

Dynamics of the real estate market: determining factors in the value of housing in Portugal

Maria J. Ferreira^{??}, Aldina Correia^{??,??}, Vanda Lima^{??}, Diogo Ribeiro^{??}

¹*CIICESI, ESTG, Politécnico do Porto Rua do Curral, Casa do Curral, Margaride, 4610–156 Felgueiras – Portugal*

²*MySense, UK*

^{a)}Corresponding author: aic@estg.ipp.pt
vlima@estg.ipp.pt, URL: <https://ciicesi.estg.ipp.pt/>
diogo.ribeiro@mysense.ai, URL: <https://www.mysense.ai>

Abstract. This study provides decision-making support in the Portuguese housing sector, serving both individual and political stakeholders. For individuals—buyers and investors alike—it offers insights for informed purchasing and investment decisions. Politically, it guides municipal through to national policymakers on housing promotion, renewal, and rehabilitation strategies. Utilizing data from PORDATA for all 308 Portuguese municipalities, a multiple linear regression model identifies key factors influencing municipal housing values. These include location, population density, local purchasing power, family size, unemployment rates, environmental quality, population aging, education levels, civil status, homeowner ratios, and housing availability. The methodical approach aims to quantify how these variables collectively impact housing market dynamics at the municipal level.

Keywords: Housing Price, Portugal, Multiple Linear regression, Decision making, Housing policy, Data analysis.

INTRODUCTION

Housing is crucial for personal and societal stability, a fact underscored by the 2007 financial crisis which highlighted the connections between housing markets and economic stability. The crisis led to tougher credit, reduced liquidity, and economic distress, affecting homeownership and increasing foreclosures. Post-crisis, ensuring the right to housing against such shocks has become a policy focus.

Societal changes increase the demand for housing that meets modern standards and personal needs, including location, sustainability, and technological needs. This demand shapes market preferences influenced by demographic shifts, economic trends, cultural changes, and policy reforms. Such insights are vital for stakeholders to address housing issues, with preferences affecting property values and market dynamics.

The study focuses on how municipal characteristics in Portugal affect housing values, filling a research gap on local-level analysis. It assesses the influence of these traits on price trends and categorizes municipalities by average transaction values. This research aids buyers, investors, and policymakers with data for better decision-making and informs housing policy development.

FACTORS INFLUENCING THE VALUE OF HOUSING

This paper includes a bibliometric analysis conducted in January 2021, using the Web of Science database to explore the prevalence of “housing prices” as a research topic. The initial search yielded 2,369 records, which, after refining to only scholarly articles, resulted in 2,003 publications for review. The earliest article dates back to 1972, with a marked increase in interest observed from 1995 and a significant uptick in research activity in the early 21st century. This growing interest likely correlates with global housing market volatility and the disparity between housing prices and income growth. As of June 30, 2021, the number of articles published in 2021 alone reached 140, representing 6.272% of all publications on this topic to date. The authors contributing most prolifically to this field include Hui, Tsai, Gupta, Song, Wen, Glaeser, Su, Zhang, Chau, and Quigley, reflecting a diverse geographic interest spanning the USA, China,

Spain, South Korea, the UK, Taiwan, Australia, Canada, and Portugal. The paper also delves into the determinants of transactional housing values, highlighting contributions from both international and regional studies. A special focus is placed on municipal-level research, which remains sparse, suggesting that this work provides both innovative and theoretical contributions. In summarizing the factors affecting housing prices, location emerges as a key determinant. It's noted that quality of life, influenced by factors such as proximity to natural resources and amenities, significantly impacts housing values. Comprehensive studies using hedonic pricing models and GIS for spatial adjustments offer insights into environmental and neighborhood attributes' effects on prices. For instance, the proximity to destinations, views, and air quality have been emphasized as pivotal in determining the willingness to pay for housing. The paper underscores the relevance of location in housing acquisition, evidenced by studies on Istanbul and Ankara's housing markets, and corroborates these findings with extensive data analysis across various Spanish provinces using STAR and GLM methodologies.

