TÉCNICO LISBOA

Immigration and Real Estate Prices in Portugal: A Panel Data Analysis

André Miguel Machado Querido

Dissertation to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisors:

Prof. Hugo Miguel Fragoso de Castro Silva

Prof. António Sérgio Constantino Folgado Ribeiro

Examination Committee:

Chairperson: Prof. Tânia Rute Xavier de Matos Pinto Varela

Supervisor: Prof. António Sérgio Constantino Folgado Ribeiro

Member of the committee: Prof. João Pereira dos Santos

November 2023

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I want to start by saying how grateful I am to both my thesis advisors, Professor Hugo Castro Silva and Professor Antonio Sérgio Ribeiro. Thank you for the constant guidance and encouragement throughout the making of this dissertation.

Now that this journey comes to an end, I cannot express how thankful I am for the day that my 18-year-old self decided to apply for Instituto Superior Técnico. What a ride. This amazing academy gifted me with moments that tested my limits and made me emerge stronger and wiser.

Here I met my amazing girlfriend, Eduarda. Your unwavering support and encouraging words during the most stressful moments of this journey have been my anchor. This dissertation, which symbolizes the culmination of my academic efforts, is as much yours as it is mine. Thank you for always being there.

Here I met friends that I will take for life, and I hope that no matter where life takes us, there will always be occasions to reminisce about the spectacular five years we spent together. To João, Leonor, Diogo, and Luiz – you have been my partners in crime, my confidants, and my companions on this incredible journey. The countless hours we spent together in the Civil Engineering Department, the shared laughter, the breaks, and the much-needed nights out are all memories I will cherish forever. It's not just about the destination; it's about the unforgettable moments we've had along the way.

Here I met brilliant professors that inspired me to stay hungry and pursue my own path, molding me into the person I proudly am today. You have instilled in me the hunger for knowledge, the drive to excel, and the courage to forge a unique path, and for that, I am deeply thankful.

Above all, here I met myself. It has been a journey of self-discovery, self-improvement, and self-realization. It's not just the knowledge I've gained, but the person I've become that I am most proud of.

Thank you to my parents, grandparents, and family. Your love and support have made all this possible. Your dedication and sacrifices to ensure I could pursue my dreams have not gone unnoticed. I can never thank you enough for the opportunities you've provided and the sacrifices you've made.

Finally, I want to thank my sister, Micaela. You've been my lifelong mentor and my guiding light. Your wisdom, guidance, and encouragement have been constants in my life ever since I can remember. Your tireless belief in my abilities has motivated me to overcome every obstacle and to be the best version of myself. I am deeply grateful for everything you have ever done for me, and I hope I can repay you someday by being a man worth calling your brother.

As I look forward to the future, I carry with me the lessons learned, the friendships forged, and the love and support. The end of this chapter is not a goodbye, but a turning point in my life. With a heart full of gratitude and a spirit full of ambition, I am excited to see where the next chapter of my journey will take me. Thank you, Instituto Superior Técnico, for this incredible adventure, and thank you to everyone who has been a part of it.

Abstract

The global economy's cornerstone, the real estate market, profoundly influences the well-being of individuals and the Portuguese economy. Over the past decade, housing prices in Portugal have surged, defying the trends of Gross Domestic Product (GDP) per capita and average wages. This surge is attributed, in part, to a complex interplay of factors, including immigration. This research aimed to examine the impact of immigration on Portuguese real estate prices and offers a comprehensive analysis of past research in the Portuguese economy, the real estate market, and housing price determinants. This analysis, grounded in econometric methodology, revealed a significant and consistent negative effect of immigration on housing prices in Portugal. This was accomplished through the examination of regional housing price data spanning from 2011 to 2020. To uncover the intricate relationship between immigration and housing prices, it was used a balanced panel data, which allowed to study the impact of immigration over time. The application of various econometric models, including fixed effects, first differences, and instrumental variable regressions, established a robust negative relationship between immigration and housing prices in Portugal. The obtained results indicate that a one-percentage-point increase in immigration inflow corresponds to housing price reductions ranging from 7.8% to 14.7. This research contributes to the existing body of knowledge, providing important insights for policymakers to address affordability concerns, investors to develop sustained investment strategies, contributes to a better comprehension of how Portuguese housing prices respond to various determinants and how immigration affects different markets around the world.

Key Words: Immigration; Housing Prices; Real Estate Market; Portuguese Economy; Instrumental Variables.

Resumo

O mercado imobiliário, exerce uma profunda influência no bem-estar das pessoas. Na última década, os preços das habitações em Portugal dispararam, desafiando as tendências do Produto Interno Bruto (PIB) per capita e dos salários médios. Este aumento é atribuído a uma complexa interação de fatores, incluindo a imigração. Esta pesquisa teve como objetivo examinar o impacto da imigração nos preços imobiliários em Portugal e oferece uma análise abrangente de pesquisas anteriores na economia portuguesa, no mercado imobiliário e nos determinantes dos preços das habitações. Esta análise, revelou um efeito negativo significativo da imigração nos preços das habitações em Portugal. Isso foi alcançado através da análise de dados regionais de preços das habitações de 2011 a 2020. Para desvendar a relação entre imigração e preços das habitações, utilizou-se um conjunto de dados em painel equilibrados, que permitiu estudar o impacto da imigração ao longo do tempo. A aplicação de vários modelos econométricos, incluindo efeitos fixos, primeiras diferenças e regressões com variáveis instrumentais, estabeleceu uma relação negativa robusta entre imigração e preços das habitações em Portugal. Os resultados obtidos indicam que um aumento de um ponto percentual na entrada de imigração corresponde a reduções nos preços das habitações que variam de 7,8% a 14,7%. Esta pesquisa fornece informações importantes para os decisores políticos abordarem problemas de acessibilidade, para investidores desenvolverem estratégias de investimento sustentadas e para uma melhor compreensão de como os preços das habitações em Portugal respondem a vários fatores e como a imigração afeta diferentes mercados em todo o mundo.

Palavras-chave: Imigração; Preço da Habitação; Mercado Imobiliário; Economia Portuguesa; Variáveis Instrumentais

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

2SLS	Two Stage Least Squares
CPI	Consumer Price Index
EU	European Union
FD	First Differences
FDI	Foreign Direct Investment
FE	Fixed-Effects
GDP	Gross Domestic Product
IMI	Imposto Municipal sobre Imóveis
IMT	Imposto Municipal sobre as Transmissões Onerosas de Imóveis
INE	Instituto Nacional de Estatística
IV	Instrumental Variables
LFS	Labour Force Survey
OLS	Ordinary Least Squares
RBI	Residence By Investment
US	United States

1. Introduction

The real estate market plays a fundamental role in the global economy, impacting the lives of more individuals than any other single commodity and it is closely tied with the well-being of economic agents (Fão, 2019; Pivar & McKenzie, 2008). Over the past decade, Portuguese housing prices have more than doubled.[1] Surprisingly, macroeconomic factors like Gross Domestic Product (GDP) per capita and average wages have not followed the same trend. One possible driver of rising housing prices is population growth, leading to an increased demand for housing. According to Eurostat data, the number of foreign-born individuals in Portugal increased by 36% since 2013. On the other hand, the native and total population have both decreased by 4.9% and 1.3% respectively during the same period.[2] However, it is not clear-cut how immigration affects housing prices, and how other factors like native out-migration and decreasing income levels can counteract the increased demand. This complex interplay of factors has led to a housing crisis in Portugal, with many individuals believing that expats and digital nomads (attracted by the climate, affordability, and a favorable tax regime) are outpricing locals and inflating housing costs. But is this really the case? Saiz (2007) inferred on the same problem related to the US market and found that immigrants were indeed increasing the housing prices. On the other hand, Sá (2015), Accetturo et al. (2014) and other researchers reported immigration had a negative effect in housing prices on the British and Italian market correspondently.

This study seeks to assess the impact of immigration on Portuguese real estate prices, while offering an in-depth analysis of the Portuguese economy, the real estate market, and housing price determinants. By examining the dynamics of the Portuguese economy and the intricate relationship between housing prices and various factors, we aim to shed light on the complex housing market in Portugal. To do this, the regional housing price data from 2011 to 2020 was analyzed by relating annual housing price changes to the annual variation in the number of foreign-born individuals relative to the initial total population. Data from the Portuguese employment survey or Labor Force Survey (LFS), conducted by the *Instituto Nacional de Estatística* (INE) were used by us in this analysis, which provided essential insights into key attributes of the Portuguese population, such as: age, gender, marital status, education, employment, and earnings. Additionally, data from the Portuguese Property Transfer Tax (IMT) and the Municipal Property Tax (IMI) databases offered comprehensive information regarding housing prices and other characteristics. This analysis used balanced panel data, revealing that a one-percentage-point increase in immigration inflow led to housing price decreases ranging from 7.8% to 14.7%. Year fixed-effects were incorporated in our models to account for macroeconomic trends and a set of control variables to address time-varying region-specific effects was included. In line with previous research, different estimators were applied, including First Differences, Fixed Effects to control for unobserved time-invariant factors, and Instrument Variables regressions to address endogeneity issues. For the Instrumental Variable (IV) regressions, it was used a widely accepted shift-share instrument,

[1] https://tradingeconomics.com/portugal/housing-index accessed on October 18, 2023.
[2] https://ec.europa.eu/eurostat/databrowser/view/migr_pop3ctb/default/table?lang=en accessed on October 18, 2023

developed by Card (2001) and based on the historical distribution of foreign-born individuals across Portugal.

While numerous studies have explored similar topics, this research provides additional knowledge on the regional impact of immigration on the Portuguese real estate market. It is the first study to make such inferences, adding a valuable piece of information to the existing literature. The research sheds light on the potential economic policy implications of immigration on the housing market. Policymakers can use this information to develop strategies for managing housing prices and addressing affordability concerns. Investors and real estate professionals can benefit from understanding how immigration affects housing prices and adjust their investment strategies based on the findings, potentially identifying opportunities in regions where immigration is stronger. By offering a comprehensive analysis of the subject, the research contributes to public awareness. It informs the general public about the complex dynamics of the real estate market and how immigration plays a role in shaping housing prices.

The remainder of this document is structured as follows: Section 2 presents a comprehensive literature review, establishing the theoretical framework for our research by reviewing existing studies on the Portuguese economy, the real estate market, housing price determinants, and the role of immigration in the housing market and its influence on housing prices. Section 3 characterizes the data used in our analysis, offering a detailed description of the datasets. Section 4 provides the theoretical background and outlines the formulation of our estimation models. In Section 5, the results of the econometric analysis are presented, concerning the impact of various factors on housing prices in Portugal. Multiple models and estimation methods to thoroughly explore the relationship between housing prices and demographic, macroeconomic, and immigration factors were employed. Finally, in Section 6, there is a summary of our work, presenting the main conclusions and offering suggestions for future research.

2. Literature Review

2.1. The Portuguese Economy

In 1974, Portugal experienced a significant political transformation through the overthrow of a 50-year military dictatorship. This upheaval resulted in the nationalization of most Portuguese banks and financial institutions. These nationalized entities adopted a lending policy characterized by fixed interest rates and a limited emphasis on promoting capital investment. Subsequently, in 1986, Portugal, in conjunction with Spain, undertook a momentous step by joining the European Union, following the third enlargement of the European Communities. Figure 2.1 illustrates that during this period, the Portuguese population underwent a process of aligning their consumption patterns more closely with the European average. It was postulated that this integration would lead to a continuation of a similar growth trajectory in the Portuguese economy during the subsequent decades, facilitated by the financial integration within the European Union (Teles Morais, 2018).

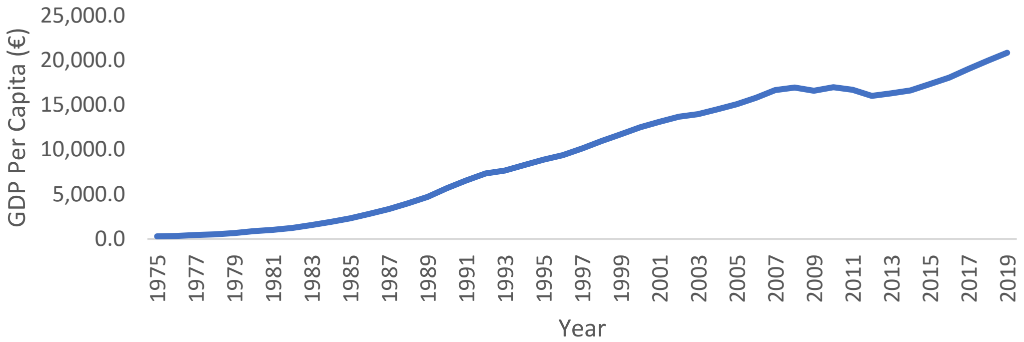

Figure 2.1: Portuguese GDP per capita Note: Data from 1970 to 2022 (Source: INE)

In the 2000s, expansionary fiscal policies, such as raising public sector's wages and pensions, resulted in significant budget deficits and increased public debt. Furthermore, during that period, there was a speculative bubble in the IT sector, which eventually burst at the turn of the millennium. Consequently, global markets experienced a slowdown as we entered a phase characterized by low consumer confidence, commonly referred to as a bear market. This era, often referred to as the "Dot-com downfall", aggravated the country's economic stability, and the macroeconomic metrics reflected the economic difficulties felt during this period. Portugal's GDP growth clearly declined after 2000, going from 3.8% to a minimum value of -0.9% in 2003[3] and represented a period of serious economic stagnation (Figure 2.2).

[3] https://databank.worldbank.org/source/world-development-indicators accessed on May 8, 2023.

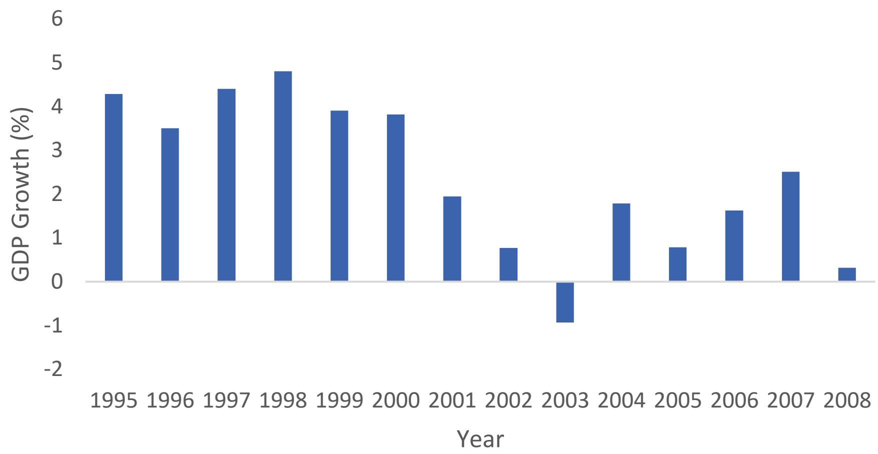

Figure 2.2: Portuguese GDP Growth Note: Data from 1995 to 2008 (Source: World Bank Data)

The 2008 financial crisis was also harshly felt in Portugal and considering it was originated from the real estate sector, it is imperative to dive into the causes of the economic downturn. In the decade of 1990, the housing sector was facing constant growth and multiple investors were turning their attention to the market. In 1994 the United States (US) government made significant revisions to the Community Reinvestment Act (CRA). These revisions encouraged banks and other financial institutions to meet the credit needs of low- and average-income households, ensuring banks provided housing credit to underserved populations. The government had the objective of providing its own dwelling to every citizen. Additionally, and following the "Dot-com" downfall, the US government was forced to lower the interest rates, leading to facilitated credit access and a considerable increase in the quantity of mortgages being taken. Financial institutions decided to bundle subprime mortgages and other mortgage loans into complex financial products, known as mortgage-backed securities (MBS) and collateralized debt obligations (CDOs) and sold them to investors. Originally, these were well built and supervised instruments, nonetheless, as the years went by, regulatory agencies failed to adequately regulate the financial institutions involved in these risky practices and allowed riskier products to thrive. In 2008, some of the US largest banks defaulted due to the volume of faulty real estate assets they owned, and it wasn't long before the crisis heavily impacted the European financial system. Portugal's economy also suffered a downturn as records show a GDP per capita growth of -3.2% in 2009 and -3.7% in 2012,[4] representing the metric's worst performance since 1975. This directly impacted the population, as the unemployment rate went from 8.0% in 2007 to 17.1% in 2013 (Figure 2.3). According to Morais (2018) the Portuguese real GDP per capita, grew less during this period, than the US in the Great Depression, and in average terms, Portugal's economic state was worse in 2012 than in 2000.

