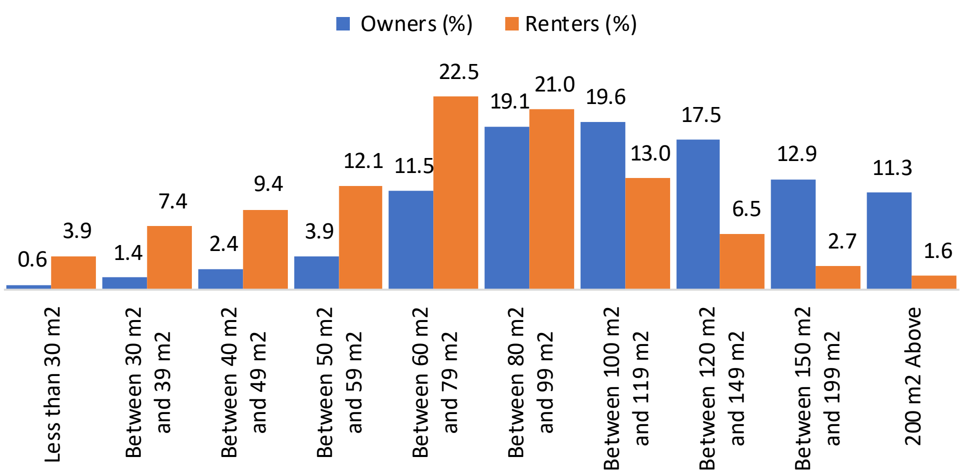
[6] These additional conditions are intended to bring the tenants' universe closer to market conditions, which landlords would have to resort to if they wanted to rent accommodation.

Figure 2 – Number of rooms in dwelling.



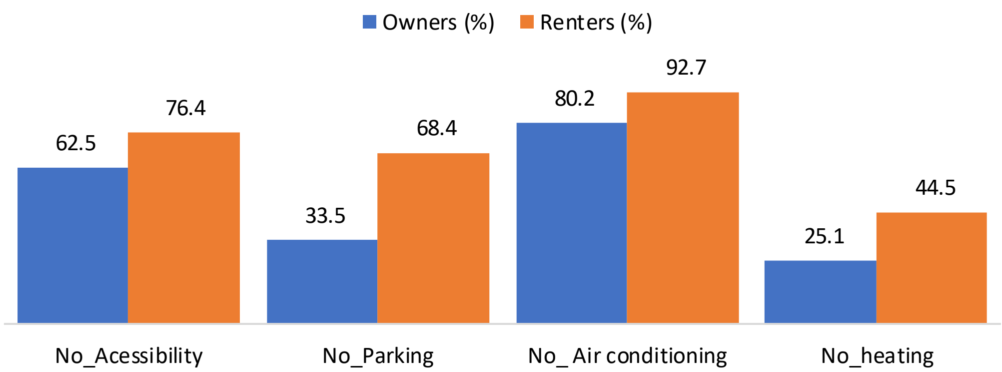
The Figure 2 above indicates that we have more rooms and with greater variability in the owner's dwellings than in the renters.

Figure 3 – Size of dwelling by floor area.



In terms of the floor area of the dwelling (Figure 3), the ranks with the smallest areas are the most representative in terms of sample coverage in the universe of renters.

Figure 4 – Access and comfort features of the accommodation.



About accessibility (Figure 4), most accommodation is not possible for a person using wheelchairs to enter independently for both owner and renter accommodation. Most of the renters' accommodation does not have car parking. As for air conditioning, most of the owners' and renters' lodgings do not have it. Most renters' accommodation does not have a heating system and they end up opting for other different types of heating.

Figure 5 – Buildings: type, construction period, state of conservation.

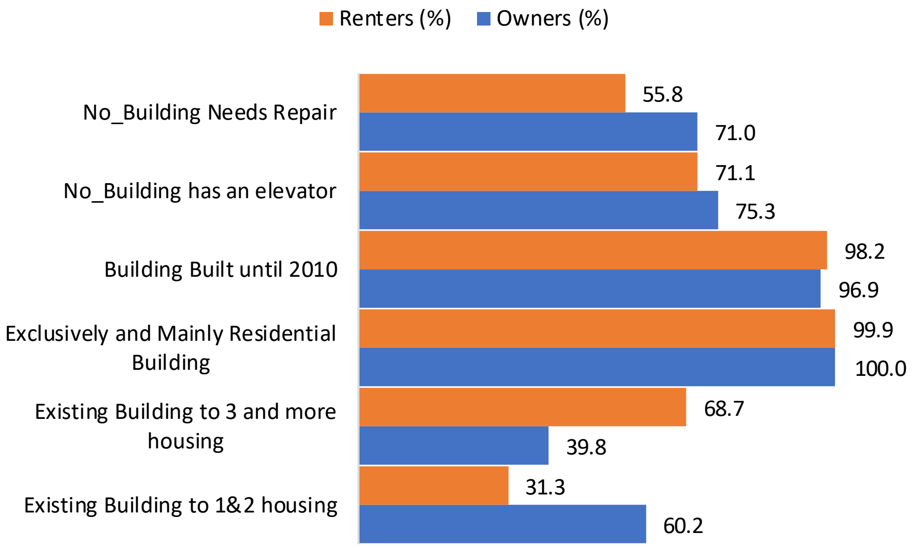


Figure 5 shows that most of the owners reside in buildings constructed for 1 or 2 dwellings while renters mostly reside in buildings constructed for 3 or more dwellings, i.e., owners reside in "detached houses " and renters reside in flats. Most owners and renters reside in exclusively residential buildings. Most of the buildings were constructed before 2011, i.e., they are old. For both owners and renters, most buildings do not have a lift. Most buildings do not require repair for both owner-occupied and tenant-occupied buildings.

Figure 6 – Derived indicators.

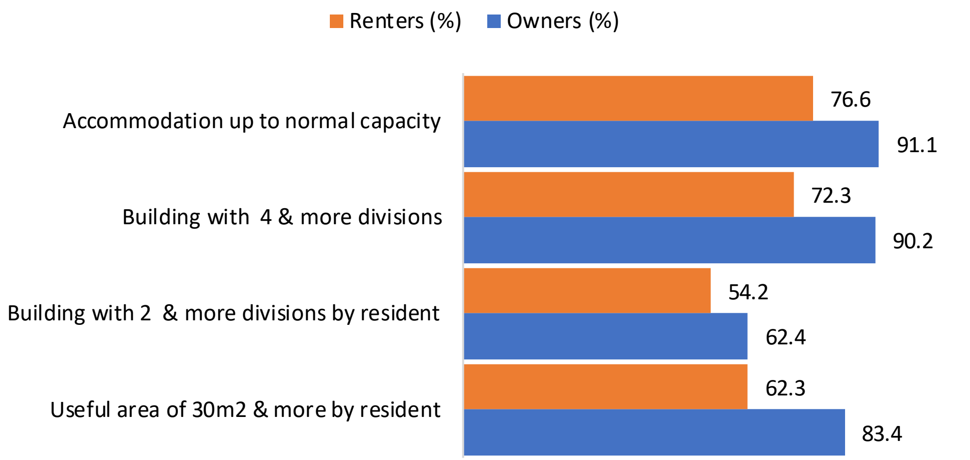


Figure 6 shows that most of the buildings for owners and renters have a floor area per resident of 30m2 and over. Most of the buildings have 2 or more rooms per resident for the owners and renters. Most of the buildings have 4 or more divisions specializing in both universes. And also, most of the accommodation has a normal occupancy rate in both universes.

Figure 7 – Lease agreement: season, owner entity and rent scale.

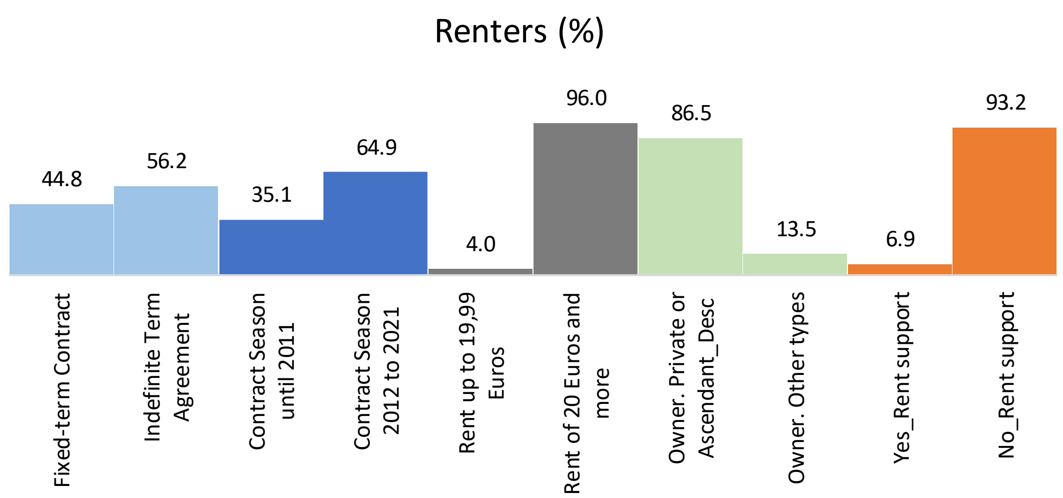


Figure 7 shows that there are more contracts concluded in the 2012 season until 2021 and most of the contracts are open-ended. There is an insignificant number of rents below 20 euros, which leads to excluding this universe from the object of this research. Most of the rented accommodation is owned by private individuals or private companies and their ascendants or descendants. A large proportion of renters do not have any type of rental support.

4. METHODOLOGY

In the previous session, our objective was to bring a comprehensive theoretical literature approach to bear on the theme of this study, however, in this session, we focus on providing a systematic and objective method in order to produce reliable and vertical knowledge in response to the specific objectives mentioned above.

4.1. HEDONIC METHOD

The hedonic regression method recognizes that heterogeneous goods can be described by their attributes or characteristics. That is, a good is essentially a bundle of (performance) characteristics[7]. In the housing context, this bundle may contain attributes of both the structure and the location of the properties. There is no market for characteristics, since they cannot be sold or even rented separately, so the prices of the characteristics are not independently observed. The demand and supply for the properties implicitly determine the characteristics' marginal contributions to the prices of the properties. Even though these prices are not observed, they can be estimated with hedonic regression techniques (Eurostat, 2013).

In the literature review, we aborted the five advantages by using log-linear functional form. In this study we use the log-linear form. The distribution of real estate rental prices is negatively skewed. The log-linear form reduces the impact of skewness in the price distribution and consequently, reduces heteroscedasticity. The log-linear form allows for parabolic relationships of the variables and for a multiplicative association between characteristics (State Secretariat for Economic Affairs, 2020).

The specification of a hedonic equation in the log-linear form is as follows:

$R_i^* = \beta' X_i + \varepsilon_i$. [1]

Where,

[7] The hedonic regression approach dates back at least to Court (1939) and Griliches (1961). Lancaster (1966) and Rosen (1974) laid down the conceptual foundations of the approach. P. F. Colwell & Dilmore (1999) argue that the first published hedonic study was a 1922 University of Minnesota master's thesis on agricultural land values.

$R_i^* = \log(R_i^*)$ is the natural logarithm of the unobserved rental price (latent variable).

i is each of the characteristics.

X_i is the vector of explanatory variables.

$\beta^{\prime}$ is the vector of unknown parameters.

ε_i is the error term.

$\varepsilon_i \sim N(0, \sigma^2)$ and $R_i^* | X_i \sim N(\beta^{\prime} X_i, \sigma^2)$.

Wooldridge *et al*. (2016) also defend that if the dependent variable is logarithmically transformed, the rental scales must also be logarithmically transformed.

