Papers in Regional Science 104 (2025) 100113

Contents lists available at ScienceDirect

Papers in Regional Science

journal homepage: www.sciencedirect.com/journal/papers-in-regional-science

Housing market investment in peripheral areas: Evidence from Italy☆

Giuseppe Pernagallo [a,*], Giampaolo Vitali [b]

[a] University of Turin, Department of Economics and Statistics "Cognetti de Martiis", Italy
[b] National Research Council of Italy - Research Institute on Sustainable Economic Growth, Turin, Italy

ARTICLE INFO

JEL Classifications:
G5
R3
R11

Keywords:
Machine Learning
Peripheral Areas
Real Estate
Regional Economics
Urban Economics

ABSTRACT

The causal relationship between living in a peripheral area and real estate investment decisions is currently unexplored in empirical literature. Using survey data on 2711 Italian respondents collected between 2022 and 2023, we show that living in a peripheral area has a negative impact on the probability of having invested or investing in real estate in the future. OLS estimates suggest that being in a peripheral area reduces the likelihood of having invested in real estate in the last 24 months by about 3–4 percent and the likelihood of investing in real estate in the next 24 months by about 4–6 percent. These results are robust, having tried different models (OLS, probit, bivariate probit, and rare events logistic regression), different control variables, and also an instrumental variable approach to control for the potential endogeneity of homeownership. In this regard, our work introduces a new instrument to tackle the endogeneity problem of homeownership: hereditary motivation to purchase a home. The instrument has all the necessary characteristics to address this problem, such as a very strong first stage. Finally, we adopt a new machine learning algorithm, ABESS, to show the importance of including residence in a peripheral area as an explanatory variable and to select the best subset. This approach could also be exploited in further real estate and regional studies.

1. Introduction

Investment and savings choices vary depending on many factors (Nagy and Obenberger, 1994; Colabella et al., 2023). Among possible investments, the housing market represents one of the oldest investments and a major component of the assets of Italian households. The role of "geography" in determining the decisions of housing investment has begun to receive greater attention in the recent years, especially after the impact of the pandemic (Van Nieuwerburgh, 2023). In this regard, some researchers have studied the factors that influence the housing market in peripheral areas (e.g., Voith, 1999; Tsai et al., 2022), but very little has been documented regarding investment decisions. Peripheral areas may become increasingly attractive to real estate investors. Several studies have documented a counter-urbanization process across various countries. This trend appears to be driven by a desire to reconnect with one's origins, improve environmental quality, or avoid the escalating cost of living in urban centers (Goryunov and Kokovin, 2016; Cui et al., 2019; Sandow & Lundholm, 2023).

Empirical evidence linking peripheral areas to real estate investment is scattered. One of the earliest studies found that geographic areas perceived as peripheral tend to attract fewer institutional investors (Adair et al., 1995). More recently, research has shown that peripheral cities negatively affect foreign direct investment, although this impact is mitigated by proximity to central urban areas (McDonald et al., 2018). In contrast, recent research suggests that peripheral areas can offer profitable opportunities. The development of coworking spaces, for example, has been shown to stimulate the real estate market in tourist regions (Vogl and Micek, 2022). More broadly, local residents can play a key role in revitalizing these markets, as investors in residential real estate often purchase property within their own communities (Wright

☆ This study was carried out within the FRES (Fondo per la ricerca in campo economico e sociale") program and received funding (FRES2021–0000016) from MUR (Ministero Università e Ricerca) and Centro di Ricerca e Documentazione Luigi Einaudi di Torino. The views and opinions expressed are, however, only those of the authors and do not necessarily reflect those of the MUR or Centro Einaudi.

The authors would like to thank Centro di Ricerca e Documentazione Luigi Einaudi di Torino and FRES (Fondo per la ricerca in campo economico e sociale) for granting access to data. The authors would also like to thank participants at the European Regional Science Association (ERSA), 63rd Congress, Azores, Portugal, 26 August 2024, and at the Associazione Italiana di Scienze Regionali (AISRe), XLV Annual Scientific Conference, Turin, 05 September 2024, for useful comments and suggestions. Finally, the authors are also grateful to three anonymous reviewers for their useful comments, and to the editor, Prof. Rosella Nicolini.

* Corresponding author.
E-mail addresses: giuseppe.pernagallo@unito.it, giuseppepernagallo@yahoo.it (G. Pernagallo).

https://doi.org/10.1016/j.pirs.2025.100113
Received 12 June 2025; Received in revised form 25 August 2025; Accepted 26 August 2025
Available online 28 August 2025
1056-8190/© 2025 The Authors. Published by Elsevier B.V. on behalf of The Regional Science Association International. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and Yanotti, 2019).

However, to our knowledge, the causal relationship between living in a peripheral area and real estate investment decisions is currently unexplored in empirical literature. Therefore, the main contribution of this paper is to fill this gap. A second contribution of this paper is methodological in nature, as we propose a new instrument in the hereditary motivation to purchase a home to address the potential endogeneity of homeownership when used as an independent variable in a regression framework (e.g., DiPasquale and Glaeser, 1999; Cocco, 2005; Yamamura, 2011; Michielsen et al., 2016: Kaas et al., 2019; Stotz, 2019; Iwaisako et al., 2022). Finally, we apply a recent machine learning tool to select the best covariates in our regressions (for some recent applications of machine learning in real estate and regional economics see, for example, Baur et al., 2023, or Kopczewska, 2021).

When it comes to peripheral cities, the literature has shown that these areas have several disadvantages compared to central cities in that they are underdeveloped, have lower economic performance, or have lower human capital accumulation (Coniglio and Prota, 2008; Kubeš and Chvojkov, 2020; Pugh and Dubois, 2021). Regarding real estate investment, we expect that living in a suburban area negatively affects residents' ability to purchase real estate. Consequently, spatial peripherality is likely to be strongly associated with economic peripherality. As such, in a regression framework, residing in a peripheral area is expected to exhibit a negative relationship with real estate investment decisions.

Investment decisions in Italy's peripheral real estate markets are pivotal due to the country's pronounced regional disparities and demographic challenges. Italy exhibits significant socio-economic contrasts between its northern and southern regions, with peripheral areas often experiencing depopulation, aging populations, and economic stagnation. Strategic investments in these areas can stimulate local economies, preserve cultural heritage, and counter urban overconcentration. Comparatively, other European countries have implemented policies to revitalize their peripheral regions. For instance, Germany has seen the emergence of coworking spaces in peripheral areas, linking real estate markets with new forms of work (Vogl and Micek, 2022). In Italy, initiatives like the National Strategy for Inner Areas (SNAI) aim to address these challenges by promoting development in areas characterized by depopulation and marginalization (Guzzardi and Morelli, 2024). However, the success of such initiatives depends on coordinated investment in infrastructure, services, and policies that make peripheral areas attractive for living and investment. Therefore, investment decisions in Italy's peripheral real estate markets are not only economically significant but also essential for fostering regional equity and sustainable development.

Istat, the Italian institute of statistics, reports[1] that Italian peripheral and ultra-peripheral municipalities (see Section 2.1 for a definition) have experienced a severe decline in resident population over the years. Between 2002 and 2014, while the population of peripheral municipalities still showed a slight growth of 0.6 %, that of ultra-peripheral municipalities had already begun a clear downward trend, amounting to −3.1 %. Then, between 2014 and 2024, the demographic decline became widespread across much of the national territory but was most marked in peripheral areas (−6.3 %) and ultra-peripheral areas (−7.7 %). The issue is particularly urgent in Italy, where a small number of urban centers are concentrating the majority of the resident population, resulting in an unsustainable economic model. Investment in the revitalization of peripheral areas would not only foster the development of economically disadvantaged regions and alleviate pressure on overpopulated urban centres, but also serve as a testing ground for concrete policy solutions that could be replicated across other European contexts.

Our empirical analysis exploits a dataset from the *Centro Einaudi Research Centre* that contains a set of variables collected through questionnaires administered to Italian savers in 2022 and 2023, with a total of 2711 respondents. The sample was collected by a professional data analysis company, and the sampling was designed to ensure representativeness of the sample and adequate statistical power.[2] We integrated this dataset of microeconomic variables with additional contextual indicators, such as the state of the regional housing market (bullish or bearish), and constructed a dummy variable, *Periphery*, to identify whether a respondent resides in a peripheral area. The methodology used to define this variable is detailed in the following sections. Using this variable and a rich set of controls, we studied the relationship between peripherality and real estate investment decisions. Our dependent variables are two dummy variables, one to test whether the respondent has purchased real estate in the last 24 months (Y_1) to measure past investment decisions and the other to test whether the respondent plans to purchase real estate in the next 24 months (Y_2) to measure future investment decisions

We studied these relationships using different models in order to ensure robustness to the relationships found, such as linear probability models (LPM), probit models, bivariate probit models, and instrumental variable models. We also used a new machine learning technique to determine the importance of peripherality among all available covariates and select the best subset. Our results indicate that living in a peripheral area decreases the probability of having invested or investing in the housing market in the near future. These effects are significant and hold in any model, so we have robust evidence in favor of a causal relationship between peripherality and real estate investment decisions.

The rest of the paper is structured as follows. Section 2 describes the data and the econometric framework. Section 3 presents the results of the analysis. Section 4 provides policy implications and concludes the paper.

2. Data and methodology

2.1. Definition of peripheral area

The definition of "peripherality" for Italy is provided by the Italian *Department of Economic Development and Cohesion*. The methodology involves two steps: in the first, urban poles, denoted by the letters A and B, are identified, then four other types of municipalities are identified according to their distance from the urban poles, where "cintura" is the area closest to the pole while the ultra-peripheral area is the most distant.

This distance is quantified on the basis of an accessibility indicator calculated in terms of minutes of travel time to the nearest pole. The mapping of the municipalities is shown in Fig. 1, where the bands used are calculated using the second and third quartiles of the distribution of the index of distance in minutes to the nearest pole, equal to approximately 20 and 40 min. The third band, over 75 min, on the other hand, is equal to the 95th percentile, to identify ultra-peripheral territories.