In the following years, the situation exhibited a continuing deterioration. In response to the escalating debt crisis, the European Union (EU) and the International Monetary Fund (IMF) were compelled to intervene in 2011, and a series of austerity measures and structural reforms were instituted under the

[4] https://databank.worldbank.org/source/world-development-indicators accessed on May 9, 2023.

program known as the Financial Assistance Programme (FAP).[5] The primary objectives of this program were to bolster Portugal's efforts in reinstating market confidence, curbing public debt, and averting a financial crisis with the potential to exert significant repercussions on the Eurozone and the broader European financial system.[6] Between 2011 and mid-2014, a collaborative financial package amounting to €78 billion was extended to the Portuguese government. During the period from 2014 to 2018, the country recorded an average annual GDP growth rate of 2.2%, concurrently witnessing a reduction in unemployment and emigration (Puig & Sánchez, 2018). In addition to these measures, aligning with strategies implemented by countries such as the United Kingdom to address short-term economic gaps, Portugal opted to introduce a Residence by Investment (RBI) program, colloquially referred to as the "Golden Visa program."

Economic growth is significantly augmented through investment, as indicated by prior research (Colen et al., 2009). The importance of investment as an essential tool used by governments to improve overall economic well-being is highlighted by this. Foreign direct investment (FDI) is an ownership stake in a foreign company or project made by an investor, company, or government from another country with the intent to manage that asset[7]. Host countries often encourage FDI by offering favorable tax regimes and grants to attract foreign investors though Residence by Investment programs. "Residence by investment" (RBI) programs are a form of FDI and consist in a group of policies that allow foreign investors to obtain a residence in the country in return for investing a certain amount of capital. Such programs take place since FDI is a way of increasing employment, import technology and know-how, promote trade, and increasing productivity levels (Alfaro, 2017). However, it is important to acknowledge the presence of disadvantages, notably the need to navigate the regulatory landscapes of multiple governments, which inevitably elevates the level of political risk inherent in such investments. A notable example is the government initiative that aimed at boosting FDI through the Portuguese Golden Visa program, often believed to have led wealthy immigrants to outprice natives in the real estate market. Surak & Tsuzuki (2021) examined the economic impact of these programs and concluded that, although the majority of the investment from these programs flows into the real estate market, the effect on housing prices is minimal. However, when analyzing the economic effect of the Golden Visa program on the Portuguese real estate market, Scherrer & Thirion (2018) state that the program generated €3.5 billion from 2012 to 2018, and the number of real estate transactions increased by more than 100% in the same period. The authors suggest that the rapid increase in applications led to a steep rise in prices. Nonetheless, it is important to consider that Portugal was recovering from a severe financial crisis, and housing transactions had been significantly low in the previous years.

Subsequent to 2015, although the FAP program formally concluded, the EU and IMF continued to vigilantly oversee Portuguese fiscal stability in the ensuing years. Portugal's GDP growth rates exhibited a sustained increase, reaching a peak of 3.5% in 2017, surpassing the EU average for that period. A parallel pattern was evident in the evolution of unemployment rates, which declined from 14.7% in 2014

---

[5] https://economy-finance.ec.europa.eu/eu-financial-assistance/euro-area-countries/financial-assistance-portugal_en accessed on May 10, 2023.
[6] https://www.imf.org/en/Countries/PRT/portugal-lending-case-study accessed on May 10, 2023.
[7] https://www.investopedia.com/terms/f/fdi.asp accessed on May 30, 2023.

to 6.7% in 2019, falling below the European average (see Figure 2.3). Furthermore, the government implemented concerted efforts to reduce public debt, resulting in a decline in the Portuguese debt-to-GDP ratio from 131.4% in 2014 to 116.6% in 2019.

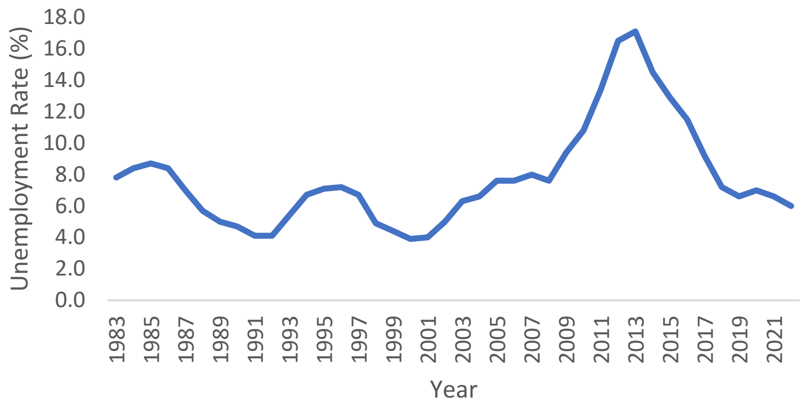

Figure 2.3: Portugal's Unemployment Rate Note: Data from 1983 to 2022 (Source: PORDATA)

2.2. Immigration and the housing market

The housing market is directly affected by the household income of the country or region. Past research, using historical data, finds a negative relationship between immigration and wage level in periods remarking to the beginning of the 20th century (Ferrie, 1996; Goldin, 1995). Immigrants also tend to set themselves in low-quality dwellings during the first years of their arrival (Michael Schill & Rosenbaum, 1998). Simply by looking at these affirmations, one could easily conclude that a new wave of immigration would lead to a downfall in housing prices. However, research has shown that it is not that simple. There is no consensus on the link between immigration and the real estate market. Multiple factors regarding scope, migrants' behavior, and the host country's housing market elasticity led researchers to fail to find consistent results when addressing such questions.

Theoretically, an increase in the overall population originated, for example, by a wave of immigration, should lead to an increase in housing demand and, consequently, to an increase in housing prices (Saiz, 2007). Saiz (2007) used a fixed effects and a Two Stage Least Squares (2SLS) regression using data from 1979 to 1983 and found that, in the short term, immigration led to an increase in housing prices and rents. (Akbari & Aydede, 2012) also found a positive correlation between immigration and real estate prices using data from the Canadian housing market from 1996, 2001, and 2006. However, the relationship is not as straightforward as one might predict, for multiple reasons. Local perceived amenities are highly important for native residents. Also, historically, immigrants tend to cluster together when moving to a new city or region, even sharing dwellings among multiple families, outpricing natives and forming well-known intra-communities within the city borders (Card, 2001). This behavior leads natives to depreciate their perception of local amenities and move out to other regions. This exodus effect offsets the inflow of immigrants, not necessarily changing the overall population size. Gonzalez & Ortega (2009) tried to infer the migration effect on the Spanish real estate market from 1998 to 2008, and according to their findings, if the regional inflow of immigrants is higher than the outflow of natives in a certain region, then there is a positive correlation between immigration and housing prices, even implying that immigration accounted for 30% of the price increase in that period. Nonetheless, the authors found that if the regional outflow of natives offsets the inflow of immigrants, then there should be no relationship between both variables.

Carter (2005) stated that in certain regions incoming immigrants are generally from a lower socioeconomic status, thus looking for poorly paid jobs, depreciating the local salary level, which leads natives to migrate. This effect is often referred to as an "Income effect," and it also offsets the inflow of immigrants into a country or region. Saiz & Wachter (2011) used census data from the years 1980, 1990, and 2000 and discovered a negative correlation between immigration and housing prices in the US. The authors highlighted the income effect, immigrant clustering, and the lower housing demands as the main drivers for this relationship. Sá (2015) studied the connection between immigration and real estate in England and Wales using regression analysis and found that house prices tended to decrease in areas with a high supply of low-skilled jobs, as migrants were clustering in such areas, forcing natives out of them. Along these lines, Zhu et al (2019) inferred on the English and Welsh Labor market and confirmed these results.

There is also another important factor that researchers consider when drawing results. In the same way immigration inflows may influence housing prices, lower housing prices also tend to influence the inflows of immigration, as immigrants tend to choose countries with lower housing costs. These kinds of problems are often referred to as reverse causality problems, and some authors suggest that it may explain the negative relationship found in some studies (Saiz & Wachter, 2011). Saiz (2007) and Saiz & Wachter (2011) used an IV regression approach to address this relationship. Additionally, Zabel (2012) used data from 277 metropolitan areas and, using a panel VAR approach, found that migrants were choosing their destinations based on housing prices and job supply.

There are multiple factors that influence the real estate market behavior. From our literature review, we could state that the predominant ones identified by researchers were immigrant clustering, education, job supply, housing supply and its elasticity, local amenities, and the immigrant perception by the native population. As stated, the results on the sign of the impact of immigration have been mixed. Saiz (2007) found that an immigration inflow of 1% relative to the city's total population represented a housing price increase of 2.9% to 3.4% in the US market. Fischer & Degen (2009) found that an inflow of 1% increased family home prices by 2.7% on the Swiss market. On the other hand, in the British market, Sá (2015) found that an immigration inflow of 1% of the initial local population led to a 1.6% decrease in housing prices. As such, the main focus of this study will be to test out this hypothesis on the Portuguese real estate market and to infer the immigration impact on the housing prices.

2.3. Real Estate Market Characterization

The real estate market constitutes a fundamental pillar of the global economy. The assets transacted on the real estate market touch more lives than any other single commodity and have a strong correlation with the quality of life of the economic agents (Fão, 2019; Pivar & McKenzie, 2008). Portugal was always characterized by a high preference for home ownership over private rentals (Moreira Braga, 2013).

Furthermore, dwelling expenses represent the highest financial cost for most households, reaching a median value of 10.5% in 2021 for the Portuguese population (INE, 2022). Apart from fixed expenses, housing services have the largest share in the consumption package of most households (Dias & Duarte, 2019).

At its essence, this market entails the processes of acquisition, disinvestment, leasing, and development of properties. These properties find utility in diverse sectors, including residential, commercial, industrial, and institutional applications. Furthermore, the real estate market is split into two distinct segments: residential properties and commercial properties, a division that arises from their discernible disparities in market dynamics. The residential market shares resemblances with fixed income markets, which are influenced by overall fluctuations in interest rates and individual credit evaluations. On the other hand, Tunaru (2017) stated that the commercial market can be seen as a combination of a bond and equity investment. The short-term lease system agreed between a landlord and a tenant resembles a steady stream of income, as bonds do, and the fact that the property may appreciate or depreciate over time resembles an equity investment, thus being characterized as a hybrid investment.

Housing prices diverge across countries, cities, and villages being determined by location, space, and consumption patterns. Additionally, a range of economic, social, and environmental factors also influence the real estate market. This includes factors such as interest rates, inflation, unemployment rates, demographic growth, technological changes, and environmental regulations. (Xu & Tang, 2014). Mankiw (2016) discussed the supply and demand model for housing, stating that demand is influenced by metrics such as income levels, and population growth, while supply is determined by state regulations and the costs of construction. Housing supply in the short term is often characterized by its inelasticity, instability, and constraints, primarily resulting from limited land availability and construction completion deadlines (Stepanyan et al., 2010). Housing supply is also influenced by a set of factors such as the availability of land, construction activities, and the local land-planning system. In theory, if there is a scarcity of available land, then it will be a bottleneck for the housing supply growth. Additionally, when construction costs of new dwellings increase it leads to an increase in the overall housing prices (Zainal et al., 2016).

Going more in-depth into the main factors that influence housing prices we will focus on a macro-perspective, ignoring specific characteristics of properties and focusing on the economic, social, and demographic factors. Égert & Mihaljek (2007) assert that comprehending the dynamics of the housing market demands a comprehensive approach, one that takes into account economic, financial, demographic, and policy elements to gauge the equilibrium between supply and demand. In the following section we will look to gain a deeper knowledge of which determinants academics proved to sort an effect on the real estate market prices and how do they impact them. This knowledge will prove pivotal as we progress in the development of our analytical model.

2.3.1 Demographic Housing Price Determinants

Demographic factors play a role in in the determination of real estate prices. While some of these factors may initially appear insignificant, Winkler & Donald Jud (2002) examined the fluctuations of housing prices in metropolitan areas across the US and found that elements as distinct as population growth, household income and construction costs, wield a substantial influence on the appreciation of real estate values. Later, Égert & Mihaljek (2007) studied the determinants of house prices in eight economies of Central and Eastern Europe, and stated that demographic factors should be considered as well when discussing housing price determinants. For instance, fluctuations in age, household size, and family structure of the population may cause variations in housing demand. An increase in household income or urbanization levels also leads to a higher housing demand, as the population has more disposable income. To illustrate, housing prices typically exhibit an upward trajectory in areas characterized by robust population growth, considering that such regions experience higher demand for real estate. However, Rehman et al. (2020) differentiate this growth by distinguishing between migration flow (in the short run) and new births (in the long run), as new births exert an influence on housing prices with a lag of approximately two decades. Another example is the population characterization of the area. If the population is composed of younger families with higher incomes, the likelihood of these individuals engaging in real estate purchases is notably higher, driving the prices upwards. Employment opportunities may also sort an effect on housing prices, since individuals relocate to areas offering enhanced job prospects. A final example could also be the ethnic and racial diversity, as a research study by Bloomberg has shown that areas with a higher diversification on their community may show lower housing prices due to discrimination.[8]

2.3.2 Macroeconomic Housing Price Determinants

Housing prices are formulated based on the population's ability to pay, which is dependent on the country's economic state. (Rehman et al., 2020). According to Luo et al. (2020) economic factors like the GDP have a positive correlation with housing prices due to the positive effect on the population's income. Also, the unemployment rate is considered the most potent determinant of housing prices, having a strong negative linear relationship with the real housing price index (Égert & Mihaljek, 2007). Chen & Hobbs (2003) stated that a prosperous economy positively affects the investment activity of a country. Economic factors such as the GDP per capita, mortgage interest rates, inflation, are correlated with the housing prices in European countries (Égert & Mihaljek, 2007; Giussani & Hadjimatheou, 1992). Monetary policies highly impact housing prices as well, buyers rely heavily on mortgage loans to acquire their properties, hence the increase of interest rates decreases the availability of capital and consequently the housing demand. (Xu & Tang, 2014) Regarding the implementation of housing market policies, there's a risk that there may be some unintended consequences coming from the implementation such as increased market volatility and magnified inequalities. In the following sub sections, we will delve into how each of the most important determinants affect housing prices.

2.3.2.1 Gross Domestic Product

The GDP represents all the goods and services a country produces in a given period. As stated by Cornell (2010), GDP variations may sort changes in earnings, which can cause fluctuations on equity

[8] https://www.bloomberg.com/graphics/2023-home-prices-racial-gap-us-cities/ accessed on May 2, 2023.

prices. A strong and growing GDP is a positive indicator of the country's economy and Hoskins et al. (2004) inferred that GDP growth was correlated with property returns as well. Kohler (2010) conducted an examination of the primary drivers influencing the real estate market in the United States and identified economic factors, such as GDP, as the most pertinent determinants. Theoretically, during phases of consistent GDP growth, businesses experience expansion, and the general population tends to allocate more funds for expenditure. When individuals possess greater disposable income to invest in real estate, this heightened demand will naturally increase property prices. On the other hand, a decrease in GDP may lead to a decrease in demand for the real estate market, resulting in lower prices. In a recession, people tend to be more budget-conscious and take fewer risks, which can lead to a reduction in demand and consequently lower prices.