R_i is observed rental price, where $R_i = k$, if $A_{k-1} \leq R_i^* \leq A_k$, where

A_{k-1} and A_k are known limits of the rent scales and ($k = 2, \ldots, 12$). Therefore:

$R_i = 2, \Rightarrow 20 \leq R_i^* < 49.99$ euros, then ($2.996 \leq log\ R_i^* < 3.912$)

$R_i = 3, \Rightarrow 50 \leq R_i^* < 74.99$ euros, then ($3.912 \leq log\ R_i^* < 4.317$)

$R_i = 4, \Rightarrow 75 \leq R_i^* < 99.99$ euros, then ($4.317 \leq log\ R_i^* < 4.605$)

$R_i = 5, \Rightarrow 100 \leq R_i^* < 149.99$ euros, then ($4.605 \leq log\ R_i^* < 5.011$)

$R_i = 6, \Rightarrow 150 \leq R_i^* < 199.99$ euros, then ($5.011 \leq log\ R_i^* < 5.298$)

$R_i = 7, \Rightarrow 200 \leq R_i^* < 299.99$ euros, then ($5.298 \leq log\ R_i^* < 5.704$)

$R_i = 8, \Rightarrow 300 \leq R_i^* < 399.99$ euros, then ($5.704 \leq log\ R_i^* < 5.991$)

$R_i = 9, \Rightarrow 400 \leq R_i^* < 499.99$ euros, then ($5.991 \leq log\ R_i^* < 6.215$)

$R_i = 10, \Rightarrow 500 \leq R_i^* < 649.99$ euros, then ($6.215 \leq log\ R_i^* < 6.477$)

$R_i = 11, \Rightarrow 650 \leq R_i^* < 999.99$ euros, then ($6.477 \leq log\ R_i^* < 6.908$)

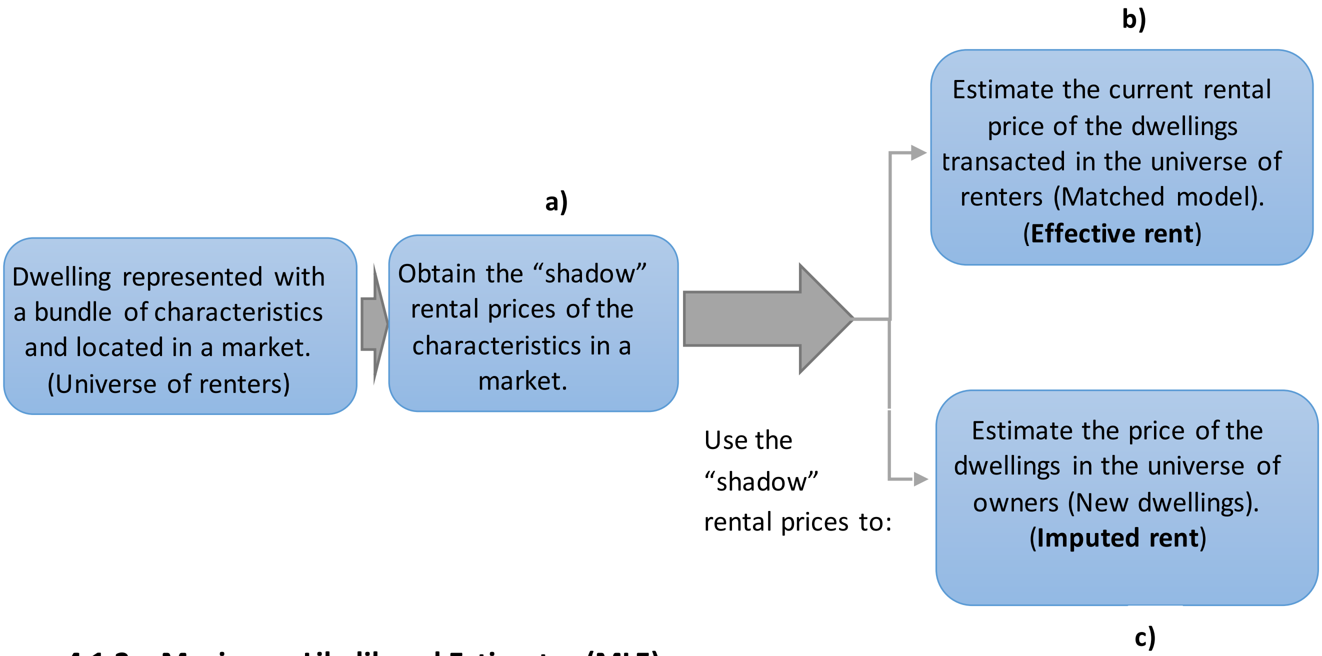
$R_i = 12, \Rightarrow R_i^* \geq 1000$ euros, then ($log\ R_i^* \geq 6.908$).

4.1.1. The imputation hedonic method

Data scientists may also want to consider using the imputations hedonic method to estimate the effective and imputed price rents. The imputations hedonic method imputes a rental price for each property or housing for each period or data universe. This method requires detailed information on housing characteristics, and it is generally appropriate if geospatial data (coordinates or location) are used (State Secretariat for Economic Affairs, 2020).

In this study, due to the nature of the data presented, which is interval rental prices, as described in the literature review, we will use the Maximum Likelihood Estimator (MLE) method to estimate: a) "shadow" rental prices, or the estimated coefficients $\hat{\beta}$; b) the effective rental price; and c) the imputed rental price. The below Figure 8 illustrate the mechanism of the imputation method in use.

Figure 8 – The mechanics of the imputation's hedonic method.



4.1.2. Maximum Likelihood Estimator (MLE).

Let´s consider the equation [1], by maximum likelihood estimation the equation will be given by:

$$L(\beta,\sigma) = \sum_{k} I[R_i = k]\left\{\Phi\left(\frac{\log(A_k) - \beta' X_i}{\sigma}\right) - \Phi\left(\frac{\log(A_{k-1}) - \beta' X_i}{\sigma}\right)\right\}$$

. Where,

$I[R_i = k]$ is the indicator function, assuming 1 when the expression inside [...] is true.

A_{k-1} and A_k are known limits of the rent scales.

Φ is the cumulative distribution function (cdf) of the standard normal respectively.

$Var[\varepsilon_i] = \sigma^2$.

In this context, through Stewart (1983) Maximum Likelihood Estimation (MLE) method, the consistent rental price estimation of the R_i^* is given by:

$$log\ (R_i^*) = E[R_i^* | A_{k-1} \leq R_i^* \leq A_k, X_i] = \beta^{\prime} X_i + \sigma \left[\frac{\emptyset\left(\frac{A_{k-1} - \beta^{\prime} X_i}{\sigma}\right) - \emptyset\left(\frac{A_k - \beta^{\prime} X_i}{\sigma}\right)}{\Phi\left(\frac{A_k - \beta^{\prime} X_i}{\sigma}\right) - \Phi\left(\frac{A_{k-1} - \beta^{\prime} X_i}{\sigma}\right)} \right].$$

Because R^* is unobserved, in practical we cannot estimate σ, therefore, the equation will be represented as below Stewart (1983):

a) **"shadow prices", or the estimated coefficients $\hat{\boldsymbol{\beta}}$**.

$E[R_i^* | A_{k-1} \leq R_i^* \leq A_k, X_i] = log\ (R_i^*) = \beta^{\prime} X_i$. [2]

From the equation [2] the "shadow" rental price or the estimated coefficients $\hat{\beta}$ will be imputed in the renter's universe to estimate the effective rental price (**matched model**), given by the below equation:

b) **Effective rental price**.

$R_i^{*\,Renter's\ Universe} = E[R_i^* | A_{k-1} \leq R_i^* \leq A_k, X_i] = \exp(\beta^{\prime} X_i)$. [3]

Finally, to estimate the (new rental price), the same coefficients from the equation [2] will be used to impute the owner's universe to estimate the imputed rental price (new dwelling), and the equation will be given by:

c) **Imputed rental price**.

$R_i^{*\,Owner's\ Universe} = E[R_i^* | A_{k-1} \leq R_i^* \leq A_k, X_i] = \exp(\beta^{\prime} X_i)$. [4]

5. RESULTS AND DISCUSSION

Six main points in this chapter we will present, such as: 1) - the construction of the final hedonic rental price model with respective characteristics variables and location; 2) - A statistical description of the variables included in the model; 3) – estimate the "shadow price", or the coefficients $\hat{\beta}$ and analyse the statistical significance; 4) – estimate the effective rental price; 5) – estimate the imputed rental price and 6) - Convergent algorithm of the hedonic rental price model.

5.1. FINAL HEDONIC RENTAL PRICE MODEL

$$R_i^* = E[R_i^*|X_i\,] = \beta_0 + \sum_{i=1}^{307} \beta_i REG_i + \sum_{i=308}^{337} \beta_i VAR_i$$

In which:

REG: are the 307 binary variables referring to the municipality where the dwelling is located.

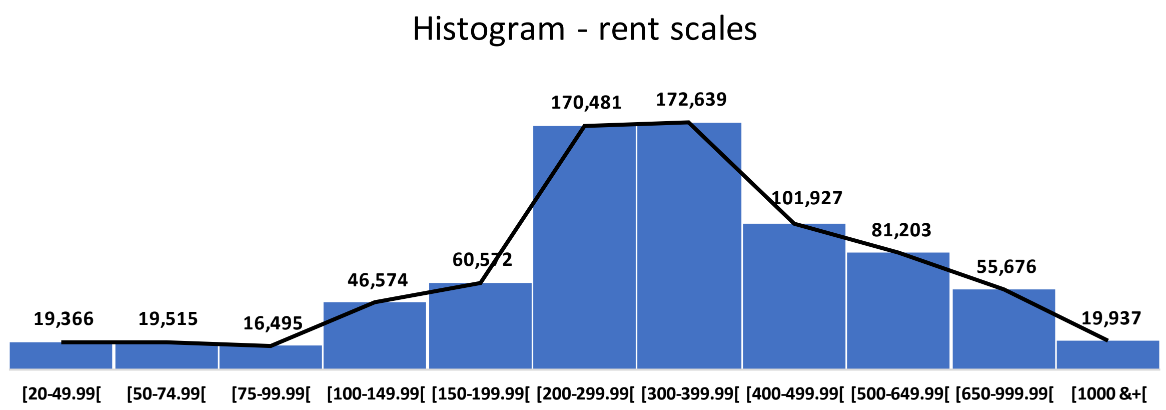
VAR: are the 30 variables associated with qualitative and quantitative dwelling characteristics.

Model restrictions:

As mentioned in the database characterization section, the model will consider rents whose contract has a "fixed or undetermined duration" and the owner of the accommodation is a "private or private entity or an ascendant or descendant" who is "renting the accommodation at a rate higher than 20 euros per month" and "does not have any support such as (social rent or supported rent, central government rent subsidy or municipal rent subsidy)", this total corresponds to about 764,385 renters.

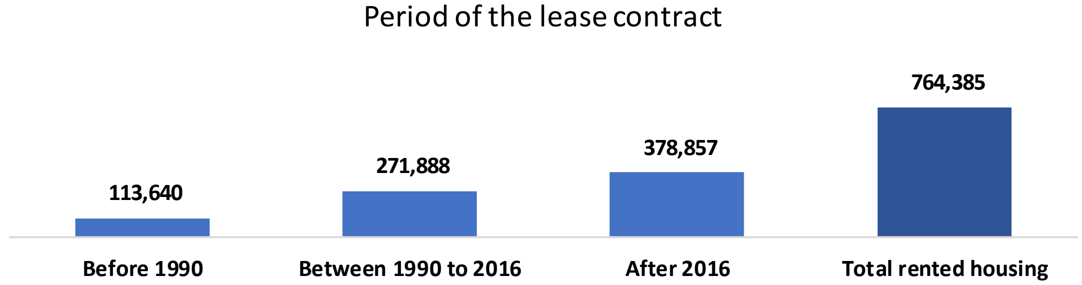
5.2. Descriptive Statistics

Figure 9 – Histogram of rent scales.



The above Figure 9, shows how the rental prices in the renter's universe are distributed. The histogram illustrates a left-skewed data distribution, meaning that we have more observations on the right side (large rental prices are in the upper interval), i.e., most housing has rental prices above or equal to 200 euros. Also, most frequent rental prices are in the interval of 200 euros to 399.99 euros.

Figure 10 – Period of the lease contract.



The bar diagram (Figure 10) above, illustrates the number of rental contracts through the period. Although, the figure shows the long bar in the period after 2016, until before 2016 the total number of contracts is about 385,528 (50.4%) and in the period after 2016 the total number is about 378,857 (49.6%) of contracts celebrations.

Table 1: Descriptive statistics of the variables associated with quantitative dwelling characteristics included in the model.

	Renters		Owners	
Dummy Variable	Mean	Std Dev	Mean	Std Dev
DIVISOES	3.16	1.20	4.16	1.55
SQRT_NDIVISOES	1.75	0.33	2.01	0.36

Renters	N	764,385
Owners	N	2,900,093

The above Table 1, shows that on average, owner-occupied housing is "bigger" than rented housing (4 against 3 rooms).

Table 2: Descriptive statistics of the variables associated with qualitative dwelling characteristics included in the model.

	Renters		Owners	
Dummy Variable	Mean	Std Dev	Mean	Std Dev
DEDIFEPOCA2	0.07	0.25	0.03	0.18
DEDIFEPOCA3	0.13	0.33	0.06	0.24
DEDIFEPOCA4	0.33	0.47	0.25	0.43
DEDIFEPOCA5	0.15	0.36	0.19	0.39
DEDIFEPOCA6	0.15	0.36	0.22	0.41
DEDIFEPOCA7	0.07	0.26	0.12	0.33
DEDIFEPOCA8	0.05	0.21	0.08	0.27
DEDIFEPOCA9	0.01	0.09	0.01	0.11
DEDIFEPOCA10	0.01	0.10	0.02	0.13
DEDIFELEV	0.30	0.46	0.25	0.43
DEDIFNECREP	0.42	0.49	0.29	0.45
DAREA2	0.07	0.26	0.01	0.12
DAREA3	0.09	0.29	0.02	0.15
DAREA4	0.12	0.32	0.04	0.19
DAREA5	0.22	0.42	0.11	0.32
DAREA6	0.21	0.41	0.19	0.39
DAREA7	0.13	0.34	0.20	0.40
DAREA8	0.07	0.25	0.17	0.38
DAREA9	0.03	0.17	0.13	0.34
DAREA10	0.02	0.13	0.11	0.32
DESTACION	0.35	0.48	0.67	0.47
DARCOND	0.08	0.27	0.20	0.40
DBOALOC	0.11	0.31	0.05	0.21
DBOALOCMOR	0.01	0.09	0.01	0.08
DAREAGRMOR	0.03	0.16	0.20	0.40
DNECREPMOR	0.15	0.35	0.16	0.37
DEDIFMOR	0.33	0.47	0.60	0.49
DARCMOR	0.02	0.15	0.11	0.32

The Table 2 above, shows that differences are more in qualitative characteristics that are less common in the housing stock (e.g., existence of air conditioning equipment, parking). Detached houses are more frequent in owner-occupied housing than in rented housing (60%

against 33%). On the other hand, the most frequent area range for the owners is 80 m2 or more while the renters live in dwellings with ranges from 40 m2 to 99 m2, that is, the renters live in smaller dwellings. And, the most frequent construction periods for owner-occupied dwellings are from 1961 to 2005 and for rented dwellings from before 1961 to 2000, i.e., renters live in old dwelling.