It should be emphasized that the peripherality described in the policy maker's methodological notes is not necessarily understood as a weakness, but rather could also represent a strength. Indeed, while it is true that distance from essential services such as education, health care or rail transportation is a deficit for a peripheral area, these areas may have points of attractiveness, for example on the tourism side due to better preserved and unspoiled landscapes. In our case, a peripheral area could represent an investment with a view to renting the property to tourists or simply to escape the chaotic life of big cities.

The areas of interest to us are municipalities classified E or F, by

[1] Istat (2024). La demografia delle aree interne: Dinamiche recenti e prospettive future, Focus Statistiche. Available at https://www.istat.it/wp-content/uploads/2024/07/STATISTICA-FOCUS-DEMOGRAFIA-DELLE-AREE-INTERNE_26_07.pdf

[2] Further information about the sampling methodology is available in "Indagine sul Risparmio e sulle scelte finanziarie degli italiani 2023," edited by Giuseppe Russo and Pier Marco Ferraresi.

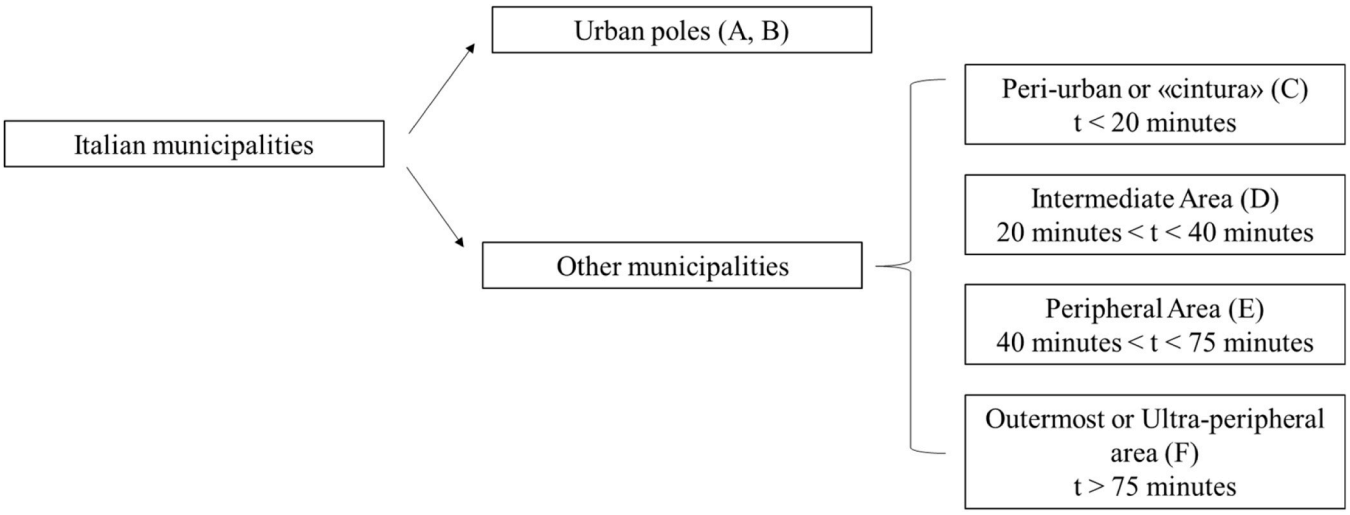


Fig. 1. Mapping of Italian municipalities according to the degree of distance from poles. The letter t indicates the travel time in minutes.

definition peripheral municipalities. Thus, using the classification provided by Istat based on the methodological note described above, we then created the dummy variable *Periphery* which is equal to 1 if the respondent's municipality is classified E or F, 0 in other cases.

2.2. Data and questionnaire

The data covered in this chapter exploit a *Centro Einaudi* dataset created through questionnaires administered to Italian savers in 2022 and 2023, with a total of 2711 respondents. The sample was collected by a professional data analysis company, and the sampling was designed to ensure representativeness of the sample and adequate statistical power. Specifically, to ensure territorial representativeness, the sample was stratified such that 30.5 percent of the interviews were conducted in northwestern Italy, 19.8 percent in northeastern Italy, 20.9 percent in central Italy and 28.8 percent in southern Italy. The target population consists of all Italian households with a bank account, totaling about 25 million households. When considering two self-excluding modes of a response (e.g., respondents "have" or "have not saved"), with a simple random sample, the confidence interval within which the percentage of responses in the universe can fluctuate (with the risk of being wrong 5 times out of 100) is given by this formula:

$$p \pm 1.96 \quad \sqrt{\frac{p(100-p)}{n}} \tag{1}$$

where p (percentage) indicates the first mode considered, and n is the sample size.

When we want to make comparisons between the data of two samples, for example, we want to compare those of 2021, with those of 2022 based, the calculation is as follows: calculation of the significance of the difference between the percentages corresponding to the same modalities in sample 2021 and sample 2022 (at the 5 percent significance level, i.e., taking the risk that 5 times out of 100 we consider significant a difference that in fact is not so) is given by

$$|p21 - p22| > 1.96 \quad \left[\frac{p21 \quad n21 + p22 \quad n22}{n21 + n22}\left(100 - \frac{p21 \quad n21 + p22 \quad n22}{n21 + n22}\right)\left(\frac{1}{n21} + \frac{1}{n22}\right)\right]^{1/2} \tag{2}$$

where $p21$ and $p22$ are the percentage of outcomes in 2021 and 2022, respectively, and $n21$ and $n22$ are the sample sizes in those years. The left member of the inequality represents the absolute value of the difference between the percentages: if it is greater than the right member, which represents the error of the sample distribution, the difference is significant; that is, sample 2021 and sample 2022 were drawn from two populations representing the difference between the two percentages.

Interviews were conducted with personal computers for CAPI (Computer Aided Personal Interviews), using a specific questionnaire divided into subject areas and consisting of numerous questions. The questionnaire was developed from 2011 by researchers at the *Centro Einaudi* and is based on relevant economic and statistical literature. Respondents are guaranteed anonymity. The answers to the questions in the questionnaire are pre-coded and the language of the questionnaire is such as to ensure its rigorousness while making it accessible to less experienced people, this in order to minimize the possibility of misunderstanding for questions of a technical nature. For the more complex questions, "name tags" have been prepared to be shown to respondents to help to answer, and in any case a glossary has been prepared to support interviewers with the more complex definitions to help them in administering the questionnaire. Finally, a verification operation of the interviewers' work is also planned, consisting of calls by the inspectors to a proportion (between 15 and 20 percent) of the respondents in order to verify the authenticity and seriousness of the interviews done.

Let us now report some background information on the sample (descriptive statistics in Table A1 in Appendix). Our respondents are divided into 1464 men and 1247 women, with an average age of 52 years and an average monthly household income of 2793 euros. There are 914 inactives/retired, while employed (employees) 1051 and self-employed 746. Specifically, the average income is 2734 euros in the poles (A + B), 2875 in the intermediate areas (C + D), and 2445 euros in the peripheral areas (E + F). The fact that average income is lower in peripheral areas is a well-known factor that disadvantages these areas (see, for example, Fortuna et al., 2001). These results seem to confirm that people living in the peripheral areas receive lower incomes. However, considering that the poles have a significantly higher cost of living, in real terms these differences may be less pronounced. Of the 2711

respondents, only 509 hold a college degree, while in terms of marital status, 582 are single, 223 widowed, 247 divorced, and the remaining married.

The questionnaire is aimed at financially independent individuals that manage the family finances. Regarding the geographical distribution, 740 individuals live in the south of Italy, 520 in the center, 553 in the northeast, and 898 in the northwest. Turning to the distribution of respondents in different areas as per the classification presented in the previous sections, most respondents live in the poles or the peri-urban areas. Specifically, about 33.5 percent live in area A or B, 48 percent in area C, 10 percent in area D, and about 8.5 percent in peripheral or ultra-peripheral municipalities (areas E or F). Table A2 in Appendix shows the distribution of our respondents by area classification. We note that most respondents live in peri-urban areas. This distribution reflects what is the distribution nationwide, in that the most populated areas are those closest to urban hubs.

At this point we need to ask whether there is any association between area of residence and our dependent variables, and whether this association is statistically significant. We can test this hypothesis by using a chi-square test on the contingency tables summarized in Table A3 in Appendix. With 95 percent confidence, we can conclude that there is a significant association between the area of municipality and past (last 24 months) real estate investment decisions while we cannot draw the same conclusion for future (next 24 months) real estate investment decisions.

The results in Table A3 while interesting have marginal relevance to our problem, which is focused only on whether one lives in a peripheral area. For this reason, we repeat the analysis by substituting the *Periphery* variable for the six-area classification. Table A4 in Appendix shows the contingency tables between the two dependents and the Peripheral variable. The chi-square test suggests that with 90 percent confidence, the association is statistically significant for both dependent variables, and even 95 percent for future investment decisions. It therefore seems legitimate to think that there may be some relationship between living in a peripheral area and making certain investment decisions.

2.3. Variables

As already anticipated, the dependent variables in our econometric models are:

- *Real estate purchase in the last 24 months* (Y_1): to measure past real estate investment decisions, obtained through the "yes or no" question "have you purchased real estate in the past 24 months?"
- *Investment in real estate in the next 24 months* (Y2): to measure future real estate investment decisions, obtained through the "yes or no" question "will you purchase a property in the next 24 months?"

We are particularly interested in the effect of the *Periphery* on our two dependent variables using a regression model. However, the variable *Homeownership* is among the most important in real estate studies, so it is also interesting to assess its effect on real estate investment. These variables are defined as follows:

- *Periphery*: dummy equal to 1 if a respondent lives in a municipality classified E or F (see Fig. 1), zero otherwise. This variable should be exogenous, since the peripherality of a city is determined by its geographical location, which is obviously assigned by nature. However, one could argue that if households sort by distance from the city center, *Periphery* will be endogenous if the sorting behavior leads investment-averse/investment-loving households to choose to live in the suburbs. This concern is not warranted, as our econometric model includes the most important variables that can explain investment decisions (such as income, risk aversion, or financial education) and may be simultaneously correlated with *Periphery*. Moreover, the different specifications proposed in this paper show that even drastically changing the number of covariates in the model yields small changes in the magnitude of the coefficient of *Periphery*, providing strong evidence in favor of its exogeneity.
- *Homeownership:* dummy equal to 1 if the respondent owns the house he or she lives in (this house is called the "first house" in Italy, and is considered the most important property, so much so that even some taxes are not due on this type of property). This variable could be endogenous as we discuss in Section 2.6.