2.3.2.2 Unemployment Rate

A country's unemployment rate represents the portion of the population that is unemployed but actively willing to work and looking for employment. Agnew & Lyons (2018) studied the effect of the unemployment rate on housing prices in Ireland and concluded that a lower unemployment rate led to higher real estate prices. On the same page, Luo et al. (2020) state that there is an inverse relationship between housing prices and unemployment (Figure 2.4).

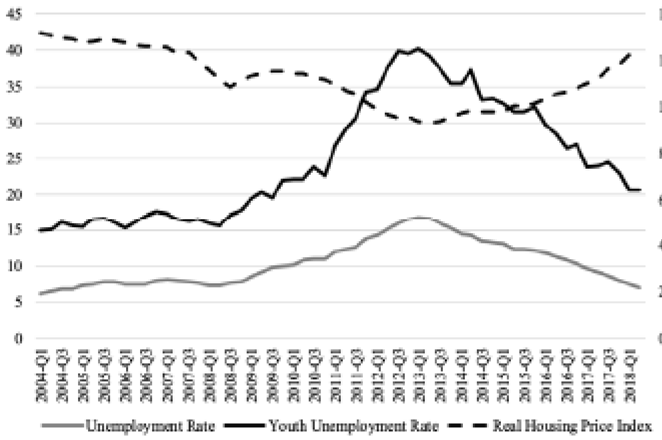

Figure 2.4: Relation between Unemployment Rate and Housing Prices (Source: Luo et al. 2020)

A higher unemployment rate directly correlates with a diminished pool of active buyers, and, according to the principles of supply and demand, this factor tends to result in a reduction in housing prices. The availability of capital significantly diminishes for the unemployed population, therefore curbing the overall demand for real estate. In contrast, when unemployment rates are lower, they are expected to positively influence the overall income of the population, consequently contributing to an increase in housing prices. From the perspective of construction firms, a higher unemployment rate serves as an indicator to scale back the construction of new dwellings, effectively reducing the supply within the market. This supply reduction, in turn, facilitates a downward adjustment in price equilibrium. On the other hand, Ismail & Nayan (2019) investigated the real estate housing prices in East-Asian countries such as Malaysia, Singapore, Thailand, and Indonesia and found a counter intuitive relationship between unemployment and housing prices. In their model the coefficient of the unemployment rate showed a positive and significant sign, which contradicts the theory behind housing determinants. The authors implied that a possible explanation would be the low correlation between unemployment rates and housing prices in East-Asian countries.

2.3.2.3 Inflation

Inflation can be defined as a sustained increase in the general price level of goods and services in an economy over time. It is often measured by the percentual change in the Consumer Price Index (CPI) or the Producer Price Index (PPI) over a certain period. The Consumer Price Index measures the monthly change in prices paid by the consumers.[9] The Producer Price Index measures the average change over time in the prices domestic producers receive for their output.[10] David Hume initiated the exploration of monetary policy and later other classical economists like John Locke, John Stuart Mill and Irving Fisher made important contributions to the understanding of periods of a general price level increase and their effects. Their theories converged in the development of the Quantity Theory of Money, which states that inflation is caused by an increase in the money supply that exceeds the growth rate of the economy (David Hume, 1777).

Regarding inflation's effect on the real estate market, (Flannery & Protopapadakis, 2002) found evidence of an inverse relationship between equity returns and inflation. De Bernardi & Rodenholm (2013) stated that higher inflation can lead to higher borrowing costs for real estate investors, thus reducing demand and negatively affecting housing prices. Furthermore, inflation also contributes to the increase of construction costs thus reducing the housing supply. Furthermore, (Cohen & Burinskas, 2020) studied the real estate market with 18 macroeconomic variables that represent the monetary policy, external and construction sectors' performance, economic growth, investment, households' earnings and inflation and found that the most statistically significant factors were inflation, bank assets and total construction. Central banks worldwide employ various measures to mitigate and control inflation, collectively referred to as "Monetary Policy." The primary goal of these central banks is to uphold a low and stable inflation rate, thereby cultivating a secure climate that encourages households and businesses to make investments. This commitment to price stability is instrumental in bolstering economic growth. Nonetheless, it is essential to recognize that the management of monetary policy is a complex endeavor and is not clear of potential unintended consequences. In some instances, the pursuit of stringent monetary policies aimed at curbing inflation may inadvertently prevent economic growth and, in extreme cases, trigger a recession.

2.3.2.4 Interest Rates

Interest rates can be defined as the amount a lender charges a borrower as a percentage of the loaned amount. Consequently, an interest rate can be seen as the "cost of money," as higher interest rates make borrowing the same amount of money more expensive.[11] Investors are often attracted to acquire a property using a home loan, as suggested by Grimes & Aitken (2010), as governments allow investors

---

[9] https://www.investopedia.com/terms/c/consumerpriceindex.asp accessed on June 10, 2023.
[10] https://www.investopedia.com/terms/p/ppi.asp accessed on June 10, 2023.
[11] https://www.investopedia.com/terms/i/interestrate.asp accessed on June 12, 2023

to deduct a portion of the interest paid on their mortgage debt from their income. As such, Xu & Tang (2014) state there is a strong relationship between housing prices and real interest rates. As previously discussed, central banks maintain a dual objective of preserving a low and stable inflation rate while striving to achieve the maximum sustainable level of employment. To achieve these objectives, regulators exercise control over the interest rate at which a consortium of banks can borrow funds from one another. An illustrative example is the Euro Interbank Offered Rate, commonly referred to as Euribor. Banks, being financial institutions, invariably adjust the interest rates at which they extend capital in response to fluctuations in the interest rate at which they secure funding. Such measures illustrate a Contractionary Monetary Policy, aiming to curtail aggregate demand, in contrast to Expansionary Monetary Policies, which seek to stimulate aggregate demand. Lacoviello & Neri (2010) used a dynamic stochastic model of the US economy to price the housing market. They inferred the transmission mechanism of monetary policy shocks to the real estate market and economy in general. As such, the authors concluded that changes in monetary policies do affect housing prices, residential investment, and consumer spending. Additionally, Lourenço & Rodrigues (2017) analyzed the adjacent factors to the Portuguese housing prices evolution during and after the 2008 financial crisis. The authors ultimately concluded that interest rates and economic growth were the main drivers of housing prices during that period. Goodhart & Hofmann (2008) state that housing prices may be seen as asset prices, being determined by the discounted future returns of property. The authors explain that a change in capital supply may have an effect on real estate prices. Although the relationship between interest rates and housing prices is typically defined as a complex one, an increase in credit availability leads to a rise in loans as households see it as an opportunity to acquire a home or move to an improved one. An increase in capital availability lowers interest rates. As a result, housing prices can increase due to the expectation of higher returns on property and a lower discount factor (Goodhart & Hofmann, 2008). Also, lower interest rates make mortgages more affordable, allowing investors to increase their purchasing power and afford new dwellings, thus increasing the housing demand and its prices.

2.4. The Portuguese Real Estate Market Evolution

In the previous sections we dove into the factors that impact housing prices and the description of the macroeconomic scenario felt during economic cycles of the last twenty years. This final section of the literature review is intended to observe the historical variation of the Portuguese real estate market prices. We will analyze the work of Lourenço & Rodrigues (2017) that studied the determinants of the Portuguese real estate prices and how these react to different economic conditions. Past sections will enable us to relate the main determinants with the housing prices' fluctuation and try to comprehend if theoretical assumptions regarding housing price behavior are indeed in line with how the market reacted.

According to Lourenço & Rodrigues (2017), Portugal's real estate prices act as expected and are related to the country's economic performance and interest rates. Consequently, housing prices exhibit a positive correlation with economic growth. Historical data further confirms these claims, revealing that Portuguese housing prices experienced an average annual increase of 1% during the period spanning from 1999 to 2006. These values align with the afore mentioned assertions, particularly since this timeframe coincided with a period of economic stagnation and gradual growth in the Portuguese economy, typified by an average GDP growth rate of 1.72%.[12] From 2007 to 2013, housing prices decreased by 4% per year on average (Lourenço & Rodrigues, 2017). As previously mentioned, this was a period of financial crisis, with concerning unemployment rates up to 17.1% and GDP growth rates reaching a minimum of -3.2%. Additionally, mortgage rates increased abruptly during this period (Figure 2.5), which also decreased the overall investment confidence. Consequently, the dynamics of the real estate market during this period are in line with expectations, displaying a decline in response to the economic downturn and the attendant volatility in consumption. Subsequently, following this challenging period, Portugal began to exhibit a more favorable economic trajectory. The impact of the EU bailout and the implementation of the financial reforms started to take hold, leading to significant improvements in economic indicators. As a result, GDP growth rates and the unemployment rate either aligned with or surpassed the European average. Notably, mortgage rates receded to notably lower levels, as observed in Figure 2.5. This reduction in mortgage rates stimulated an upsurge in housing loans and, as a direct consequence, an increased demand for housing. Consequently, housing prices experienced an average annual growth of 4% after 2013, reflecting the positive turn in Portugal's economic performance.

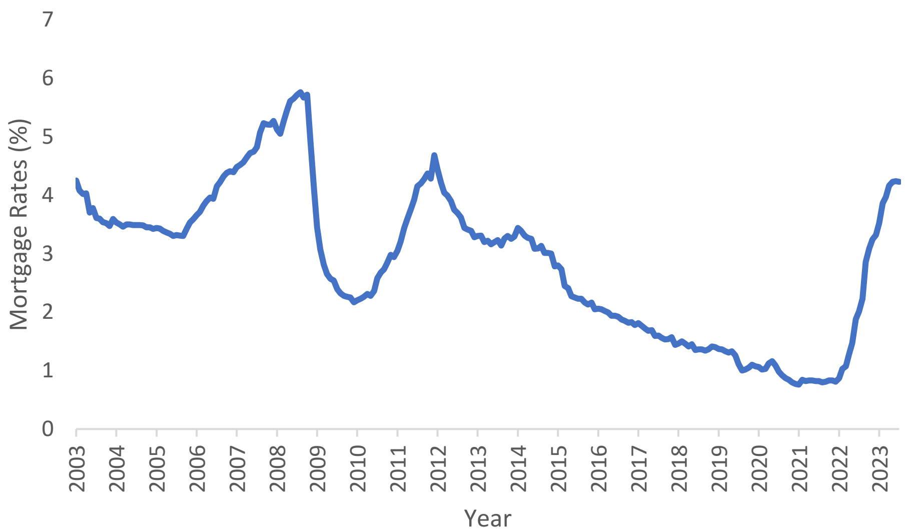

Figure 2.5: Change in Portuguese mortgage rates overtime (Source: Banco de Portugal)

[12] https://tradingeconomics.com/portugal/gdp accessed on June 12, 2023

3. Data Characterization

The data used in our study originates from three distinct databases: the Portuguese Labor Force Study (LFS), the IMT database and the IMI database, all obtained from the *Instituto Nacional de Estatistica* (INE), in Portugal. The LFS database is used to gather demographic information about the population, the IMT database contains details about housing prices and housing characteristics, as for the IMI database, this one complemented the IMT database in terms of house characteristics. The data available for econometric analysis is limited to the time span between the year 2011 and 2020 because it is the period where we had available data from all of the databases. This section is organized as follows: Section 3.1 describes the LFS, Section 3.2 presents the IMT and IMI databases Section 3.3 describes the data sampling process.

3.1. The Labour Force Survey

INE in Portugal conducts an important household survey known as the Labour Force Survey (LFS). It is quarterly in nature and provides quasi-longitudinal insights into the dynamics of the workforce in the country. The LFS's main goal is to present a complete picture of the Portuguese labor market, with a focus on the changes between employment and unemployment. Official statistics on the employment situation and other aspects of the Portuguese population connected to the labor market are derived from the LFS. These include the sector and professions of the economy, educational attainment and professional preparation, job searching, and career routes. It allows for the cross-referencing of various variables due to the quantity of personal information it contains, which helps us better comprehend the state of the nation. Additional characteristics, such as location, gender, age, and family structure, are also accessible. Being regularly carried out, it not only offers details on the structure of these occurrences but also makes it possible to analyze quarterly variations.

Going into detail about how the LFS sample was chosen and put together in this section, the following explanation was heavily influenced by the work of Correia & Lima (2006). The scope of the LFS survey encompasses all individuals whose primary residence is a family home in Portugal. INE constructed a "Mother-sample" based on the 2001 Portuguese housing census. Furthermore, they selected households from the mother-sample as the sampling units for the LFS. Additionally, people who reside in collective households, such as the military or students living in boarding schools may have some influence on the job market. Nonetheless, people residing in hotels, pensions, mobile households, and other similar collective households are dismissed from the survey. On average, 22500 households are inquired per trimester, with data collection conducted through direct interviews using the Computer Assisted Personal Interviewing (CAPI) system with all household members. In cases where one of the household members is unavailable to respond to the questionnaire, another member will answer on their behalf, resulting in a proxy response. Nonetheless, not all the survey's interviews are conducted in the same manner. The first interview involves an INE visit to the household and a face-to-face interview. As for the remaining five interviews, they are conducted via phone calls, provided the household agrees to this method.

We worked with panel data, which requires multiple observations of the same individuals at various points in time. Figure 3.1 is a graphical representation of the sample rotation of the LFS, each household undergoes interviews for six consecutive quarters, resulting in the sample being divided into six subsamples or rotations. A reference week is assigned to each household, and throughout each quarter, interviews are done one to two weeks after the reference week. Every quarter, one of the rotations completes their final interview and exits the sample, making way for a new rotation. Additionally, given that five of the six rotations remain consistent across consecutive quarters, this repetitive pattern enables us to track the evolution of the parameters rather than merely comparing two independent values. Furthermore, this scheme allows for a reduction in respondent burden, which can have a negative impact on the quality of the provided information.

	2018				2019				2020			
Quarter	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q	1Q	2Q	3Q	4Q
Set of Households: A	6	5	4	3	2	1						
B		6	5	4	3	2	1					
C			6	5	4	3	2	1				
D				6	5	4	3	2	1			
E					6	5	4	3	2	1		
F						6	5	4	3	2	1	
G							6	5	4	3	2	1

Figure 3.1: Rotation of the Labour Force Survey (Source: INE)

The process starts by selecting a sample that represents the population as a whole. By gathering data from this sample, estimates are calculated to apply to the entire population. This method relies on the assumption that each member of the sample represents a specific subset of the population with similar characteristics. Weighting parameters are used to ensure accuracy and fairness in representing different groups within the larger population. Calculating these weights requires considering the sample's design, corrections for non-responses, and the population's changing characteristics over time, which increases the difficulty of estimating population values. These weights are given to every unit in the sample, whether it be a family or an individual. A correction factor to take into account non-responses and lessen their impact on sample size, a calibration factor, and an initial weight based on the sample's design are all necessary for estimating values. The calibration factor uses a method called "margin adjustment" to match the sample with additional information obtained from the survey. This adjustment considers a variety of demographic factors, including gender, age, location of residency, and the likelihood that a certain residence will be chosen.