5.3. COEFFICIENT ESTIMATES AND STATISTICAL SIGNIFICANCE

Table 3: Coefficient estimates of the variables associated with qualitative and quantitative dwelling characteristics included in the model[8].

Truncated Ordered Probit Model (Tobit Model): Coefficient estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
CONSTANT β_0	1	5.510671	0.021586	255.29	<.0001
DBOALOC	1	0.272077	0.006173	44.07	<.0001
DEDIFEPOCA2	1	0.084148	0.007435	11.32	<.0001
DEDIFEPOCA3	1	0.136347	0.006864	19.86	<.0001
DEDIFEPOCA4	1	0.215400	0.006506	33.11	<.0001
DEDIFEPOCA5	1	0.368181	0.007010	52.53	<.0001
DEDIFEPOCA6	1	0.409407	0.007202	56.85	<.0001
DEDIFEPOCA7	1	0.448869	0.007989	56.19	<.0001
DEDIFEPOCA8	1	0.507241	0.008683	58.42	<.0001
DEDIFEPOCA9	1	0.568813	0.014001	40.63	<.0001
DEDIFEPOCA10	1	0.662080	0.013812	47.93	<.0001
DEDIFNECREP	1	-0.085805	0.002955	-29.03	<.0001
DAREA2	1	0.049455	0.007389	6.69	<.0001
DAREA3	1	0.073132	0.007177	10.19	<.0001
DAREA4	1	0.113461	0.007007	16.19	<.0001
DAREA5	1	0.171521	0.006750	25.41	<.0001
DAREA6	1	0.242820	0.006927	35.06	<.0001
DAREA7	1	0.314335	0.007333	42.87	<.0001
DAREA8	1	0.380439	0.008076	47.11	<.0001
DAREA9	1	0.432847	0.011224	38.57	<.0001
DAREA10	1	0.497215	0.013853	35.89	<.0001
DIVISOES	1	0.067174	0.006271	10.71	<.0001
SQRT NDIVISOES	1	-0.123574	0.022708	-5.44	<.0001
DARCOND	1	0.175281	0.005319	32.95	<.0001
DESTACION	1	0.189889	0.002980	63.72	<.0001
DEDIFELEV	1	0.188424	0.003170	59.45	<.0001
DEDIFMOR	1	-0.134832	0.003943	-34.20	<.0001
DAREAGRMOR	1	0.098550	0.011618	8.48	<.0001
DBOALOCMOR	1	-0.041858	0.014656	-2.86	0.0043
DNECREPMOR	1	-0.127132	0.005033	-25.26	<.0001
DARCMOR	1	0.030512	0.009443	3.23	0.0012

[8] (See Table A.1, Appendix I) the dummy variables description and in (Table A.4, Appendix I) there is a table with coefficient estimates of the location dummy variables and in (Figure A.1, Appendix I) there is a SAS Proc QLIM command with the interval known limits.

Table 3 above, refers to the estimated coefficients of the variables associated with the qualitative and quantitative characteristics of housing and in the (Table A.4, Appendix I) there is a table of the estimated coefficients of the location variables. In general, all the variables have the expected signs, and the coefficients of the variables associated with the qualitative and quantitative characteristics are statistically significant[9], as well as the coefficients of the location variables, apart from the coefficients of the locations referring to: Aljezur, Vila do Bispo and Santa Cruz da Graciosa (see Table A.4, Appendix I).

All statistically significant coefficients were maintained in the final model and although there were three location coefficients that were not statistically significant, after being removed from the initial model, the results obtained presented insignificant differences, however, it was decided to keep them for the interest of the researcher, since they play a fundamental role in the location of the lodging.

The overall model was convergent (see Figure A.2, Appendix I), and the addition of the interaction variables (DBOALOCMOR, DNECREPMOR and DARCMOR), and the square root variable (SQRT_NDIVISOES), proved to be statistical significance (see Table 3).

5.4. EFFECTIVE RENTAL PRICE ESTIMATION

Given the equation [3], the final equation to estimate the effective rental price is presented by:

$$est\ R_i^{*Renter's\ Universe} = \exp\{\beta_0 + \sum_{i=1}^{307} \beta_i REG_i + \sum_{i=308}^{337} \beta_i VAR_i\}$$

[9] All statistics test in this study were developed considering 5% significance level.

Therefore, the below Table 4 present the estimation results[10]:

Table 4: General results by region of effective rental price.

Region	Sample Size	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Std Dev
PORTUGAL	**764,385**	**71.19**	**224.20**	**300.79**	**404.57**	**2,504.10**	**165.12**
AZORES	8,520	99.93	236.91	299.50	388.37	1,135.26	137.43
ALENTEJO	41,295	81.88	184.59	241.65	318.30	1,052.35	107.81
ALGARVE	38,142	114.59	269.04	350.42	457.00	1,462.62	153.19
CENTER	128,324	74.68	204.38	272.61	365.89	1,277.46	120.57
LISBON	287,978	113.35	274.75	351.85	467.07	2,504.10	194.07
MADEIRA	10,034	104.13	227.54	324.40	449.19	1,178.37	142.42
NORTH	250,092	71.19	185.89	255.63	354.06	1,731.52	124.52

The results presented in Table 4 shows that of the total 764,385 rented dwellings in PORTUGAL, the municipality of LISBON presents in sample terms, the highest number of rented dwellings with 287,978 dwellings, the second position is represented by NORTH region with 250,092 dwellings and the third and fourth position is occupied by CENTER with 128,324 dwellings and ALENTEJO with 41,295 dwellings respectively.

In terms of minimum rental value, in general, the minimum monthly rent registered is 71.19 euros. NORTH region presents the lowest rent value which in turn is equal to the minimum value of the country, this gives an indication that the minimum rent value was registered in NORTH. In the same Table 4, it is possible to verify that LISBON presents the highest minimum rent value, giving an indication that is possibly that LISBON is the one who presents the highest rent in the country.

It is possible to see in the same Table 4 that in PORTUGAL, 25% of the lodgings have a monthly rent of up to 224.20 euros and in LISBON, around 25% of the lodgings have a monthly rent of up to 274.75 euros and the second position is represented by ALGARVE with around 25% of the lodgings being rented at monthly amounts of up to 269.04 euros followed in third

[10] The estimation results of the models were developed by the SAS program, through the PROC QLIM command (QLIM -Qualitative Limited Dependent Variable Model).

position by AZORES, with around 25% of the lodgings being rented at amounts of up to 236.91 euros.

Another interesting piece of data in Table 4 is that 50% of the dwellings in PORTUGAL had rental values of up to 300.79 euros and also showing LISBON as the largest with around 50% of dwellings being rented with monthly values of up to 351.85 euros followed by ALGARVE with around 50% of dwellings being rented with monthly values of up to 350.42 euros and in third position is MADEIRA with around 50% of dwellings being rented with monthly values of up to 324.40 euros, respectively.

The data for the upper quartile of the same Table 4 shows that 75% of the accommodation in PORTUGAL has an average monthly rent of up to 404.57 euros and once again LISBON is in the first position with approximately 75% of the accommodation to be rented with monthly rents of up to 467.07 euros, followed by ALGARVE with 75% of the accommodation to be rented with monthly rents of up to 457.00 euros and in third position is MADEIRA with approximately 75% of the accommodation with monthly rents of up to 449.19 euros, respectively.

In terms of maximum rental value, it is important to highlight that Table 4 illustrates a maximum monthly rental value in the country of approximately 2,504.10 euros and this maximum value corresponds to that registered by LISBON and the second highest rental value is registered in NORTH with a maximum monthly rental of 1,731.52 euros and the third is ALGARVE with a maximum monthly rental of 1,462.62 euros, respectively.

Finally, Table 4 also illustrates a standard deviation value of the average monthly rent registered in PORTUGAL with around 165.12 euros, that is, the average variability around the average monthly rent in the country is 165.12 euros. This deviation shows the dispersion of the rents in the country, and it must be noted that LISBON, despite having the highest monthly rent in comparison to the other regions, has the highest standard deviation of the monthly rent with a value of approximately 194.07 euros, followed by ALGARVE with a standard deviation of 153.19 euros and in third position is MADEIRA with a standard deviation of 142.42 euros, respectively.

Table 5: Summary outputs of effective rental price.

Region	Sample Size	Average monthly price	Annual effective rental value (10^6 euros)
PORTUGAL	**764,385**	**334.57**	**3,068.88**
AZORES	8,520	331.02	33.84
ALENTEJO	41,295	263.67	130.66
ALGARVE	38,142	379.80	173.83
CENTER	128,324	295.67	455.30
LISBON	287,978	402.18	1,389.82
MADEIRA	10,034	345.82	41.64
NORTH	250,092	281.16	843.79

The results in Table 5, represent the summary of the average monthly rent for PORTUGAL and the respective regions represented by their sample sizes and the estimated annual values of the total effective rents and by regions. The same Table 5 illustrates that the annual value of the effective rent in PORTUGAL, based on the sample under study is 3,068.88 billion euros with monthly average price of 334.57 euros.

Figure 11 –Rental price contribution by region.

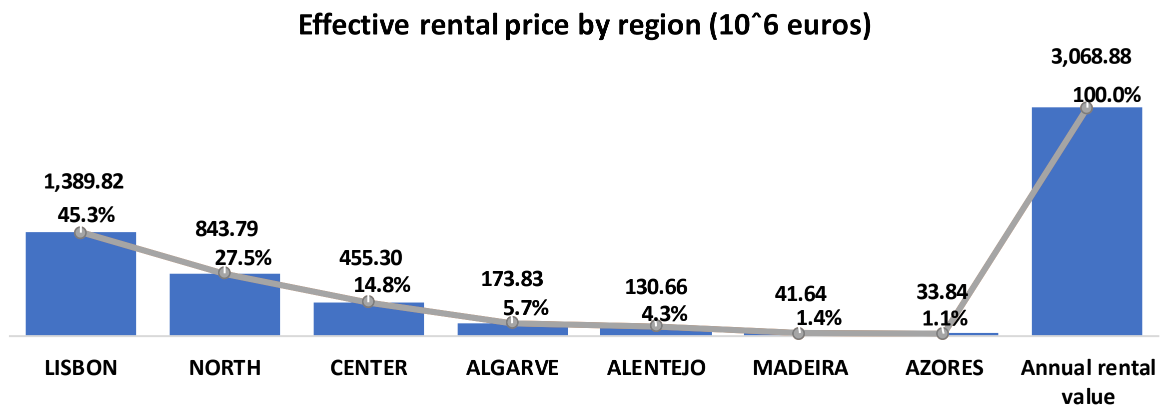


Figure 11 above, clearly shows the level of distribution of the effective rent by regions and the respective percentages. And as we can see, the first position is occupied by LISBON and the last is AZORES.

5.5. IMPUTED RENTAL PRICE ESTIMATION

Given the equation [4], the final equation to estimate the imputed rental price is presented by:

$$est\ R_i^{*Owner's\ Universe} = \exp\{\beta_0 + \sum_{i=1}^{307} \beta_i REG_i + \sum_{i=308}^{337} \beta_i VAR_i\}$$

Therefore, the below Table 6 present the estimation results:

Table 6: General results by region of imputed rental price.

Region	Sample Size	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Std Dev
PORTUGAL	**2,900,093**	**67.46**	**268.97**	**367.44**	**489.29**	**3,146.31**	**194.91**
AZORES	63,401	98.32	280.39	378.99	506.31	1,648.37	165.15
ALENTEJO	213,978	78.35	242.26	326.25	435.96	1,827.51	141.62
ALGARVE	129,301	96.61	332.58	445.30	595.13	1,888.27	188.01
CENTER	701,937	72.34	235.85	321.92	428.84	1,800.83	140.78
LISBON	758,598	117.80	358.28	470.82	646.21	3,146.31	244.55
MADEIRA	70,526	88.09	253.57	352.76	463.56	1,323.59	138.20
NORTH	962,352	67.46	244.35	331.94	433.96	2,356.93	142.25

In the effective rental's analysis, our research objective was centered on trying to answer the question related with the estimation of the current rental value of the dwellings that are on the market, represented by the variables associated with qualitative and quantitative characteristics, based on a dependent variable presented in scale intervals.

In the imputed rental's analysis, our object of investigation is centered on trying to answer the question related to the estimation of the value that the dwelling lived in by the owners may come to cost in the market, given that these dwellings have variables associated with the qualitative and quantitative characteristics, as well as location.