A negative (and significant) sign of the *Periphery* variable would indicate that living in a peripheral area has negative effects on real estate investment decisions (and capabilities), and thus spatial peripherality also translates into economic peripherality. Note that while we are interested in short-term real estate investment decisions, the *Homeownership* variable refers to the respondent's entire life span, so it could be an important predictor of past or future short-term investment decisions. To mitigate the endogeneity problem, we also use an instrumental variables approach, using two instruments based on the empirical literature (DiPasquale and Glaeser, 1999; Yamamura, 2011; Michielsen et al., 2016: Kaas et al., 2019; Stotz, 2019; Iwaisako et al., 2022):

- *Index*: A regional housing price index constructed using monthly average sales prices (in euros) per square meter of housing by region over the past 3 years (data source: *Idealista*[3]).
- *Inheritance*: a dummy variable equal to 1 if the respondent believes that buying a house is important because it is a legacy to leave to heirs.

We discuss in detail the strategy of instrumental variables in Section 2.6. In addition, we have a numerous set of control variables useful for mitigating any omitted variable bias. These variables are widely used in the empirical literature cited in this article and are also the result of the joint effort of researchers at the *Centro Einaudi* and the professional data analysis company that designed the questionnaire:

- *Risk aversion*: a Likert scale variable from 1 (high aversion) to 5 (low aversion) derived from a question asking the respondent to rate his or her risk aversion in financial investments.
- *Financial education*: self-rated by the respondent on a scale of 1–7.
- *South*, *Center*, *North-East*: dummy variables for the geographical area where the respondent is living, benchmark is *North-West.*
- *Bullish.* Dummy variable that is equal to 1 if the respondent lives in a region where the housing market is bullish. This variable was derived by collecting sales prices per square meter of housing by region for the past 3 years using monthly data from the *Idealista* website. Using the price history, we derived the national average and classified the market as bullish if the region where the respondent lives has an average price over the 3-year period that is higher than the national average. A map of prices by region is shown in Fig. 2. The choice to construct this variable at the regional level follows a literature that models housing behavior at the regional level (see, for example, Tsai, 2019).
- *Household components*: number of household components.
- *Gender*: dummy equal to 1 for males.
- *University*: dummy equal to 1 if the respondent has a degree.
- *Age*: age of the respondent.
- *Unmarried*, *Divorced*, *Widowed*: dummy variables with benchmark *Married/Cohabitant*.
- *Net monthly family income*: expressed in euros.
- *Employed, Inactive/retired*: dummy variables, benchmark is *Independent worker*.

[3] https://www.idealista.it/sala-stampa/report-prezzo-immobile/vendita/storico/

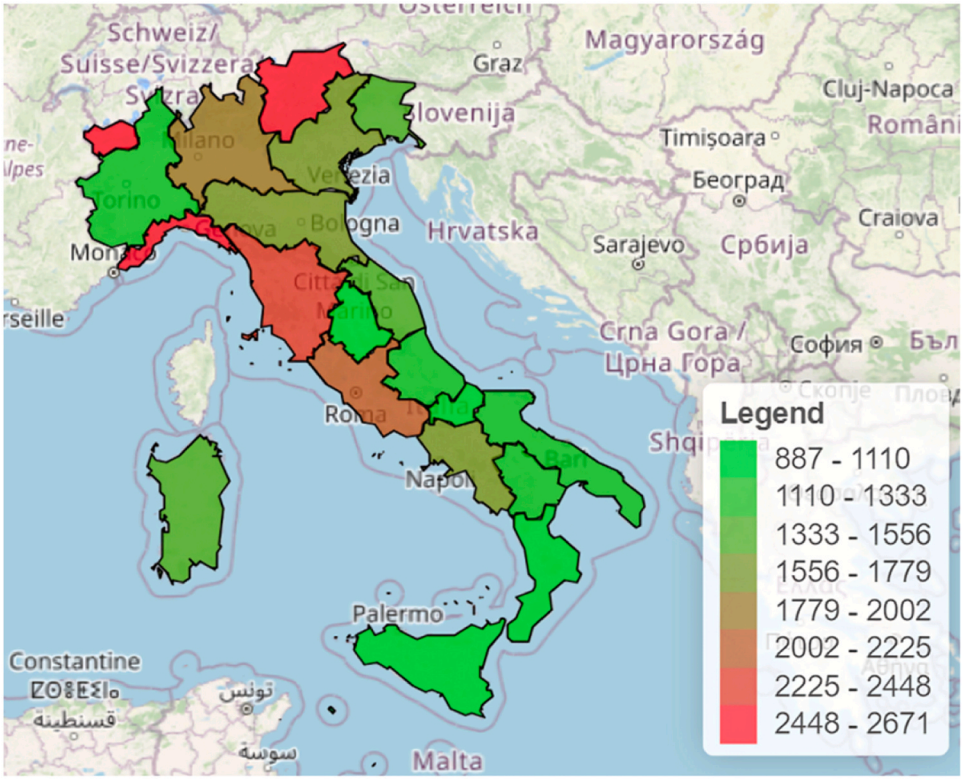


Fig. 2. Average real estate sales price per square meter in the last three years (2021-2023) by region. Our elaboration on *Idealista* data.

- *Age of financial independence*: the age at which the respondent achieved financial independence (our respondents are sampled only if they are financially independent).
- *Saved last year*: dummy variable equal to 1 if the respondent managed to save the year before.
- *Life insurance*: 1 if the respondent has signed a life insurance contract.

Fig. A1 in Appendix visualizes the correlation matrix between the dependent and the control variables, while Table A1 shows some descriptive statistics to give a preliminary overview of the dataset.

2.4. Best subset selection

Using all available covariates in regression models can certainly prove useful in order to mitigate the bias from omitted variables. However, the inclusion of variables that are not statistically significant or redundant causes the estimator to lose efficiency. Thus, finding the set of variables that best explain the dependent is very important for inference purposes. If the optimal set of covariates includes the *Periphery* variable, this would provide supporting evidence that spatial peripherality plays a significant role in shaping real estate investment decisions. In machine learning there are several approaches for selecting the so-called best subset, that is, the optimal set of explanatory variables given a dependent variable.

There are several possible approaches, in this paper we will focus on one of the most recent and effective methodologies, namely ABESS or *Adaptive Best Subset Selection* (Zhu et al., 2020; Zhang et al., 2023). ABESS dynamically selects the appropriate model size adaptively, eliminating the need to select regularization parameters. The method is designed primarily for general linear regression problems; therefore, it will be reliable especially when intended to drive OLS estimates and is characterized by its polynomial-time complexity and property of providing unbiased and consistent estimates. The ABESS algorithm does not guarantee to find the best subset for any regression problem, but it is nevertheless useful in that it allows us to objectively determine a subset that averages between the omitted variable and redundant variable problems, thus providing robustness to our empirical analysis. The algorithm is implemented through the *R* programming language.

2.5. Econometric models

In order to ensure robustness of the analysis, we used three different econometric models. First, a linear probability model (LPM), which is an OLS with a dummy dependent variable. The general form of this model can be described by the usual equation:

$$\mathrm{E}(Y_i \mid \textit{Periphery}, \textit{Homeownership}, \boldsymbol{X}) = \mathrm{P}(Y_i = 1 \mid \textit{Periphery}, \textit{Homeownership}, \boldsymbol{X}) =$$

$$\beta_0 + \alpha\, \textit{Periphery} + \gamma\, \textit{Homeownership} + \beta_1 X_1 + \ldots + \beta_k X_k,\ i = 1,2 \quad (3)$$

where we are interested mainly in the coefficients of *Periphery* and *Homeownership*, while **X** is the matrix of the *k* control variables (X_1, X_2, …,X_k), the betas are the unknown parameters of these variables, and the equality between the expected value of Y conditional on X and the probability that Y is equal to 1 conditional on X results from taking Y as a Bernoullian variable. This model presents two problems: 1) it is inherently heteroskedastic; 2) since it is a straight line it can predict probability values outside the range [0,1].

Because of the first problem the usual standard errors of OLS are erroneous and therefore robust standard errors for heteroskedasticity should be used in their place. The second problem becomes relevant if

the goal of the model is to provide predictions, while LPM is still adequate in case one is only interested in inference of estimates, which is the case our study falls into.

A more natural model for studying dichotomous dependent variables is the probit model, which is a nonlinear regression model in which instead of using a straight line, a probit function, i.e., the *cdf* (cumulative distribution function) of the standard normal distribution, is used, thus ensuring predictions in the range [0,1]. The regression equation in this case takes the following form

$$P(Y_i = 1 \mid Periphery, Homeownership, X) =$$

$$\Phi(\beta_0 + \alpha\, Periphery + \gamma\, Homeownership + \beta_1 X_1 + \ldots + \beta_k X_k),\ i = 1,2 \quad (4)$$

where Φ is the standard normal distribution *cdf*.

A potential problem in our case is that the two dependent variables may be correlated, in the sense that past real estate investment decisions may be associated with future ones and vice versa. In order, therefore, to test whether the existence of this correlation may affect our results, we proceed as well to estimate a bivariate probit model, i.e., a probit model in which two regression equations, one for Y_1 and one for Y_2, are jointly estimated, also taking into account their correlation.

2.6. Instrumental variable approach

The variable *Homeownership* is one of the most important and used in real estate studies, but its inclusion is generally problematic with regard to endogeneity (DiPasquale and Glaeser, 1999; Yamamura, 2011; Michielsen et al., 2016: Kaas et al., 2019; Stotz, 2019; Iwaisako et al., 2022). The same concern can be expressed for our estimation strategy, as *Homeownership* may be correlated with the unobserved components of the outcome variable equation that ends up in the error term. Expected future labor income from households may be a good example of an omitted factor that may bias our estimates (e.g., Cocco, 2005, or Michielsen et al., 2016), given that it can be correlated with both the dependent variable and *Homeownership*, because households with higher expected future incomes are more likely to purchase a new property and at the same time already own a home. To overcome this problem, we use an instrumental variable (IV) approach.