Zhou et al.[?] identified a significant positive correlation between housing prices and population density in 283 Chinese cities, positing that higher housing prices and denser populations are pivotal for increased productivity. Rappaport[?] highlighted the influence of population density on U.S. housing prices, observing various endogenous factors that co-vary with it, such as a slight decrease in wages, moderate house price increases, and sharp land price ascensions. High quality of living seems to be capitalized in land and house prices rather than wages. Calcagno et al. (2009) found that Italian household consumption correlates with housing capital gains, especially in younger households, suggesting different impacts of housing price increases on rental expectations. In Portugal, improved accessibility to housing markets contrasts with the community averages, despite a tendency towards higher effort rates, as indicated by Alves and Pinheiro (2017). They credit the job market's recovery for bolstering household incomes, thereby invigorating the real estate market. McGreal and Taltavull de La Paz (2013) emphasized income, accessibility, and structural features as pivotal to housing prices. Fuller et al. (2020) discuss the relationship between housing prices, wealth indices, and financial asset price changes in Western Europe, predicting a significant impact on future generations. According to INE, household mobility demands are shaped by changes in family size and demographic shifts, with Varão (2019) attributing house price variability to demographic factors such as immigration and birth rates. Archer et al. (2010) found family size to be a dominant factor affecting housing prices in the Chicago market. Johnes and Hyclak (1999) related average industry wages, unemployment rates, workforce strength, and average house prices in urban areas, finding interplays between local housing and labor markets, with unemployment and workforce strength affecting house prices in specific U.S. regions. These insights collectively formulate research hypotheses investigating the relationship between population density, purchasing power, family size, and unemployment with housing values, providing a multifaceted view of the housing market's undercurrents.

This review synthesizes findings from multiple studies on the impact of external factors on real estate values. Xiao et al. (2020) employed a hedonic pricing model and spatial econometric model, finding that noisy public square dancing in Hangzhou, China, depreciates nearby housing prices by 5.8% to 13%. In contrast, Zambrano-Monserrate's (2016) study in Machala, Ecuador, emphasized the significance of public services like water supply and waste collection for rental apartment markets. In Madrid, Chasco Yrigoyen and Sánchez Reyes (2012) observed that air pollution negatively affects housing prices across all quantiles, with noise pollution only impacting luxury homes' prices. Fitch Osuna et al. (2013) explored real estate dynamics in San Nicolás de los Garza, Mexico, noting an increase in housing prices in noisier areas, suggesting demographic and location preferences outweigh other factors. Ermisch (1996) highlighted the implications of aging demographics on housing demand in Britain, suggesting a substantial slowdown in growth rates due to population aging. Bayet et al. (1991) placed housing expenses as a primary concern for younger and older age groups in family budgets. Eichholtz and Lindenthal (2014) attributed high education, good health, and high income to sustained housing demand among aging populations in Europe. Conversely, Brasington et al. (2015) and Gyourko & Linneman (1996) found that higher educational attainment increases housing demand, suggesting a strong correlation between market values and educational levels. These diverse studies collectively support the hypothesis that environmental quality, demographic shifts, and educational levels significantly influence housing values. This integration of global research presents a nuanced understanding of real estate economics, highlighting the complex interplay between urban externalities and market behaviors. In his research conducted in China, Xu (2017) indicates that when the housing supply is at or near maximum capacity, there is increased pressure on supply, which can indeed drive up home prices. Katz and Rosen (1987) empirically demonstrated, using a large dataset on housing in the San Francisco Bay Area of California, that construction levies appear to have a significant impact on home prices. Their regression model's results suggest that home prices are between 17% and 38% higher in communities where growth moratoria and/or growth control plans are in place. The widespread use of such fees, in many communities, restricts the available supply response. The spread of these growth regulatory techniques to metropolitan areas outside

California may have substantial negative effects on housing affordability (Katz and Rosen 1987).

Theoretical Framework for Multivariate Linear Regression Using Ordinary Least Squares

Ordinary Least Squares Estimation

Maroco[?] refers to Ordinary Least Squares (OLS) estimation technique has the cornerstone for conducting multivariate linear regression analysis. It estimates the parameters of a linear model that predicts a single response variable from multiple explanatory variables. The primary goal is to minimize the sum of the squares of the differences between the observed dependent variable and those predicted by the linear model.