Notice that some of the dwellings have characteristics similar to those on the rental market, as well as there are recent dwellings in the owners' universe with characteristics that are also new, but it is important to stress that the dependent variable is not observed in this universe of owners, because these dwellings have never been rented before.

Based on the above explanation, below follows the interpretation of the results obtained in the estimation of imputed rents.

According to the rental imputation results presented in Table 6 above, we can see that the universe of owner-occupied dwellings is 2,900,093 dwellings in the whole of PORTUGAL. Within this universe, it can be seen from the same Table 6 that NORTH region presents in sample terms, the largest number of dwellings occupied by the owners with 962,352 dwellings, the second position is represented by LISBON with 758,598 dwellings and the third and fourth positions are occupied by CENTER with 701,937 dwellings and ALENTEJO with 213,978 dwellings, respectively.

In terms of minimum value in the imputation results, in general, the minimum monthly rent imputed is 67.46 euros (a lower value than that registered in the effective rent results). NORTH region presents the lowest rental value, which in turn is equal to the minimum value of the country, this indicates that the minimum value of the imputed rent was registered in NORTH region. In the same Table 6, it is possible to see that LISBON presents the highest of the minimum imputed rent values, giving an indication that is possibly to be the one with the highest imputed rent in the country, this indication is similar to the one registered in the analysis of the effective rents.

It is possible to see in the same Table 6 that in Portugal, for the imputation exercise, 25% of the owner-occupied dwellings have an imputed monthly rental of up to 268.97 euros (a figure that is higher than the one estimated in the effective rent) and LISBON occupies the first position with approximately 25% of the dwellings presenting a monthly imputed rent of up to 358.28 euros (a result also higher than that found in the effective rent), the second position is also represented by ALGARVE with around 25% of the accommodation with monthly imputed rent of up to 332.58 euros (higher than that found in the effective rent) and the third position is occupied by AZORES with around 25% of the accommodation with monthly imputed rent of up to 280.39 euros (higher than that found in the effective rent).

Another interesting data in Table 6 is that 50% of the dwellings of the owners in PORTUGAL have imputed rents of up to 367.44 euros (higher value found in the effective rentals results) and also presenting LISBON as the highest with around 50% of the dwellings with imputed monthly rents of up to 470.82 euros (also higher than that found in the effective rentals), followed ALGARVE with around 50% of the dwellings with monthly imputed rentals up to 445.30 euros (higher than the result found in the actual rentals) and in third position is AZORES with around 50% of the dwellings with monthly imputed rentals up to 378.99 euros (higher than the result found in the actual rentals).

The data for the top quartile in the same Table 6 shows that 75% of the accommodation in PORTUGAL has an average monthly imputed rent of up to 489.29 euros (higher than the effective rent) and once again LISBON occupies the first position with around 75% of the accommodation with monthly imputed rent of up to 646.21 euros (higher than the result found in the effective rent), followed by ALGARVE with 75% of the accommodation with monthly imputed rent of up to 595.13 euros (a result superior to that found in the effective rent) and third position occupied by AZORES with around 75% of the accommodation with monthly imputed rent of up to 506.31 euros (replacing the position of MADEIRA in the effective rent results but with a result superior to that found in the effective rent).

In terms of maximum imputed rent, it is important to highlight that Table 6 illustrates a maximum monthly imputed rent in the country of around 3,146.31 euros (higher than that found in the effective rent) and this maximum rent corresponds to that registered in LISBON (consistent position but with higher results than those found in the effective rent) and the second highest rent is registered in NORTH with a maximum monthly imputed rent of 2,356.93 euros (consistent position but with higher results than those found for actual rentals) and the third is ALGARVE with the maximum monthly imputed rent of 1,888.27 euros (also consistent position but with higher results than those found in the effective rent).

Finally, Table 6 also shows the standard deviation of the average monthly imputed rent in PORTUGAL for the sample of owners, which is 194.91 euros (standard deviation greater than that registered in the effective rent), that is, the average variability around the average monthly imputed rent in the country for the sample of owners is 194.91 euros. This deviation shows the dispersion of the imputed rent amount in the country, and it is worth noting that LISBON, despite of having the highest imputed monthly rental amount in comparison with the other regions, as can be seen in the effective rent, also has the highest standard deviation of the imputed rent, with a value of approximately 244.55 euros (higher than that found in the effective rent), this is followed by ALGARVE with a standard deviation of the imputed rent of approximately 188.01 euros (a consistent position but with results higher than those found in the effective rent) and in third position is AZORES with a standard deviation of the imputed rent of approximately 165.15 euros (replacing the position of MADEIRA, but with results higher than those found in the actual rent).

Table 7: Summary outputs of imputed rental price.

Region	Sample Size	Average monthly price	Annual imputed rental value (10^6 euros)
PORTUGAL	**2,900,093**	**402.96**	**14,023.36**
AZORES	63,401	407.63	310.13
ALENTEJO	213,978	350.27	899.40
ALGARVE	129,301	475.44	737.70
CENTER	701,937	342.33	2,883.53
LISBON	758,598	532.65	4,848.85
MADEIRA	70,526	364.97	308.88
NORTH	962,352	349.39	4,034.87

Table 7 represents the summary of the average monthly imputed rent for PORTUGAL and the respective regions represented by their sample sizes of the universe of landlords and the annual values of the total and per region imputed rents. The same Table 7 illustrates that the annual value of the imputed rent in PORTUGAL, based on the sample in study, is 14,023 billion euros with monthly average imputed price of 402.96 euros.

Figure 12 – Imputed price contribution by region.

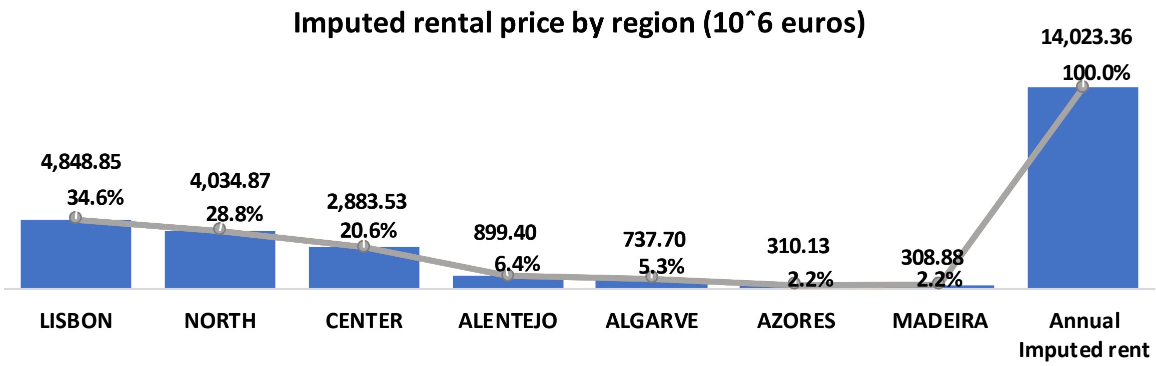


Figure 12 clearly shows the level of distribution of imputed rent by regions and the respective percentages. And as we can see, the first position is occupied by LISBON and the last is MADEIRA.

6. CONCLUSION, LIMITATIONS AND FURTHER RESEARCHS

The need to collect information on the prices of rented properties in the market to estimate the total value of current rent and then estimate the value of an owner-occupied dwelling, leads the National Institute of Statistics to combine efforts in the production of these results, since they are used for the purposes of harmonising the Gross National Income, which plays a fundamental role in the analysis and coordination of economic policies.

However, during the development of the research through the application of the method of imputation of hedonic price models, it is possible to conclude that the study was divided into the universes of owners with a total of 2,900,93 dwellings and renters utilised as the basis of estimation, which represent a total of 764,385 dwellings.

The owners live in housing with greater floor area than renters and most of accommodations has no accessibility for a person using wheelchairs to enter. In both owner and renters' accommodation most of the housing does not have air conditioning, lifts and was built before 2011. In the renter's universe, 50.4% (385,528 housings) of the housing contracts were celebrated before the year 2017 and the remaining 49.6% (378,857 housings) were celebrated from the year 2017 onwards.

The hedonic regression method recognizes that heterogeneous goods can be described by their attributes or characteristics and location. The functional forms used in the hedonic function is log-linear and the distribution of the real estate rental prices is negatively skewed. The dependent variable is interval rental price and the model used was Truncated Ordered Probit model (or Tobit model) with Maximum Likelihood Estimator ion (MLE) method.

In the hedonic function, the universe of renters was used to define the “shadow prices” or implicit rental prices and subsequently obtain the imputed rents of the universe of owners and effective rents from the same universe. All the coefficients associated with housing characteristics in the model have the expected sign and the overall Truncated Ordered Probit model was convergent.

The model includes interactions variables, squared roots variables and external variable of the coefficient of location. The average monthly rental price in the renter’s universe is 334.57 euros and the annual effective rental price estimation is 3,069 billion euros. In terms of region, LISBON is the first with about 45.3% (1,390 billion euros by year). For the imputed rental price, the average monthly rental price in the owner's universe is 402.96 euros and the annual imputed rental price is 14,023 billion euros. In terms of region, also LISBON is the first with about 34.6% (4,849 billion euros by year).

In terms of limitations, it should be noted that because the nature of our dependent variable is income in interval brackets, we faced technical difficulties in estimating the appropriate model and for this reason we opted to use the methodology of hedonic indices based on logistic regression and through maximum likelihood estimators.

For future studies, we recommend the use of the combination of alternative administrative information to obtain the global information of the dependent variable data, namely: 1- Model 2 of the stamp duty[11]; 2- The electronic receipts of income[12]; and finally, 3)- Model 44 of the IRS[13]. These sources combined allow us to obtain the total universe of data of the dependent variable rent values in the form of continuous data, so that it is possible to apply hedonic prices via multiple linear regression and through the ordinary least squares method (OLS).

We also recommend exploring alternative external databases for the addition of other independent variables such as age coefficient or quality coefficient and also the number of year that the family live in the house for better understanding of the average price obtained in the model. In this paper we use an external IMI database and include the independent

[11] Where the description of new lease contracts, or changes to existing contracts, is made available, with variables such as the period of the contract, the value of the rent, the identification of the article/accommodation, as well as the lessor and the tenant.

[12] Residential property owners who have their properties rented must issue and submit to the Portuguese Tax and Customs Authority (AT), an electronic receipt regarding the rental period. It is possible, through the variables included in the receipts, to identify, among other aspects, the lease agreement, the article/accommodation, the amount of the rent, or the parties to the agreement ('landlord' and 'tenant').

[13] This is related to the identification of all IRS taxpayers who may not issue electronic rent receipts, but who quantify the income arising from rental contracts earned in the previous year. In addition to identifying the landlord, the rented article/accommodation or the amount of the rents received, the tenants are also identified.

variable coefficient of location of (IMI) equal to or greater than 2. However, in addition to this variable, other sources of independent variables may be explored.

7. REFERENCES

Abraham, J. M., & Hendershott, P. H. (1996). Bubbles in Metropolitan Housing Markets. *Journal of Housing Research*, *7*(2), 191–208.

Ahmed, N. M. (2018). Limited dependent variable modelling (truncated and censored regression models) with application. *The Scientific Journal of Cihan University Sulaimanyia*, *2*(2), 82–96.

Ambrose, B. W., & Nourse, d H. O. (1993). Factors Influencing Capitalization Rates. *Journal of Real Estate Research*, *8*(2), 221–237.

Amemiya, T. (1980). Selection of Regressors. *International Economic Review*, 331–354.

Amemiya, T. (1984). Tobit models. *Journal of Econometrics*, *24*(1–2), 3–61.

Anas, A., & Eum, J. (1984). Hedonic Analysis of a Housing Market in Disequilibrium. *Journal of Urban Economics*, *15*.

Awan, K., Smee, O., & Whitehead, M. (1982). Household Attributes and the Demand for Private Rental Housing. *Economica*, *49*, 183–200.

Blomquist, G., & Worley, L. (1982). Specifying the Demand for Housing Characteristics: The Exogeniety Issue. *The Economics of Urban Amenities*.

Bourassa, S. C., Hamelink, F., Hoesli, M., & MacGregor, B. D. (1999). Defining Housing Submarkets. *Journal of Housing Economics*, 160–184.

Bowden, R. J. (1978). *The Econometrics of Disequilibrium*.

Boyle, M., & Kiel, K. (2001). A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities. *Journal of Real Estate Literature*, *9*(2), 117–144.

Burgess, J. F., & Harmon, O. R. (1991). Specification Tests in Hedonic Models. *Journal of Real Estate Finance and Economics*, *4*(4), 375–393.

Burridge, J. (1981). A Note on Maximum Likelihood Estimation for Regression Models Using Grouped Data ". *Journal of the Royal Statistical Society*, *43*(B), 41–45.

Butler, R. V. (1982). The Specification of Hedonic Indexes for Urban Housing. *Land Economics*, *58*(1), 96–108.

Capozza, D. R., Green, R. K., & Hendershott, P. H. (1996). *Taxes, Mortgage Borrowing and Residential Land Prices. In Henry Aaron and William Gale (eds.), Economic Effects of Fundamental Tax Reform*.

Capozza, D. R., Green, R. K., & Hendershott, P. H. (1997). *Income Taxes and House Prices*.