3.2. IMT and IMI Databases

To get access to housing prices and real estate information we used another pair of INE databases. INE has prepared some files with public information known as Public Use Files (PUFs)". These files (data and metadata) contain anonymized records, processed and prepared in such a way that the unit of observation cannot be identified directly or indirectly, except when it concerns individual statistical data about the Public Administration. They are freely accessible and comply with the principle of statistical secrecy and personal data protection. As in our case, the academic community has specific needs regarding statistical information, particularly for research projects and the preparation of Master's dissertations and doctoral theses. In this context, INE has established a Protocol with the Ministry of Education and Science, with the aim of facilitating access for researchers to the statistical information they require for their work. Additionally, free access to statistical microdata for research purposes is only possible for accredited researchers, in accordance with the Protocol established with INE.

Taxation plays an integral role in real estate transactions, impacting the financial aspects and feasibility of property acquisitions. In order to explain why these databases were created we must explain what is the IMT and IMI. In Portugal, IMT (*Imposto Municipal sobre Transmissões Onerosas de Imóveis*) and IMI (*Imposto Municipal sobre Imóveis*) are the two key taxes tied to property transactions. As important components of Portugal's fiscal landscape, these taxes affect both property buyers and owners. IMT is a tax on the transfer value of real estate and is paid by the purchaser. IMT is calculated as a one-time tax and varies depending on various factors related to the property being acquired. To calculate the IMT, multiple variables are considered, such as the declared value of the property. IMI, on the other hand, is an annual property tax that applies to property owners in Portugal. Unlike IMT, which is paid during property acquisition, IMI is an ongoing financial obligation for property owners. In the computation of IMI, the dwelling's characteristics, particularly its size, assume great importance. Nevertheless, the geographical location of the property also factors into the tax rate calculation.

INE has compiled data pertaining to IMT and IMI tax rates. At the time of this writing, IMT data spans from 2007 to 2021. This dataset contains information to calculate the IMT tax rate, along with additional details that must be provided by the parties involved in a transaction. When constructing our models and drawing statistical inferences, we selected variables that we considered relevant. The property information has been aggregated on a NUTS II level (Nomenclature of Territorial Units for Statistics)[13], each observation belonging to one of the following regions: North, Center, Lisbon, *Alentejo*, South, Azores, and *Madeira*. We extracted data concerning the overall property value, average property size, location, and the number of properties in each county. Regarding the IMI data, it covers the period from 2010 to 2021 at the time of this writing. Unlike the IMT database, this dataset places a stronger emphasis on the individual characteristics of dwellings. While we did not utilize all available variables, we did have access to information such as the number of floors, typology, land prices, and more. Similar to the IMT database, this information has been aggregated at the county level (NUTS II).

[13] Portugal is comprised of 308 municipalities and these group themselves in 25 NUTS III, 7 NUTS II and 3 NUTS I.

3.3. Sample Construction and Characterization

The LFS comprehends the responses of 292,165 individuals, resulting in 1,281,959 observations during the years of 2011 until 2020. As previously explained, there are multiple observations for each individual, as each person answers the survey up to six times. Since not every individual answered the survey the same amount of time, taking into account all the responses would result in unbalanced and biased statistics. Therefore, we solely kept the first response of each individual leaving our sample with 292,165 observations. This sample will only be used when describing the population and calculating the summary statistics, as we found it would give a more realistic vision of the population when compared to the collapsed form used to compute the models. Similarly to the IMT and IMI observations, the LFS database also contains information regarding the individual's location, and as such, it is also characterized at the NUTS II level. Moreover, there was a castaway of observations from the regional areas of *Região Autónoma dos Açores* and *Região Autónoma da Madeira* which represent the two Portuguese archipelagos, as the IMT and IMI databases didn't contemplate these regions.

The IMT database originally consisted of 4,628 observations, with one observation for each county and year. In contrast, the IMI database had a smaller time span and was initially comprised of 3,704 observations. Similarly to our approach with the LFS data, we applied a different database for the data description phase since using the final database, composed of 50 observations, to describe the sample would result in an unrepresentative sample characterization. Like for LFS database, some minor data treatments were applied to the described database. The IMT database contained transactions with exorbitant values, which did not accurately reflect real market conditions. Consequently, we performed an outlier removal treatment. Given that retaining these values in the data characterization step would be conflicting, considering their subsequent elimination, we opted to exclude them in order to maintain consistency between the described sample and the final database. Additionally, the raw data was aggregated at the municipality level, as such, we had to convert the observations into a broader regional one so we could later merge the databases with the LFS database. To facilitate this process, we employed a correspondence table to link each municipality to its respective region (NUTS II). Municipalities that lacked correspondence in the table were removed from the database. Accounting for these observations in our sample description would have been unwarranted, and we chose to remove them for the same reasons as the outliers.

3.4. LFS Summary Statistics

In this section we look to describe the LFS sample, representing the overall Portuguese population. We take the first observation of each individual to account for double counting problems. The sample is composed of 292,165 individuals, 270,675 native observations and 21,481 immigrant observations, therefore the immigrant population represents approximately 7.4% of the total sample. All the information is presented in Table 3.1.

3.4.1 Age

The average age of the individuals within the sample stands at 45.1 years. It is noteworthy that the average age of the immigrant population is notably lower, with an average of 41.4 years. Figure 3.2 contains a visual representation of the age groups distribution. When dissecting the age distribution, it becomes evident that a higher percentage of the sample is concentrated within the older age groups, with 53% comprising individuals aged 45 years and above. A comparison between the native and immigrant populations reveals a distinct pattern, where the number of immigrant individuals in the age groups below 16 and above 65 is significantly lower. This distribution may be attributed to the common practice of residing in a foreign country during one's working-age years and subsequently returning to the home country upon retirement. This stands in contrast to the native population, wherein 25% of the sample consists of individuals aged over 65 years.

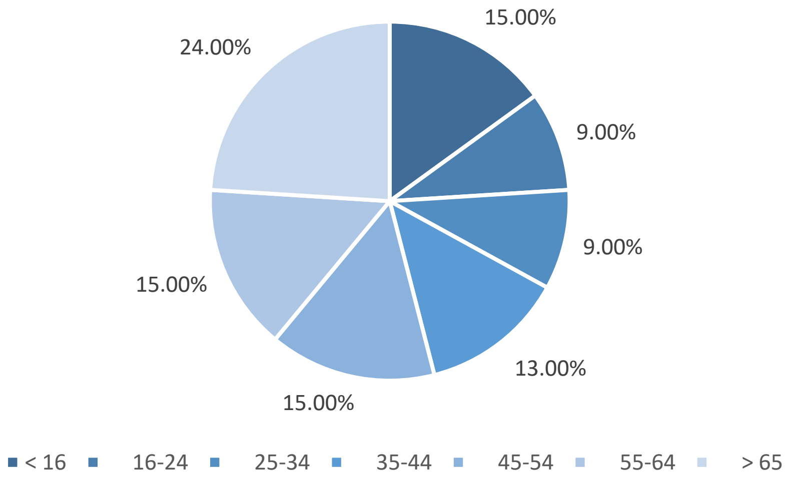

Figure 3.2: Portugal's Age Groups Distribution

3.4.2 Gender

Regarding the gender composition of the population, approximately 47% of the sampled individuals are male. However, when differentiating between natives and immigrants, a slightly lower proportion of males is observed in the native population, accounting for 45%, as opposed to the 48% of immigrant males within the sample.

3.4.3 Education

In the domain of education, the sample population can be categorized into distinct groups, including those with no education, basic education, high school education, and college education. Figure 3.3 contains a visual representation of the population's education distribution. The data underscores a significant aspect: the Portuguese population appears to exhibit a relatively lower level of educational attainment. Approximately 66% of the sample possess at most a basic education, 18% have completed high school, and a mere 15% hold a college degree. A notable and somewhat counterintuitive observation arises when comparing both the native and immigrant populations. Surprisingly, about 59%

of the immigrant population consists of high school or college graduates, in stark contrast to the 31% within the native population. This highlights a divergence in educational profiles between the two groups.

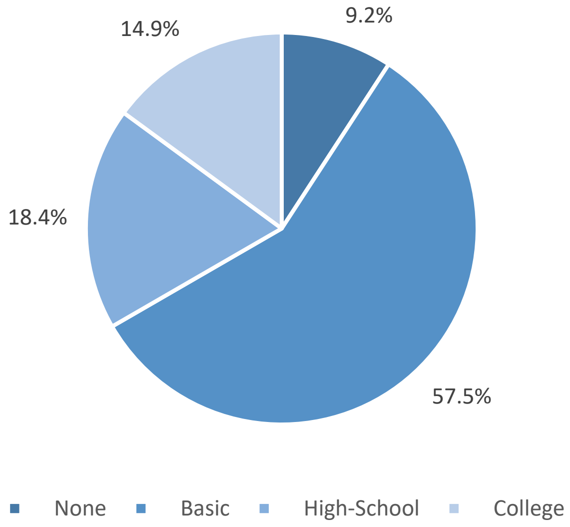

Figure 3.3: Portugal's Education Distribution

3.4.4 Employment

In terms of employment status within the sampled population, the following conditions prevail. The active population comprises individuals who are either employed or actively seeking employment, while the remaining segment of the sample consists of individuals who are below working age or retired. Of the entire population, 43% are employed, and an additional 6% are actively seeking employment, representing a total active population of 51%. A comparison between the native and immigrant populations reveals two distinct yet anticipated scenarios. Among immigrants, the percentage of employed individuals is notably higher, exceeding that of the native population by 16%. Furthermore, the non-active population within the immigrant group stands at approximately 31%, which is considerably lower compared to the 53% observed in the native population. These disparities may be attributed to the higher percentage of immigrants within the working-age group and their aspiration to improve their living conditions.

3.4.5 Earnings

In the analysis of monthly earnings distribution within the sampled population, earners were classified into five distinct groups based on their salary thresholds, corresponding to the 10^{th}, 50^{th}, 90^{th} and 99^{th} percentile of the wage distribution percentiles of the wage distribution. It is important to note that due to some data gaps (with approximately 70% of the sample either being unemployed or not providing income information), the percentages were proportionally scaled to provide a clearer view of the distribution. When comparing the two groups, the native population demonstrates a slightly higher proportion of individuals on the lower end of the wage spectrum, with 41% falling below the 50th percentile, compared to 48% within the immigrant population. On the higher end of the income scale, the immigrant population exhibits 9.6% of the sample falling between the 90th and 99th percentiles, and

1.3% above the 99th percentile. In contrast, the native population comprises 8.7% and 0.9%, respectively, within these higher income percentiles.

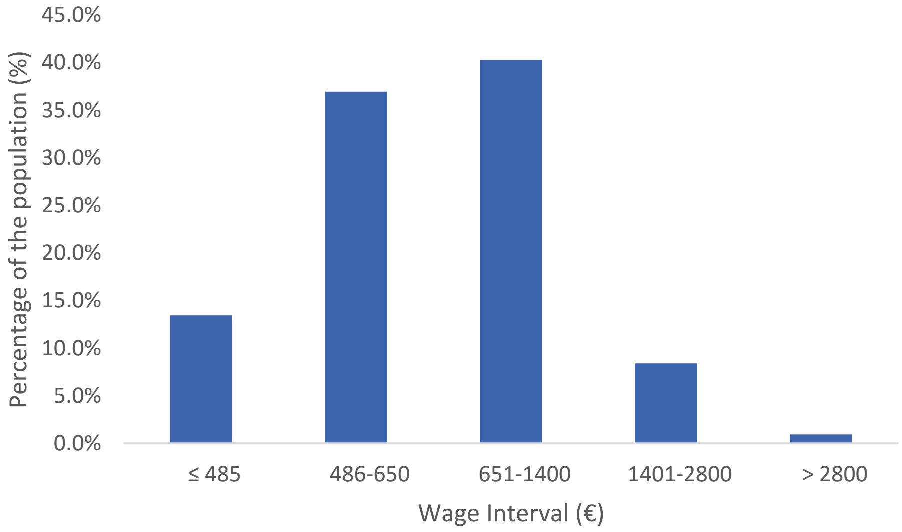

Figure 3.4: Portugal's Wage Groups Distribution

3.4.6 Demographics

As for the demographic statistics of our sample, the average percentage of immigrants per region is 7.19%. Lisbon and the South of Portugal represent the highest percentages, at 11.4% and 12.7% correspondently, and *Alentejo* represents the lowest percentage at 4.04%. The average immigrant inflow is positive, at 0.13%, which in nominal values translates itself into about 3266 immigrants. Conversely, the native population experiences a negative net flow at an average rate of -0.5%. Over various years, immigration flows exhibit fluctuations in response to global macroeconomic conditions. For instance, in 2011, the immigrant flow stood at 0.1%, and it consistently increased over the subsequent years, reaching 0.74% in 2020, representing the highest recorded value during this period.

Table 3.1 Native and Immigrant Population - Descriptive Statistics

	All		**Natives**		**Immigrants**	
	mean	std. dev	mean	std. dev	mean	std. dev
Immigrant	0.074	0.261				
Age (years)	45.1	23.3	45.4	23.7	41.4	16.8
< 16	0.15	0.36	0.15	0.36	0.07	0.27
16-24	0.09	0.29	0.09	0.29	0.1	0.3
25-34	0.09	0.29	0.09	0.29	0.15	0.35
35-44	0.13	0.34	0.12	0.33	0.26	0.44
45-54	0.15	0.35	0.15	0.35	0.22	0.41
55-64	0.15	0.35	0.15	0.36	0.11	0.32
> 65	0.24	0.430	0.25	0.43	0.09	0.28
Male	0.47	0.5	0.45	0.5	0.48	0.5
Married	0.47	0.5	0.48	0.5	0.46	0.5
Residence Location						

	All		Natives		Immigrants	
	mean	std. dev	mean	std. dev	mean	std. dev
North	0.32	0.47	0.33	0.47	0.19	0.38
South	0.13	0.33	0.12	0.32	0.22	0.42
Center	0.19	0.39	0.19	0.4	0.15	0.36
Lisbon	0.22	0	0.21	0.40	0.36	0
Alentejo	0.14	0.35	0.15	0.35	0.08	0.266
Education						
Did not answer	0.13		0.13		0.04	
None	0.09	0.28	0.10	0.28	0.03	0.16
Basic	0.57	0.5	0.59	0.5	0.39	0.48
High-School	0.18	0.37	0.17	0.35	0.33	0.46
College	0.15	0.34	0.14	0.32	0.26	0.43
Employment Condition						
Employed	0.43	0.5	0.42	0.5	0.58	0.49
Unemployed	0.06	0.23	0.05	0.5	0.11	0.31
Other	0.51	0.5	0.53	0.22	0.31	0.46
Monthly Earnings (€)						
≤ 485 (p.10%)	0.13		0.145		0.120	
486-650 (p.10% - p.50%)	0.37		0.362		0.361	
651-1400 (p.50% - p.90%)	0.40		0.398		0.409	
1401-2800 (p.90 - p.99%)	0.08		0.087		0.096	
> 2800 (p.99%)	0.01		0.009		0.013	
Immigrant Pop. (% of Regional Pop.)	7.19	3.6				
North	4.24	0.4				
South	12.70	0.8				
Center	6.03	0.6				
Lisbon	11.40	0.9				
Alentejo	4.04	0.61				
Immigrant Flow (Annual) (%)	0.13	0.7				
Immigrant Flow (Annual) (n)	3266.30	18283.1				
Native Flow (%)	-0.5	0.72				
Immigrant Flow (2012) (%)	-0.100	0.408				
Immigrant Flow (2020) (%)	0.740	0.46				
Number of Observations	292.156		270.675		21.481	

3.5. IMT and IMI Summary Statistics

In this section, we describe the summary statistics of the IMT and IMI databases. The housing prices have been adjusted to real values according to the Portuguese CPI to account for inflation.

3.5.1 Housing Prices

The primary variable of interest in this database is real estate prices. As it is statable in Figure 3.5 real estate properties exhibit significant annual price fluctuations. In 2011, the average price was 176,201.5€, and over the course of the next decade, it experienced a substantial increase of 31.5%, ultimately reaching a value of 231,595.6€ in 2020. This equates to an average yearly nominal increase of 6,282.9€ and an average yearly percentage increase of 67.3%. These figures underscore the dynamic and evolving nature of the Portuguese real estate market during the studied period. The presented housing prices values have undergone an outlier treatment, inflation adjustment and a weighted mean prices computation in order to reduce the weight of singular transactions, nonetheless, they still display a 720% average price change in the year 2014.

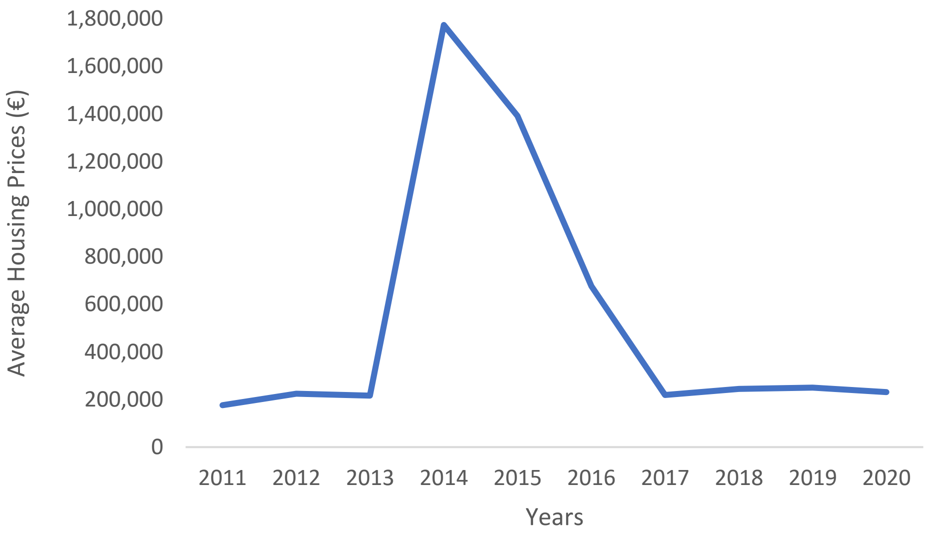

Figure 3.5: Average Portuguese Housing Prices throughout the years.

The average price per region has evident differences between each region. As expected, the Lisbon Metropolitan Area represents the region with the highest housing prices, with an average value of 1,243,585€. On the other hand, it was not particularly expected that the central region showed the lowest results in terms of housing prices, recording a value of 269,992.9€.

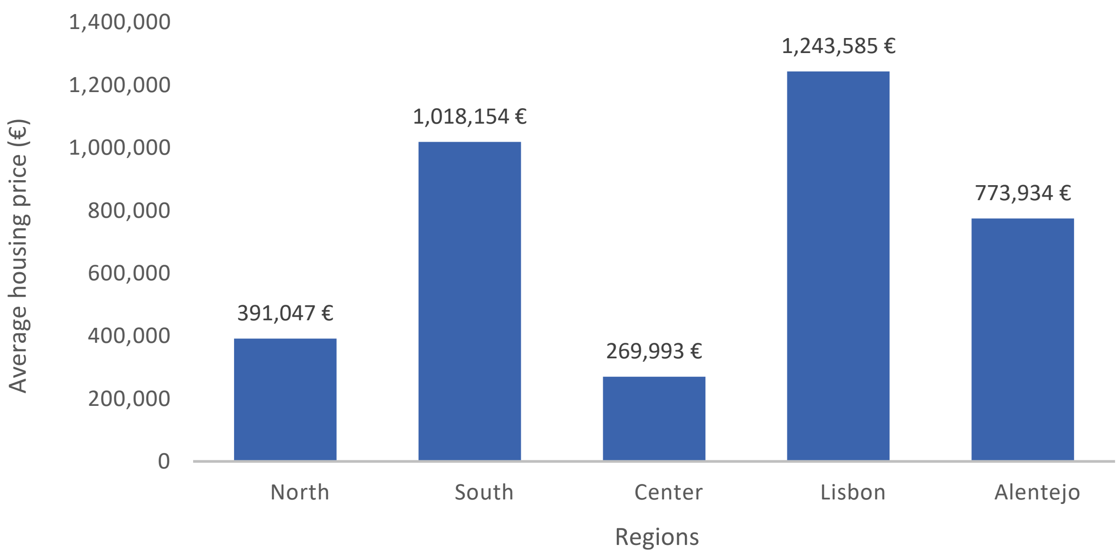

Figure 3.6: Portugal's Average Housing Price Per Region

3.5.2 Property Characteristics

Within the dataset, two key property characteristics merit our attention: Property Age and Interior Area. Beginning with Property Age, it is evident that the region where age is most pronounced is *Alentejo*, with an average property age of 42.7 years. Conversely, the remaining regions exhibit relatively similar property ages, except for the Center region, which boasts an average age of 34.9 years.

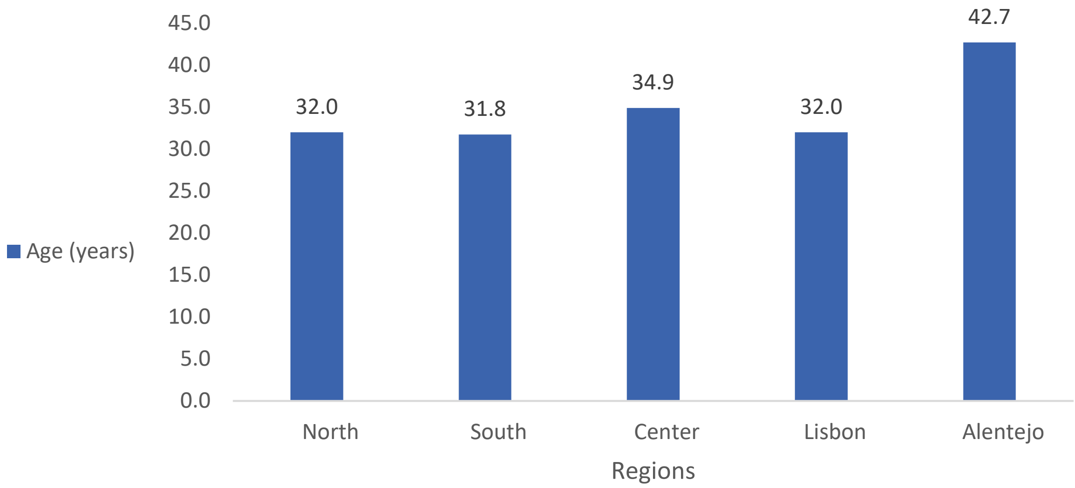

Figure 3.7: Portugal's Average Building Age per Region

Shifting the focus to the total Interior Area, we can discern that the Center and North regions feature the most spacious houses, with average sizes of 137.5 m2 and 134.1 m2, respectively. In contrast, the Metropolitan Area of Lisbon and the South region lag behind in this regard, with the smallest average interior areas of 115.4 m2 and 105.5 m2, respectively. These variations in property age and interior area provide valuable insights into the diversity of housing stock across the different regions.

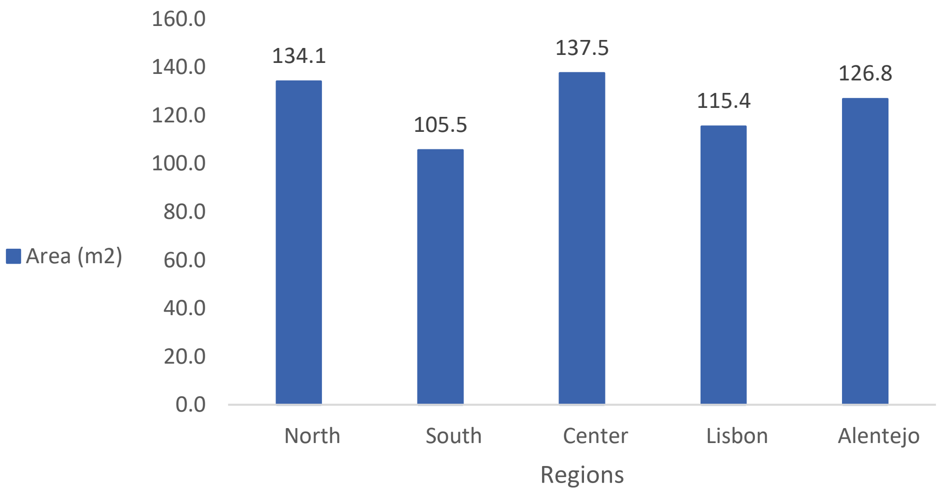

Figure 3.8: Average Interior Area per Region

Table 3.2: Housing Prices and Characteristics – Descriptive Statistics

	All	
	mean	std. dev
Avg, House Price (2011) (€)	176201.5	112060.3
Avg, House Price (2020) (€)	231595.6	142282.3
House Price increase (annual) (€)	6282.97	740033.4
House Price increase (annual) (%)	67.29	243.0908
House Price increase (2012) (%)	29.41	19.4
House Price increase (2013) (%)	-0.33	12.6
House Price increase (2014) (%)	719.5	306.7
House Price increase (2015) (%)	45.04	172.72
House Price increase (2016) (%)	-44.67	30.23
House Price increase (2017) (%)	-68.02	11.3
House Price increase (2018) (%)	24.63	22.97
House Price increase (2019) (%)	6.62	17.57
House Price increase (2020) (%)	-9.71	8.33
Avg. House Price per zone (€)		
North	391046.7	367635.7
South	1018154	798,830
Center	269992.9	236592.8
Lisbon	1243585	1229934
Alentejo	773933.9	1004893
Total Transactioned Houses (2011) IMT	75027	
Total Transactioned Houses (2020) IMT	27850.8	
Avg. Property Age (years)		
North	32.0	10.2
South	31.8	12.4
Center	34.9	8.5
Lisbon	32.0	8.5
Alentejo	42.7	12.2
Avg Property Area (m2)		
North	134.1	70.1
South	105.5	46.3
Center	137.5	336.3
Lisbon	115.4	53.0
Alentejo	126.8	57.9

3.6. Model Estimation Sample

In this section, we will provide a detailed account of the process leading to the construction of our final database, which consists of 50 observations. While this sample size may not be ideal for generating robust and consistent estimates, it represented the most practical approach for aggregating the three distinct databases at our disposal.

The initial challenge lay in the LFS database, which was fragmented into 10 separate yearly databases. To create a cohesive dataset spanning from 2011 to 2020, we diligently merged these segments. This dataset was structured at the NUTS II level, with observations assigned to one of the following seven regions: North, Center, Lisbon, *Alentejo*, South, Azores, and *Madeira*. The IMI database encompassed data from 2007 to 2021, while the IMT database covered the period from 2010 to 2021. As the LFS data was available exclusively for the years 2011 to 2020, observations from other years were systematically excluded. Both the IMI and IMT databases initially employed a municipal organizational structure, necessitating a conversion to the NUTS II level for compatibility with the LFS data. This conversion was executed using a correspondence table, with the observations lacking correspondence being meticulously removed. It is worth noting that the IMI and IMT databases lacked information on housing prices in the Portuguese archipelagos (Azores and *Madeira*), resulting in the exclusion of LFS observations related to these regions.

Despite the LFS data being divided into quarterly intervals, the IMI and IMT data were recorded on an annual basis. This disparity precluded the possibility of enhancing data granularity through finer time intervals. Ultimately, we aggregated all variables at the NUTS II and annual levels, resulting in a final database comprising data from five regions over a span of ten years, yielding a total of 50 observations.

Furthermore, housing prices were adjusted to real values and additional data treatments were applied to address anomalies in the original database. As mentioned earlier, the original IMT and IMI database contained unrealistic transaction values that were adversely impacting our estimates. These outliers, which implied an approximate 1300% average housing price increase in a specific year at the national level, were identified and rectified. To achieve this, we calculated Z-values for each observation using standard deviations and subsequently removed observations with an absolute Z-value exceeding 3. This corrective action resulted in the removal of 42 observations in 2014, 6 observations in 2015, and 3 observations in 2016 from the original IMT databases. In order to further attenuate the impact of singular transactions we decided to compute a weighted average of each NUTS II's housing prices. This way the municipalities with fewer houses had less impact in the NUTS II average price.

3.7. Variable Description

In the following section we look to explain the variables we will utilize in our model formulation. Our choice of variables has been guided by prior research and an extensive review of the existing literature. While not all of these variables will feature prominently in our final model formulations, they have played a central role throughout the research process. Some were essential in testing our models, while others assisted in identifying the most suitable set of control variables. Collectively, these variables have provided valuable insights and merit special attention in our data description.

Table 3.3: Variables Description

Variables	Description
Housing Prices	Deflated average housing price in a specific region and year (€);
Employment	Individual employment state. This variable is divided into the following categories: 1 - 4) working or unemployed, 5) student, 6) home worker, 7) retired, 8) inactive;
Active population	Dummy variable. Takes the value of 1 if the individual is part of the active population, 0 otherwise;
Unemployment rate	Percentage of the population not part of the labour force in the year t-1 (%);
Total population	Total number of residents or inhabitants living within Portugal. Computed from the LFS weights (n);
Regional population	Total number of residents or inhabitants living within each region (n);
Regional immigrant rate	Percentage of foreign-born individuals living within each region (n);
Education	Individual level of education. This variable is divided into the following categories: 1) Uneducated, 2 - 4) Basic education, 5-6) High School education, 7-10) College education;
Immigrant College graduates	Total number of foreign-born individuals with college education (n);
High Income Immigrants	Percentage of foreign-born individuals whose income is part of the 90^{th} percentile within each region (%);
Wealthy country immigrant	Percentage of foreign-born individuals whose native country is considered “High Income” (%);[14]
Immigration flow	Annual percentual change in the number of foreign-born individuals (%);
Native flow	Annual percentual change in the number of native individuals (%);
Income	Average regional income per capita (€);

[14] https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups According to the “World Bank Data” 2022 classification. Accessed on June 11, 2023

Variables	Description
Housing supply	Average number of IMT transactions. Used as a *proxy* variable for housing supply (n);
Building Age	Average building age (n);
Year dummies	Set of yearly dummy variables. The year set ranges from 2011 to 2020 and the variable is intended to capture the year fixed-effects;
Shares Based Instrument	Shift-share instrumental variable (%);

4. Methodology

From the literature review section, it is understandable that there is a relationship between immigration inflows and real estate prices. In the following sections we will present the methodology used to quantify this effect on the Portuguese scenarios.

4.1. Theoretical Background

4.1.1 Ordinary Least Squares

The ordinary least squares (OLS) method is the most common statistical technique used in regression analysis. Its purpose is to estimate the relationship between a dependent variable and one or more explanatory variables. OLS is a foundational method in econometrics and serves as source for more advanced regression methods. Many authors along the immigration literature use OLS as a term of comparison in their approaches. The model provides a baseline for evaluating model performance and allows authors to assess if the additional complexity of their models improves the explanatory power and outperforms OLS. Additionally, the models are easily interpretable making it simple to explain relationships between different variables. OLS is a technique used to find the line of best fit for a set of data points by minimizing the residual (the difference between observed and predicted values). This involves estimating the coefficients of a linear regression model by minimizing the sum of squared differences between the observed values of the dependent variable and the predicted values of the model. The generic equation of OLS can be written as follows:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{4.1}$$

Where β_0 represents the intercept, β_1 is the coefficient of determination of x_1, and u represents the error term of the equation, in other words, represents the other variables that sort an effect on y but are not considered in the regression. The resulting line is called the regression line, which represents the best fit for the data and can be mathematically defined as:

$$Min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.2}$$

Where y_i is the observed value and $\hat{y}_i$ is the predicted value. Nonetheless, the OLS method relies on several assumptions to be valid.