Model Specification

The general form of the multivariate linear regression model is expressed as:

$$\mathbf{Y} = \mathbf{X}\mathbf{\beta} + \boldsymbol{\epsilon} \quad (1)$$

where \mathbf{Y} is an $n \times 1$ vector of the dependent variable, \mathbf{X} is an $n \times k$ matrix of independent variables for n observations across k predictors, $\mathbf{\beta}$ is a $k \times 1$ vector of regression coefficients to be estimated, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of the residuals or errors of the model.

Estimation and Properties

The OLS estimator is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2)$$

This estimator is known to be the Best Linear Unbiased Estimator (BLUE), provided that the Gauss-Markov theorem's assumptions are satisfied. These include the absence of perfect multicollinearity, homoscedasticity, and uncorrelated error terms with a zero conditional mean.

Goodness-of-Fit

The fit of the regression model is usually quantified by the coefficient of determination R^2 , and the adjusted R^2 which provides a measure adjusted for the number of predictors. An F-test is used to determine the overall significance of the model parameters.

Assumptions Verification

Verifying the assumptions underlying the OLS estimator is crucial for the reliability of the model. This involves the examination of residuals for normality, constant variance (homoscedasticity), and independence. The Variance Inflation Factor (VIF) is used to assess the degree of multicollinearity among predictors.

Implications

The regression coefficients estimated from OLS regression reflect the average effect of one unit change in the independent variable on the dependent variable, with other variables held constant. The interpretation of these coefficients should consider the study design and data characteristics, especially since the presence of bias can arise from any violations of the OLS assumptions.

Generalization to Other Models

While OLS is traditionally applied to continuous and normally distributed response variables, the regression framework can be extended to models with non-normal response distributions or non-linear relationships. Such generalizations are termed Generalized Linear Models (GLM), which accommodate various types of data by linking a function of the mean of the response variable to the linear predictor through a link function.

The GLM is expressed as:

$$g(\mathbb{E}[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta} \quad (3)$$

where $g(\cdot)$ is a monotonic and differentiable link function. This broadens the scope of regression analysis to include binary, count, and other types of outcomes by using appropriate link functions and distributional assumptions.

Implications

The interpretation of estimated coefficients in the context of GLMs must account for the link function used. For instance, in logistic regression—a special case of GLM with a binary outcome—the coefficients represent the log odds of the outcome per unit change in the predictor, holding other predictors constant. As with OLS, the accurate interpretation of these models necessitates careful consideration of model assumptions and potential estimation biases.

DATA AND RESULTS

Data

The population dynamics of Portugal, as reported by the National Statistics Institute (INE) in 2017, reveal a gendered count of 10,291,027 with a declining trend since 2010, slightly mitigated in recent years. An aging demographic is evident with a rise in the median age from 42.7 to 44.2 years between 2012 and 2017.

This study utilizes data from the PORDATA database, specifically from the year 2017, which is the latest available year for most variables relevant to the research literature. Data covers all 308 municipalities in Portugal, including mainland Portugal, the Autonomous Region of the Azores, and the Autonomous Region of Madeira. The data's spatial distribution is summarized as follows:

TABLE 1. Distribution of Municipalities by NUTS I

NUTS I	Frequency	Percentage
Continente	278	90.3
Região Autónoma da Madeira	11	3.6
Região Autónoma dos Açores	19	6.2
Total	308	100.0

The variables related to location are converted into ordinal variables, labeled as follows for NUTS I, II, and III, indicating their respective continental and regional distributions. The assignment of ordinal values to these spatial categories aids in the structured analysis of location-based trends within the dataset.

For NUTS II, the distribution is provided in the subsequent table:

TABLE 2. Distribution of Municipalities by NUTS II

NUTS II	Frequency	Percentage
Alentejo	58	18.8
Algarve	16	5.2
Área Metropolitana de Lisboa	18	5.8
Centro	100	32.5
Norte	86	27.9
Região Autónoma da Madeira	11	3.6
Região Autónoma dos Açores	19	6.2
Total	308	100.0

The study's dependent variable, housing value, is represented by the Average Value of Property Transactions VMPT across municipalities. Additionally, a range of determinants such as population density, purchasing power, family size, unemployment, environmental quality, aging, education level, marital status, home ownership, and existing housing are considered, all of which are quantified according to data from the 2011 Census and PORDATA resources.

All variables, except location (NUTS I, II, and III), are continuous quantitative variables, reflecting the data's nature and the analytical approaches employed.

TABLE 3. Descriptive Statistics of the Variables

Variables	Observations	Mean	Standard Deviation	Minimum	Maximum
VMPT	308	54.426	49.597	3.980	436.176
DP	308	299.34	828.14	5.1	7.363
PCPC	308	80.11	18.36	55.20	219.60
DMF	308	2.569	0.227	2.1	3.60
DIEFP	308	5.851	2.932	0	15.80
IQA	308	0.064	0.461	0.002	0.528
IE	308	216.93	106.73	40	733.30
PES	308	0.092	0.042	0.028	0.311
PCAS	308	0.490	0.038	0.375	0.561
PFHP	308	0.803	0.077	0.507	0.939
PFH	308	0.001	0.001	0.000	0.005

Factors that affect the value of housing in Portugal

Regression Analysis of Housing Price Determinants in Portugal

We refined our regression model to elucidate the determinants of housing prices in Portuguese municipalities in 2017. Initially, we addressed multicollinearity by excluding the NUTS I and III location variables, given their redundancy with NUTS II within the model. The refined model leverages R and RStudio for computation and adheres to the following specification:

$$\begin{aligned} \ln(\text{VMPT}) = & \beta_0 + \beta_1 \times \text{NUTS II} + \beta_2 \times \text{DP} + \beta_3 \times \text{PCPC} \\ & + \beta_4 \times \text{DMF} + \beta_5 \times \text{DIEFP} + \beta_6 \times \text{IQA} + \beta_7 \times \text{IE} \\ & + \beta_8 \times \text{PCAS} + \beta_9 \times \text{PFHP} + \beta_{10} \times \text{PFH} + u_i \end{aligned} \quad (4)$$

Where $\ln(\text{VMPT})$ is the natural logarithm of the median property transaction value. The independent variables in the model include NUTS II, unemployment (DP), purchasing power per capita (PCPC), and other regional demographic factors. The adjusted R-squared value of 0.7121 suggests that our model explains approximately 71.21% of the variance in housing prices.

A correlation analysis identified a high correlation between PCPC and the existence variable PES, leading to the latter's exclusion to avoid multicollinearity. The subsequent Ordinary Least Squares (OLS) estimation model was statistically significant with a p -value less than $2.2e - 16$, indicating that at least one independent variable contributes significantly to explaining the dependent variable. Despite the overall model significance, DP and IQA variables were not statistically significant at a 10% significance level.

To confirm the validity of the OLS estimators, we tested for normal distribution of residuals, absence of error autocorrelation, and homoscedasticity. The Breusch-Pagan test for homoscedasticity reported a p -value of 0.1269, indicating no evidence of heteroscedasticity. However, the Shapiro-Wilk test indicated a departure from normality (p -value = 0.002582), which is mitigated by the sample size allowing for normal approximation. The Durbin-Watson test showed evidence of autocorrelation with a p -value less than 0.002.

Given these findings, we considered a Generalized Least Squares (GLS) model for robust estimation, which validated the absence of autocorrelation and homoscedasticity concerns. The estimated GLS model coefficients were presented as follows.

The interpretation of the model suggests that location (NUTS II), purchasing power, family size, and ownership status significantly affect housing prices. The results are in line with literature and provide a robust foundation for understanding the housing market in Portugal.

CONCLUSIONS

The analysis presented in this paper provides a detailed exploration of the factors influencing housing prices across municipalities in Portugal, with a distinct emphasis on the period of transition from growth to stabilization in 2017. Through a refined multiple linear regression approach, we have delineated the impact of geographical, demographic,

TABLE 4. GLS Estimation Model Coefficients

Coefficient	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.74110	0.894	15.149	< 0.0001 * **
NUTSII	-0.12690	0.220	-9.670	< 0.0001 * **
DP	0.00003	0.045	0.572	0.5674
PCPC	0.00454	0.242	4.070	< 0.0001 * **
DMF	0.33442	0.217	4.616	< 0.0001 * **
DIEFP	-0.01575	0.113	-1.655	0.0989.
IQA	0.35309	0.658	-1.111	0.2686
IE	-0.00021	0.503	-1.803	0.0724.
PCAS	0.41720	1.150	-4.746	< 0.0001 * **
PFHP	-1.56822	0.562	-5.825	< 0.0001 * **
PFH	62.04535	40.900	2.167	0.031*

Significance codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1

Residual standard error: 0.3441522

Degrees of freedom: 308 total; 297 residual

economic, and educational variables on property values within local contexts. Our findings reveal a pronounced geographic determinant in housing valuation, asserting the premium placed on coastal and urban locations. The size of the average household emerges as a significant predictor, reflecting the market's response to family housing demands. Economic health, as measured by local unemployment rates, exhibits a predictable yet critical inverse correlation with housing prices, indicating the sensitivity of the real estate market to job availability. Moreover, the positive association between educational levels and housing prices emphasizes the value attributed to intellectual capital in residential desirability. Notably, the study has highlighted the nuanced effect of homeownership rates on the market, contesting the straightforward view of ownership as a mere indicator of market heat. Instead, it is shown to interact with other factors in a way that could point to market maturity or saturation, suggesting a more complex scenario than previously understood. These results carry implications for stakeholders in the housing market. Potential homebuyers can benefit from understanding how these variables interconnect, potentially guiding more informed decision-making. For policymakers, the insights offer concrete evidence to inform strategic planning, economic development initiatives, and educational investment to foster balanced growth and maintain market vitality. The static nature of the data from a single year is a recognized limitation of this study, suggesting the need for ongoing research. As new census data becomes available, it will be imperative to re-evaluate these findings and observe how the interplay of these factors evolves over time. This continued analysis will be vital for adapting policies to the shifting realities of Portugal's diverse municipalities and ensuring that interventions remain relevant and effective in stabilizing and stimulating the housing market.

ACKNOWLEDGMENTS

This work has been supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

REFERENCES

- [] J. Rappaport, *Regional Science and Urban Economics* **38**, 533–552 (2008).
- [] J. Marôco, ReportNumber, Lda (2010).
- [] D. Duffie, *Journal of Economic Perspectives* **33**, 81–106 (2019).
- [] P. Kotler and G. Amstrong, *Introdução ao marketing* (Livros Tecnicos e Cientificos, 2000).
- [] E. Hoffman, *The right to be human: A biography of Abraham Maslow* (JP Tarcher Los Angeles, CA, 1988).
- [] T. T. L. Chong and X. Li, *Economic and Political Studies* 1–20 (2019).
- [] A. E. R. Guerreiro, *Determinantes no preço de um imóvel: uma análise hedónica no contexto algarvio*, Master's thesis (2012).
- [] M. S. G. Moreira, *repositorio-aberto.up.pt* (2000).

- [] M. M. G. F. P. Varão, “The determinants of housing prices in Portugal between 2000 and 2015,” Ph.D. thesis 2019.
- [] A. S. Adair, J. N. Berry, and W. S. McGreal, *Journal of Property Research* **13**, 67–83 (1996).
- [] Idealista, Nos últimos 2 anos notamos uma tendência muito maior para os reembolsos antecipados, 2019, accessed: 2020-01-21.
- [] W. Peng, “Real estate indicators in Hong Kong,” in *Conference on Real Estate Indicators and Financial Stability*, Vol. 27 (2003) p. 28.
- [] K. Tsatsaronis and H. Zhu, *BIS Quarterly Review*, March (2004).
- [] R. Portuguesa, *Constituição da República Portuguesa, VII Revisão* (Lisboa, 1976).
- [] B. de Portugal, *Relatório de estabilidade financeira* (2019) accessed: 2020-01-21.
- [] N. Winston and M. P. Eastaway, *Social Indicators Research* **87**, 211–221 (2008).
- [] A. Kaya and M. Atan, *Journal of Business Economics and Finance* **3**, 313–327 (2014).
- [] L. Alkana *et al.*, *International Journal of Strategic Property Management* **19**, 13–26 (2015).
- [] S. McGreal and P. Taltavull de La Paz, *Urban Studies* **50**, 2024–2043 (2013).
- [] S. McGreal and P. Taltavull de La Paz, *Urban Studies* **50**, 2024–2043 (2013).
- [] Ine – Instituto Nacional de Estatística, 2019-20.
- [] R. Calcagno, E. Fornero, and M. Rossi, *The Journal of Real Estate Finance and Economics* **39**, p. 284 (2009).
- [] A. e Pinheiro, O Setor Imobiliário em Portugal, (2017), accessed: 2020-01-15.
- G. W. Fuller, A. Johnston, and A. Regan, *West European Politics* **43**, 297–320 (2020).
- W. R. Archer, D. C. Ling, and B. C. Smith, *The Journal of Real Estate Finance and Economics* **40**, 41–61 (2010).
- G. Johnes and T. Hyclak, *The Annals of Regional Science* **33**, 33–49 (1999).
- M. T. Robson, *The Manchester School* **71**, 132–155 (2003).
- C. E. Borio and P. McGuire, *BIS Quarterly Review* (2004).
- M. A. Zambrano-Monserrate, *Cuadernos de Economía* **39**, 12–22 (2016).
- C. Chasco Yrigoyen and B. Sánchez Reyes, *Revista Galega de Economía* **21**, 1–21 (2012).
- J. M. Fitch Osuna, K. Soto Canales, and R. Garza Mendiola, *Estudios demográficos y urbanos* **28**, 383–428 (2013).
- J. Ermisch, *Economica* 383–404 (1996).
- A. Bayet, Y. Guegano, J.-M. Hourriez, and C. Chambaz, (1991).
- P. Eichholtz and T. Lindenthal, *Journal of Housing Economics* **26**, 19–32 (2014).
- D. M. Brasington, D. Hite, and A. Jauregui, *Journal of Regional Science* **55**, 442–467 (2015).
- Y. Liu and Z. Li, *Journal of Asian Architecture and Building Engineering* **17**, 299–305 (2018).
- J. Gyorko and P. Linneman, *Journal of Urban Economics* **39**, 318–341 (1996).
- A. Chevan, *American Journal of Sociology* **87**, 1133–1149 (1982).
- G. W. Gau and K. Wang, *The Journal of Real Estate Finance and Economics* **8**, 71–85 (1994).
- T. Xu, *Research in Economics and Management* **2**, 30–37 (2017).
- L. Katz and K. T. Rosen, *The Journal of Law and Economics* **30**, 149–160 (1987).
- J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis*, 7 (New Jersey Pearson, 2010).
- G. Calmasur, Determining factors affecting housing prices in Turkey with hedonic pricing model, in *International Conference on Business and Economics Studies, Washington DC, USA* (2016), pp. 255–269.
- Y. Xiao, E. C. Hui, and H. Wen, *Urban Forestry Urban Greening* **54**, p. 126769 (2020).
- B. de Portugal, <https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref%2006%2020%20pt.pdf> *Relatório de estabilidade financeira* (2020) accessed: 2020-12-12.
- S. Yanan, The impact of demographic structure on housing price: Evidence from Macau, in *Proceedings of the 2017 International Conference on Economics, Finance and Statistics (ICEFS 2017)*, edited by A. Press (2017), pp. 268–272.
- Y. Huang and W. A. V. Clark, *Urban Studies* **39**, 7–32 (2002).
- E. C. Hui, C. K. Chau, L. Pun, and M. Law, *Building and Environment* **42**, 2333–2343 (2007).
- I. C. Tsai and Peng, *Habitat International* **35**, 379–390 (2011).
- R. Gupta and A. Kabundi, *Journal of Economic Studies* (2010).
- Y. Song and G.-J. Knaap, *Journal of Urban Economics* **54**, 218–238 (2003).
- Q. Zhou, Q. Shao, X. Zhang, and J. Chen, 10.1016/j.landusepol.2019.104410 *Land Use Policy* **91**, p. 10441002 (2020).
- E. Reis, (2001).