Chay, K. Y., & Powell, J. L. (2001). Semiparametric censored regression models. *Journal of Economic Perspectives*, *15*(4), 29–42.

Christensen, R., Jorgenson, D., & Lau, L. (1973). Transcendental Lograrithmic Production Frontiers. *Review of Economics and Statistics*, *55*, 28–45.

Chung, F., & Goldberger, A. (1984). *Proportional Projections in Limited Dependent Variable Models* (2nd ed., Vol. 52).

Colwell, P. (1999). Who Was First? An Examination of an Early Hedonic Study. *Land Economics*, *75*(4), 620–626.

Colwell, P. F., & Dilmore, G. (1999). Who was First? An Examination of an Early Hedonic Study. *Land Economics*, *75*(4), 620–626.

Court, A. (1939). *Hedonic Price Indexes With Automotive Examples. The Dynamics of Automobile Demand, New York, General Motors*.

Dempster, P., Laird, M., & Rubin, B. (1977). " Maximum Likelihood from Incomplete Data via the EM Algorithm ". *Journal of the Royal Statistical Society*, *39*(B), 1–38.

Diamond, D., & Smith, B. (1985). Simultaneity in the Market for Housing Characteristics. *Journal of Urban Economics*, *17*, 280–292.

Dreiman, M., & Follain, J. (2000). *Drawing Inferences about Housing Supply Elasticity from House Price Responses to Income Shocks*.

Dulberger, E. R. (1989). “The Application of a Hedonic Model to a Quality-Adjusted Price Index for Computer Processors”, in: Dale W. Jorgenson and Ralph Landau (eds.).

Technology and Capital Formation, Cambridge, MA: Massachusetts Institute of Technology Press, 37–75.

Dusse, N., & Jones, C. (1998). A hedonic price model of office rent. *Journal of Property Valuations and Investment*, *16*(3), 297–312.

Eurostat. (2013). *Handbook on Residential Property Prices Indices (RPPIs) Handbook on Residential Property Prices Indices (RPPIs) 2013 edition 2013 edition Handbook on Residential Property Prices Indices (RPPIs)*. http://publications.europa.eu/others/agents/index_en.htm.

Fair, C. (1977). " A Note on the Computation of the Tobit Estimator ". *Econometrica*, *45*, 1723–1727.

Fair, R. C., & Jaffee, D. M. (1972). *Methods of Estimation for Markets in Disequilibrium* (Econometrica, Vol. 40).

Flaih, A., Guardiola, J., Elsalloukh, H., & Akmyradov, C. (2019). Statistical inference on the ESEP Tobit regression model. *Journal of Statistics Applications and Probability Letters*, *6*(1), 1–9.

Follain, J., & Jimenez, E. (1985). The Demand for Housing Characteristics in Developing Countries. *Urban Studies*, *5*(22), 421–432.

Follain, J., & Malpezzi, S. (1980). Estimates of Housing Inflation for Thirty-Nine SMSAs: An Alternative to the Consumer Price Index. *Annals of Regional Science*, 41–56.

Follain, J., & Malpezzi, S. (1981). Another Look at Racial Differences in Housing Prices. *Urban Studies*, *18*(2), 195–203.

Freeman M. (1979). Hedonic prices, property values and measuring environmental benefits: A survey of the issues. *Scandinavian Journal of Economics*, *81*, 154–171.

Gabriel, A. (1984). A Note of Housing Market Segmentation in an Israeli Development Town. *Urban Studies*, 189–194.

Gatzlaff, D. H., & Haurin, D. R. (1997). Sample Selection Bias and Repeat-Sales Index Estimates. *Journal of Real Estate Finance and Economics*, *14*, 33–50.

Gillard, Q. (1981). *The effect of environment amenities on house values: The example of a view lot* (Vol. 33).

Goldberger, A. S. (1968). The Interpretation and Estimation of Cobb-Douglas Functions. *Econometrica*.

Goodman, A. (1998). Court and the Invention of Hedonic Price Analysis. *Journal of Urban Economics*, *44*(2), 291–308.

Goodman, J. L., & Ittner, J. B. (1992). The Accuracy of Home Owners' Estimates of House Value. *Journal of Housing Economics*, *2*(4), 339–357.

Greene, W. H. (2003). *Econometric analysis*. Prentice Hall.

GRI, E. (2023, September 12). *The most senior and private capital and equity marketplace for making real estate deals*.

Grigsby, W., Baratz, M., & Maclennan, D. (1985). *The Dynamics of Neighborhood Change and Decline*.

Griliches, Z. (1961). *Hedonic Price Indexes for Automobiles: An Econometric Analysis for Quality Change*. Harvard University Press.

Griliches, Z. (1971). On an Index of Quality Change. *Journal of the American Statistical Association*, *56*, 535–548.

Gross, D. J. (1986). Designing a Suitable Project: Integration of a Demand Module into a Supply Side Planning Model. *World Bank, Water Supply and Urban Development Department Discussion Paper No. UDD-103, 1986*.

Haas, G. (1922). Sales Prices as a Basis for Farm Land Appraisal. *The University of Minnesota Agricultural Experiment Station*.

Halverson, R., & Pollakowski, H. (1981). Choice of Functional Form for Hedonic Price Equations. *Journal of Urban Economics*, *1*(10), 37–49.

Halvorsen, R., & Palmquist, R. (1980). The Interpretation of Dummy Variables in Semilogrithmic Regressions. *American Economic Review*, *70*, 474–475.

Harold, W. (1964). *An Introduction to the Theory of Binary Variables (Or, All About Dummies)*.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 1251–1272.

Hausman, J. A., & Wise, D. A. (1977). SOCIAL EXPERIMENTATION, TRUNCATED DISTRIBUTIONS, AND EFFICIENT ESTIMATION . *Econometrica*, *45*(4), 919–938.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica*, *45*, 153–162.

Hicks, J. (1939). Value and Capital. *Oxford: Clarendon Press*.

Hicks, J. (1960). Linear Theory. *Economic Journal*, *70*(280), 671–709.

Hill, R. (2004). Constructing Price Indexes Across Countries and Time: The Case of the European Union. *American Economic Review*, *5*(94), 1379–1410.

Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, *32*, 1–49.

Kain, J. F., & Quigley, J. M. (1972). Note on Owners' Estimates of Housing Value. *Journal of the American Statistical Association*, *67*, 803–806.

Kennedy, P. E. (1981). *Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations*. *71*, 801–801.

King, A. T. (1975). *The Demand for Housing: Integrating the Roles of Journey-to-Work, Neighborhood Quality, and Prices. In Nestor Terlecky, Household Production and Consumption*.

Kumar, & Ranjit. (2011). *Research methodology a step-by-step guide for beginners*. www.sagepublications.com

Kwong, C. (2003). *A Critical Review of Literature on the Hedonic Price Model*. https://www.researchgate.net/publication/255726402.

Khoshnoud, M., Sirmans, G. S., & Zietz, E. N. (2023). The Evolution of Hedonic Pricing Models. In Journal of Real Estate Literature (Vol. 31, Issue 1, pp. 1–47). Taylor and Francis Ltd. https://doi.org/10.1080/09277544.2023.2201020.

Lancaster, J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*, 132–157.

Lancaster, J. (1971). Consumer Demand: A New Approach. *Columbia University Press*.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference With Nonexperimental Data* (Wiley).

Li, M., & Brown, J. (1980). *Micro-neighbourhood externalities and hedonic housing prices, Land Economics: Vol. 56 2*.

Linneman, P. (1981). The Demand for Residence Site Characteristics. *Journal of Urban Economics*, 129–148.

Maclennan, D. (1977). Some Thoughts on the Nature and Purpose of Hedonic Price Functions. *Urban Studies*, *14*, 59–71.

Maclennan, D., & Tu, Y. (1996). *Economic Perspectives on the Structure of Local Housing Systems* (3rd ed., Vol. 13).

Maddala, G. S. (1983). Limited Dependent and Qualitative Variables in Econometrics. *Cambridge University Press*.

Malpezzi, S. (1998). Welfare Analysis of Rent Control With Side Payments: A Natural Experiment in Cairo, Egypt. *Regional Science and Urban Economics*, *28*(2), 773–796.

Malpezzi, S. (2000). *Housing. In Margaret Grosh and Paul Glewwe (eds.), Designing Household Survey Questionnaires for Developing Countries: Lessons from Fifteen Years of the Living Standards Measurement Study*.

Malpezzi, S. (2003). *Hedonic Pricing Models: Selective and Applied Review*. 1–57.

Malpezzi, S., Ozanne, L., & Thibodeau, T. (1980). Characteristic Prices of Housing in 59 SMSAs. *The Urban Institute*.

Mayo, S. K. (1981). Theory and Estimation in the Economics of Housing Demand. *Journal of Urban Economics*, 95–116.

Mayo, S. K., & Gross, D. J. (1987). Sites and Services -- And Subsidies: The Economics of Low Cost Housing in Developing Countries. *World Bank Economic Review*, *1*(2), 301–335.

Mills, S., & Simenaue, R. (1996). New Hedonic Estimates of Regional Constant Quality Housing Prices. *Journal of Urban Economics*, *39*(2), 209–215.

Ohta, M., & Griliches, Z. (1975). *"Automobile Prices Revisited: Extensions of the Hedonic Price Hypothesis." In Household Production and Consumption. Studies in Income and Wealth* (N. Terleckyj, Ed.; Terleckyj, Vol. 40). University of Chicago Press, for the National Bureau of Economic Research.

Olsen, E. O. (1987). *The Demand and Supply of Housing Services: A Critical Review of the Empirical Literature. In Edwin S. Mills (ed.)* (Vol. 2). Elsevier.

Ozanne, L., & Malpezzi, S. (1985). The Efficacy of Hedonic Estimation with the Annual Housing Survey: Evidence from the Demand Experiment. *Journal of Economic and Social Measurement*, *13*(2), 153–1572.

Pakes, A. (2003). "A Reconsideration of Hedonic Price Indexes with an Application to PCs." *American Economic Review*, *93*(5), 1578–1596.

Phillips, R. S. (1988). Unraveling the Rent-Value Puzzle: An Empirical Investigation. *Urban Studies*, *25*, 487–496.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 35–55.

Rothenberg, J., Galster, G., Butler, R. V., & Pitkin, J. K. (1991). *The Maze of Urban Housing Markets: Theory, Evidence and Policy*.

SECO. (2020). *Residential Property Index*.

Sedeeq, B., & Meran, B. (2022, November 20). The Use of Tobit and Logistic Regression Models to Study Factors that Affect Blood Pressure in Cardiac Patients. *Cihanuniversity-Erbil*, 133–140.

Shaw, D. (1988). "On-Site Samples'' Regression: Problems of Non-Negative Integers, Truncation and Endogenous Stratification." *Journal of Econometric*, *37*, 211–223.

Shell, K., & Fisher, F. (1971). Taste and Quality Change in the Pure Theory of the True Cost of Living Index. *Harvard University Press*.

Sheppard, S. (1999). *Hedonic Analysis of Housing Markets* (Vol. 3). Elsevier.

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. In *Source: Journal of Real Estate Literature* (Vol. 13, Issue 1). https://www.jstor.org/stable/44103506.

Sirpal, R. (1994). Empirical modelling of the relative impacts of various sizes of shopping centres on the value of surrounding residential properties. *Journal of Real Estate Research*, *9*(4), 487–505.

Stewart, M. B. (1983). *On Least Squares Estimation when the Dependent Variable is Grouped* (Vol. 50, Issue 4). https://about.jstor.org/terms.

Straszheim, M. (1975). *An Econometric Analysis of the Urban Housing Market*.

Struyk, R. (1980). The Performance Funding System Inflation Factor: Analysis of Predictive Ability of Candidate Series and Models. *Urban Institute Working*, 1436–1502.

Suits, D. B. (1984). Dummy Variables: Mechanics vs. Interpretation. *Review of Economics and Statistics*, *66*, 177–180.

Suits, D. B., Mason, A., & Chan, L. (1978). Spline Functions Fitted by Standard Regression Methods. *Review of Economics and Statistics*, *60*, 132–139.

Taylor, L. O. (2008). Theoretical Foundations and Empirical Developments in Hedonic Modeling. In *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation* (pp. 1–12). Springer New York. https://doi.org/10.1007/978-0-387-76815-1_1.

Teekens, R., & Koerts, J. (1972). “Some Implications of the Log Transformation of Multiplicative Models.” *Econometrica*, *40*(5), 793–819.

Teixeira, A., Evangelista, R., & Moreira, H. (2019). On the derivation of hedonic rental price indexes for commercial properties : Rent price determinants , data quality problems and evidence from the use of administrative data for the Portuguese market. *Economics*.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*, 24–36.

Triplett, J. E. (2006). *Handbook on hedonic indexes and quality adjustments in price indexes : special application to information technology products*. OECD.

Vaz, E., Damásio, B., Bação, F., Shaker, R. R., & Penfound, E. (2023). Urban habitats and food insecurity: Lessons learned throughout a pandemic. *Habitat International*, *135*, 102779. https://doi.org/10.1016/J.HABITATINT.2023.102779.

Walden, L. (1990). Magnet schools and the differential impact of quality on residential property values. *Journal of Real Estate Research*, *5*, 221–230.