Some instruments have been proposed in the literature in contexts similar to that of our study, so we draw from these works to construct our IV estimates. A popular strategy is to use regional house prices or a regional house price index in the current year and/or the year of house purchase (e.g., Yamamura, 2011; Michielsen et al., 2016; Chetty et al., 2017). The validity of this instrument means that regional house price changes are orthogonal to all other factors that may influence housing investment decisions, but regional house price indices may be correlated with housing investment decisions through channels other than homeownership. For this reason, the literature suggests taking into account macroeconomic conditions, such as the regional unemployment rate in the current year and/or the year of purchase (Michielsen et al., 2016; Chetty et al., 2017). Unfortunately, we do not know the year of purchase, so we will only control for the current year's regional unemployment rate. This will not affect our estimates too much, given the large number of covariates included, the many robustness checks performed, and the fact that the coefficient of *Unemployment* turned out to be non-significant. We constructed a regional house price index, *Index*, using average sales prices per square meter of housing by region over the past 3 years from the *Idealista* website.

In addition, Michielsen et al. (2016) note that selection effects might be relevant if the decision to buy a house at a given price level is related to unobserved household characteristics, such as the degree of risk aversion. This is not a problem in our work, as we have a variable to measure the degree of risk aversion of our respondents and many other household characteristics.

However, the use of the price index instrument does not guarantee a solution to the problem because it is weakly correlated with the potential endogenous variable in our sample (see section 3.2). Therefore, one of the main contributions of this paper is the introduction of a second instrument. We propose as an instrument *Inheritance*, a dummy variable (yes/no) obtained by asking respondents whether they would buy property to leave an inheritance to their children or relatives. We expect this variable to be correlated with homeownership, since one reason for owning a home is to pass it on to future generations. Moreover, *Inheritance* exploits its effect on dependent variables only through *Homeownership*. Finally, we expect it to be uncorrelated with the error term, since our specification includes many of the variables that may be correlated with *Inheritance*, such as degree of risk aversion, financial education, long-term investment decisions such as life insurance, or number of household members. The empirical literature presents examples of using inheritance-related instruments to account for the endogeneity of home ownership (e.g., Kaas et al., 2019), so our choice seems sound. We apply the IV procedure directly to the probit model for reasons of synthesis and because it is the most natural model, since both the dependent and endogenous variables are dichotomous.

3. Results

3.1. OLS, probit and bivariate probit

Best subset analysis using ABESS, allows us to select the ideal set of covariates to explain our dependent variables. The choice of the best model can be made using several criteria, in this case we will focus on the Akaike information criterion (AIC). The choice of the AIC rather than the Bayesian information criterion (BIC) is justified by the fact that the BIC tends to be a criterion that selects more parsimonious models (thus with fewer variables) exposing the estimates more likely to the problem of omitted relevant variables. The results of the ABESS algorithm are shown in Fig. 3.

Looking at Fig. 3 we learn that the optimal number of covariates for the first dependent variable is 6, while for the second dependent variable it is 9, because for these two subsets the minimum value of the AIC is reached. The algorithm also tells us precisely what these variables are.

- Best subset for Y_1: *Periphery*, *Homeownership, Risk aversion, University, Widowed, Life insurance.*
- Best subset for Y_2: *Periphery*, *Homeownership, Financial Education, South, North-East, Age, University, Saved last year, Life insurance*

We note that for both variables the best subset includes *Periphery*, which along with *Homeownership*, turns out to be a key variable in explaining real estate investment decisions. We therefore estimated different LPMs and probits: one version with only the two most relevant variables (i.e., *Periphery* and *Homeownership*), one with all variables, and one with the best subset. The results are shown in Table 1 (LPM) and Table 2 (probit). We report estimates only of the coefficients of *Periphery* and *Homeownership* being the variables of greatest interest to our study (as well as the most important). The interested reader can also find complete estimates for the other variables in the Appendix (Tables A5-A8).

Table 1 estimates through OLS that living in a peripheral area decreases the probability of having invested or investing in real estate. The estimated coefficient is robust to Y_1 in that it varies very little between models. Specifically, *ceteris paribus*, living in a peripheral area decreases the probability of having invested in real estate in the last 24 months by

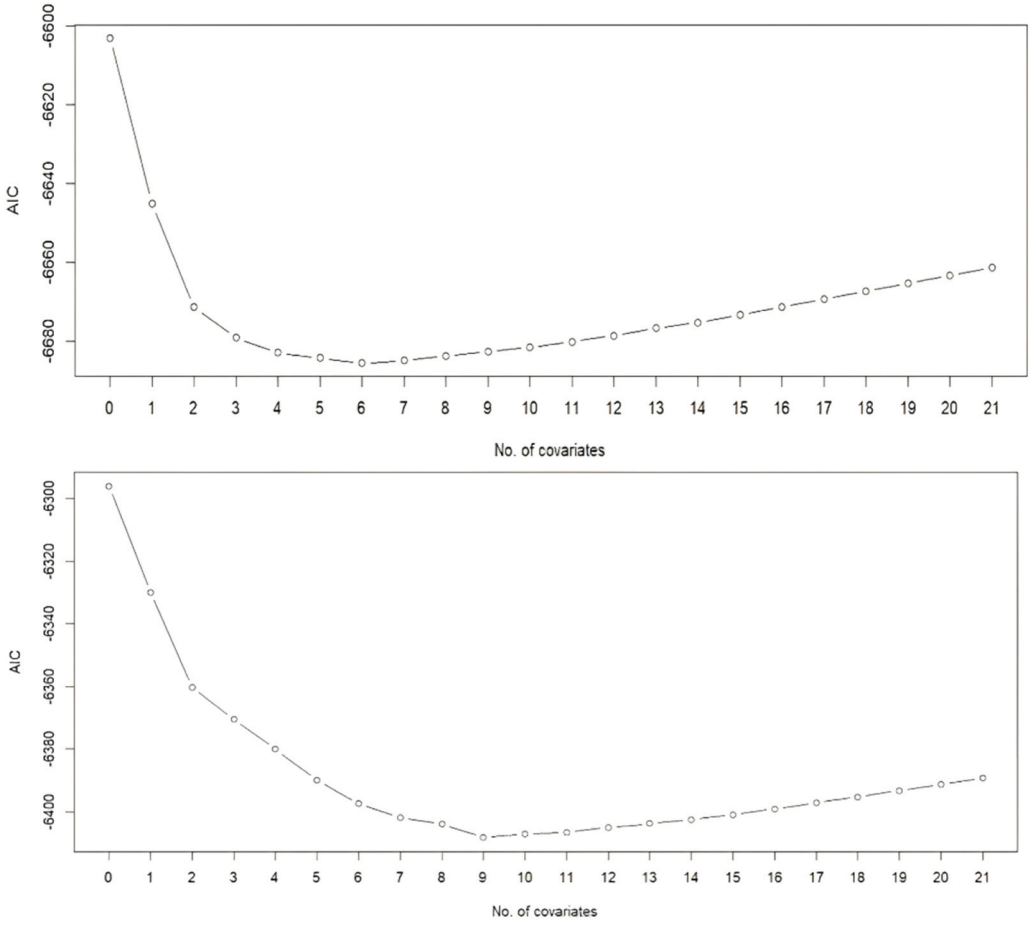


Fig. 3. Results of the ABESS algorithm. The upper panel shows the optimal number of covariates for Y_1 while the lower panel for Y_2.

about 4 percent. As for Y_2 once the control variables are entered, the regression coefficient is about –5 percent, so living in a peripheral area reduces the probability of investing in real estate in the next 24 months by about 5 percent, all other things being equal. The significance of the results is very high, as the *Periphery* coefficient is always significant with 90, 95 and even 99 percent confidence in some models. In contrast, already owning a house is positively associated with past real estate investment and negatively with future real estate investment with 99 percent confidence. The result may mean that real estate investment is mainly put in place to procure the primary residential home, and once this investment is completed the probability of reinvestment drops, for example, for speculative purposes.

Table 2 confirms the OLS estimates using a probit model. All model estimates are statistically significant at the 95 percent confidence level. We find again a negative effect of *Periphery* on past and future real estate investment decisions, and a positive effect of *Homeownership* on past choices and a negative effect on future choices. It should be noted that the interpretation of the coefficients is different here. If we look, for example, at the coefficient of *Periphery* in Model 9, we have that for those living in a peripheral area the z-score (or probit index) decreases by 0.36, all other things being equal. A z-score further to the left in the standard normal distribution is associated with a lower *cdf* and thus a lower probability. We avoid reporting the slopes at mean of the probit models since they are very similar to OLS (this is common, as also documented in other empirical literature, e.g., Pernagallo et al., 2021).

Finally, Table 3 shows the results of the bivariate probit. Even considering the correlation between the two dependent variables, the estimates obtained are very similar to those in Table 2, thus confirming that the negative effect of *Periphery* on real estate investment decisions found in this study is robust.

3.2. Instrumental variable regression

Tables 4 and 5 present the results of the IV estimation procedure using the two instruments discussed in Section 2.6. The models include all available covariates (such as the regional unemployment rate, as recommended in the literature) as well as the optimal subset of variables previously identified using the ABESS algorithm. If we look at the first stage of the *Index* instrument (Table A9 and A10 in Appendix), we see that the instrument is weak as the absolute Z value of its coefficient in the first stage is only 3.46 when using all covariates (Table A9), 0.37 for Y_1 and 1.12 for Y_2, respectively, when using the best subset (Tabe A10).

On the other hand, the *Inheritance* instrument is very strong as the absolute Z value of its coefficient in the first stage is over 4.5 when using all covariates (Table 4), and over 7 when using the best subset (Table 5). In light of these results, the IV estimates with *Inheritance* are the most reliable and we will focus only on these in our comments (Tables 4 and 5), but for completeness we also report the estimates with *Index* (Tables A9 and A10). The complete estimates also for control variables are provided in Appendix (Table A11).

Table 1
OLS estimates with robust standard errors.

	Model (1)				**Model (2)**			
	Y1				Y2			
	Estimate	SE	t	p-value	Estimate	SE	t	p-value
Const	0.0174	0.0049	3.547	0.0004	0.1265	0.0129	9.823	0.0000
Periphery	−0.0351	0.0151	−2.326	0.0201	−0.0418	0.0159	−2.638	0.0084
Homeownership	0.0810	0.0081	10.03	0.0000	−0.0398	0.0142	−2.804	0.0051
Control variables	No				No			
	n	2711			n	2711		
	AIC	436.31			AIC	980.31		
	Model (3)				**Model (4)**			
	Y1				Y2			
	Estimate	SE	t	p-value	Estimate	SE	t	p-value
Const	−0.0196	0.0516	−0.381	0.7030	0.0901	0.0515	1.750	0.0803
Periphery	−0.0379	0.0194	−1.953	0.0509	−0.0582	0.0188	−3.093	0.0020
Homeownership	0.0610	0.0115	5.312	0.0000	−0.0694	0.0169	−4.118	0.0000
Control variables	All				All			
	n	1967			n	1967		
	AIC	295.77			AIC	578.10		
	Model (5)				**Model (6)**			
	Y1				Y2			
	Estimate	SE	t	p-value	Estimate	SE	t	p-value
Const	−0.0410	0.0109	−3.736	0.0002	0.1036	0.0237	4.360	0.0000
Periphery	−0.0390	0.0158	−2.468	0.0136	−0.0540	0.0154	−3.499	0.0005
Homeownership	0.0623	0.0078	7.898	0.0000	−0.0691	0.0142	−4.851	0.0000
Control variables	Best subset				Best subset			
	n	2503			n	2509		
	AIC	208.89			AIC	581.66		

Table 2
Probit estimates with robust standard errors.

	Model (7)				**Model (8)**			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.1543	0.1248	−17.26	0.0000	−1.1384	0.0629	−18.08	0.0000
Periphery	−0.2936	0.1485	−1.977	0.0480	−0.3102	0.1421	−2.183	0.0291
Homeownership	0.8671	0.1302	6.661	0.0000	−0.2242	0.0744	−3.012	0.0026
Control variables	No				No			
	n	2711			n	2711		
	AIC	1393.67			AIC	1668.59		
	Model (9)				**Model (10)**			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.3980	0.4088	−5.865	0.0000	−1.5467	0.3192	−4.845	0.0000
Periphery	−0.3624	0.1770	−2.047	0.0407	−0.5120	0.1963	−2.608	0.0091
Homeownership	0.7610	0.1630	4.668	0.0000	−0.4774	0.1034	−4.615	0.0000
Control variables	All				All			
	n	1967			n	1967		
	AIC	990.04			AIC	1096.96		
	Model (11)				**Model (12)**			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.6876	0.1579	−17.01	0.0000	−1.3408	0.1632	−8.218	0.0000
Periphery	−0.3908	0.1648	−2.371	0.0177	−0.4779	0.1788	−2.673	0.0075
Homeownership	0.8181	0.1407	5.816	0.0000	−0.4595	0.0884	−5.195	0.0000
Control variables	Best subset				Best subset			
	n	2503			n	2509		
	AIC	1167.29			AIC	1323.87		

Table 3
Bivariate probit estimates with robust standard errors.

	Model (13)				Model (14)			
	Y1				Y1			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.3982	0.4088	−5.867	0.0000	−2.6872	0.1580	−17.01	0.0000
Periphery	−0.3626	0.1771	−2.047	0.0407	−0.3901	0.1648	−2.367	0.0180
Homeownership	0.7609	0.1630	4.667	0.0000	0.8184	0.1407	5.816	0.0000
Control variables	All				Best subset			
	Y2				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−1.5470	0.3194	−4.843	0.0000	−1.3352	0.1630	−8.193	0.0000
Periphery	−0.5117	0.1962	−2.608	0.0091	−0.4798	0.1787	−2.686	0.0072
Homeownership	−0.4775	0.1035	−4.615	0.0000	−0.4621	0.0885	−5.219	0.0000
Control variables	All				Best subset			
	n	1967			n	2503		
	AIC	2088.99			AIC	2491.98		

Table 4
IV probit estimates with robust standard errors, with *Inheritance* as the instrumental variable. Control variables also include *Unemployment* (regional unemployment rate in current year).

	Model (15)				Model (16)			
	Y1				Y2			
First stage (dependent is Homeownership)					**First stage (dependent is Homeownership)**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Inheritance	0.0866	0.0189	4.59	0.000	0.0866	0.0189	4.59	0.000
	Y1				Y2			
IV estimates					**IV estimates**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−0.6623	0.6084	−1.09	0.276	−1.091	0.6171	−1.77	0.077
Periphery	−0.1665	0.1525	−1.09	0.275	−0.4627	0.2009	−2.30	0.021
Homeownership	−1.1682	0.5990	−1.95	0.051	−0.7410	1.0286	−0.72	0.471
Control variables		All				All		
	n	1967			n	1967		
	AIC	2927.71			AIC	3041.57		

Table 5
IV probit estimates with robust standard errors, with *Inheritance* as the instrumental variable. The best subset is the same as in the previous section.

	Model (17)				Model (18)			
	Y1				Y2			
First stage (dependent is Homeownership)					**First stage (dependent is Homeownership)**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Inheritance	0.1441	0.0167	8.61	0.000	0.1225	0.0168	7.30	0.000
	Y1				Y2			
IV estimates					**IV estimates**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−1.5157	0.5206	−2.91	0.004	−1.0744	0.3430	−3.13	0.002
Periphery	−0.2688	0.1542	−1.74	0.081	−0.4220	0.1812	−2.33	0.020
Homeownership	−0.4762	0.4615	−1.03	0.302	−1.0296	0.5722	−1.80	0.072
Control variables	Best subset				Best subset			
	n	2503			n	2509		
	AIC	3976.73			AIC	4039.10		

Starting with past investment decisions (Y_1), we can see that the *Periphery* coefficient retained its sign in all IV estimates, but its magnitude was generally slightly lower than the OLS and probit estimates. When the best subset is used, this coefficient is also statistically significant at 90 % confidence (Table 5). The sign of *Homeownership* is not clear because it changes depending on the instrument used. With the *Index* instrument, we get the same sign discovered in previous estimates, whereas when using *Inheritance*, the sign changes. This is probably due to the endogeneity issues discussed earlier. When using all covariates and *Inheritance* as an instrument, *Homeownership* is statistically significant at almost 95 percent (Table 4). This robustness check further confirms that residing in a peripheral area is associated with a lower

Table 6
Rare events logistic regression estimates with robust standard errors.

	Model (19)				Model (20)			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−4.0982	0.3189	−12.85	0.000	−1.9182	0.1183	−16.21	0.000
Periphery	−0.5540	0.3183	−1.74	0.082	−0.5951	0.3041	−1.96	0.050
Homeownership	1.8918	0.3269	5.79	0.000	−0.4364	0.1428	−3.06	0.002
Control variables	No				No			
	n	2711			n	2711		
	Model (21)				**Model (22)**			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−4.4402	0.9376	−4.74	0.000	−2.8067	0.6036	−4.65	0.000
Periphery	−0.6588	0.3767	−1.75	0.080	−0.9801	0.4129	−2.37	0.018
Homeownership	1.5699	0.3962	3.96	0.000	−0.9178	0.1988	−4.61	0.000
Control variables	All				All			
	n	1967			n	1967		
	Model (23)				**Model (24)**			
	Y1				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−5.0952	0.3610	−14.11	0.000	−2.3483	0.3234	−7.26	0.000
Periphery	−0.7160	0.3457	−2.07	0.038	−0.9074	0.3839	−2.36	0.018
Homeownership	1.7314	0.3432	5.04	0.000	−0.8934	0.1731	−5.16	0.000
Control variables	Best subset				Best subset			
	n	2503			n	2509		

probability of having invested in real estate, and this effect appears to be relatively robust. However, the analysis does not allow us to draw definitive conclusions regarding the direction of the relationship between homeownership and past investment decisions.

The effect of residence in a peripheral area on future investment decisions (Y_2) is clear. The sign and magnitude of the IV estimates for *Periphery* are close to those found with previous estimates and always significant with 95 percent confidence. The sign of *Homeownership* is also preserved in almost all estimations but the only significant one with 90 percent confidence results from using the *Inheritance* instrument and the best subset. We have evidence that homeownership negatively affects the probability of investing in real estate, but this evidence is not very robust given the results of the IV regression. It is also worth noting that the coefficient on inheritance exhibits some sensitivity to model specification, showing slight variation between the full set of control variables and the best subset identified through variable selection.

3.3. Rare events logistic regression

As pointed out by one of the referees, the results may be influenced by the fact that only a few respondents purchased properties and among them only a minority were located in peripheral areas. This situation is sometimes referred to as "rare events data," which is a situation in which binary dependent variables have tens to thousands of times fewer ones (events) than zeros ("non-events"). To address this problem, we adopt the framework developed by King and Zeng (2017), using a logistic regression for rare events data. This regression applies a correction that accounts for the fact that ones are underrepresented in the sample. For a detailed technical discussion of this issue, we refer the reader to the work of King and Zeng.

The results of this methodology are presented in Table 6. We avoid reporting the full list of estimates in the appendix, as the results are similar to those presented previously. We can see that not only the sign and significance of the *Periphery* coefficient are preserved, but also the magnitude seems larger than in previous models. This shows that although our sample may be unbalanced in the distribution of zeros and ones, our models are robust and clearly indicate the negative effect of residing in peripheral areas on real estate investment decisions.

4. Conclusions

Peripherality can act on economic development in different ways. The concept of peripherality is the subject of several fields of study, suffice it to mention the vast literature in tourism economics (e.g., Pearce, 2002 or Ianioglo and Rissanen, 2020) or economic geography (e.g., Gatrell, 1999, Pugh and Dubois, 2021, or Nilsen et al., 2023). The idea that peripherality negatively affects economic development is the most logical but not necessarily the most agreed upon, and this interpretation responds to the idea of the Italian policy maker, who in his methodological notes also emphasizes the strengths that peripheral areas can present. A peripheral area, for example, may be more pristine and thus present tourist attractiveness, or quieter areas may present better living conditions in terms of viability, air quality and cost of living.

The link between living in a peripheral area and real estate investment has been investigated only marginally in the literature, and generally the results indicate a negative effect of suburbia on investment decisions (e.g., Adair et al., 1995, or McDonald et al., 2018). Our work adds to this strand of literature, as we found a negative relationship between peripherality and past or future real estate investment decisions. OLS estimates suggest that being in a peripheral area reduces the likelihood of having invested in real estate in the last 24 months by

about 3–4 percent and the likelihood of investing in real estate in the next 24 months by about 4–6 percent. The sign of this relationship is also confirmed using different models (e.g., probit or bivariate probit), changing the set of control variables, or even taking into account the endogeneity of homeownership through the instrumental variables approach. In the latter regard, our work introduces a new instrument to address the endogeneity of homeownership in hereditary motivation to purchase a home. This instrument has a very strong first stage and is plausibly uncorrelated with the error term, having also been used in other empirical work (Kaas et al., 2019).

The effect of homeownership on real estate investment decisions is ambiguous for past investment decisions (IV probit estimates partly contradict OLS and probit estimates), while it is clearer for future investment decisions. Homeownership decreases the probability of investing in real estate in the future. Although we are not primarily interested in the causal effect of homeownership on future real estate investment, the result is nonetheless interesting, as it may indicate that Italians are not interested in real estate as a source of return, but as a way to gain the security offered by living in a home they own.

The adopted econometric framework may have some shortcomings. We have strong evidence in favor of the exogeneity of the *Periphery*, since the regression models include most of the variables needed to explain investment decisions and the magnitude of the coefficient of *Periphery* changes only slightly if we change the set of covariates. The major limitation of this study is that it may not have a large degree of external validity, as the timing and national context limit its application to other settings. Nevertheless, it is still an important survey for at least two reasons. First, since it was conducted after the pandemic, it shows the investment behavior of Italians after such a disruptive shock. Second, although the concept of peripherality depends strictly on the country under investigation, there are, at least in Europe, many countries closely similar to Italy, so we expect the results to be valid for these countries as well.

Importantly, these results indicate that living in a peripheral area strongly affects the investment opportunities of its residents, so we document further evidence in favor of the thesis that being peripheral has a negative impact on investment. Italian policy makers have addressed the existence of disadvantaged areas, and in some regions, policies have been introduced to support real estate in peripheral areas. The most frequent policies are aimed at stimulating the housing market in suburban areas with economic incentives in order to attract residents back. This is the case in Trentino, where a policy to revitalize areas at risk of depopulation was introduced in 2025. The initiative is designed to address not only structural challenges but also social and economic dimensions of development in vulnerable territories. It aims to foster a sense of community identity in fragile areas by ensuring access to affordable housing and encouraging stable, long-term residency. This measure is part of a broader provincial strategy, launched following the "General Assembly of the Mountains," which prioritizes tackling housing issues and depopulation in mountainous regions, one of the key objectives of the provincial government.

The program targets individuals who either own or intend to acquire ownership or usage rights of residential properties in eligible municipalities. Funding can be requested for up to three housing units. Beneficiaries are required to establish their residence in the municipality or to rent the property at a moderate rate to individuals who commit to moving their official residence to the area for at least ten years. Notably, applicants must not already be residents of the municipality where the property is located, except in the case of individuals under 45, for whom this restriction does not apply. The financial incentives include contributions for renovation, covering up to 40 % of expenses in historic centers and 35 % in other areas, with a maximum grant of €80,000 (based on a maximum eligible cost of €200,000), and up to €20,000 for property purchases.

Another type of policy aims to stimulate the housing market through incentives for lending. The Valle d'Aosta region offers subsidized home loans to support first-time homeownership through new construction, purchase, or renovation, including property expansion. Eligible beneficiaries must reside in the region, have at least five years of residency (not necessarily consecutive), and an economic situation of the family not exceeding €60,000. The maximum loan amount is €180,000, raised to €200,000 for properties with energy class A or higher. Loans are repayable for up to 30 years with interest rates ranging from 0.50 % to 1.20 %, depending on income level and family composition. This initiative aims to improve residential access and energy efficiency for local households.

While initiatives such as subsidized housing loans and energy-efficiency incentives in Italy's peripheral areas are commendable, they address only a fraction of the broader challenges these regions face. Without concurrent and substantial investments in the labor market, these policies risk fostering only short-term or cyclical growth. Research indicates that European Structural Funds had limited impact on employment and population retention in Southern Italy between 2007 and 2013, highlighting the need for more integrated approaches that combine housing support with labor market development (Ciani and de Blasio, 2015). Furthermore, Italy's regional disparities, particularly in the labor market, necessitate place-sensitive territorial development policies focused on providing opportunities to residents in disadvantaged areas (Fina et al., 2021). Without such comprehensive strategies, housing policies alone may not suffice to reverse depopulation trends or stimulate sustainable economic growth in these regions. Nonetheless, stimulating real estate investment is a first important step as it could be a valuable channel for the economic recovery of these areas. In this sense, our study offers a methodology that can be exploited by other researchers to investigate other nations, although we expect that in general our results will be confirmed.

CRediT authorship contribution statement

Giuseppe Pernagallo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Giampaolo Vitali:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix (to be published with the paper)

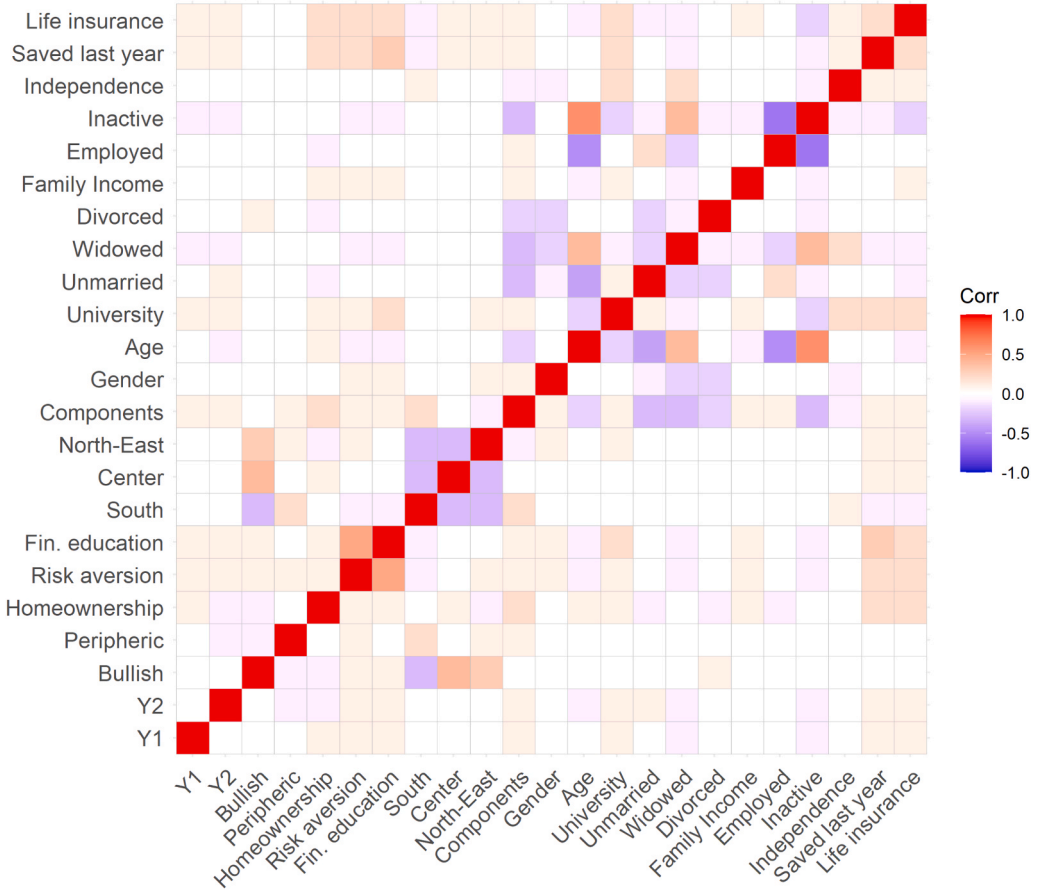


Figure A1. Pearson correlation matrix.

Table A1
Descriptive statistics.

	Mean	SD	Min	Max
Y1	0.0756	0.2644	0	1
Y2	0.0929	0.2904	0	1
Peripheric	0.0844	0.2781	0	1
Homeownership	0.7558	0.4296	0	1
Risk aversion	1.9793	1.0374	1	5
Financial education	3.4821	1.5813	1	7
South	0.2729	0.4455	0	1
Center	0.1918	0.3937	0	1
North-East	0.2039	0.4030	0	1
Bullish	0.6938	0.4609	0	1
Household components	2.4621	1.1624	1	7
Gender	0.5400	0.4984	0	1
University	0.1877	0.3905	0	1
Age	51.540	15.771	18	85
Unmarried	0.2151	0.4110	0	1
Divorced	0.0913	0.2881	0	1
Widowed	0.0824	0.2750	0	1
Net family income	2894.2	5948.9	200	222222
Employed	0.3876	0.4873	0	1
Inactive/retired	0.3371	0.4728	0	1
Age of financial independence	22.737	7.0883	10	83
Saved last year	0.5308	0.4991	0	1
Life insurance	0.1980	0.3986	0	1

Table A2
Distribution of respondents by classification of municipality of residence.

Classification	Frequencies	% Frequencies
A	814	0.3003
B	96	0.0354
C	1299	0.4792
D	273	0.1007
E	181	0.0668
F	48	0.0177
Sum	2711	1

Table A3
Contingency tables for the two dependent variables and municipality of residence classification. Row frequencies in parentheses.

	A	B	C	D	E	F
Y1						
0	758 (0.30)	83 (0.03)	1197 (0.48)	250 (0.10)	175 (0.07)	43 (0.02)
1	56 (0.27)	13 (0.06)	102 (0.50)	23 (0.11)	6 (0.03)	5 (0.02)
Y2						
0	741 (0.30)	83 (0.03)	1173 (0.48)	245 (0.10)	170 (0.07)	47 (0.02)
1	73 (0.29)	13 (0.052)	126 (0.50)	28 (0.11)	11 (0.044)	1 (0.004)

The p-values of the chi-square test are 0.0489 for Y_1 and 0.1623 for Y_2, respectively. *Expected frequency assumption* is met for both tests.

Table A4
Contingency tables for the two dependent variables and the *Periphery* variable.

	Periphery	
Y1	0	1
0	2288 (0.91)	218 (0.09)
1	194 (0.95)	11 (0.05)
Y2	0	1
0	2242 (0.91)	217 (0.09)
1	240 (0.95)	12 (0.05)

The p-values of the chi-square test are 0.0989 for Y_1 and 0.0272 for Y_2, respectively. *Expected frequency assumption* is met for both tests.

Table A5
OLS estimates with robust standard errors.

	Model (I)				**Model (II)**			
	Y1				Y1			
	Estimate	SE	t	p-value	Estimate	SE	t	p-value
Const	−0.0196	0.0516	−0.381	0.7030	−0.0410	0.0109	−3.736	0.0002
Periphery	−0.0379	0.0194	−1.953	0.0509	−0.0390	0.0158	−2.468	0.0136
Homeownership	0.0610	0.0115	5.312	0.0000	0.0623	0.0078	7.898	0.0000
Risk aversion	0.0292	0.0073	3.975	0.0001	0.0287	0.0058	4.919	0.0000
Financial education	0.0034	0.0044	0.768	0.4424				
South	0.0061	0.0170	0.354	0.7229				
Center	−0.0096	0.0193	−0.496	0.6194				
North-East	−0.0089	0.0192	−0.460	0.6453				
Bullish	−0.0005	0.0155	−0.035	0.9717				
Household components	0.0039	0.0070	0.566	0.5711				
Gender	−0.0022	0.0134	−0.166	0.8680				
University	0.0294	0.0210	1.398	0.1623	0.0346	0.0165	2.102	0.0357
Age	−0.0005	0.0006	−0.931	0.3519				
Unmarried	0.0001	0.0205	0.006	0.9951				
Divorced	−0.0037	0.0210	−0.178	0.8582				
Widowed	−0.0269	0.0189	−1.423	0.1549	−0.0363	0.0097	−3.729	0.0002
Net family income	0.0000	0.0000	0.231	0.8172				
Employed	−0.0122	0.0182	−0.674	0.5003				
Inactive/retired	0.0022	0.0215	0.104	0.9167				
Age of financial independence	−0.0001	0.0010	−0.055	0.9555				
Saved last year	0.0113	0.0122	0.925	0.3549				
Life insurance	0.0323	0.0195	1.650	0.0992	0.0574	0.0166	3.446	0.0006

(*continued on next page*)

Table A5 (*continued*)

		Model (I)		Model (II)
Control variables	All		Best subset	
	n	1967	n	2503
	AIC	295.77	AIC	208.89

Table A6
OLS estimates with robust standard errors.

	Model (III)				Model (IV)			
	Y2				Y2			
	Estimate	SE	t	p-value	Estimate	SE	t	p-value
Const	0.0901	0.0515	1.750	0.0803	0.1036	0.0237	4.360	0.0000
Periphery	−0.0582	0.0188	−3.093	0.0020	−0.0540	0.0154	−3.499	0.0005
Homeownership	−0.0694	0.0169	−4.118	0.0000	−0.0691	0.0142	−4.851	0.0000
Risk aversion	−0.0009	0.0075	−0.121	0.9033				
Financial education	0.0143	0.0049	2.922	0.0035	0.0137	0.0036	3.733	0.0002
South	0.0407	0.0192	2.115	0.0345	0.0312	0.0136	2.290	0.0221
Center	0.0143	0.0211	0.676	0.4988				
North-East	−0.0243	0.0208	−1.168	0.2428	−0.0198	0.0139	−1.421	0.1553
Bullish	−0.0050	0.0175	−0.288	0.7734				
Household components	0.0075	0.0074	1.015	0.3105				
Gender	−0.0009	0.0141	−0.069	0.9445				
University	0.0407	0.0236	1.724	0.0850	0.0467	0.0194	2.404	0.0163
Age	−0.0011	0.0006	−1.861	0.0630	−0.0012	0.0003	−4.240	0.0000
Unmarried	0.0283	0.0231	1.222	0.2220				
Divorced	0.0028	0.0230	0.124	0.9013				
Widowed	0.0079	0.0209	0.377	0.7060				
Net family income	0.0000	0.0000	1.041	0.2980				
Employed	−0.0146	0.0197	−0.743	0.4573				
Inactive/retired	−0.0251	0.0214	−1.173	0.2408				
Age of financial independence	0.0002	0.0006	0.324	0.7455				
Saved last year	0.0609	0.0135	4.492	0.0000	0.0602	0.0115	5.206	0.0000
Life insurance	0.0505	0.0198	2.547	0.0110	0.0609	0.0171	3.548	0.0004
Control variables	All				Best subset			
	n	1967			n	2509		
	AIC	578.10			AIC	581.66		

Table A7
Probit estimates with robust standard errors.

	Model (V)				Model (VI)			
	Y1				Y1			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.3980	0.4088	−5.865	0.0000	−2.6876	0.1579	−17.01	0.0000
Periphery	−0.3624	0.1770	−2.047	0.0407	−0.3908	0.1648	−2.371	0.0177
Homeownership	0.7610	0.1630	4.668	0.0000	0.8181	0.1407	5.816	0.0000
Risk aversion	0.2013	0.0467	4.305	0.0000	0.2146	0.0396	5.414	0.0000
Financial education	0.0349	0.0359	0.972	0.3310				
South	0.0565	0.1269	0.445	0.6559				
Center	−0.0732	0.1418	−0.516	0.6055				
North-East	−0.0817	0.1477	−0.553	0.5799				
Bullish	−0.0299	0.1145	−0.261	0.7935				
Household components	0.0219	0.0482	0.454	0.6493				
Gender	−0.0285	0.0965	−0.295	0.7676				
University	0.1406	0.1221	1.151	0.2495	0.1840	0.0957	1.921	0.0547
Age	−0.0055	0.0046	−1.177	0.2390				
Unmarried	−0.0121	0.1467	−0.083	0.9338				
Divorced	−0.0236	0.1731	−0.136	0.8914				
Widowed	−0.5431	0.2632	−2.063	0.0391	−0.6715	0.2374	−2.828	0.0047
Net family income	0.0000	0.0000	0.085	0.9315				
Employed	−0.0777	0.1100	−0.706	0.4799				
Inactive/retired	0.0075	0.1459	0.051	0.9588				
Age of financial independence	−0.0007	0.0109	−0.067	0.9464				
Saved last year	0.1037	0.0996	1.041	0.2978				
Life insurance	0.1661	0.1075	1.545	0.1223	0.3243	0.0899	3.604	0.0003

(*continued on next page*)

Table A7 (*continued*)

		Model (V)		Model (VI)
Control variables	All		Best subset	
	n	1967	n	2503
	AIC	990.04	AIC	1167.29

Table A8
Probit estimates with robust standard errors.

	Model (VII)				Model (VIII)			
	Y2				Y2			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−1.5467	0.3192	−4.845	0.0000	−1.3408	0.1632	−8.218	0.0000
Periphery	−0.5120	0.1963	−2.608	0.0091	−0.4779	0.1788	−2.673	0.0075
Homeownership	−0.4774	0.1034	−4.615	0.0000	−0.4595	0.0884	−5.195	0.0000
Risk aversion	0.0143	0.0462	0.311	0.7556				
Financial education	0.1002	0.0332	3.012	0.0026	0.1070	0.0265	4.031	0.0001
South	0.2706	0.1168	2.316	0.0206	0.2244	0.0898	2.497	0.0125
Center	0.1002	0.1355	0.739	0.4596				
North-East	−0.1712	0.1510	−1.134	0.2570	−0.1409	0.1055	−1.335	0.1819
Bullish	−0.0403	0.1128	−0.357	0.7206				
Household components	0.0719	0.0416	1.727	0.0842				
Gender	−0.0174	0.0917	−0.190	0.8493				
University	0.2002	0.1099	1.821	0.0686	0.2493	0.0930	2.681	0.0073
Age	−0.0084	0.0040	−2.101	0.0357	−0.0108	0.0023	−4.669	0.0000
Unmarried	0.2084	0.1277	1.632	0.1027				
Divorced	0.0052	0.1580	0.033	0.9733				
Widowed	−0.0008	0.2314	−0.003	0.9971				
Net family income	0.0000	0.0000	1.514	0.1300				
Employed	−0.0851	0.1034	−0.823	0.4103				
Inactive/retired	−0.2213	0.1433	−1.544	0.1226				
Age of financial independence	0.0013	0.0061	0.218	0.8274				
Saved last year	0.4472	0.0997	4.483	0.0000	0.4433	0.0889	4.984	0.0000
Life insurance	0.3111	0.1034	3.007	0.0026	0.3696	0.0888	4.160	0.0000
Control variables	All				Best subset			
	n	1967			n	2509		
	AIC	1096.96			AIC	1323.87		

Table A9
IV probit estimates with robust standard errors, with *Index* as the instrumental variable. *Bullish* variable is omitted to avoid collinearity with *Index*. Control variables also include *Unemployment* (regional unemployment rate in current year).

	Model (IX)				Model (X)			
	Y1				Y2			
First stage (dependent is Homeownership)					**First stage (dependent is Homeownership)**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Index	−0.0001	0.00003	−3.46	0.001	−0.0001	0.00003	−3.46	0.001
	Y1				Y2			
IV estimates					**IV estimates**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−1.6015	0.8390	−1.91	0.056	−1.1789	0.6734	−1.75	0.080
Periphery	−0.2852	0.2016	−1.42	0.157	−0.4783	0.2159	−2.22	0.027
Homeownership	0.1578	1.3251	0.12	0.905	−0.5562	1.2875	−0.43	0.666
Control variables		All				All		
	n	1967			n	1967		
	AIC	2948.09			AIC	3057.18		

Table A10
IV probit estimates with robust standard errors, with *Index* as the instrumental variable. The best subset is the same as in the previous section.

	Model (XI)				Model (XII)			
	Y1				Y2			
First stage (dependent is Homeownership)					**First stage (dependent is Homeownership)**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Index	−0.00001	0.00002	−0.37	0.709	−0.00002	0.00002	−1.12	0.261
	Y1				Y2			
IV estimates					**IV estimates**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−2.2590	11.6778	−0.19	0.847	−1.5033	1.0809	−1.39	0.164
Periphery	−0.3479	1.2469	−0.28	0.780	−0.5033	0.2200	−2.29	0.022
Homeownership	0.2678	13.3650	0.02	0.984	0.0404	4.1008	0.01	0.992
Control variables	Best subset				Best subset			
	n	2503			n	2509		
	AIC	4053.07			AIC	4089.82		

Table A11
IV probit estimates with robust standard errors, with *Inheritance* as the instrumental variable.

	Model (XIII)				Model (XIV)			
	Y1				Y2			
First stage (dependent is Homeownership)					**First stage (dependent is Homeownership)**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Inheritance	0.0866	0.0189	4.59	0.000	0.0866	0.0189	4.59	0.000
	Y1				Y2			
IV estimates					**IV estimates**			
	Estimate	SE	Z	p-value	Estimate	SE	Z	p-value
Const	−0.6623	0.6084	−1.09	0.276	−1.091	0.6171	−1.77	0.077
Periphery	−0.1665	0.1525	−1.09	0.275	−0.4627	0.2009	−2.30	0.021
Homeownership	−1.1682	0.5990	−1.95	0.051	−0.7410	1.0286	−0.72	0.471
Risk aversion	0.1393	0.0524	2.66	0.008	0.0151	0.0461	0.33	0.743
Financial education	0.0536	0.0293	1.83	0.068	0.1042	0.0365	2.85	0.004
South	0.8152	0.2496	3.27	0.001	0.8787	0.3171	2.77	0.006
Center	0.1376	0.1295	1.06	0.288	0.1381	0.1789	0.77	0.440
North-East	−0.2575	0.1284	−2.00	0.045	−0.2911	0.1626	−1.79	0.074
Bullish	−0.0875	0.1057	−0.83	0.408	0.0082	0.1394	0.06	0.953
Household components	0.0954	0.0448	2.13	0.033	0.0858	0.0597	1.44	0.151
Gender	−0.0952	0.0811	−1.17	0.241	−0.0212	0.1061	−0.20	0.842
University	0.2194	0.1001	2.19	0.028	0.2235	0.1223	1.83	0.068
Age	0.0060	0.0054	1.11	0.268	−0.0070	0.0077	−0.91	0.363
Unmarried	−0.1191	0.1200	−0.99	0.321	0.1824	0.1461	1.25	0.212
Divorced	−0.2498	0.1535	−1.63	0.104	−0.0313	0.2013	−0.16	0.876
Widowed	−0.5685	0.2150	−2.64	0.008	0.0012	0.2598	0.00	0.996
Net family income	0.0000	0.0000	1.24	0.216	0.0000	0.0000	1.19	0.233
Employed	−0.2019	0.0892	−2.26	0.024	−0.1066	0.1328	−0.80	0.422
Inactive/retired	−0.1376	0.1236	−1.11	0.265	−0.2470	0.1655	−1.49	0.136
Age of financial independence	−0.0009	0.0085	−0.11	0.916	0.0016	0.0062	0.26	0.797
Saved last year	0.2549	0.0858	2.97	0.003	0.4724	0.1378	3.43	0.001
Life insurance	0.3121	0.0919	3.39	0.001	0.3198	0.1480	2.16	0.031
Unemployment	−0.0849	0.0256	−3.31	0.001	−0.0654	0.0322	−2.03	0.042
Control variables		All				All		
	n	1967			n	1967		
	AIC	2927.71			AIC	3041.57		

References

Adair, A., Berry, J., McGreal, S., 1995. Property investment in peripheral regions. J. Prop. Financ. 6 (2), 43–55.

Baur, K., Rosenfelder, M., Lutz, B., 2023. Automated real estate valuation with machine learning models using property descriptions. Expert Syst. Appl. 213 (Part C), 119147.

Chetty, R., Sandor, L., Szeidl, A., 2017. The effect of housing on portfolio choice. J. Financ. 72 (3), 1171–1212.

Ciani, E., de Blasio, G., 2015. European structural funds during the crisis: evidence from Southern Italy. IZA J. Labor Policy 4, 20 (Article number:).

Cocco, J., 2005. Portfolio choice in the presence of housing. Rev. Financ. Stud. 18 (2), 491–533.

Colabella, A., Guglielminetti, E., Rondinelli, C., 2023. The distribution and use of Italian households' savings after the pandemic. Quest. di Econ. e Finanz. (Occas. Pap.)no. 797.

Coniglio, N.D., Prota, F., 2008. Human capital accumulation and migration in a peripheral EU region: the case of basilicata. Pap. Reg. Sci. 87 (1), 77–96.

Cui, C., Wang, Z., He, P., Yuan, S., Niu, B., Kang, P., Kang, C., 2019. Escaping from pollution: the effect of air quality on inter-city population mobility in China. Environ. Res. Lett. 14, 124025.
DiPasquale, D., Glaeser, E.L., 1999. Incentives and social capital: are homeowners better citizens? J. Urban Econ. 45 (2), 354–384.
Fina, S., Heider, B., and Prota, F. (2021). Unequal Italy: Regional socio-economic disparities in Italy. Friedrich-Ebert-Stiftung – Politics for Europe, available at ⟨https://feps-europe.eu/wp-content/uploads/downloads/publications/210712%20unequal%20italy%20report.pdf?utm_source=chatgpt.com⟩.
Fortuna, M.J.A., Dentinho, T.P., Vieira, J.A.C., 2001. The costs of peripherality. Eur. Parliam. Dir. Gen. Res. REGI 111, 01–2001 (EN).
Gatrell, J.D., 1999. Re-thinking economic development in peripheral regions. Soc. Sci. J. 36 (4), 623–639.
Goryunov, M., Kokovin, S., 2016. Vanishing cities': can urban costs explain deindustrialization? Pap. Reg. Sci. 95 (3), 633–652.
Guzzardi, D., Morelli, S., 2024. A new geography of inequality: top incomes in Italian regions and inner areas. World Inequal. Lab Work. Pap. no. 2024/19.
Ianioglo, A., Rissanen, M., 2020. Global trends and tourism development in peripheral areas. Scand. J. Hosp. Tour. 20 (5), 520–539.
Iwaisako, T., Ono, A., Saito, A., Tokuda, H., 2022. Disentangling the effect of home ownership on household stockholdings: evidence from Japanese micro data. Real. Estate Econ. 50 (1), 268–295.
Kaas, L., Kocharkov, G., Preugschat, E., 2019. Does homeownership promote wealth accumulation? Appl. Econ. Lett. 26 (14), 1186–1191.
King, G., Zeng, L., 2017. Logistic regression in rare events data. Political Anal. 9 (2), 137–163.
Kopczewska, K., 2021. Spatial machine learning: new opportunities for regional science. Ann. Reg. Sci. 68, 713–755.
Kubeš, J., Chvojkov, A., 2020. Back to peripheries based on remoteness. Human capital in the peripheral municipalities of south bohemia. J. Rural Stud. 79, 116–124.
McDonald, C., Buckley, P.J., Voss, H., Cross, A., Chen, C., 2018. Place, space, and foreign direct investment into peripheral cities. Int. Bus. Rev. 27 (4), 803–813.
Michielsen, T., Mocking, R., van Veldhuizen, S., 2016. Home ownership and household portfolio choice. CESIFO Work. Pap. 5705 (No.).
Nagy, R.A., Obenberger, R.W., 1994. Factors influencing individual investor behavior. Financ. Anal. J. 50 (4), 63–68.
Nilsen, T., Grillitsch, M., Hauge, A., 2023. Varieties of periphery and local agency in regional development. Reg. Stud. 57 (4), 749–762.
Pearce, D.G., 2002. Tourism and peripherality: perspectives from Asia and the south pacific. Tour. Hosp. Res. 3 (4), 295–309 (Special issue on tourism in peripheral areas).
Pernagallo, G., Torrisi, B., Bennato, D., 2021. A classification algorithm to recognize fake news websites. In: Mariani, P., Zenga, M. (Eds.), Data Science and Social Research II. DSSR 2019. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Cham.
Pugh, R., Dubois, A., 2021. Peripheries within economic geography: four "problems" and the road ahead of us. J. Rural Stud. 87, 267–275.
Sandow, E., Lundholm, E., 2023. Leaving the city: counterurbanisation and internal return migration in Sweden. Eur. J. Popul. 39 (7).
Stotz, O., 2019. The perception of homeownership utility: Short-term and long-term effects. J. Hous. Econ. 44, 99–111.
Tsai, I., 2019. Relationships among regional housing markets: evidence on adjustments of housing burden. Econ. Model. 78, 309–318.
Tsai, I., Chiang, Y., Lin, S., 2022. Effect of COVID-19 lockdowns on city-center and suburban housing markets: evidence from hangzhou, China. J. Asian Econ. 83, 101544.
Van Nieuwerburgh, S., 2023. The remote work revolution: impact on real estate values and the urban environment: 2023 AREUEA presidential address. Real. Estate Econ. 51 (1), 7–48.
Vogl, T., Micek, G., 2022. Why do they go to the peripheries? Studying the relations between the real estate market and coworking spaces in the peripheral areas of Germany. J. Prop. Invest. Financ. 40 (5), 479–492.
Voith, R., 1999. The suburban housing market: effects of city and suburban employment growth. Real. Estate Econ. 27 (4), 621–648.
Wright, D., Yanotti, M.B., 2019. Home advantage: the preference for local residential real estate investment. Pac. Basin Financ. J. 57, 101167.
Yamamura, E., 2011. Comparison of the effects of homeownership by individuals and their neighbors on social capital formation: evidence from Japanese general social surveys. J. SocioEcon. 40 (5), 637–644.
Zhang, Y., Zhu, J., Zhu, J., Wang, X., 2023. A splicing approach to best subset of groups selection". In: INFORMS Journal on Computing, 35. INFORMS, pp. 104–119.
Zhu, J., Wen, C., Zhu, J., Zhang, H., Wang, X., 2020. A polynomial algorithm for best-subset selection problem. Proc. Natl. Acad. Sci. 117 (52), 33117–33123.