The key assumptions are the following:

- Linearity: There must be a linear relationship between the dependent variable and the independent variables.
- Independence: The observations must be independent of each other.
- Homoscedasticity: The variance of the residuals should be constant across all levels of the independent variables.
- Normality: The residuals / errors should be normally distributed.
- No multicollinearity: The independent variables should not be highly correlated with each other.

- No endogeneity: The independent variables are exogenous and not affected by the error term.
- No autocorrelation: The errors should not be correlated with each other over time or across observations.

When using OLS regression, it is important to evaluate these assumptions. Violations of these assumptions can lead to biased or inefficient parameter estimates and affect the validity of hypothesis tests and confidence intervals.

4.1.2 Fixed-Effects Regression

Fixed Effects (FE) regression is an estimation technique used in the presence of panel data. This type of data involves observations from a set of e.g., individuals, firms, entities over different time periods. The Fixed effects regressions allows us to control for entity-specific heterogeneity or time-invariant unobserved characteristics that may be correlated with the independent variable.

Taking equation 4.3 as an unobserved effects model and equation 4.4, where the variables $\overline{u_i}, \overline{a_i}$ and $\bar{x}_i$ represent the average over time for individual i.

$$y_{it} = \beta_1 x_{it} + a_i + u_{it} \tag{4.3}$$

$$\overline{y_i} = \beta_1 \overline{x_i} + \overline{a_i} + \overline{u_i} \tag{4.4}$$

Since $a_i = \bar{a}_i$ and it represents the time-invariant individual effect, subtracting $\overline{y_i}$ to y_{it} eliminates a_i and thus eliminates its effect using the within transformation. The result is the equation 4.6.

$$y_{it} - \overline{y_i} = \beta(x_{it} - \overline{x_i}) + u_{it} - \overline{u_i} \tag{4.5}$$

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \ddot{u}_{it} \tag{4.6}$$

After this process the model satisfies the classical assumptions, and we can estimate the fixed effects coefficient by regressing $\ddot{y}_{it}$ on $\ddot{x}_{it}$ through OLS. This method is powerful for addressing issues of endogeneity and unobserved heterogeneity in panel data analysis, however, the within transformation of the fixed effects method has the disadvantage of excluding the presence of time-invariant independent variables in the regression because they are deleted along with the fixed unobserved component. Additionally, if the time series dimension is limited, parameter estimations are likely to be imprecise.

4.1.3 First Differences Estimation

Similarly to the fixed effects regression, the First Differences (FD) is an alternative estimation technique that is used to address the problem of omitted variable bias in econometrics and statistics by using panel data. This bias arises when relevant independent variables are excluded from the regression. Hence, it is important to control for these unobserved effects to obtain reliable estimates.

The objective of the FD estimator is to eliminate firm-specific effects that do not vary over time. This is achieved by taking the first difference of the variables within each cross-section. Take the following original models, represented by equations 4.7 and 4.8 correspondent to period t and t-1, with the variable a_i representing the equation's fixed effects.

$$y_{it} = \beta_1 x_{it} + a_i + u_{it} \tag{4.7}$$

$$y_{it-1} = \beta_1 x_{it-1} + a_i + u_{it-1} \tag{4.8}$$

The first-difference transformation is applied as follows:

$$\Delta y_{it} = y_{it} - y_{it-1} = \beta_1(x_{it} - x_{it-1}) + (a_i - a_i) + (u_{it} - u_{it-1}) \iff$$

$$\iff \Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it} \tag{4.9}$$

The transformation is similar to the fixed effects estimation, since a_i is time-invariant, it is eliminated when subtracting y_{it-1} to y_{it}. The resulting differenced data contains information on how each variable changes over time within each entity. Finally, we regress Δx_{it} on Δy_{it} through an OLS, obtaining estimates that are not affected by time-constant unobserved effects.

4.1.4 Instrumental Variables and Two Stages Least Squares

Many linear regression models assume the independent variables to be uncorrelated with the error term. When this is not the case, using OLS no longer provides unbiased and consistent estimates. This situation happens when there are unobserved factors that influence both the independent and the dependent variables. In such cases, we call the independent variable an endogenous variable. Take the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{4.10}$$

In this case, x_1 is an endogenous variable and is correlated with the error term u. As such, if we use OLS in equation 4.10, we will get inconsistent and biased estimates of β_2. To consistently estimate this equation, an instrumental variable z must be found. An instrumental variable must satisfy two conditions: a relevance condition, in which z is correlated with x_1, and a validity condition, in which z is uncorrelated with the error term u. Although the latter condition cannot be tested, we can test the validity condition by regressing x_1 on z. This step is what we call the first stage of the 2SLS estimation. Consider the following model:

$$x_1 = \pi_0 + \pi_1 z + \pi_2 x_2 + v \tag{4.11}$$

In this first stage we want to test the hypothesis $H_0: \pi_1 \neq 0$ in order to understand if the validity condition is fulfilled. Then, we run equation 4.11 using OLS obtaining the estimated coefficients $\hat{x}_1 = \pi_0 + \pi_1 z$. Proceeding to the second stage of the regression, we then replace x_1 with $\hat{x}_1$ on the original model:

$$y = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 x_2 + u \tag{4.12}$$

And after we achieve equation 4.12, we estimate the coefficients using OLS. Summing up, in a 2SLS regression, the estimated values of the problematic predictor are first calculated using instrumental variables that are uncorrelated with the error terms, and the estimated values are then used to compute an estimated linear regression model of the dependent variable. The outcomes of the two-stage model are consistent since the computed values are based on factors that are uncorrelated with the errors. As such, Instrumental Variables and 2SLS regression are valuable tools for dealing with endogeneity in regression analysis, allowing researchers to estimate causal relationships between variables even when there are concerns about omitted variables or simultaneous causality.

Nonetheless, there is no utility in this method if we cannot test the endogeneity of our variables. As previously discussed, a variable is classified as endogenous when it is correlated with the error term u, which means, it is affected by unobservable factors. Although there is no possible way to measure these factors, we can proxy them. In equation 4.11 we can see that x_1 depends on v, the error term. Therefore, we can obtain the predicted residuals of equation 4.11 and include them as additional variables in equation 4.10. If the residuals are statistically significant, it means that the unobserved factors that influence x_1 also have an effect on y. In mathematical terms it translates in the following test: We must calculate $\hat{v}$, which is the same as $\hat{v} = x_1 - \hat{x}_1$ and include them in equation 4.10 as an additional variable. In the end we get the following:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \delta \hat{v} + u \tag{4.13}$$

Finally, we simply regress by OLS and if the estimated residual's coefficient $\hat{\delta}$ is statistically significant then x_1 is classified as an endogenous variable, otherwise, it is not.

Another step of Instrument Variable regressions is to test for instrumental strength. Weak instruments are instruments that have low correlation with the endogenous independent variable they are intended to instrument. In IV estimation, in the presence of weak instruments, estimates are often led to be biased and inefficient, undermining the validity of the results. Some methods to infer instrumental strength have been developed, nonetheless, there is still no universal test, the number of endogenous variables, instrumental variables, and the amount of bias the researcher is willing to tolerate lead us to try different procedures.

There are certain limitations to using metrics such as R-square to assess the instrumental strength as an extremely high R-square might mean over-fitting. Staiger & Stock (1997) developed an improved procedure. Their test would reject the null hypothesis of weak instruments when the (Cragg & Donald, 1993) statistic exceeds the value of 10. The Cragg and Donald statistic also corresponds to the first-stage F statistic when we are dealing with only one instrumental variable. Then Stock & Yogo (2005) developed a new way of instrument testing, the authors realized that the value of 10 was not wide enough to cover all possible scenarios, as such, they developed numerous tables to which one can compare the values of the first-stage F statistic. According to the authors, different values of endogenous and instrumental variables correspond to different thresholds, additionally, we must state how much Bias are we willing to tolerate, as different thresholds are assigned to different bias percentages,

Nonetheless, Antoine & Lavergne (2014) found that there could still be weak instruments with high F-statistics in the presence of heteroskedasticity. The (Kleibergen & Paap, (2006) (K-P) statistic calculates the minimum eigenvalue based on the differences between the two-step efficient generalized method of moments estimators using different weighting matrices and is more robust to heteroskedastic errors. Our understanding of the instruments' overall strength depends on the outcome of the K-P test. If the test statistic is greater than a threshold determined by the chi-squared distribution, it shows that the instruments are reliable enough to estimate IV. The validity of the IV results can be trusted if the test is significant. Sanderson & Windmeijer (2016) later developed a broader test to compute the instrumental strength in the presence of two or more endogenous variables. In the case of having only one endogenous variable the value of the Sanderson-Windmeijer (S-W) F-stat is equivalent to the Kleibergen-Paap F stat. Since our Stata software displays both values, taking in consideration the broader test we will discuss the values of the Sanderson-Windmeijer F-stat in the following section.

4.2. Modeling

In this section, we developed a set of theoretical models to assess the causal effect of immigration on housing costs. Although the results are not consistent across the available literature, authors tend to maintain a consistent approach when formulating their models for addressing this issue. This section provides context regarding the models prevalent in our field of study.

Saiz (2007) employed a model estimation technique based on long differences, utilizing both OLS and 2SLS methods. In his model, the dependent variable is the logarithm of housing rents, with the primary independent variable being the change in the immigrant population relative to the total population. In a similar study, Gonzalez & Ortega (2009) adopted a first-difference approach and employ OLS and 2SLS regressions in their modeling efforts. Their dependent variable consists of changes in the average price of a square meter, and, akin to Saiz (2007), their principal independent variable is the alteration in the foreign-born population divided by the total population from the previous year. Sá (2015) similarly applied OLS and IV regression techniques in her study, implementing the model in first differences to remove time-invariant factors. Like Gonzalez & Ortega (2009), Sá's model features the same independent variable: the change in the foreign-born population divided by the lagged total population. The dependent variable in her analysis is the alteration in the logarithm of real estate prices. In a comparable fashion, Accetturo et al. (2014) aligned with the methodologies of prior research papers. Although the authors explored various hypotheses and research questions, they adopted a 2SLS approach when examining the effects of immigration on housing prices. This consistency underscores our tendency to follow this research path. Based on the extant literature, we developed the following model to estimate the impact of immigration on house prices:

$$\ln(P_{it}) = \beta_1 \frac{\Delta FB_{it}}{POP_{it-1}} + \beta_2 X_{it} + \gamma + \alpha_i + \varepsilon_{it} \tag{4.14}$$

Where $\ln(P_{it})$ is the log of the house prices in region i and time period t . The main independent variable is the annual inflow of foreign-born divided by the total population on year $t-1$. The coefficient β is interpreted as the percentage change in real estate prices corresponding to a one percentage-point increase in the inflow of foreign-born individuals as a proportion of the total population of the previous year. X_{it} denotes a vector of control variables capturing time-varying region-specific effects. γ is a set of year dummies intended to capture national trends. α_i and ε_{it} represent the time-invariant and time-variant error terms, respectively. We also developed a simpler version of model 4.14 which does not contemplate X_{it}.

$$\ln(P_{it}) = \beta_1 \frac{\Delta FB_{it}}{POP_{it-1}} + \gamma + \alpha_i + \varepsilon_{it} \tag{4.15}$$

Based in our literature review, it became apparent that researchers generally adopted similar methodologies with slight variations to address the problem at hand. One notable concern is that the OLS estimator is prone to producing inconsistent estimates due to omitted variable bias. Furthermore, as discussed in section 2.6, the issue of reverse causality looms large when investigating the impact of immigration on housing prices. Immigrants do not relocate to random locations but rather choose where to settle based on local amenities or international settlements (Card, 2001). This raises causality concerns and underscores the importance of considering alternative approaches beyond a simple OLS regression. In light of the insights from the literature and recognizing the need to mitigate omitted variable bias and account for time-invariant factors, we employed a Fixed-Effects Regression as one of our estimation techniques. Additionally, we used a First-Differences Regression as another estimation technique to further enhance the robustness of our results. In the literature, many authors manually compute the dependent variables as a first difference $\Delta\ln(P_{it})$ or ΔP_{it} for example. In order to have one unique independent variable for all models, and since we tested both the fixed-effects and the first-differences regression, we left our dependent variable in the raw logarithmic form ($\ln(P_{it})$).

In order to handle the issue of endogeneity in the immigrant inflow variable, we exploited an instrumental variable approach that employs a widely utilized shift-share instrument. The instrument was first proposed out of the real estate market context, by (Card, 2001) and was later adopted by many researchers in the immigration field. The instrument is based on the historical share of immigrants of each country across different regions in Portugal (taking the year 2011 as base) and takes these shares to predict the present flows of immigration in each region. Taking a practical example, if in 2011 the Spanish immigrants were distributed as follows: 40% in Lisbon and 60% in the North. If in 2020 the national inflow of Spanish immigrants was a positive net value of 1000 individuals, then according to our instrument we would predict the regional inflow as an increase of 400 Spanish individuals in Lisbon and 600 in the North region. According to (Bartel (1989) immigrants tend to settle in regions with historical settlements of people from the same native country and the instrument exploits that assumption. To establish the instrument's validity, we need to rely on two assumptions. Firstly, we assumed that recent economic fluctuations in Portugal's regions are not correlated with the historical presence of immigrants. If the prior immigrant settlements were founded a significant time ago, their spatial distribution should be unrelated to the current distribution of housing demand shocks at the regional level. The second assumption concerns the exogeneity of yearly migrant inflows into various regions, as the number of immigrants in Portugal is significantly influenced by political laws and decisions.

The equation of the instrument of the current stock of foreign-born population in region i and year t is computed as follows:

$$Z_{it} = \frac{\sum_c \left(\frac{FB_{cit_0}}{FB_{ct_0}}\right) \cdot \Delta FB_{ct}}{POP_{it-1}} \tag{4.16}$$

Where FB_{cit_0} is the number of individuals foreign-born in country c residing in region i in the base year t_0. FB_{ct_0} is the number of individuals foreign-born in country c residing across all regions in year t_0. Therefore, the term in parentheses is the share of individuals born in c that resided in each region. ΔFB_{ct} is the change in foreign-born in country c in year t and is a time-variant term. Another time-variant term is the denominator of the equation, which is the population of region i in the year t-1. While recent literature, exemplified by Adão et al. (2018) and Goldsmith-Pinkham et al. (2020) explore new variations of the shift-share instrument identifying a valid and robust instrument is often the most challenging aspect of an IV approach. Therefore, opting for the version of the instrument that has been extensively employed in various research contexts and has demonstrated its utility in estimating diverse immigration impacts instills confidence in its reliability.

Panel data, or longitudinal data, represents multidimensional information collected over time, encompassing observations of various phenomena across multiple time periods for a consistent set of entities. Time series and cross-sectional data can be thought of as subsets of longitudinal data with measurements in only one dimension. Panel data yields two distinct types of insights: the cross-sectional component, which highlights differences among individual subjects, and the time series component, reflecting changes within a single entity over time. Analyzing panel data can be intricate, yet its flexibility is advantageous due to the wealth of unique data points it provides. Panel data is characterized by increased variability, reduced collinearity between variables, greater degrees of freedom, and enhanced efficiency. Nevertheless, like any analytical approach, panel data analysis has its limitations, emphasizing the importance of selecting appropriate methods based on the data's specifications.

In model construction, researchers frequently encounter situations where the working data deviates from the assumption of independent and identically distributed (i.i.d.) observations. Instead, units may exhibit correlations within groups while being independent across groups. Deviations from the i.i.d. assumption can lead to issues in estimating parameters reliant on this assumption, such as standard errors and variances. Consequently, it is common practice to cluster standard errors to account for within-group correlation in the dataset. However, when testing our results using clustered standard errors, we encountered alerts indicating that our Variance-Covariance Estimator (VCE) did not have a sufficient rank to perform certain model tests, and we obtained no F-statistic in some estimations. This issue arises because the theoretical foundation for standard error calculation is asymptotic in the number of clusters, while our dataset contains only five clusters. As a rule of thumb, we should estimate as many parameters as we have clusters. Given that we are estimating more than five parameters, we encountered the aforementioned error. To address this issue, we opted to disregard the clustering of standard errors in models where these errors occurred. These different model specifications allow us to explore the relationship between immigration and housing prices from various angles, considering both long-term and short-term effects and addressing potential sources of bias and endogeneity.

5. Results and Discussion

We now present the results of our econometric analysis regarding the impact of immigration on Portuguese real estate prices, accompanied by an in-depth discussion of the key findings. Multiple models have been employed, utilizing various estimation methods and combinations of variables to comprehensively examine this issue.

The dependent variable in our models is the logarithmic representation of housing prices, adjusted for inflation. Our primary independent variable is the change in the number of foreign-born individuals divided by the total number of immigrants in the previous year. In the IV regressions, we instrumented this main independent variable using a widely accepted shift-share instrument. This approach was essential for addressing the endogeneity problem associated with the main independent variable and the reverse causality issue between immigration and housing prices. Furthermore, all models included year fixed effects to account for macroeconomic trends. While the significance of variables such as GDP and mortgage rates in determining housing prices was discussed in Section 2.3, it is worth noting that the macroeconomic trends are already captured by the year dummies. Introducing additional macroeconomic control variables would lead to the exclusion of some year dummies due to collinearity issues. Therefore, aside from the year dummy variables, we have not included national-level economic controls in our models.

It is important to mention that the number of observations varies across our models, ranging from 40 in the first differences models to 45 in the remaining fixed effects models. This deviation from the initially expected 50 observations is due to the nature of our chosen estimation methods, as discussed in Sections 4.1.2 and 4.1.3. Specifically, these methods require the computation of changes, resulting in the omission of the first year of observations. Additionally, the first differences regression leads to the loss of one additional year of data.

In Table 5.1, all the models represent the baseline specification, based on equation 4.15, which consists exclusively of the dependent variable, the main independent variable, and the inclusion of year dummy variables. These models are designed to capture the relationship between these key variables while accounting for time-specific effects. Model 1 is a standard fixed effects regression, wherein we incorporated fixed effects to control for unobserved time-invariant factors. This model allowed us to examine the impact of immigration on housing prices while mitigating the influence of these unobservable factors. In Model 2, we employed a first differences regression approach. This technique involves differencing the variables over time, effectively capturing changes in the variables from one period to the next. This approach helped us analyze the short-term effects of immigration on housing prices. Model 3 is a 2SLS instrumental variable regression, where we use an additional fixed effects estimation method. We also went through different endogeneity tests to confirm that the 2SLS IV regression approach was valid. Instrumental variables are employed to address endogeneity concerns in the main independent variable. By combining this approach with fixed effects, we further refine our analysis, considering both unobservable factors and endogeneity. Model 4 is also a 2SLS instrumental variable regression, however, this time, it incorporates a first differences estimation method. Here, we used instrumental variables to address endogeneity while focusing on the short-term dynamics of immigration's impact on housing prices. Analyzing the results of Table 5.1, we observe that both Model 2 and Model 4 exhibit statistical significance at the 5% and 10% significance levels, respectively. In Model 2, the coefficient associated with the main independent variable is -0.099. This implies that, all else being equal, a one-percentage-point increase in immigration inflow corresponds to a substantial 9.9% decrease in housing prices. Model 4 reveals a similar pattern, indicating that, under identical conditions, a one-percentage-point increase in immigration inflow results in a 7.8% reduction in housing prices. As for Model 1 and Model 3, we did not find a statistically significant relationship between immigration and housing prices. However, it is noteworthy that both models exhibited negative coefficients (although not statistically significant), reinforcing the findings of Models 2 and 4.

Additionally, we conducted weak instrument tests to assess the instrumental variables' strength. When instrumental variables are weak, the IV estimator may introduce significant bias, potentially exceeding that of the OLS estimator. To evaluate instrument strength, we compared the S-W F Stat values to critical values based on Stock & Yogo (2005). The results of the weak instrument tests for Model 3 and Model 4 yield values of 11.94 and 12.23, respectively. In comparison to the lowest critical value of 16.38, which allows for up to a 10% bias relative to OLS, our instrument could be considered weak. However, it is important to note that the second lowest critical value is 8.96, permitting a 15% bias relative to OLS, and we surpassed this threshold. Moreover, the instrument we employed is well-established in various fields of research, instilling confidence in its validity. Therefore, while our weak instrument testing results may not be optimal, we believe they are sufficiently robust to yield significant and efficient coefficient estimates.

Table 5.1: Coefficient Estimates of Baseline Specification

	FE **(1)**	**FD** **(2)**	**2SLS – FE IV** **(3)**	**2SLS – FD IV** **(4)**
Immigration Inflow (%)	-0.060 (0.057)	-0.099** (0.042)	-0.051 (0.108)	-0.078* (0.041)
Constant	12.554*** (0.106)		12.555*** (0.106)	
Year Dummies	Yes	Yes	Yes	Yes
Adjusted R^2 / Within R^2	0.94	0.88	0.94	0.88
S-W F Stat			11.94	12.23
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the logarithm of house prices. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*

Moving on to the results in Table 5.2, these models are based on equation 4.14, and in contrast to the previous specifications, they include a vector of control variables aimed at capturing time-varying region-specific effects. Specifically, these control variables encompass the lagged unemployment rate, the percentage of the active population, average per capita income, the logarithm of the number of immigrant college graduates, the percentage of immigrants originating from "high-income" countries, and the average age of dwellings within each region.

Among these models, only Model 5 and Model 6, which represent traditional estimation techniques, yielded statistically significant coefficient estimates. Their interpretations are as follows: Model 5 exhibits significance at the 10% level, indicating that a one-percentage-point increase in immigrant inflow corresponds to an average 12.7% decrease in housing prices. Conversely, Model 6 demonstrates the highest level of significance, at 1%, signifying that a one-percentage-point increase in immigrant inflow is associated with an average 14.7% decrease in housing prices. However, for both Model 7 and Model 8, we do not discern a statistically significant relationship between immigration and housing prices. It is noteworthy that these models still display negative coefficients (though not statistically significant), which aligns with the findings observed in Models 1, 2, 5, and 6. Consequently, we can assert that, based on our dataset and models, there exists a negative effect of immigration on Portuguese real estate prices.

In comparison to prior research, as discussed in Section 2, the consensus on the direction of the immigration-housing prices relationship has been elusive. For instance, Sá (2015) reported that a 1% immigration inflow increase relative to the initial local population led to a mere 1.6% decline in housing prices within the British market, which contrasts starkly with our findings in the Portuguese context. Similarly, Accetturo et al. (2014) identified a negative association between immigration and housing prices in the Italian market, and in line with previous studies, their coefficient estimates were notably lower than our results.

In its majority, our control variables do not present statistically significant coefficients. However, the ones that do are counterintuitive when compared to our literature review and logical expectations. Nonetheless, these are complex relationships and empirical findings may vary depending on the specific context and data. The sign of the coefficients should always be analyzed with caution and considering local economic conditions and the characteristics of the Portuguese housing market. Moreover, the presence of time dummies means that a significant portion of the variability in the dependent variable is already explained by these variables. When a substantial portion of the variance in the dependent variable is already accounted for by time dummies, there may be less room for other variables, such as income per capita or unemployment rate, to explain additional variance.

The coefficient for the unemployment rate lacks statistical significance in all models except for model 8, where it reveals a positive coefficient. This suggests a seemingly counterintuitive relationship: as the unemployment rate in a region rises, housing prices tend to increase. However, there are plausible explanations for this unexpected effect. One plausible explanation is that individuals tend to migrate to regions marked by robust economic growth. In such cases, a growing local economy can trigger an influx of job seekers, leading to higher unemployment rates. Paradoxically, this surge in economic activity can stimulate housing demand, resulting in a price increase. Additionally, the variable is the lagged unemployment rate, this means that the unemployment rate could be a reflection of past economic conditions, and we may be observing the effects of adjustments that have happened since that time. For example, if the economy contracted in the previous year, the lagged unemployment rate might be high because of that contraction, but the economy could be recovering now. Another intriguing observation pertains to the negative coefficient associated with the "income per capita" variable. Conventional economic theory posits that higher income levels should correspond to higher housing prices. Nevertheless, there are instances where an increase in income per capita may contribute to elevated living and housing costs, rendering housing less affordable and driving down prices. Government policies can also play a role, local authorities, for example, might implement regulations to mitigate housing price inflation in high-income areas. Furthermore, among the remaining significant control variables, the variable representing the percentage of wealthy country immigrants stands out. It exhibits a negative coefficient, which may initially seem counterintuitive, as one might expect wealthy immigrants to bolster housing prices. However, several factors can elucidate this phenomenon. Wealthy immigrants may possess distinct market preferences, inclining them towards renting or investing in alternative assets rather than real estate. Alternatively, they may favor high-end properties, with limited impact on lower-priced segments of the housing market.

It is also important to address the lack of statistical significance in our IV regressions, which might suggest potential instrument weakness. However, the weak instrument tests convey a different story for Models 7 and 8. In contrast to the results observed in Models 3 and 4, Models 7 and 8 exhibit significantly improved weak instrument test statistics. Model 5 approaches but falls just short of surpassing the critical value established by Stock & Yogo (2005) with a recorded value of 15.96. On the other hand, Model 6 securely exceeds this critical threshold, boasting a substantial S-W F Stat of 20.83. This reinforces our rationale for selecting the instrumental variable used in our analysis.

Table A.1 reports the first-stage estimates corresponding to the IV estimates of Tables 5.1 and 5.2. In a 2SLS regression, the first stage regression is used to test the validity and strength of the IV used to address endogeneity in the main regression model. The primary objective of the first stage is to ensure that the chosen instrument is relevant or correlated with the endogenous independent variable. In all the models presented, the coefficients of the predicted stock of foreign-born individuals relative to the initial population exhibit a high level of statistical significance (at the 1% significance level). This significant result confirms the existence of a relationship between the instrument and the endogenous variable.

To further enhance the robustness of our findings, we conducted an analysis without the outlier treatment, retaining the original housing price data. For presentation purposes, we have chosen to display the results in Table A.2, which is structured based on equation 4.14, representing the most comprehensive model. We concluded that the coefficients remains negative and exhibits similar results to the previous estimations. We also observed an intriguing relationship in terms of the significance of income per capita and the unemployment rate. In Table 5.2, it is evident that when income per capita is statistically significant, the unemployment rate variable loses its significance. Despite their correlation coefficient of approximately -0.3, we conducted model estimations in Table A.3 without the unemployment rate variable and in Table A.4 without the income per capita variable. In terms of the results, the coefficients remained relatively consistent across all regressions, with no relevant changes, yielding no significant implications from this experiment.

Table 5.2: Coefficient Estimates of Full Specification

	FE (5)	FD (6)	2SLS – FE IV (7)	2SLS – FD IV (8)
Immigration Inflow (%)	-0.127* (0.065)	-0.147*** (0.045)	-0.123 (0.110)	-0.050 (0.041)
Unemployment Rate (t-1)	5.358 (5.110)	9.465 (5.250)	5.430 (5.304)	10.449* (6.034)
Active Population	-0.902 (6.187)	0.348 (6.602)	-0.855 (6.255)	-3.395 (4.948)
Income per capita	-3.416* (1.711)	-2.436* (1.354)	-3.379* (1.863)	0.301 (1.130)
Percentage of Immigrant College graduates (%)	0.319 (0.392)	0.096 (0.334)	0.311 (0.422)	0.129 (0.218)
Percentage of wealthy country immigrants (%)	-0.029* (0.017)	-0.044** (0.018)	-0.037* (0.019)	-0.041* (0.022)
Building Age	0.038 (0.037)	0.012 (0.027)	0.037 (0.038)	-0.010 (0.042)
Constant	7.790 (24.112)		7.348 (25.635)	
Year Dummies	Yes	Yes	Yes	Yes
Adjusted R^2 / Within R^2	0.95	0.90	0.95	0.90
S-W F Stat			15.96	20.83
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the logarithm of house prices. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*

To conclude, we conducted an in-depth econometric analysis to assess the impact of immigration on Portuguese real estate prices. We utilized multiple models, employing various estimation methods and a wide range displayed and non-displayed control variables to comprehensively explore this relationship. Our findings revealed a consistent negative effect of immigration on housing prices in Portugal, with statistically significant results in specific models indicating that a one-percentage-point increase in immigration inflow corresponds to a significant reduction in housing prices, with decreases from 7.8% to 14.7%, respectively. These results, compared to some prior research, align with the results in other European countries and suggest that immigration exerts a downward pressure on the Portuguese real estate market. As we discussed in the literature review, there are some factors that make this a sustained result. Immigrants tend to settle in low-quality dwellings after arriving in a new environment. Additionally, immigrants cluster in certain regions, since it helps with integration and cluster in dwellings, even sharing them among multiple families, in order to reduce housing costs (Card, 2001). Moreover, since immigrants usually come from lower socioeconomic statuses, they tend to cluster in regions with a high supply of low-skilled jobs and eventually depreciate the average wage level (Carter, 2005). Ultimately, these factors and others tend to depreciate the native's comfort and perception of local amenities and leads them to leave for other regions thus offsetting the inflow of immigration and potentially leading to a decrease in housing demand.

Additionally, our weak instrument tests, while not ideal, provided confidence in the validity of the instrumental variable used in our analysis. Despite the presence of several control variables, their significance in explaining housing price variance was limited, reflecting the complex and context-dependent nature of these relationships. Overall, our findings shed light on the intricate dynamics between immigration and real estate prices in Portugal, emphasizing the importance of considering local economic conditions and unique market characteristics when interpreting these relationships.

6. Conclusions

Portugal has experienced significant shifts in immigration patterns over time. In recent decades, Portugal has witnessed a transformation in its immigration landscape. With its stable political environment, pleasant climate, and economic prospects, Portugal has become an attractive destination for immigrants from various regions, including Europe, Africa, Brazil, and Asia. The European Union's expansion, Portugal's membership in the Schengen Area, and its growing reputation as a welcoming nation have all played a role in this shift. As a result, Portuguese cities, particularly Lisbon and Porto, have seen increased cultural diversity and a growing immigrant population. This demographic change has led the population to claim that immigrants were outpricing natives and sparked interest in understanding the relationship between immigration and various aspects of Portuguese society, including the housing market.

Past research has uncovered a noteworthy relationship between immigration and various socioeconomic aspects, including wage levels and initial housing conditions. Notably, these studies have often highlighted a negative relationship between immigration and wage levels, with immigrants frequently residing in low-quality housing upon their arrival. Nevertheless, it is essential to acknowledge that the relationship between immigration and the real estate market is intricate and subject to multifaceted influences. The impact of immigration on housing prices remains uncertain and exhibits inconsistencies across different studies. While some researchers have reported immigration as a driver of housing price increases in countries such as the United States and Switzerland, others have observed a decline in housing prices associated with immigration in places like Spain.

The data characterization process undertaken in this study entailed an analysis of various databases, including the Labor Force Survey (LFS), the *Imposto Municipal sobre as Transmissões Onerosas de Imóveis* (IMT) and the *Imposto Municipal sobre Imóveis* (IMI) databases. The LFS dataset afforded insights into key attributes of Portugal’s population, such as age, gender, marital status, education, employment, and earnings. Meanwhile, the IMT and IMI databases delivered comprehensive information regarding housing prices, which were adjusted for inflation to accurately portray real market conditions. Data treatments, including outlier removal, were diligently applied to ensure the accuracy and reliability of the dataset.

The core of this study revolved around conducting a robust econometric analysis to uncover the impact of immigration on Portuguese real estate prices. Multiple models were meticulously constructed, employing diverse estimation methods and combinations of variables. The empirical results underscored the substantial and statistically significant negative effect of immigration inflow on housing prices. Specifically, the findings revealed that a one-percentage-point increase in immigration inflow corresponded to a noteworthy decrease in housing prices, ranging from 7.8% to 14.7%. By exploring how immigration affects the housing market, this study offers insights for policymakers to address housing affordability. Investors and real estate professionals can adapt strategies based on these findings, potentially identifying promising areas with strong immigration trends. Given the scarcity of literature on immigration's influence on real estate prices in Portugal and the divergent findings in existing studies, this research serves as a valuable addition, contributing to a better comprehension of how Portuguese housing prices respond to various determinants and how immigration affects different markets around the world.

Nonetheless, it is important to recognize the limitations of this study. The sample size employed for model estimation remains relatively small. Consequently, the robustness and generalizability of the results may be limited. Future research can substantially enhance the reliability of findings by expanding the sample size by, for example, replacing the LFS database with a database aggregated at the municipal level instead of NUTS II level. Another hypothesis would be using quarterly housing prices data thus increasing the number of observations. Furthermore, exploring additional factors such as housing supply, native outflow, demand dynamics and immigrant's perception by the natives could enrich our understanding of the Portuguese real estate market. Moreover, the adoption of alternative estimation methods like Spatial Correlation model considers how the relationships between regions are interconnected. However, we didn't find it was reasonable considering the reduced number of regions in our database. Another alternative estimation method is the Limited Information Maximum Likelihood (LIML) but on the same page Limited Information Maximum Likelihood (LIML) assumes a sufficiently large sample size. With a large enough database these models hold potential for generating alternative estimates worth exploring. In addition, although it was not possible to experiment due to lack of computational power, employing a wild bootstrap approach as an alternative to the conventional standard errors clustering method could offer improvements and merits exploration in subsequent research. This method can address issues like heteroscedasticity, autocorrelation, and model misspecification and it doesn't rely on specific data structures or clustering. Instead, it resamples the residuals from your model to simulate the data's variability under different error structures.

In conclusion, this thesis has provided important insights into how the Portuguese economy, the real estate market, and housing prices are connected. The findings underscore the pronounced influence of demographic and macroeconomic factors in shaping the real estate landscape, alongside the significant impact of immigration on housing prices. The present-day housing crisis facing Portugal underscores the urgency of addressing these issues. We sincerely hope this research can be a starting point for future efforts to improve the situation.

References

Accetturo, A., Manaresi, F., Mocetti, S., & Olivieri, E. (2014). *Don't stand so close to me: The urban impact of immigration*. *Regional Science and Urban Economics, 2014, vol. 45, issue C, 45-56*

Adão, R., Kolesár, M., & Morales, E. (2018). *Nber Working Paper Series Shift-Share Designs: Theory And Inference. NBER Working Paper No. 24944.*

Agnew, K., & Lyons, R. C. (2018). The impact of employment on housing prices: Detailed evidence from FDI in Ireland. *Regional Science and Urban Economics*, *70*, 174–189.

Akbari, S., & Aydede, Y. (2012). Effects of immigration on house prices in Canada. *Applied Economics*, *44*(13), 1645–1658.

Alfaro, L. (2017). Gains from Foreign Direct Investment: Macro and Micro Approaches. *The World Bank Economic Review*, *30*(Supplement_1), S2–S15.

Antoine, B., & Lavergne, P. (2014). Conditional moment models under semi-strong identification. *Journal of Econometrics*, *182*(1), 59–69.

Bartel, A. P. (1989). Where Do the New U.S. Immigrants Live? *Journal of Labor Economics*, *7*(4), 371–391.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics*, *19*(1), 22–64.

Carter, T. (2005). The influence of immigration on global city housing markets: The Canadian perspective. *Urban Policy and Research*, *23*(3), 265–286.

Chen, J., & Hobbs, P. (2003). Global Real Estate Risk Index. *The Journal of Portfolio Management Special Real Estate Issue*.

Cohen, V., & Burinskas, A. (2020). *The Evaluation of the Impact of Macroeconomic Indicators on the Performance of Listed Real Estate Companies and Reits. Ekonomika (Economics), 2020, vol. 99, issue 1, 79-92*

Colen, L., Maertens, M., & Swinnen, J. (2009). *Foreign Direct Investment as an Engine for Economic Growth and Human Development: A Review of the Arguments and Empirical Evidence*.

Cornell, B. (2010). Economic growth and equity investing. *Financial Analysts Journal*, *66*(1), 54–64.

Correia, M. J., & Lima, F. (2006). *Estatísticas do Emprego-1º trimestre de 2006 6. TEMA EM ANÁLISE O Inquérito ao Emprego: o que é e para que serve?*

Cragg, J. G., & Donald, S. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory*, *9*(2), 222–240.

David Hume. (1777). *Hume, D. (1777). Essays Moral, Political, Literary (LF ed.). Liberty Fund.* (D. Hume, Ed.).

De Bernardi, D., & Rodenholm, R. (2013). *Macroeconomic effects on securitized real estate markets A comparative study of Sweden and Switzerland*. KTH School of Architecture and the Built Environment

Dias, D. A., & Duarte, J. B. (2019). Monetary Policy, Housing Rents and Inflation Dynamics. *International Finance Discussion Paper*, *2019*(1248), 1–26.

Égert, B., & Mihaljek, D. (2007). *Determinants of house prices in central and eastern Europe Monetary and Economic Department*. No 236, BIS Working Papers

Fão, M. (2019). *O Efeito Do Sentimento Económico Nos Preços Do Mercado Imobiliário*. ESTG

Ferrie, J. (1996). *Immigrants and Natives: Comparative Economic Performance in the U.S. NBER Working Paper No. h0093*

Fischer, A., & Degen, K. (2009). *Immigration and Swiss House Prices* (Issue 7583).

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *The Review of Financial Studies*, *15*(3), 751–782.

Giussani, B., & Hadjimatheou, G. (1992). House prices: An econometric model for the U.K. *Netherlands Journal of Housing and the Built Environment*, *7*(1), 31–58.

Goldin, C. (1995). The U-Shaped Female Labor Force Function in Economic Development and Economic History. In T. P. Schultz (Ed.), *Investment in Womens Human Capital and Economic Development* (pp. 61–90). University of Chicago Press.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, *110*(8), 2586–2624.

Gonzalez, L., & Ortega, F. (2009). *Immigration and Housing Booms: Evidence from Spain*. *Journal of Regional Science, 2013, vol. 53, issue 1, 37-59*

Goodhart, C., & Hofmann, B. (2008). *House Prices, money, credit and the macroeconomy*. *No 888, Working Paper Series*

Grimes, A., & Aitken, A. (2010). Housing Supply, Land Costs and Price Adjustment. *Real Estate Economics*, *38*(2), 325–353.

Hoskins, N., Higgins, D., & Cardew, R. (2004). Macroeconomic variables and real estate returns: an international comparison. *The Appraisal Journal*, *122*, 8.

INE. (2022). *Inquérito às Condições de Vida e Rendimento*.

Ismail, N. H., & Nayan, S. (2019). *Determinants of housing price Evidence from panel data*. Working Paper

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Kohlert, D. (2010). The determinants of regional real estate returns in the United Kingdom: a vector error correction approach. *Journal of Property Research*, *27*(1), 87–117.

Lacoviello, M., & Neri, S. (2010). *Housing Market Spillovers: Evidence from an Estimated DSGE Model*. *No 659, Temi di discussione (Economic working papers)*

Lourenço, R. F., & Rodrigues, P. M. M. (2017). *Preços da habitação em Portugal-uma análise pós-crise*. Banco de Portugal

Luo, J., Pereira, R., & Dias, Á. (2020). An Empirical Investigation of the Portuguese Housing Prices (2004-18). In *Journal of Reviews on Global Economics* (Vol. 9).

Mankiw, G. (2016). Macroeconomics (9th Edition). (Cengage)

Michael Schill, S. F., & Rosenbaum, E. (1998). The Housing Conditions of Immigrants in New York City. *Journal of Housing Research*, *9*(2), 201–235.

Moreira Braga, F. (2013). *Mercado Imobiliário em Portugal*.

Pivar, W. H., & McKenzie, D. J. (2008). *California real estate license preparation*. Thomson/South-Western.

Puig, S., & Sánchez, Á. (2018). *Portugal versus Spain: An Iberian economic saga*. *Real Instituto Elcano Royal Institute, ARI 96/2018, 1-23.*

Sá, F. (2015). Immigration and House Prices in the UK. *Economic Journal*, *125*(587), 1393–1424.

Saira Khalil ur Rehman, Nuno Filipe Lopes Moutinho, & Jorge Manuel Afonso Alves. (2020). *The Relationship between Portuguese Economy Indicators and Housing Prices*.

Saiz, A. (2007a). Immigration and housing rents in American cities. *Journal of Urban Economics*, *61*(2), 345–371.

Saiz, A., & Wachter, S. (2011). Immigration and the Neighborhood. *American Economic Journal: Economic Policy*, *3*(2), 169–188.

Sanderson, E., & Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212–221.

Scherrer, Amandine., & Thirion, Elodie. (2018). *Citizenship by investment (CBI) and residency by investment (RBI) schemes in the EU : state of play, issues and impacts.*

Staiger, D., & Stock, J. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*(3), 557–586.

Stepanyan, V., Poghosyan, T., & Bibolov, A. (2010). House Price Determinants in Selected Countries of the Former Soviet Union. *International Monetary Fund, IMF Working Papers*, *10*.

Stock, J., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews (Ed.), *Identification and Inference for Econometric Models* (pp. 80–108). Cambridge University Press.

Surak, K., & Tsuzuki, Y. (2021). Are golden visas a golden opportunity? Assessing the economic origins and outcomes of residence by investment programmes in the EU. *Journal of Ethnic and Migration Studies*, *47*(15), 3367–3389.

Teles Morais, L. (2018). *Portugal's economic prospects: was the political miracle responsible for the economic miracle? Real Instituto Elcano Royal Institute*

Tunaru, R. S. (2017). *Real-Estate Derivatives* (1st ed.). *Oxford*

Winkler, D. T., & Donald Jud, G. (2002). The Dynamics of Metropolitan Housing Prices. In *Journal of Real Estate Research* (Vol. 23, Issue 2).

Xu, L., & Tang, B. (2014). *On The Determinants Of UK House Prices*. *International Journal of Economics and Research, 5(2), 57-64*

Zabel, J. (2012). Migration, housing market, and labor market responses to employment shocks. *Journal of Urban Economics*, *72*, 267–284.

Zainal, R., Chai Teng, T., & Mohamed, S. (2016). International Journal of Economics and Financial Issues Construction Costs and Housing Prices: Impact of Goods and Services Tax. *International Journal of Economics and Financial Issues* |, *6*(S7), 16–20.

Zhu, J., Pryce, G., & Brown, S. (2019). Immigration and house prices under various labour market structures in England and Wales. *Urban Studies*, *56*(9), 1801–1817.

Appendix

Table A.1: First-stage regressions

	2SLS – FE IV Regression (3)	2SLS – FD IV Regression (4)	2SLS – FE IV Regression (7)	2SLS – FD IV Regression (8)
Instrumental Variable	0.847*** (0.245)	1.034*** (0.296)	0.971*** (0.243)	1.079*** (0.237)
Unemployment Rate (t-1)			-28.29** (12.21)	-34.35** (6.034)
Active Population			-16.97 (14.73)	-14.65 (26.06)
Income			-5.172 (3.952)	-8.657** (4.125)
Immigrant College graduates			2.125** (0.887)	1.952** (0.889)
Wealthy country immigrant			-0.082** (0.039)	-0.068 (0.046)
Building Age			-0.014 (0.921)	-0.018 (0.089)
Constant	0.045 (0.286)		135.54** (55.60)	
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2 / Within R^2	0.45	0.40	0.68	0.65
S-W F Stat			15.96	20.83
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the change is the change in immigrants divided by the initial population. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*

Table A.2: Model estimation w/ outliers

	FE (5)	FD (6)	2SLS – FE IV (7)	2SLS – FD IV (8)
Immigration Inflow	-0.089	-0.112**	-0.077	-0.051
	(0.072)	(0.043)	(0.115)	(0.073)
Unemployment Rate (t-1)	2.875	8.188	3.073	8.032
	(5.600)	(6.021)	(5.815)	(8.010)
Active Population	-1.132	-1.180	-1.004	-4.437
	(6.780)	(6.931)	(6.858)	(4.046)
Income	-2.609	-1.787	-2.507	0.655
	(1.875)	(1.569)	(2.042)	(1.397)
Immigrant College graduates	0.127	-0.329	0.105	-0.253
	(0.429)	(0.351)	(0.463)	(0.324)
Wealthy country immigrant	-0.018	-0.033	-0.037*	-0.033
	(0.019)	(0.023)	(0.019)	(0.029)
Building Age	0.032	0.034	0.030	0.017
	(0.041)	(0.032)	(0.042)	(0.040)
Constant	14.745		13.533	
	(26.422)		(28.104)	
Year FE	Yes	Yes	Yes	Yes
Adjusted R2 / Within R2	0.96	0.91	0.96	0.94
S-W F Stat			15.96	20.83
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the logarithm of house prices. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*

Table A.3: Model estimation w/o unemployment rate

	FE (5)	FD (6)	2SLS – FE IV (7)	2SLS – FD IV (8)
Immigration Inflow	-0.142**	-0.172***	-0.094	-0.047
	(0.064)	(0.042)	(0.119)	(0.065)
Income	-3.674**	-3.457**	-3.260*	-1.216***
	(1.697)	(1.445)	(1.917)	(0.265)
Constant	30.907***		29.317***	
	(9.782)		(10.424)	
Year FE	Yes	Yes	Yes	Yes
Adjusted R2 / Within R2	0.95	0.90	0.95	0.91
S-W F Stat			10.88	12.89
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the logarithm of house prices. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*

Table A.4: Model estimation w/o income per capita

	FE (5)	FD (6)	2SLS – FE IV (7)	2SLS – FD IV (8)
Immigration Inflow	-0.082	-0.128**	-0.075	-0.048
	(0.065)	(0.046)	(0.100)	(0.045)
Unemployment Rate (t-1)	6.820	11.714**	6.929	9.878**
	(5.340)	(5.082)	(5.465)	(4.204)
Constant	7.790		7.348	
	(24.112)		(25.635)	
Year FE	Yes	Yes	Yes	Yes
Adjusted R^2 / Within R^2	0.95	0.90	0.95	0.91
S-W F Stat			19.20	13.79
Observations	45	40	45	40

*Coefficients estimates and standard errors in parentheses (Robust standard errors in the FD regressions). Dependent variable is the logarithm of house prices. Adjusted R^2 for FD reg. and Within R^2 for FE. The S-W F-stat is the Sanderson and Windmeijer (2016) multivariate F-test of excluded instruments for weak identification of each endogenous regressor separately; * significant at 10%, **significant at 5%, and *** significant at 1%*