Wallace, H. (1926). Comparative Farmland Values in Iowa. *Journal of Land and Public Utility Economics*, 385–392.

Whitehead, C. M. E. (1999). *Urban Housing Markets: Theory and Policy. In Paul Chesire and Edwin S. Mills (eds.)* (Elsevier, Vol. 3).

Witte, A., Sumka, H., & Erekson, J. (1979). *An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen’s Theory of Implicit Markets.* (Econometrica, Vol. 47).

Wolynetz, S. (1979). " Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data ". *Applied Statistics*, *28*, 196–206.

Wooldridge, J. M., Brazil, A. •, Mexico, •, & Singapore, •. (2016). *Introductory econometrics*. www.cengage.com/highered.

APPENDIX I

Table A.1: Description of Dummy variables associated with qualitative and quantitative dwellings characteristics.

Dummy variables associated with qualitative and quantitative dwelling characteristics.	
Dummy variable	**Descriptive**
CONSTANT	Constant (β_0).
DBOALOC	Dummy that assumes value 1 when the location coefficient (IMI) is greater than or equal to 2.
DEDIFEPOCA2	Dummy that assumes value 1 if the building is built between 1919 and 1945.
DEDIFEPOCA3	Dummy assuming value 1 if the building was built between 1946 and 1960.
DEDIFEPOCA4	Dummy assuming value 1 if the building was built between 1961 and 1980.
DEDIFEPOCA5	Dummy assuming value 1 if the building was built between 1981 and 1990 .
DEDIFEPOCA6	Dummy assuming value 1 if the building was built between 1991 and 2000 .
DEDIFEPOCA7	Dummy assuming value 1 if the building was built between 2001 and 2005.
DEDIFEPOCA8	Dummy assuming value 1 if the building was built between 2006 and 2010.
DEDIFEPOCA9	Dummy assuming value 1 if the building was built during the period from 2011 to 2015 .
DEDIFEPOCA10	Dummy assuming value 1 if the building was built between 2016 and 2021.
DEDIFNECREP	Dummy that assumes value 1 when the building is in need of repair.
DAREA2	Dummy assuming value 1 if the dwelling has a floor area between 30 m2 and 39 m2 .
DAREA3	Dummy assuming value 1 if the dwelling has floor area between 40 m2 and 49 m2 .
DAREA4	Dummy with value 1 if the dwelling has floor area between 50 m2 and 59 m2 .
DAREA5	Dummy taking value 1 if the dwelling has floor area between 60 m2 and 79 m2 .
DAREA6	Dummy taking value 1 if the dwelling has floor area between 80 m2 and 99 m2 .
DAREA7	Dummy assuming value 1 if the dwelling has floor area between 100 m2 and 119 m2 .
DAREA8	Dummy assuming value 1 if the dwelling has floor area between 120 m2 and 149 m2 .
DAREA9	Dummy assuming value 1 if the dwelling has floor area 150 m2 to 199 m2 .
DAREA10	Dummy assuming value 1 if the dwelling has floor area of 200 m2 or more.
DIVISOES	Number of rooms in dwelling.
SQRT NDIVISOES	Square root of the number of rooms in the dwelling.
DARCOND	Dummy assuming value 1 if the dwelling has air conditioning.
DESTACION	Dummy assuming value 1 if the dwelling has parking facilities.
DEDIFELEV	Dummy assuming value 1 if the dwelling is served by a lift.
DEDIFMOR	Dummy assuming value 1 if the building where the dwelling is located has 1 or 2 dwelling units.
DAREAGRMOR	Dummy that takes value 1 when DEDIFMOR =1 and the floor area of the dwelling is 150 m2 or more.
DBOALOCMOR	Interaction dummy between DEDIFMOR and DBOALOC.
DNECREPMOR	Interaction dummy between DEDIFMOR and DEDIFNECREP.
DARCMOR	Interaction Dummy between DEDIFMOR and DARCOND.

Table A.2: Descriptions of 308 Dummy variables associated with dwellings location.

Dummy variables associated with location.		Dummy variables associated with location.		Dummy variables associated with location.	
Dummy varioable	**Descriptive**	**Dummy varioable**	**Descriptive**	**Dummy varioable**	**Descriptive**
D1	Águeda	D21	Almodôvar	D41	Guimarães
D2	Albergaria-a-Velha	D22	Alvito	D42	Póvoa de Lanhoso
D3	Anadia	D23	Barrancos	D43	Terras de Bouro
D4	Arouca	D24	Beja	D44	Vieira do Minho
D5	Aveiro	D25	Castro Verde	D45	Vila Nova de Famalicão
D6	Castelo de Paiva	D26	Cuba	D46	Vila Verde
D7	Espinho	D27	Ferreira do Alentejo	D47	Vizela
D8	Estarreja	D28	Mértola	D48	Alfândega da Fé
D9	Santa Maria da Feira	D29	Moura	D49	Bragança
D10	Ílhavo	D30	Odemira	D50	Carrazeda de Ansiães
D11	Mealhada	D31	Ourique	D51	Freixo de Espada à Cinta
D12	Murtosa	D32	Serpa	D52	Macedo de Cavaleiros
D13	Oliveira de Azeméis	D33	Vidigueira	D53	Miranda do Douro
D14	Oliveira do Bairro	D34	Amares	D54	Mirandela
D15	Ovar	D35	Barcelos	D55	Mogadouro
D16	São João da Madeira	D36	Braga	D56	Torre de Moncorvo
D17	Sever do Vouga	D37	Cabeceiras de Basto	D57	Vila Flor
D18	Vagos	D38	Celorico de Basto	D58	Vimioso
D19	Vale de Cambra	D39	Esposende	D59	Vinhais
D20	Aljustrel	D40	Fafe	D60	Belmonte

Dummy variables associated with location.		Dummy variables associated with location.		Dummy variables associated with location.	
Dummy varioable	**Descriptive**	**Dummy varioable**	**Descriptive**	**Dummy varioable**	**Descriptive**
D61	Castelo Branco	D81	Oliveira do Hospital	D101	Vila Viçosa
D62	Covilhã	D82	Pampilhosa da Serra	D102	Albufeira
D63	Fundão	D83	Penacova	D103	Alcoutim
D64	Idanha-a-Nova	D84	Penela	D104	Aljezur
D65	Oleiros	D85	Soure	D105	Castro Marim
D66	Penamacor	D86	Tábua	D106	Faro
D67	Proença-a-Nova	D87	Vila Nova de Poiares	D107	Lagoa
D68	Sertã	D88	Alandroal	D108	Lagos
D69	Vila de Rei	D89	Arraiolos	D109	Loulé
D70	Vila Velha de Ródão	D90	Borba	D110	Monchique
D71	Arganil	D91	Estremoz	D111	Olhão
D72	Cantanhede	D92	Évora	D112	Portimão
D73	Coimbra	D93	Montemor-o-Novo	D113	São Brás de Alportel
D74	Condeixa-a-Nova	D94	Mora	D114	Silves
D75	Figueira da Foz	D95	Mourão	D115	Tavira
D76	Góis	D96	Portel	D116	Vila do Bispo
D77	Lousã	D97	Redondo	D117	Vila Real de Santo Antóni
D78	Mira	D98	Reguengos de Monsara	D118	Aguiar da Beira
D79	Miranda do Corvo	D99	Vendas Novas	D119	Almeida
D80	Montemor-o-Velho	D100	Viana do Alentejo	D120	Celorico da Beira

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D121	Figueira de Castelo Rodrigo
D122	Fornos de Algodres
D123	Gouveia
D124	Guarda
D125	Manteigas
D126	Mêda
D127	Pinhel
D128	Sabugal
D129	Seia
D130	Trancoso
D131	Vila Nova de Foz Côa
D132	Alcobaça
D133	Alvaiázere
D134	Ansião
D135	Batalha
D136	Bombarral
D137	Caldas da Rainha
D138	Castanheira de Pêra
D139	Figueiró dos Vinhos
D140	Leiria

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D141	Marinha Grande
D142	Nazaré
D143	Óbidos
D144	Pedrógão Grande
D145	Peniche
D146	Pombal
D147	Porto de Mós
D148	Alenquer
D149	Arruda dos Vinhos
D150	Azambuja
D151	Cadaval
D152	Cascais
D154	Loures
D155	Lourinhã
D156	Mafra
D157	Oeiras
D158	Sintra
D159	Sobral de Monte Agraçc
D160	Torres Vedras

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D161	Vila Franca de Xira
D162	Amadora
D163	Odivelas
D164	Alter do Chão
D165	Arronches
D166	Avis
D167	Campo Maior
D168	Castelo de Vide
D169	Crato
D170	Elvas
D171	Fronteira
D172	Gavião
D173	Marvão
D174	Monforte
D175	Nisa
D176	Ponte de Sor
D177	Portalegre
D178	Sousel
D179	Amarante
D180	Baião

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D181	Felgueiras
D182	Gondomar
D183	Lousada
D184	Maia
D185	Marco de Canaveses
D186	Matosinhos
D187	Paços de Ferreira
D188	Paredes
D189	Penafiel
D190	Porto
D191	Póvoa de Varzim
D192	Santo Tirso
D193	Valongo
D194	Vila do Conde
D195	Vila Nova de Gaia
D196	Trofa
D197	Abrantes
D198	Alcanena
D199	Almeirim
D200	Alpiarça

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D201	Benavente
D202	Cartaxo
D203	Chamusca
D204	Constância
D205	Coruche
D206	Entroncamento
D207	Ferreira do Zêzere
D208	Golegã
D209	Mação
D210	Rio Maior
D211	Salvaterra de Magos
D212	Santarém
D213	Sardoal
D214	Tomar
D215	Torres Novas
D216	Vila Nova da Barquinha
D217	Ourém
D218	Alcácer do Sal
D219	Alcochete
D220	Almada

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D221	Barreiro
D222	Grândola
D223	Moita
D224	Montijo
D225	Palmela
D226	Santiago do Cacém
D227	Seixal
D228	Sesimbra
D229	Setúbal
D230	Sines
D231	Arcos de Valdevez
D232	Caminha
D233	Melgaço
D234	Monção
D235	Paredes de Coura
D236	Ponte da Barca
D237	Ponte de Lima
D238	Valença
D239	Viana do Castelo
D240	Vila Nova de Cerveira

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D241	Alijó
D242	Boticas
D243	Chaves
D244	Mesão Frio
D245	Mondim de Basto
D246	Montalegre
D247	Murça
D248	Peso da Régua
D249	Ribeira de Pena
D250	Sabrosa
D251	Santa Marta de Penaguião
D252	Valpaços
D253	Vila Pouca de Aguiar
D254	Vila Real
D255	Armamar
D256	Carregal do Sal
D257	Castro Daire
D258	Cinfães
D259	Lamego
D260	Mangualde

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D261	Moimenta da Beira
D262	Mortágua
D263	Nelas
D264	Oliveira de Frades
D265	Penalva do Castelo
D266	Penedono
D267	Resende
D268	Santa Comba Dão
D269	São João da Pesqueira
D270	São Pedro do Sul
D271	Sátão
D272	Sernancelhe
D273	Tabuaço
D274	Tarouca
D275	Tondela
D276	Vila Nova de Paiva
D277	Viseu
D278	Vouzela
D279	Calheta (R.A.M.)
D280	Câmara de Lobos

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D281	Funchal
D282	Machico
D283	Ponta do Sol
D284	Porto Moniz
D285	Ribeira Brava
D286	Santa Cruz
D287	Santana
D288	São Vicente
D289	Porto Santo
D290	Vila do Porto
D291	Lagoa (R.A.A.)
D292	Nordeste
D293	Ponta Delgada
D294	Povoação
D295	Ribeira Grande
D296	Vila Franca do Campo
D297	Angra do Heroísmo
D298	Vila da Praia da Vitória
D299	Santa Cruz da Graciosa
D300	Calheta (R.A.A.)

Dummy variables associated with location.	
Dummy varioable	**Descriptive**
D301	Velas
D302	Lajes do Pico
D303	Madalena
D304	São Roque do Pico
D305	Horta
D306	Lajes das Flores
D307	Santa Cruz das Flores
D308	Corvo

Table A.3: Census 2021 variables type and description.

Joint Database - Censo 2021			
Seq.	**Variable**	**Type**	**Descriptive**
1	MUNICIPIO	Char	Concatenated particle code from DT_COD and MN_COD. Basis of the split.
2	ALOJAMENTO	Char	Concatenated code of all particles from DT_COD to N_ALOJ
3	CENSO	Char	Year of the Census: 2021
4	ALOJAMENTO_CENSITARIO	Char	It identifies census dwellings: 1 - Census dwelling; 2 - Fictitious census dwelling; 3 - Non-census dwelling
5	DT_COD	Char	District code
6	MN_COD	Char	Municipality code
7	FR_COD	Char	Borough code
8	SEC_COD	Char	Section code
9	SS_COD	Char	Subsection code
10	N_EDIF	Char	Number of building within the subsection
11	N_ALOJ	Char	Number of accommodation in the building
12	N1_COD	Char	NUT1 Code
13	N2_COD	Char	NUT2 Code
14	N3_COD	Char	NUT3 code
15	NUTS1	Char	Region code NUT1
16	NUTS2	Char	Concatenated code of NUT1 and NUT2
17	NUTS3	Char	Concatenated code of NUT1, NUT2 and NUT3
18	FREGUESIA	Char	Concatenated code of all particles from DT_COD to FR_COD
19	SECCAO	Char	Concatenated code of all particles from DT_COD to SEC_COD
20	SUBSECCAO	Char	Concatenated code of all particles from DT_COD to SS_COD
21	EDIFICIO	Char	Concatenated code of all particles from DT_COD to N_EDIF
22	LG_COD	Char	Location Code
23	GRID_COD	Char	Internal GRID code used in the FNA
24	GRD_INSPIR	Char	Identification code for each cell, from the 2011 version of the European GRID (defined according to the EU Regular No. 1089/2010).
25	GRD_ID2021	Char	Identification code for each cell, from version 1.2 (16/03/2021) of the European Reference GRID.
26	GRD_ID2021_OFICIAL	Char	GRD_ID2021 code with the prefix _EN, required under point 3 of "Article 3 (Technical specifications of the 1 km2 reference grid) of COMMISSION IMPLEMENTING REGULATION (EU) 2018/1799"
27	FNA_EDIF_COD	Char	FNA Building Code
28	CANAL_RESPOSTA	Char	Response Channel
29	TALOJ	Char	Accommodation typology
30	TALOJFAM	Char	Type of family accommodation
31	OCUPFAM	Char	Housing situation of the family
32	TALOJCOL	Char	Type of collective accommodation
33	OCUPCOL	Char	Situation of collective accommodation
34	RESIDEM_PESSOAS	Char	People living in the dwelling
35	COLETIVO_ADP	Char	Private household (in collective dwelling)
36	COLETIVO_AI	Char	Institutional Aggregate
37	COLETIVO_TP_AGR	Char	Type of household (in collective dwelling)
38	NAGR_ADP_ALOJ	Num	Number of private households (ADP) in dwelling
39	NAGR_ALOJ	Num	Total number of households in the dwelling
40	NRES_ALOJ	Num	Number of residents in the dwelling

41	ESCAREA_ALOJ	Char	Code/class according to the floor area of the dwelling
42	NDIVISOES	Num	No. of rooms in the accommodation
43	ESTACIONAMENTO	Char	Parking space or garage
44	ACESSO	Char	Accessibility
45	AR_COND	Char	Air conditioning
46	AQUECIMENTO	Char	Most frequent type of heating
47	ANOS_OCUP	Num	Years of residence in the dwelling
48	COND_OCUP	Char	Accommodation occupancy conditions
49	ENCARGOS	Char	Financial charges
50	ESCENCARGOS	Char	Monthly scale of financial charges for accommodation
51	ARREND_DURACAO	Char	Duration of the rental contract
52	ARREND_EPOCA	Char	Season of the lease contract
53	ESCARREND	Char	Monthly rent scale for the dwelling
54	ARREND_APOIO	Char	Type of rent support
55	PROPR	Char	Entity that owns the accommodation (only those that do not belong to the owner)
56	DIM_AGLO	Char	Derivative: Size of agglomeration indicator
57	AREAU_RESID	Char	Derived: Code according to the floor area per resident
58	NDIV_ESP	Char	Derivative: Specialised divisions
59	NDIV_RESID	Char	Derived: Code according to the number of rooms per resident
60	TLOT	Char	Derivative: allotment index
61	ENT_PROP	Char	Derived: Proprietary entity
62	IND_TITULAR	Char	Code of the person who is the representative/owner of the accommodation (only for dwellings)
63	EDIFICIO_CENSITARIO	Char	"Allows the identification of classic census buildings: 1 - Classic census building ; 2 - Fictitious census building; 3 - Non-census building"
64	EST_CONST_EDIF	Char	Construction status of the building
65	NALOJ_EDIF	Num	Total number of actual census dwellings in the building
66	TEDIF	Char	Building type
67	TUTILEDIF	Char	Type of usage
68	EPOCA	Char	Construction period
69	NPAV	Num	Number of floors
70	ELEVADOR	Char	Lift
71	NECREPARACAO	Char	Repair requirements

Table A.4: Coefficient estimates of dwelling locations.

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D1	1	-0,522667	0,021393	-24,43	<.0001
D2	1	-0,549694	0,029595	-18,57	<.0001
D3	1	-0,524774	0,029137	-18,01	<.0001
D4	1	-0,570243	0,040969	-13,92	<.0001
D5	1	-0,333602	0,013043	-25,58	<.0001
D6	1	-0,751446	0,041065	-18.30	<.0001
D7	1	-0,393556	0,018995	-20,72	<.0001
D8	1	-0,481585	0,029371	-16.40	<.0001
D9	1	-0,612444	0,013215	-46,35	<.0001
D10	1	-0.354230	0,020964	-16.90	<.0001
D11	1	-0,457228	0,034632	-13.20	<.0001
D12	1	-0,459583	0,055323	-8,31	<.0001
D13	1	-0,627801	0,016695	-37.60	<.0001
D14	1	-0,537126	0,030109	-17,84	<.0001
D15	1	-0,461152	0,018368	-25,11	<.0001
D16	1	-0,594241	0,022432	-26,49	<.0001
D17	1	-0,587521	0.059260	-9,91	<.0001
D18	1	-0,440306	0,036741	-11,98	<.0001
D19	1	-0,644417	0,030874	-20,87	<.0001
D20	1	-0,412951	0,065484	-6,31	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D21	1	-0,491095	0,063984	-7,68	<.0001
D22	1	-0,504714	0,107095	-4,71	<.0001
D23	1	-0.798000	0,158949	-5,02	<.0001
D24	1	-0,375335	0,026396	-14,22	<.0001
D25	1	-0,261563	0,065659	-3,98	<.0001
D26	1	-0,601003	0,083389	-7,21	<.0001
D27	1	-0,505047	0,057004	-8,86	<.0001
D28	1	-0,502464	0,079624	-6,31	<.0001
D29	1	-0,522468	0,045392	-11,51	<.0001
D30	1	-0,219289	0,025068	-8,75	<.0001
D31	1	-0,514529	0,075369	-6,83	<.0001
D32	1	-0,497577	0,051926	-9,58	<.0001
D33	1	-0,456383	0.067140	-6.80	<.0001
D34	1	-0,679121	0,040132	-16,92	<.0001
D35	1	-0,528836	0.017510	-30.20	<.0001
D36	1	-0,563224	0,010044	-56,08	<.0001
D37	1	-0,809361	0.042720	-18,95	<.0001
D38	1	-0,877527	0,040663	-21,58	<.0001
D39	1	-0,378666	0,027896	-13,57	<.0001
D40	1	-0,766906	0.017860	-42,94	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D41	1	-0,708406	0,010387	-68.20	<.0001
D42	1	-0,728104	0,033244	-21.90	<.0001
D43	1	-0,658878	0,125063	-5,27	<.0001
D44	1	-0,678028	0,081576	-8,31	<.0001
D45	1	-0,631923	0,012974	-48,71	<.0001
D46	1	-0,657695	0,026903	-24,45	<.0001
D47	1	-0,782537	0,023082	-33.90	<.0001
D48	1	-0,761594	0,090751	-8,39	<.0001
D49	1	-0,653397	0,023094	-28,29	<.0001
D50	1	-0,714937	0,083088	-8.60	<.0001
D51	1	-0,690742	0,086984	-7,94	<.0001
D52	1	-0,846758	0,043319	-19,55	<.0001
D53	1	-0,796108	0,071687	-11,11	<.0001
D54	1	-0,625548	0,030422	-20,56	<.0001
D55	1	-0,722029	0,063067	-11,45	<.0001
D56	1	-0,740147	0,063381	-11,68	<.0001
D57	1	-0.744340	0.059290	-12,55	<.0001
D58	1	-0,697444	0,125015	-5,58	<.0001
D59	1	-0,780369	0,114951	-6,79	<.0001
D60	1	-0,835337	0,046491	-17,97	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D61	1	-0,616945	0,018062	-34,16	<.0001
D62	1	-0,711279	0.016450	-43,24	<.0001
D63	1	-0,680736	0,027414	-24,83	<.0001
D64	1	-0,699821	0,071144	-9,84	<.0001
D65	1	-0,717658	0,097591	-7,35	<.0001
D66	1	-0,774069	0,089975	-8.60	<.0001
D67	1	-0,815733	0,073693	-11,07	<.0001
D68	1	-0,669985	0,044137	-15,18	<.0001
D69	1	-0,645503	0,107988	-5,98	<.0001
D70	1	-0,497583	0,100585	-4,95	<.0001
D71	1	-0,644762	0,051685	-12,47	<.0001
D72	1	-0,491205	0,029078	-16,89	<.0001
D73	1	-0,421567	0,009802	-43,01	<.0001
D74	1	-0,390865	0,037379	-10,46	<.0001
D75	1	-0,422628	0,017844	-23,68	<.0001
D76	1	-0,651716	0,102939	-6,33	<.0001
D77	1	-0.720210	0,029135	-24,72	<.0001
D78	1	-0,486077	0.051450	-9,45	<.0001
D79	1	-0,654738	0,046931	-13,95	<.0001
D80	1	-0,511895	0,044696	-11,45	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates

Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D81	1	-0,744878	0,033068	-22,53	<.0001
D82	1	-0,654195	0,108027	-6,06	<.0001
D83	1	-0,663373	0,074381	-8,92	<.0001
D84	1	-0,542313	0,089178	-6,08	<.0001
D85	1	-0,459188	0,049675	-9,24	<.0001
D86	1	-0,719769	0,052122	-13,81	<.0001
D87	1	-0,688061	0,054713	-12,58	<.0001
D88	1	-0,762251	0,074895	-10,18	<.0001
D89	1	-0,536124	0,055037	-9,74	<.0001
D90	1	-0,686481	0,051257	-13,39	<.0001
D91	1	-0,569447	0,032138	-17,72	<.0001
D92	1	-0,252931	0,015893	-15,91	<.0001
D93	1	-0,495107	0,027665	-17.90	<.0001
D94	1	-0,644898	0,068583	-9.40	<.0001
D95	1	-0,655132	0.106890	-6,13	<.0001
D96	1	-0,619884	0,069183	-8,96	<.0001
D97	1	-0,549152	0,047338	-11.60	<.0001
D98	1	-0,413277	0,050161	-8,24	<.0001
D99	1	-0,463486	0,036558	-12,68	<.0001
D100	1	-0,516369	0,064053	-8,06	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates

Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D101	1	-0,692391	0,042917	-16,13	<.0001
D102	1	-0,087793	0,016723	-5,25	<.0001
D103	1	-0,523348	0,149957	-3,49	0,0005
D104	1	-0,045996	0,053463	-0,86	0,3896
D105	1	-0,341028	0,064048	-5,32	<.0001
D106	1	-0,250398	0.013150	-19,04	<.0001
D107	1	-0,233651	0,027329	-8,55	<.0001
D108	1	-0,131203	0,021539	-6,09	<.0001
D109	1	-0,196103	0,013715	-14.30	<.0001
D110	1	-0,428545	0,055428	-7,73	<.0001
D111	1	-0,381382	0,019087	-19,98	<.0001
D112	1	-0,271447	0,014934	-18,18	<.0001
D113	1	-0,222785	0,038929	-5,72	<.0001
D114	1	-0,312583	0,020742	-15,07	<.0001
D115	1	-0,211416	0,025712	-8,22	<.0001
D116	1	-0,072648	0,054738	-1,33	0,1844
D117	1	-0,392549	0,029688	-13,22	<.0001
D118	1	-0,848362	0,088819	-9,55	<.0001
D119	1	-0,775868	0,067702	-11,46	<.0001
D120	1	-0,784901	0,069206	-11,34	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates

Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D121	1	-0.732840	0,083463	-8,78	<.0001
D122	1	-0,834302	0,085638	-9,74	<.0001
D123	1	-0,844587	0,042198	-20,01	<.0001
D124	1	-0,638603	0,021006	-30.40	<.0001
D125	1	-0.737200	0,091831	-8,03	<.0001
D126	1	-0,815499	0,084148	-9,69	<.0001
D127	1	-0,863831	0,071618	-12,06	<.0001
D128	1	-0,770101	0,073188	-10,52	<.0001
D129	1	-0,827962	0,031021	-26,69	<.0001
D130	1	-0,769921	0,068116	-11.30	<.0001
D131	1	-0,730039	0,053886	-13,55	<.0001
D132	1	-0,492281	0,020006	-24,61	<.0001
D133	1	-0.717320	0,089533	-8,01	<.0001
D134	1	-0,613729	0,056498	-10,86	<.0001
D135	1	-0,408231	0,041185	-9,91	<.0001
D136	1	-0,481171	0,035541	-13,54	<.0001
D137	1	-0,383055	0,016165	-23.70	<.0001
D138	1	-0,723652	0,082568	-8,76	<.0001
D139	1	-0,685124	0,067977	-10,08	<.0001
D140	1	-0.427040	0,012664	-33,72	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates

Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D141	1	-0,487035	0,020875	-23,33	<.0001
D142	1	-0,440449	0,027978	-15,74	<.0001
D143	1	-0,317874	0,044025	-7,22	<.0001
D144	1	-0.539020	0,095849	-5,62	<.0001
D145	1	-0,412592	0,025013	-16.50	<.0001
D146	1	-0,510202	0,023376	-21,83	<.0001
D147	1	-0.611020	0,030424	-20,08	<.0001
D148	1	-0,369579	0,020096	-18,39	<.0001
D149	1	-0,304968	0,034507	-8,84	<.0001
D150	1	-0.330530	0,029698	-11,13	<.0001
D151	1	-0,548466	0,045346	-12.10	<.0001
D152	1	0,065492	0.008850	7.40	<.0001
D154	1	-0,227829	0,009048	-25,18	<.0001
D155	1	-0,389368	0,026169	-14,88	<.0001
D156	1	-0,121384	0,013903	-8,73	<.0001
D157	1	0,045733	0,009962	4,59	<.0001
D158	1	-0,151019	0,008204	-18,41	<.0001
D159	1	-0,289447	0,036549	-7,92	<.0001
D160	1	-0,244372	0,015004	-16,29	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D161	1	-0,298216	0,011268	-26,47	<.0001
D162	1	-0.195130	0.008980	-21,73	<.0001
D163	1	-0,176273	0,010235	-17,22	<.0001
D164	1	-0,666028	0,100918	-6.60	<.0001
D165	1	-0,670148	0,082408	-8,13	<.0001
D166	1	-0,701208	0,083765	-8,37	<.0001
D167	1	-0,497387	0.048010	-10,36	<.0001
D168	1	-0,449792	0,073535	-6,12	<.0001
D169	1	-0,593331	0,104076	-5.70	<.0001
D170	1	-0,617047	0,024928	-24,75	<.0001
D171	1	-0,626002	0,092758	-6,75	<.0001
D172	1	-0.561050	0,100368	-5,59	<.0001
D173	1	-0,794255	0,070108	-11,33	<.0001
D174	1	-0.716810	0,096831	-7.40	<.0001
D175	1	-0,562692	0,077886	-7,22	<.0001
D176	1	-0,564725	0,034088	-16,57	<.0001
D177	1	-0.573920	0,024975	-22,98	<.0001
D178	1	-0,698175	0,071228	-9.80	<.0001
D179	1	-0.815160	0,019298	-42,24	<.0001
D180	1	-0,762461	0,035821	-21,29	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D181	1	-0,827957	0,017244	-48,01	<.0001
D182	1	-0.534250	0.010800	-49,47	<.0001
D183	1	-0,747652	0,021155	-35,34	<.0001
D184	1	-0,387435	0,011636	-33.30	<.0001
D185	1	-0,840845	0,019959	-42,13	<.0001
D186	1	-0,324934	0.010220	-31,79	<.0001
D187	1	-0,772578	0,018285	-42,25	<.0001
D188	1	-0,693061	0,015484	-44,76	<.0001
D189	1	-0,745396	0,017993	-41,43	<.0001
D190	1	-0,182338	0,008378	-21,77	<.0001
D191	1	-0,407383	0,016522	-24,66	<.0001
D192	1	-0,682715	0,016229	-42,07	<.0001
D193	1	-0,465436	0,013073	-35.60	<.0001
D194	1	-0,400074	0,015674	-25,52	<.0001
D195	1	-0,430462	0.008650	-49,77	<.0001
D196	1	-0,527957	0,020666	-25,55	<.0001
D197	1	-0,648729	0,023713	-27,36	<.0001
D198	1	-0.598870	0,039521	-15,15	<.0001
D199	1	-0,509501	0,031199	-16,33	<.0001
D200	1	-0,495909	0,058024	-8,55	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D201	1	-0,326632	0,022111	-14,77	<.0001
D202	1	-0,417933	0,026945	-15,51	<.0001
D203	1	-0,710374	0,044747	-15,88	<.0001
D204	1	-0,737997	0,072971	-10,11	<.0001
D205	1	-0,627527	0,036999	-16,96	<.0001
D206	1	-0,534914	0,026842	-19,93	<.0001
D207	1	-0,575942	0,069283	-8,31	<.0001
D208	1	-0,515997	0,062052	-8,32	<.0001
D209	1	-0.584320	0,076799	-7,61	<.0001
D210	1	-0,470617	0,029019	-16,22	<.0001
D211	1	-0,417384	0,036816	-11,34	<.0001
D212	1	-0,474359	0,016652	-28,49	<.0001
D213	1	-0,688564	0,100966	-6,82	<.0001
D214	1	-0,457165	0,022152	-20,64	<.0001
D215	1	-0,505125	0,024219	-20,86	<.0001
D216	1	-0,733779	0,050647	-14,49	<.0001
D217	1	-0,616939	0,022602	-27.30	<.0001
D218	1	-0.536250	0,041334	-12,97	<.0001
D219	1	-0,226447	0,026303	-8,61	<.0001
D220	1	-0.218380	0,009677	-22,57	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D221	1	-0,337946	0,012757	-26,49	<.0001
D222	1	-0,271625	0,039211	-6,93	<.0001
D223	1	-0,391041	0,014555	-26,87	<.0001
D224	1	-0,290278	0,015381	-18,87	<.0001
D225	1	-0,293219	0,016632	-17,63	<.0001
D226	1	-0,283639	0.028280	-10,03	<.0001
D227	1	-0,210759	0,011384	-18,51	<.0001
D228	1	-0,247374	0,018887	-13.10	<.0001
D229	1	-0,269183	0,011785	-22,84	<.0001
D230	1	-0,211824	0,034466	-6,15	<.0001
D231	1	-0,661496	0,035293	-18,74	<.0001
D232	1	-0,467418	0,031959	-14,63	<.0001
D233	1	-0,597011	0,074542	-8,01	<.0001
D234	1	-0,684209	0,038627	-17,71	<.0001
D235	1	-0,653444	0.069470	-9,41	<.0001
D236	1	-0,824765	0.047610	-17,32	<.0001
D237	1	-0,574612	0,027399	-20,97	<.0001
D238	1	-0,515647	0,034866	-14,79	<.0001
D239	1	-0,473478	0,014646	-32,33	<.0001
D240	1	-0,408485	0,052093	-7,84	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D241	1	-0,620019	0,060888	-10,18	<.0001
D242	1	-0,735299	0,135992	-5,41	<.0001
D243	1	-0,756544	0,023979	-31,55	<.0001
D244	1	-0,688776	0,085934	-8,02	<.0001
D245	1	-0,635823	0,066717	-9,53	<.0001
D246	1	-0,748172	0,108873	-6,87	<.0001
D247	1	-0,736773	0,092474	-7,97	<.0001
D248	1	-0,632425	0,034244	-18,47	<.0001
D249	1	-0,598164	0,107666	-5,56	<.0001
D250	1	-0,662357	0,106766	-6.20	<.0001
D251	1	-0,672011	0,094327	-7,12	<.0001
D252	1	-0.843590	0,048937	-17,24	<.0001
D253	1	-0,657092	0,060576	-10,85	<.0001
D254	1	-0,506447	0.020400	-24,83	<.0001
D255	1	-0,758023	0,088999	-8,52	<.0001
D256	1	-0,695398	0,060673	-11,46	<.0001
D257	1	-0,755376	0,066009	-11,44	<.0001
D258	1	-0,710896	0,040997	-17,34	<.0001
D259	1	-0,698845	0,030495	-22,92	<.0001
D260	1	-0,681218	0.035520	-19,18	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D261	1	-0,890673	0,057694	-15,44	<.0001
D262	1	-0,648705	0,060317	-10,75	<.0001
D263	1	-0,678285	0,044798	-15,14	<.0001
D264	1	-0,567846	0,056899	-9,98	<.0001
D265	1	-0,770916	0,089341	-8,63	<.0001
D266	1	-0,670783	0,145331	-4,62	<.0001
D267	1	-0,773432	0,059803	-12,93	<.0001
D268	1	-0,650575	0,057175	-11,38	<.0001
D269	1	-0,657752	0,073764	-8,92	<.0001
D270	1	-0,586862	0.042490	-13,81	<.0001
D271	1	-0,836613	0,063867	-13.10	<.0001
D272	1	-0,818644	0,107763	-7.60	<.0001
D273	1	-0,880105	0.074720	-11,78	<.0001
D274	1	-0,761982	0,063731	-11,96	<.0001
D275	1	-0,582876	0,036236	-16,09	<.0001
D276	1	-0,845691	0,097855	-8,64	<.0001
D277	1	-0,497316	0,013413	-37,08	<.0001
D278	1	-0,683181	0,084054	-8,13	<.0001
D279	1	-0,416861	0,061725	-6,75	<.0001
D280	1	-0,391012	0,037654	-10,38	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D281	1	-0,367075	0,013901	-26,41	<.0001
D282	1	-0,323368	0,056562	-5,72	<.0001
D283	1	-0,311811	0,064542	-4,83	<.0001
D284	1	-0,614158	0,129307	-4,75	<.0001
D285	1	-0,319976	0,066258	-4,83	<.0001
D286	1	-0.390250	0,028941	-13,48	<.0001
D287	1	-0,631621	0,124112	-5,09	<.0001
D288	1	-0,447573	0,099897	-4,48	<.0001
D289	1	-0.264360	0,065673	-4,03	<.0001
D290	1	-0.338060	0,065532	-5,16	<.0001
D291	1	-0,192154	0,054402	-3,53	0,0004
D292	1	-0,547169	0,084952	-6,44	<.0001
D293	1	-0,147297	0,020227	-7,28	<.0001
D294	1	-0,387423	0,066992	-5,78	<.0001
D295	1	-0,330196	0,037758	-8,74	<.0001
D296	1	-0,336132	0,066541	-5,05	<.0001
D297	1	-0,212915	0,027352	-7,78	<.0001
D298	1	-0,316915	0,040277	-7,87	<.0001
D299	1	-0,185742	0,105158	-1,77	0,0773
D300	1	-0,587369	0,086633	-6,78	<.0001

Ordered Probit Model (Tobit Model): Coefficients estimates					
Variables	DF	Parameter Estimates	Standard Error	t Value	Approx Pr > |t|
D301	1	-0,429462	0,070096	-6,13	<.0001
D302	1	-0,535269	0,089461	-5,98	<.0001
D303	1	-0,252266	0,057099	-4,42	<.0001
D304	1	-0,369636	0,095247	-3,88	0,0001
D305	1	-0,267889	0.041160	-6,51	<.0001
D306	1	-0,329426	0,125852	-2,62	0,0089
D307	1	-0.245650	0.108530	-2,26	0,0236
D308	1	-0,398324	0,183578	-2,17	0.0300

Figure A.1: The basic PROC QLIM algorithm of (Truncated Ordered Probit Model or Tobit Model) in SAS.

```
PROC QLIM DATA=RENDAS OUTEST= aaaa;
 MODEL yyyy = %INDEP / DISCRETE(D=PROBIT) LIMIT1=VARYING;
  RESTRICT  LIMIT1 = xxxx, _LIMIT2= xxxx, _LIMIT3= xxxx,
 _LIMIT4= xxxx, _LIMIT5= xxxx, _LIMIT6= xxxx, _LIMIT7= xxxx,
 _LIMIT8= xxxx, _LIMIT9= xxxx, _LIMIT10= xxxx;

RUN;
```

known limits of the rent scales

Truncated and censured
dependent variable rental price

Regarding the independent variables, the following were selected:

%macro *Indep*;

D1--D152 D154--D308 DBOALOC

DEDIFEPOCA2 DEDIFEPOCA3 DEDIFEPOCA4 DEDIFEPOCA5 DEDIFEPOCA6 DEDIFEPOCA7 DEDIFEPOCA8 DEDIFEPOCA9 DEDIFEPOCA10

DEDIFNECREP DAREA2--DAREA10 DIVISOES SQRT_NDIVISOES DARCOND DESTACION DEDIFELEV DEDIFMOR DAREAGRMOR DBOALOCMOR DNECREPMOR DARCMOR

%mend Indep;

The specified model features variables related to:

- location at municipality level.
- micro-location from IMI stock information.
- the time of construction of the buildings.
- Conservation status of the accommodation/buildings.
- useful area.

Figure A.2: Convergence algorithm of Truncated Ordered Probit Model.

Discrete Response Profile of ESCY		
Index	Value	Total Frequency
1	1	19366
2	2	19515
3	3	16495
4	4	46574
5	5	60572
6	6	170481
7	7	172639
8	8	101927
9	9	81203
10	10	55676
11	11	19937

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	ESCY
Number of Observations	764385
Log Likelihood	-1659882
Maximum Absolute Gradient	4.95774
Number of Iterations	280
Optimization Method	Dual Quasi-Newton
AIC	3320440
Schwarz Criterion	3324343

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	-103E3	2 * (LogL - LogL0)
Upper Bound of R (U)	3.22E6	- 2 * LogL0
Aldrich-Nelson	-0.155	R / (R+N)
Cragg-Uhler 1	-0.144	1 - exp(-R/N)
Cragg-Uhler 2	-0.146	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	-0.141	1 - (1-R/U)^(U/N)
Adjusted Estrella	-0.142	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	-0.032	R / U
Veall-Zimmermann	-0.192	(R * (U+N)) / (U * (R+N))
McKelvey-Zavoina	0.1642	
N = # of observations, K = # of regressors		

Algorithm converged.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa