Should Governments Subsidize Homeownership?

A Quantitative Analysis of Spatial Housing Policies

Claudio Luccioletti*

December 16, 2024

Abstract

Should governments promote homeownership? Although such policies are widespread, their welfare implications are not straightforward. While subsidies can overcome financial frictions and, by increasing homeownership, insure against rent volatility, they can also reduce internal migration to productive locations. To address this question, I build a dynamic spatial equilibrium model with coresidence, homeownership, internal migration, and savings decisions. Homeownership provides utility and insurance against aggregate rental price risk but reduces migration due to the transaction costs associated with selling the property. Migration decisions, in turn, affect homeownership. In particular, non-migrant workers can coreside with their parents, which allows them to save and buy a house earlier than migrants. I develop a new strategy to solve dynamic spatial models with aggregate uncertainty, which models agents' expectations about local endogenous prices and wages using lower-rank factors. The model is estimated for Spain and validated using quasi-experimental evidence from recent place-based homeownership subsidies. I find that mortgage interest deduction policies reduce wealth inequality and are welfare-increasing, despite reducing migration to productive locations. Targeting high-wage locations leads to lower welfare gains and does not increase homeownership due to higher house prices.

Keywords: Housing Policies; Coresidence; Homeownership; Internal Migration; Estimation of Dynamic Spatial Models.

JEL Classification: R13; R21; R23; R28.

---

*Bank of Italy, claudio.luccioletti@cemfi.edu.es. I thank my advisor Nezih Guner for invaluable guidance and support. I am also grateful to Manuel Arellano, Florian Oswald, Eduardo Morales, Josep Pijoan-Mas, Diego Puga, Esteban Rossi-Hansberg, and seminar participants at CEMFI and the University of Chicago for their comments and suggestions. Special thanks to Laura Crespo and Sandra García-Uribe for facilitating access to the restricted version of the EFF data. I acknowledge financial support from Grant PRE2020-094296 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future", the María de Maeztu Unit of Excellence (CEMFI MDM-2016-0684), funded by MCIN/AEI/10.13039/501100011033, and CEMFI. The views expressed in this paper are personal and do not necessarily reflect the position of the Bank of Italy.

1 Introduction

It is very common for governments to incentivize homeownership through tax benefits and subsidies. These policies may overcome financial frictions in the mortgage market and, by increasing homeownership, insure against aggregate rent risk and reduce wealth inequality. However, their welfare implications are not obvious. Subsidizing young homebuyers in low-productivity locations may tie them to areas with limited job opportunities, since homeowners are less likely to relocate in the future, partly due to the transaction costs associated with selling their property. On the other hand, providing subsidies in productive locations where house prices are already high can contribute to price increases by further stimulating housing demand. The resulting higher housing costs restrict the ability of lower-income people to access labor markets that offer higher-paying and more stable employment opportunities, and can undermine the original policy objective.[1]

This paper investigates whether governments should promote homeownership and, if so, whether housing policies should be targeted to specific population groups and locations. To address these questions, I develop a dynamic spatial equilibrium model with coresidence, homeownership, internal migration, and saving decisions. Throughout their lifecycle, agents have a set of locations to choose from as their residence, each offering distinct housing prices, wages, and amenities. Additionally, once they make their location choice, they can either buy a house, rent one, or, if they currently live in their birthplace, coreside with their parents. Financial frictions in the credit market imply that agents who want to get a mortgage to become homeowners must first gather sufficient funds for an initial down-payment. Dynamic decisions are taken by forming expectations over local *endogenous* housing prices and wages, which fluctuate in response to multiple aggregate shocks to local housing construction costs and productivity. These shocks are uninsurable for non-homeowners, while homeowners are insured against rental price volatility. Moreover, agents face uninsurable idiosyncratic income and unemployment risk that varies across locations.

Solving dynamic spatial equilibrium models with aggregate shocks is a long-standing challenge in urban economics. Standard tools to solve macroeconomic models with aggregate uncertainty, e.g., Krusell and Smith (1998), cannot be easily applied to high-dimensional settings with geography, where aggregate price and wage dynamics might differ across many locations. I develop a new tractable strategy to solve this class of models by combining insights from Krusell and Smith (1998) and the econometrics literature on factor models (Bai 2009). In particular, I use lower-rank aggregate factors to model agents' expectations about endogenous prices and wages across space, thus reducing the problem's dimensionality. I estimate the agents' forecast rule outside the model, minimizing the computational burden.

[1] A list of the main housing policies implemented in OECD countries is reviewed in OECD (2021). Beyond financial frictions in the mortgage market (Engelhardt 1996) paired with homeownership's insurance against (otherwise uninsurable) aggregate rent risk (Paz-Pardo 2023), rationales for these policies may include the existence of positive externalities associated with homeownership such as crime reduction (DiPasquale and Glaeser 1999, Disney et al. 2023) or policy objectives to reduce wealth inequality (Kaas, Kocharkov and Preugschat 2019, OECD 2021).

The forecast rule is accurate, as the prediction of the factor model aligns very closely with the observed time series for local prices and wages, which can be perfectly matched in the benchmark equilibrium. In counterfactual exercises, in contrast, the agents' forecast rule is updated to align with the new equilibrium prices and wages.

In the model, the housing and location decisions are interconnected. Homeownership provides direct utility and insurance against rental price volatility, but makes it more costly to migrate, as agents need to sell their house before moving and bear the associated transaction cost. On the other hand, migration decisions themselves influence homeownership outcomes. Young workers who choose not to migrate from their birthplace (*natives*) may forgo potential labor market opportunities elsewhere, but can coreside with their parents, saving on living expenses. This allows them to accumulate resources for a down-payment and secure a mortgage earlier than *migrants*. Moreover, agents receive housing bequests with some probability over their lifecycle, and become homeowners in their birthplace if they are native while are forced to sell if they are migrants.

I estimate the model using data from Spain, a country with high homeownership, yet low rates of internal migration despite large disparities in income and unemployment risk across locations. Some parameters, including those governing income and unemployment transition probabilities, are estimated externally. The remaining parameters are calibrated to match key moments in the data, such as the lifecycle evolution of homeownership, coresidence, and migration rates, as well as cross-sectional moments which include the median wealth to income ratio.

I use quasi-experimental evidence from a recent policy that subsidized young homebuyers in small cities to validate the model. The place-based design of the policy allows me to find a proper control group in the data, i.e. young residents of slightly larger municipalities where the population was just above the threshold to qualify for the subsidy. In line with the model's prediction, I find that the policy increased homeownership and reduced out-migration rates: new homeowners decrease their annual migration by around 2 percentage points, a substantial effect given the average migration rate of 0.8%. When simulating the same place-based policy in the model, I obtain an untargeted migration elasticity with respect to changes in homeownership status that is very close to the one estimated in the data.

The model also perform well with respect to other non-targeted moments. These include the lifecycle homeownership gap between natives and migrants, the lifecycle migrant income premium, and the observed negative relationship between local homeownership rates and Gini wealth inequality across locations. I find that the option to coreside with parents is the main reason behind the higher homeownership rate observed among natives. Coresidence offers natives a way to overcome frictions in the mortgage market, by allowing them to save on housing costs and accumulate funds for the down-payment earlier than migrants, despite earning less on average. Moreover, the negative relationship between local homeownership and wealth inequality is explained by the fact that a higher homeownership rate tends to disproportionally increase wealth at the bottom of the distribution.

Next, I use the model to study the implications of an array of housing policies in terms of welfare, measured by a consumption compensating variation for newborn agents. I also assess the majority support for each policy by computing the share of newborn individuals who benefit from it compared to the benchmark scenario without it. Policies in the model are implemented in a fiscal neutral fashion, ensuring a balanced government budget through the adjustment of income taxes. Moreover, prices and wages in each location adjust in equilibrium in response to changes in local housing demand and migration induced by the policy. Agents' expectations about local endogenous prices and wages are consistent with their new stochastic steady state.

First, I consider mortgage interest deductions. This policy, which was in place in Spain until 2013, allows homeowners to annually deduct 1,300 euros in mortgage interests from their labor income taxes. As a result, I find that homeownership increases by 0.8 percentage points, which leads to a reduction in both wealth inequality (-0.95%) and internal migration (-0.86%). Overall welfare increases by 1.64% and the policy has majority support, although it reduces the population in productive locations (-0.2%). If mortgage interest deductions are only targeted to residents in productive urban locations, then spatial housing price dispersion and wealth inequality increase, and welfare gains are lower with respect to the untargeted policy (0.6%). Due to higher house prices in equilibrium, homeownership increases only marginally. Targeting unproductive rural locations, instead, barely affects welfare. The targeted policy leads to an increase in supply in low-wage labor markets which, in turn, increases spatial income dispersion by 0.4%.

Second, I study the effect of a 3,000 euros rent subsidy to workers younger than 35 and earning less than a low-income threshold (19,500 euros annually). The policy, introduced in 2018, is found to decrease welfare by 1.34%, despite increasing internal migration (0.3%) and the population in productive locations (0.1%) following the decrease in homeownership (-1.7%). Few agents are better off with the policy (15%), whereas all workers need to pay higher taxes to finance it. Targeting rent subsidies to productive urban or unproductive rural locations marginally mitigates the negative welfare effect of the policy. Support for the targeted policies, however, is even lower than for the untargeted rent subsidies.

Finally, subsidizing the down-payment makes coresiding less attractive and increases migration. This occurs because less people need to stay in their birthplace in order to live with their parents and overcome the financial friction in the mortgage market. Nonetheless, the welfare gains associated with the policy are small (0.3%). Although the partial equilibrium gains are substantial (2.5%), most of these benefits disappear as taxes adjust, and due to the large increase in house prices following the rise in the homeownership rate.

The presence of aggregate uncertainty in the economy has important implications for housing policies. In counterfactual simulations without aggregate shocks, the welfare gains from untargeted mortgage interest deductions would be six times lower (0.03%), whereas the policy targeted to rural locations would lead to welfare losses (-0.8%). In the presence of aggregate shocks, pro-homeownership policies such as mortgage interest deductions are particularly valued by risk-averse agents, because owning a house provides insurance against rental price volatility. Homeowners are less likely to self-insure against negative income shocks by migrating to less affected locations, but the net insurance value of owning rather than renting is positive.

Related Literature This paper contributes to the recent literature studying homeownership in the context of quantitative spatial models (Giannone et al. 2023, Greaney 2023, Oswald 2019). While prior works have emphasized the negative influence of homeownership on internal migration, I contribute by providing new causal evidence for this channel using a place-based housing policy. I also study the role of coresidence and housing bequests as novel mechanisms affecting homeownership through migration decisions. Finally, this is the first spatial equilibrium model with aggregate shocks that analyzes the welfare and aggregate implications of housing policies. I find that a model without aggregate uncertainty would vastly underestimate the welfare gains from pro-homeownership policies. The focus on the evaluation of policies in a model with geography, with a particular emphasis on their effects on the spatial allocation of labor, relates to works by Eeckhout and Guner (2017), Fajgelbaum et al. (2019), Ganong and Shoag (2017), and Hsieh and Moretti (2019). A large macro literature, surveyed in Davis and Van Nieuwerburgh (2015) and Piazzesi and Schneider (2016), models homeownership in general equilibrium.[2] Within this literature, Floetotto, Kirker and Stroebel (2016), Kaas et al. (2021) and Sommer and Sullivan (2018) study the welfare effects of housing policies, while Kaplan, Mitman and Violante (2020) focus on aggregate price fluctuations and their macroeconomic effects. None of these papers, however, model homeownership within a dynamic spatial equilibrium model with aggregate shocks.

Estimating this class of dynamic models with aggregate uncertainty has been a longstanding challenge in urban economics. For example, "dynamic hat algebra" strategies require the economy to approach a stationary equilibrium in which aggregate variables are constant over time, and hence do not allow for (non unexpected) aggregate shocks in equilibrium (Artuç, Chaudhuri and McLaren 2010, Caliendo, Dvorkin and Parro 2019, Kleinman, Liu and Redding 2023). Other dynamic spatial equilibrium models (Desmet and Rossi-Hansberg 2014, Desmet, Nagy and Rossi-Hansberg 2018), instead, make the problem effectively static, because agents are either myopic or the future does not affect their optimal decisions. Finally, Bilal (2023) introduces the "Master Equation" representation of the economy, a concept developed in the mathematics mean field games literature (Cardaliaguet et al. 2019), which Bilal and Rossi-Hansberg (2023) use to study aggregate shocks in a dynamic spatial setting with first-order perturbations around the steady state. I develop a new tractable strategy to solve this class of models by combining insights from Krusell and Smith (1998) and the econometrics literature on factor models (Bai 2009). In particular, I use lower-rank aggregate factors to model agents' expectations about local endogenous prices and wages, which reduces the problem's dimensionality while keeping decisions dynamic and allowing for aggregate fluctuations.[3] Aggregate shocks are estimated from the data and can potentially be large. This

[2]Sinai and Souleles (2005) emphasize the value of homeownership as a hedge against rent risk – despite the fact that homeowners suffer from house price risk in case they decide to relocate. Coherently with my results, they find that rent risk empirically dominates price risk for most U.S. households.

[3]The estimation of the agents' forecast rule involves principal component analysis (Bai 2009). Relatedly, Fernández-Villaverde, Hurtado and Nuño (2023) solve a model with aggregate financial risk and use machine

contrasts with the method proposed by Bilal (2023), which requires the assumption that aggregate shocks are small.

This paper is also related to the literature on the influence of internal migration on lifecycle earnings (Bilal and Rossi-Hansberg 2021, De la Roca and Puga 2017, Díaz, Jáñez and Wellschmied 2023, Kennan and Walker 2011). I contribute by investigating new sources of home-bias in migration choices, namely coresidence and housing bequests. Relative to Bilal and Rossi-Hansberg (2021), heterogeneity in the birthplace is an additional channel that affects location decisions of financially constrained agents. Indeed, people born in productive locations can enjoy high earnings and no housing costs while they coreside with parents, whereas those born in unproductive locations face a migration trade-off between cheap cities with low labor market returns and expensive cities with better income and employment opportunities. Moreover, I study housing policies that, by affecting housing tenure decisions, can change the incentives to stay in the birthplace, thus exerting a significant influence on labor earnings. Differently from Zerecero (2021), home-bias preferences are partly endogenous to the economic environment, and can thus be affected by the policy-maker.[4] Residual exogenous preferences for the birthplace are estimated to be small, and counterfactual exercises reveal that the coresidence channel encompasses most of the benefits of staying enjoyed by Spanish natives.

In most OECD countries, there has been a increase in the share of young adults coresiding with parents over the past few decades (Esteve and Reher 2021, Srinivas 2019). Related to this paper, Kaplan 2012 and Rosenzweig and Zhang 2019 study the link between coresidence and individual savings behavior. They highlight two opposing channels. On the one hand, coresident children are negatively selected in terms of income and have lower precautionary savings motives. On the other hand, sharing of the housing public good during coresidence increases savings. I incorporate these channels and contribute to the literature by adding the coresidence decision into a spatial model where only natives can live with parents. This introduces a new trade-off, whereby coresidents must forgo the benefits of internal migration. Moreover, by also including the homeownership choice, which is subject to a down-payment requirement, I highlight another channel through which savings accrued during coresidence increase lifecycle wealth accumulation.[5]

Finally, this paper connects to the literature examining the relationship between homeownership and wealth inequality (Kaas, Kocharkov and Preugschat 2019, Kindermann and Kohls 2018, Paz-Pardo 2023). I provide new evidence focusing on a single country exploiting variation across locations and, taking advantage of the panel dimension of my data, within households, rather than relying on cross-country comparisons. The literature emphasizes that widespread access to homeownership lifts the wealth of the income poor relatively more, and tends to decrease wealth inequality. I evaluate the impact of housing policies on wealth

learning techniques to estimate households' forecast rules.

[4]For example, the option to coreside with parents becomes less attractive as housing costs decline (e.g. as a result of a policy).

[5]Relatedly, Grevenbrock, Ludwig and Siassi 2023 examine the differences in homeownership rates between Germany and Italy in a quantitative model with coresidence.

inequality in a model where the allocation of homeowners and renters across space affects the spatial dispersion of prices, which has an additional influence on the housing wealth distribution that can either reinforce or counteract the previous channel.

The remainder of the paper is organized as follows. In Section 2, I present the data and facts. Section 3 provides details on the model's economy. In Section 4, I describe how the model is estimated. Section 5 validates the model. In Section 6, I assess the welfare and distributional implications of housing policies. Finally, Section 7 concludes.

2 Facts on Homeownership and Internal Migration

In this section, I discuss the data and document some key facts about the relationship between homeownership and internal migration in Spain. First, people that become homeowners are less likely to migrate in the future. Second, I show that non-migrants are more likely to be homeowners, and explore the channels of coresidence with parents and housing bequests as candidate explanations for this homeownership gap.

2.1 Geography and Data

Geographic locations are defined as combinations of NUTS-1 regions in peninsular Spain and groups of municipalities within a region classified as either rural or urban.[6] Peninsular Spain comprises six distinct NUTS-1 regions, resulting in a total of 12 locations when combined with urban and rural areas within each region. A map of the locations and more information on their construction is given in Appendix Figure A1.

The empirical analysis focuses on individuals aged between 25 and 64 who are Spanish-born citizens and currently active, i.e. employed or unemployed. A *migrant* is defined as an individual currently residing in a Spanish location that is not their birthplace. A *native* refers to a person who currently resides in their birthplace. As such, migrants and natives should not be seen as fixed types, but rather as the result of choices that may change period by period. Finally, *coresidence* refers to the living arrangement of individuals residing with their parents.[7]

The ideal dataset would be a large panel with comprehensive information on five key dimensions: location (current and birthplace), homeownership, coresidence, income, and wealth. While no available dataset perfectly meets these criteria, I use four different data sources to measure specific features in the data. Appendix B gives more information on these four main and additional datasets, and assess their comparability. The main datasets used in the analysis are listed below:

[6]The NUTS classification is a hierarchical system developed by the European Union to divide its territory for statistical purposes. The NUTS-1 refers to major socio-economic regions, with average population size between 3 and 7 millions.

[7]The phenomenon of sharing a flat with a roommate or partner is not considered. Although such living arrangements can result in some housing cost savings due to economies of scale, living with parents often leads to substantially higher savings (Rosenzweig and Zhang 2019).

1. Continuous Work History Sample (*Muestra Continua de Vidas Laborales*, or MCVL): This administrative dataset covers the years 2005-2019 and includes 5.25 million observations. It is an individual-level panel with data on location (current and birthplace), coresidence and income.

2. European Union Statistics on Income and Living Conditions (EU-SILC): This survey spans the years 2004-2019 and contains 74,000 Spanish households. It is a household-level panel that includes information on (current) location, homeownership, coresidence, household-level income, and household members' employment status.

3. Survey of Household Finances (*Encuesta Financiera de las Familias*, or EFF): This survey covers the years 2005-2020, comprises 15,000 households, and provides a household-level panel that covers all the important data dimensions. In particular, it is the only dataset with wealth information. However, it is relatively small and has restricted access to location information when the number of local observation is too low, which makes it unsuitable for some estimation procedures.[8]

4. Census of Population and Housing: This dataset includes 1.3 million observations for the year 2011 with detailed information on location, homeownership, and coresidence. However, it is a cross-sectional dataset that does not provide information on internal migration, income, or wealth.

2.2 People Are Less Likely to Migrate After Buying a House

Using a panel regression approach and a diff-in-diff analysis that leverages quasi-experimental variation from a place-based policy, I find that homeownership is associated with a lower probability of future internal migration. Both estimation strategies, which are conducted independently, reach very similar conclusions. The probability of moving reduces by around 1.9 percentage points for homeowners, a substantial effect given the average annual migration rate of 0.82%.

Along the lifecycle, homeownership tends to increase and internal migration rates tend to decline (see Appendix Figure A2). However, as shown in column (1) of Appendix Table A1, I find that homeowners are less likely to migrate than renters even after controlling for age and other observables. While I include year-region fixed-effects to account for local time trends, other unobservables may be biasing the migration elasticity. Therefore, I exploit the panel dimension of the EU-SILC by also including household fixed-effects. I show that household heads that become homeowners are less likely to migrate in the future: the probability of migrating decreases by approximately 1.92 percentage points for homeowners (see column

[8]Moreover, the EFF does not have individual-level information on coresidents' income. The Bank of Spain, as the EFF data provider, has authorized remote access to restricted geographic information (current and birthplace) in compliance with privacy guidelines. The 2005-2006 wave lacks data on birthplace. Whenever this piece of data is needed for the analysis, I use the 2008-2020 version of the panel.

2 of Table A1), which is significant considering the average annual migration rate of 0.82%.[9] However, the presence of time-varying unobservable factors, not captured by year-region and household fixed-effects, may influence the estimated elasticity.

To identify the causal effect of homeownership on migration rates, I take advantage of the quasi-experimental nature of a 2018 policy that subsidized homeownership for individuals younger than 35 residing in municipalities with less than five thousands inhabitants. The place-based design of the policy allows me to deal with the omitted variable concern by finding a proper control group, i.e. young people belonging to the same age group but living in slightly larger municipalities, where the population was just above the threshold to qualify for the subsidy. Further details on the policy and the data are given in Section 5.1.

I plot in Figure 1 the event studies representing the impact of the policy on homeownership (Figure 1a) and migration rates (Figure 1b).[10] I find that the subsidy increased homeownership among the treated group (relative to the control), which I interpret as the first stage effect of the policy, and decreased out-migration, which I interpret as the reduced form effect. As can be seen in columns (1) and (2) of Appendix Table F5, the homeownership rate among the treated increased by 0.115 and the annual migration rate decreased by 0.002 on average.[11] When combining the first stage and reduced form estimates, I obtain a migration elasticity with respect to changes in homeownership of -1.83 pp., which is very close to the elasticity estimated with the panel regression.[12]

The absence of significant pre-trends in the event studies, as shown in Figures 1a and 1b, supports the conditional exogeneity assumption of the treatment. As additional exogeneity checks, I run two placebo event studies limiting the sample to individuals aged 37-40. These individuals, just above the age eligibility threshold, could not have accessed the subsidy in any post-treatment years of the event study. Consistent with the exclusion restrictions, the placebo treatment does not significantly affect the outcomes, as shown in Appendix Figures F15a and F15b. I also estimate the migration event study by focusing on young individuals *born* in smaller municipalities, using young people born in marginally larger municipalities as the control. This birthplace-based treatment, arguably more exogenous than one based on current residence, yields estimates comparable to the baseline regression (Appendix Figure F16). Further robustness checks are performed in Section 5.1.

The reduction in internal migration resulting from homeownership has important implications, as changes in location can substantially impact lifetime earnings. Using a migration event-study design, I show that internal migrants experience persistent income gains after moving, especially when college-educated and when moving to urban locations (see Appendix Figure A3). Additionally, internal migrants earn about 3% more than local obser-

[9]Homeownership is the single most important observable explaining internal migration rates. The only other statistically significant coefficients, although with a lower estimated magnitude, are those associated with age and college education.

[10]The event studies are estimated using specification (14) in Section 5.1

[11]These estimates come from the first stage and reduced form difference-in-differences version of the event study regressions (14). More details on this diff-in-diff specification can be found in Appendix F.1.

[12]The panel elasticity is -1.92 when using the full set of controls (column 2 of Table A1) and -1.86 when only including the controls used in the event studies of Figure 1 (column 1 of Table 4).

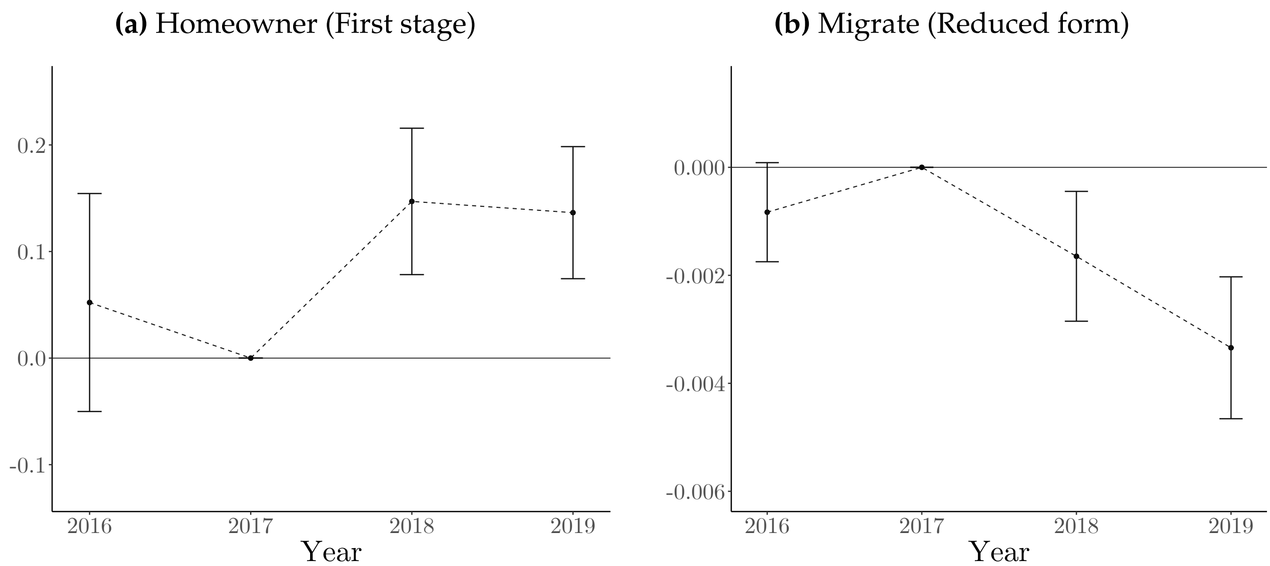


Figure 1: Treated: People aged less than 35 living in small cities (<10k inhabitants). Control group: same age living in slightly larger cities (10k-20k). Treatment year: 2018. Migration across regions (NUTS-1) and rural-urban areas (rural <10k, rural >10k or urban >40k). Included controls and fixed-effects: gender, age, age squared, region, and region-year. Clustered (locations) standard errors. Data: EPF, EPC, EVR 2016-2019.

vationally equivalent natives in the cross-section, as can be seen in column (1) of Appendix Table A2. These important data features on local income dynamics are reproduced in the model, where wages are determined locally and are endogenously higher in urban locations.

2.3 Natives Are More Likely to be Homeowners

One might expect that internal migrants, who earn on average higher incomes than observationally equivalent natives, would experience higher rates of homeownership. Yet, I document that natives are more likely to be homeowners than internal migrants. As can be seen in Figure 2a, the homeownership gap after controlling for observable characteristics and adjusting for cohort effects stands at 20 percentage points at younger ages and stabilizes at around 10 pp. along the lifecycle.[13,14]

There are three main candidate explanations for the native-migrant homeownership gap. First, migrants are more likely to change residence again in the future (see Appendix Figure A5), and so may be less willing to settle down and buy a house at younger ages. While self-selection of stayers into homeownership is expected to play a role, it is unlikely to fully account for the extent of the homeownership gap. Notably, the gap persists even at older ages, despite migration events becoming relatively infrequent after age 40.

Second, natives are more likely to have inherited the house where they currently live (see Figure 2b). While migrants are equally likely to receive housing bequests as natives, they may be forced to sell the inherited property (and incur the associated transaction costs) if

[13] The relatively high observed homeownership rate at younger ages should be interpreted taking into account that we are conditioning on individuals not living with parents (25% of the population at age 25).

[14] The reference group is the 1975 cohort. The procedure to adjust for cohort effects is discussed in Appendix A as a comment to Figure A4.

they do not wish to move to their parents' former residence. However, although housing bequests may partially explain the native-migrant homeownership gap in older age groups, when the observed difference in the share of people living in inherited houses is larger, they are unlikely to account for the gap among younger individuals.

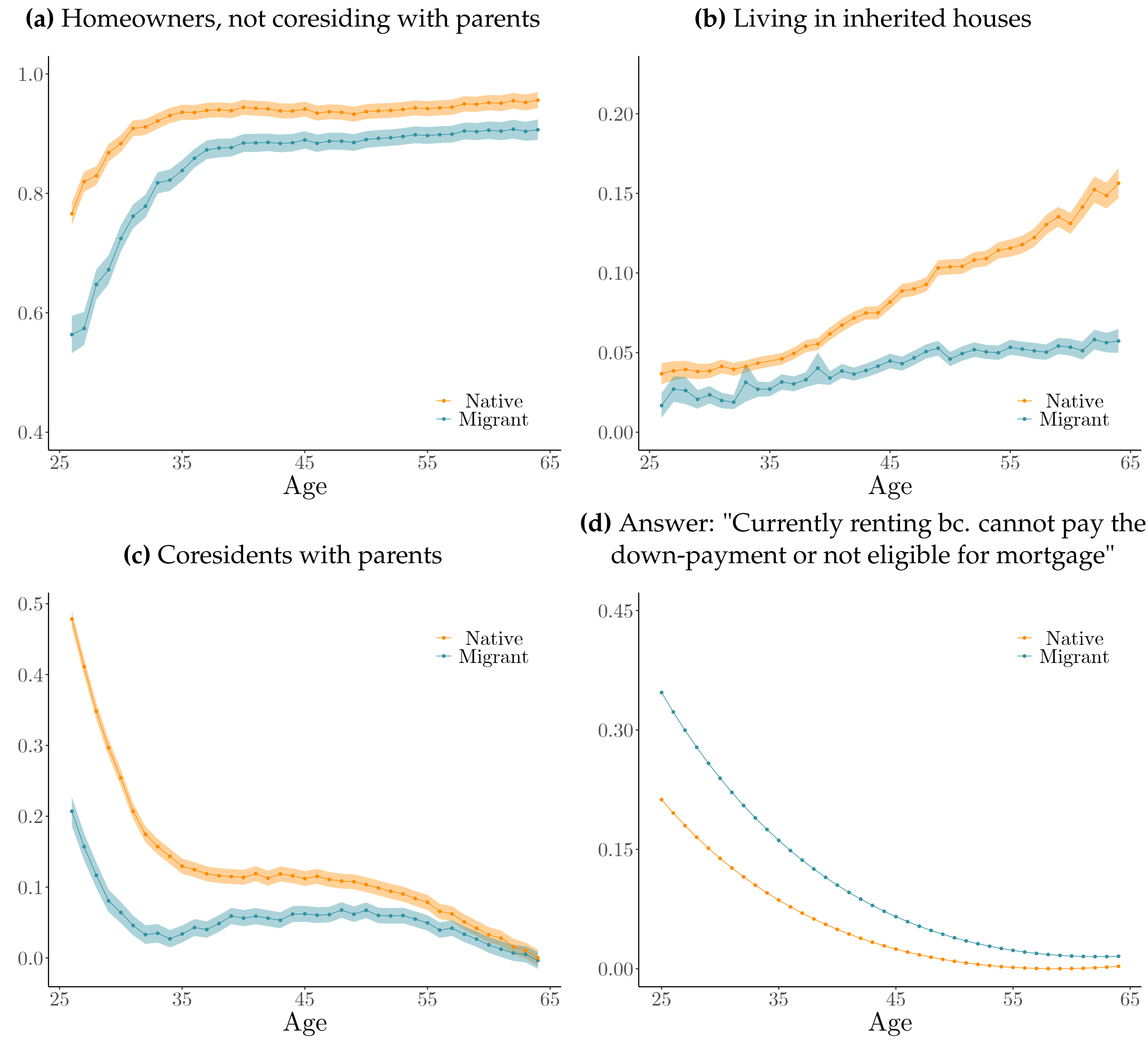


Figure 2: The figures plot age fixed-effects (panels **a** to **c**) and an age polynomial function (panel **d**). Other included fixed-effects are: location, gender, college-educated, married, parent, employed (reference group: male, non-college, single, not parent, employed). Panel **d**: share of surveyed people who say they are currently renting because "I am not eligible for a mortgage" or "I would not be able to afford the down-payment on the house". Confidence intervals at 95% level with heteroskedasticity-robust standard errors are plotted in panels **a** to **c**; not available in panel **d** due to privacy restrictions. Data: Census 2011 (panels **a** to **c**), EFF 2020 (panel **d**).

Finally, natives are more likely to live in the same location as their parents and have the option to coreside with them. As plotted in Figure 2c, the share of young native coresidents is substantially higher than among migrants.[15] By not having to pay housing costs during the coresidence period, it can be easier for natives to save for the mortgage down-payment

[15] Some migrants live in the same location as their parents and can coreside with them, often because the entire family migrated when the individual was under 25 years of age.

and become homeowners earlier than migrants. Back-of-the-envelope calculations using expenditure data suggest that natives who are younger than 35 save, on average, between 2,300 and 3,000 euros more in annual housing costs than observationally equivalent migrants (see Appendix Figure A6). As explained in Appendix A, this is also in line with the evidence on the effect of coresidence on children's personal savings described in Rosenzweig and Zhang (2019). Lower income individuals are more likely to coreside with parents, as can be seen in column (3) of Appendix Table A2. Even after accounting for individual fixed-effects, together with age and other observables, there is a negative relationship between income and coresidence (see column (4) in Appendix Table A2). This suggests that the option to live with parents can insure against negative lifetime income shocks.

Survey data in the EFF also reveal that observationally equivalent migrants are less likely to be homeowners than natives due to financial frictions in the mortgage market. Indeed, even after controlling for observables, a higher share of them report being currently renting because they cannot afford to pay the down-payment or are not eligible for a mortgage, although they would like to get one. As can be seen in Figure 2d, this is especially true among young people. The fact that potential migrant homebuyers are more financially-constrained than natives is puzzling if one considers that, as previously noted, there exists a migrant wage premium in the data. To account for this fact, I investigate the role of coresidence in explaining the native-migrant homeownership gap in a spatial model where the homeownership decision is subject to a down-payment constraints and only natives have the option to coreside.[16] The model also features the other two channels proposed as candidate explanations, i.e. the self-selection behavior of stayers into homeownership and the availability of housing bequests.

International Comparison The existence of a homeownership gap between observationally equivalent natives and internal migrants is a new fact documented in this paper. This fact is not unique to Spain: it is also observed in France, Italy, and the United States, as shown in panels (a), (c), and (e) of Appendix Figure A7. The choice of these countries is driven by the availability of Census data on coresidence, homeownership, and both current and birthplace locations within OECD countries, together with data on the observables used as controls. More details on the definition of locations in each country can be found in Appendix A.

As can be seen in panels (b), (d), and (f) of Appendix Figure A7, natives in each of these countries also exhibit a higher propensity to coreside with parents than migrants do. This suggests that, like in Spain, coresidence can be an important driver behind the observed homeownership gap between natives and migrants. Coresidence in France and the United States is less prevalent than in Italy or Spain. Nevertheless, it remains a significant phenomenon among young people, interesting around 15% of individuals aged between 25 and 35 in the U.S. and France (the share is around 40% in Italy and Spain).

[16]Some migrants do coreside in the data, as explained in footnote 15. These individuals effectively behave as natives in the model, under the assumption that they treat the location where they lived at age 25 as their birthplace. Refer to Section 4.1.3 and Appendix E.4.4 for further discussion.

2.4 Summary of Facts

Two main facts have been presented. First, people that become homeowners are less likely to migrate in the future, and migration, in turn, positively affects labor earnings. Second, natives are more likely to be homeowners than internal migrants, despite earning lower incomes on average. Three potential explanations emerge for the second fact. First, migrants might relocate again, making them more hesitant to purchase homes. Second, natives have higher chances to live in inherited houses. Finally, natives often live with parents, thereby saving resources which can be directed towards the mortgage down-payment and homeownership. These mechanisms are featured and analyzed in the model.

3 A Spatial Model With Coresidence and Homeownership

This section describes the theoretical framework, a spatial lifecycle model with coresidence, homeownership, migration, and saving decisions. Every period, agents have a set of locations to choose from as their residence, each offering distinct housing prices and wages, which are determined in equilibrium, and fixed amenities. Additionally, they can either buy a house, rent one, or, if they are native, coreside with their parents. Financial frictions in the credit market imply that individuals who need a mortgage to become homeowners must first accumulate enough savings for an initial down-payment. Moreover, agents who buy or sell a house incur some transaction costs.

Individual decisions are interconnected. Homeownership makes it endogenously more costly to migrate, as agents need to sell their house before moving and bear the associated transaction cost, and migrating away from birth location does not allow to coreside with parents. The decisions are also dynamic, and agents form expectations about the future taking into account uninsurable local unemployment risk and labor income risk, as well as *endogenous* housing prices and wages which fluctuate in response to multiple aggregate shocks to the model's primitives.

3.1 Environment

The economy is populated by a unit-mass of finitely-lived agents with age $j \in \{1, ..., J\}$ that can live in one of D location, indexed by $d \in \{1, ..., D\}$. Agents also choose their housing status h and a consumption good c, which is freely tradable across locations and acts as the numeraire. Labor income and employment status follow individual-levels stochastic processes. There exist aggregate shocks to labor productivity and housing construction primitives that affect each location. Housing and labor markets are competitive and clear in each location and period. Local housing prices and wages reflect supply and demand, as well as aggregate shocks to primitives. The interest rate r on liquid assets, instead, is fixed and exogenously given. Finally, the government taxes labor income, pays means-tested transfers, and implements housing policies, all while balancing the budget in each period.

3.2 Individual Problem

Individual States Agents are born with three fixed types: birthplace location $d_0 \in \{1, ..., D\}$, education level $e \in \{N, E\}$ (i.e., non-college or college), and migration type $\tau \in \{1, 2\}$ (i.e., non-stayer or stayer). Age j evolves deterministically over the lifecycle. Two individual states, instead, evolve following processes that vary by age, education, and location. These are employment status $l \in \{1, 2\}$ (i.e., unemployed or employed) and individual-level productivity ϖ, which has fixed, persistent, and deterministic components θ_e, z_{ej}, and Υ_{edj}, respectively. Let Ψ_{edj} denote the distribution of l' conditional on l, i.e., $l' \sim \Psi_{edj}(l)$. Likewise, let Φ_{edj} denote the distribution of ϖ' conditional on ϖ', i.e. $\varpi' \sim \Phi_{edj}(\varpi)$. Finally, choices in the previous period, i.e., assets a, location d, and housing status h (which are drawn stochastically in the first period of agents), are also part of the individual state vector, denoted by $\mathbf{x} = (j, d_0, e, \tau, \varpi, l, a, d, h)$.

Choices In each period, agents are faced with three distinct decisions. First, they choose their preferred location d', which may or may not align with their birthplace d_0. This choice determines whether the agent is classified as native ($d' = d_0$) or as migrant ($d' \neq d_0$). Then, they choose their housing tenure h', which, in case they are natives, can either be coresident ($h' = 0$), renter ($h' = 1$), or homeowner ($h' = 2$). In case they are migrants, however, their choices are limited to renting or buying, i.e. $h' \in \{1, 2\}$. Finally, they make consumption decisions c, which determines next period's assets a'.

Aggregate States Various exogenous aggregate shocks hit every location each period. These local shocks include overall and skill-specific labor productivity, and housing construction costs. The definition of these aggregate shocks, which I denote by $\boldsymbol{\mathcal{Z}}$, is postponed to Section 3.3. Shocks evolve exogenously following the aggregate process $\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}})$. Let μ denote the distribution of agents across individual states $\mathbf{x}$. I collect the aggregate states into $\Omega = (\boldsymbol{\mathcal{Z}}; \mu)$. The distribution μ evolves according to the equilibrium law of motion $\mu' \sim \Gamma_{\mu}(\Omega)$, which reflects agents' optimal decisions together, and in response to, exogenous processes. Prices and wages in each location vary in response to aggregate shocks to primitives $\boldsymbol{\mathcal{Z}}$, and also reflect demand and supply, which depend on Ω. Therefore, I can define the vector of local housing prices and wages by education as $\mathbf{p}(\Omega) = \{p_1(\Omega), ..., p_D(\Omega)\}$ and $\mathbf{w_e}(\Omega) = \{w_{e1}(\Omega), ..., w_{eD}(\Omega)\}$. Future prices and wages in each location, $\{\mathbf{p}(\Omega'), \mathbf{w_e}(\Omega')\}$, can be inferred from $\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}})$ and $\mu' \sim \Gamma_{\mu}(\Omega)$.

Frictions There are financial frictions in the mortgage market. Agents that require a mortgage to buy a house must pay the initial down-payment, a fraction χ of the housing value, up front. They can finance the rest with a mortgage with a linear repayment schedule and fixed interest rate r^h. Moreover, individuals who buy or sell a house incur some transaction costs, which are fixed fractions ϕ_b and ϕ_s of the housing value. Additionally, there are location-specific monitoring costs in the local rental markets. Spatial frictions also exists: agents who want to change location face migration costs.

Utility Function Within period utility in location d' at age j is given by

$$u_j(c,h',d',\boldsymbol{x_j}) = \frac{c^{1-\gamma}}{1-\gamma} + \overbrace{\eta_1 \mathbb{1}\{h'=2\}}^{\text{Utility from being homeowner}} - \overbrace{\xi_j \mathbb{1}\{h'=0\}}^{\text{Disutility from being coresident}}$$
$$- \underbrace{\delta_{e\tau j}\mathbb{1}\{d' \neq d_{j-1}\}}_{\text{Migration costs}} + \underbrace{\eta_2 \mathbb{1}\{d'=d_0\}}_{\text{Home-bias}} + \underbrace{A_{d'}}_{\text{Amenities}}$$

where γ is the degree of relative risk aversion. Agents value city amenities and derive additional utility from living in their birthplace. This allows for an additional home-bias channel (Zerecero 2021), that complements the ones introduced in this paper, namely coresidence and easier access to bequests for natives. Moreover, agents experience direct utility from homeownership, which captures the psychological sense of stability that comes with owning and maintaining their own house. This assumption is widely made in the literature to explain the observed high rate of homeownership.[17]

The disutility from coresiding with parents, a living arrangement that is only available for natives, captures lack of independence (Kaplan 2012) and is given by

$$\xi_j = \xi_0 + \xi_1 j.$$

Migration costs, taking place when the chosen location is different from the one in the previous period, are given by

$$\delta_{e\tau j} = \delta_0 + \delta_e \mathbb{1}\{e=E\} + \delta_\tau \mathbb{1}\{\tau=2\} + \delta_1 j + \delta_2 \log(j).$$

The two utility costs are allowed to vary by age, to capture the fact that both migration and coresidence decrease steeply over the lifecycle (see Appendix Figure A2). Migration costs can also vary by education e, since college-educated people are more likely to migrate in the data (see column (2) in Appendix Table A1). Finally, migration costs are allowed to vary by τ (stayer type), since a fraction of the population may have prohibitively high migration costs in all states (Kennan and Walker 2011).

I first state the problem of agents entering the period as non-homeowners (coresidents and renters). Next, I state the problem of homeowners. Then, I show how the probability of receiving housing bequests changes the problem. Finally, I describe the problem of agents in their last period of life. Timing in the model is such that labor income is gained from d, the agent's location when entering the period, whereas housing costs are paid in the chosen location d'. These two locations coincide for people who decide not to migrate.

Dynamic Problem: Coresidents and Renters Let $V_j^n(d';\mathbf{x},\Omega)$ denote the value function of agents who enters the period as non-homeowners, conditional on choosing location d'. These agents choose between coresiding, renting, and buying a house by solving

$$V_j^n(d';\mathbf{x},\Omega) = \max\{V_j^{c,n}(d';\mathbf{x},\Omega), V_j^{r,n}(d';\mathbf{x},\Omega), V_j^{o,n}(d';\mathbf{x},\Omega)\}.$$

[17] See for example Kaas et al. (2021), Oswald (2019), Paz-Pardo (2023). Housing is the only non-tradable good in the model. This assumption is supported by empirical evidence highlighting that housing is the key driver behind local price differences across cities, both in Spain (Forte-Campos, Moral-Benito and Quintana 2021) and in the U.S. (Moretti 2013).

Agents who choose to rent in d' solve

$$V_j^{r,n}(d';\mathbf{x},\Omega) = \max_{c>0} \left\{ u_j(c, h' = 1, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{\mathcal{Z}}} \left[\overline{V}_{j+1}^{n}(\mathbf{x}', \Omega') \right] \right\}$$

subject to

$$a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c - \underbrace{\kappa_{d'} p_{d'}(\Omega) \overline{h}_{ed'1}}_{\text{Rent}}) \geq 0,$$
$$\varpi' \sim \Phi_{edj}(\varpi), \;\; l' \sim \Psi_{edj}(l),$$
$$\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \;\; \mu' \sim \Gamma_{\mu}(\Omega).$$

where r is the exogenous interest rate for liquid assets, $y(w_{ed}(\Omega), \varpi, l)$ is labor income (after tax and transfers), which depends on local wages, as well as individual productivity and employment status, $p_{d'}(\Omega)$ is housing price per square meter, $\kappa_{d'}$ is the rent to price ratio in location d', and $\overline{h}_{ed'1}$ is the fixed housing quantity (in square meters) demanded by renters with education e living in d'. Agents who start next period as non-homeowners have value function $\overline{V}_{j+1}^{n}(\mathbf{x}', \Omega')$, which takes into account expected migration choices and is described below.

The value function $V_j^{c,n}(d';\mathbf{x},\Omega)$ for non-homeowners who decide to coreside is equal to the value function for new renters $V_j^{r,n}(d';\mathbf{x},\Omega)$, with three differences. First, the chosen location must be the birthplace ($d' = d_0$), as only natives can coreside. Second, coresidents do not pay housing costs.[18] Therefore, their budget constraint is given by

$$a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c) \geq 0.$$

Finally, they suffer a disutility costs from coriseding, which is captured in $u_j(c, h' = 0, d', \mathbf{x})$. Notice that agents who decide to rent and coreside cannot borrow. The full expression for $V_j^{c,n}(d';\mathbf{x},\Omega)$, omitted here for brevity, can be found in Appendix C.1.

Finally, non-homeowners who choose to buy in d' have value function $V_j^{o,n}(d';\mathbf{x},\Omega)$, reported in Appendix C.1. New homeowners have period utility $u_j(c, h' = 2, d', \mathbf{x})$ and budget constraint:

$$a' = (1 + r\mathbb{1}_{a\geq 0} + r^h \mathbb{1}_{a<0})(a + y(w_{ed}(\Omega), \varpi, l) - c - (1 + \underbrace{\phi_b}_{\substack{\text{Transaction cost}\\ \text{of buying}}}) p_{d'}(\Omega) \overline{h}_{ed'2}),$$
$$a' \geq -(1-\chi) p_{d'}(\Omega) \overline{h}_{ed'2} \;\; \text{if } j<J,$$
$$a' \geq 0 \;\; \text{if } j=J.$$

To buy a house, agents need to pay the sunk transaction cost of buying, a fraction ϕ_b of the housing value $p_{d'}(\Omega)\overline{h}_{ed'2}$, and at least the down-payment requirement, a fraction χ of $p_{d'}(\Omega)\overline{h}_{ed'2}$, up front.[19] They can finance the rest, up to $(1-\chi)p_{d'}(\Omega)\overline{h}_{ed'2}$, with mortgage debt

[18] The evidence discussed in Appendix A as a comment to Figure A6 is consistent with this assumption (Rosenzweig and Zhang 2019).

[19] Fixed housing quantity $\overline{h}_{ed'2}$ for homeowners is allowed to be different from renters', to capture the fact that the size of homeowners' dwellings is typically larger. It also changes by location and education, as college-educated individuals and rural residents typically live in larger houses.

with fixed interest rate r^h. Agents cannot exit the model with debt and cannot default on their mortgage.[20] The mortgage has a linear repayment schedule, as detailed below. Since home-buyers start next period as homeowners, the expected value function now contains term $\overline{V}^h_{j+1}(\mathbf{x}', \Omega')$, which is defined below.

Coming up with the money for the down-payment and the transaction costs of buying is easier if agents have coresided before and saved on housing rent. However, coresiding implies utility costs ξ_j and, possibly, the opportunity cost of living in a birthplace with low wages $w_{ed_0}(\Omega)$, high unemployment risk $\Psi_{ed_0 j}$ or low individual productivity growth $\Phi_{ed_0 j}$.

Given the value of the optimal housing tenure choice for non-homeowners, $V^n_j(d'; \mathbf{x}, \Omega)$, the migration decision of agents entering the period as non-homeowners is determined by

$$V^n_j(\mathbf{x}, \Omega, \varepsilon) = \max_{d' \in D} \left\{ V^n_j(d'; \mathbf{x}, \Omega) + \varepsilon_{d'} \right\}, \quad \varepsilon \overset{\text{iid}}{\sim} \text{Standard Gumbel}.$$

From the Standard Gumbel assumption on the vector of preference shocks $\varepsilon = (\varepsilon_1, ..., \varepsilon_D)$ for location d' at age j, there exists a simple expression for $\overline{V}^n_{j+1}(\mathbf{x}', \Omega')$ (McFadden 1973, Rust 1987), i.e.,

$$\begin{aligned}\overline{V}^n_{j+1}(\mathbf{x}', \Omega') &= \mathbb{E}_\epsilon[V^n_{j+1}(\mathbf{x}', \Omega', \varepsilon)] \\ &= \overline{\gamma} + \log\left(\sum_{d_k=1}^{D} \exp(V^n_{j+1}(d_k; \mathbf{x}', \Omega'))\right),\end{aligned}$$

where $\overline{\gamma}$ is the Euler-Mascheroni constant.

Dynamic Problem: Homeowners Let $V^h_j(d'; \mathbf{x}, \Omega)$ denote the value function of agents who enter the period as homeowners, conditional on choosing location d'. These agents choose between selling the house, by either coresiding or renting, and staying homeowners, either in the current or in a different location. Homeowners who want to migrate, always need to sell their house first. Full details on the value functions for each housing tenure option of homeowners can be found in Appedix C.1.

The budget constraint of current homeowners who choose to coreside in $d' = d_0$ is given by:

$$a' = (1 + r)(a + y(w_{ed}(\Omega), \varpi, l) - c + (1 - \underbrace{\phi_s}_{\substack{\text{Transaction cost}\\ \text{of selling}}})p_d(\Omega)\overline{h}_{ed2}) \geq 0.$$

When agents sell their house, they bear the associated transaction costs $\phi_s p_d(\Omega)\overline{h}_{ed2}$ and re-

[20]Differently from other countries, defaulting on mortgage debt is considered illegal in Spain. Individuals who are unable to repay their (potentially restructured) mortgage debt can, as a last resort, cancel their debt by transferring their house to the creditor bank through a process known as *dación en pago* (foreclosure). However, it is worth noting that this circumstance is extremely rare in practice, affecting only 1,000 to 2,000 properties annually between 2015 and 2019 (Source: *Registradores de España*). In the estimated model, the yearly evolution of housing prices is such that no agent finds themselves underwater with their mortgage debt, so that, even if they were allowed to do so, there would be no incentives for strategic defaults. In the event that individuals are unable to repay the debt, they always have the option to sell their house and retain positive assets (see also footnote 21).

ceive the remaining fraction of the housing value net of the mortgage debt in cash.[21]

Those who choose to rent, instead, face budget constraint:

$$a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c + (1-\phi_s)p_d(\Omega)\overline{h}_{ed2} - \kappa_d p_{d'}(\Omega)\overline{h}_{ed'1}) \geq 0.$$

Homeowners who choose to keep their house in $d' = d$ (non-migrants), instead, have budget constraint:

$$\begin{aligned}
&a' = (1 + r\mathbb{1}_{a\geq 0} + r^h \mathbb{1}_{a<0})(a + y(w_{ed}(\Omega), \varpi, l) - c),\\
&a' \geq a\left(\frac{(J-1)-j}{J-j}\right)\mathbb{1}_{a<0} \text{ if } j<J,\\
&a' \geq 0 \text{ if } j=J.
\end{aligned}$$

These agents need to repay their mortgage debt following a linear repayment schedule with prepayment option, which means that each year they have the option to pay more than what they are due to extinguish their debt faster. In case they exercise the prepayment option at age $j' > j$, the mortgage debt maturity automatically changes to $J-j'$, and the linear repayment schedule adjusts with it. This formulation of the budget constraint allows to generate realistic repayment schedules without adding further state variables. Notice that homeowners who choose to keep their house are not exposed to price risk uncertainty. In other words, housing price $p_{d'}(\Omega)$ does not enter their budget constraint.

Finally, homeowners who choose to migrate and to buy a house in d' face the constraint:

$$\begin{aligned}
&a' = (1 + r\mathbb{1}_{a\geq 0} + r^h \mathbb{1}_{a<0})(a + y(w_{ed}(\Omega), \varpi, l) - c + (1-\phi_s)p_d(\Omega)\overline{h}_{ed2} - (1+\phi_b)p_{d'}(\Omega)\overline{h}_{ed'2}),\\
&a' \geq -(1-\chi)p_{d'}(\Omega)\overline{h}_{ed'2} \text{ if } j<J,\\
&a' \geq 0 \text{ if } j=J.
\end{aligned}$$

This budget constraint reveals a key trade-off in the model. Unlike coresidents and renters, homeowners who want to migrate need to sell their house, and lose a fraction ϕ_s of housing wealth. Since non-homeowners do not suffer these costs, they are more likely to migrate. Therefore, the benefits of homeownership (direct exogenous utility and insurance against rental price volatility) must be weighted against the gains from migration, i.e. possible increases in labor income through $w_{ed}(\Omega)$, Φ_{edj}, and Ψ_{edj}.

Given $V_j^h(d'; \mathbf{x}, \Omega)$, the migration decision of agents entering the period as homeowners is determined by

$$V_j^h(\mathbf{x}, \Omega, \varepsilon) = \max_{d' \in D}\left\{V_j^h(d'; \mathbf{x}, \Omega) + \varepsilon_{d'}\right\}, \quad \varepsilon \overset{\text{iid}}{\sim} \text{Standard Gumbel}.$$

[21] The estimated stochastic process for the aggregate shocks and the model parameters are such that, in equilibrium, no agent has $a + y(w_{ed}(\Omega), \varpi, l) + (1-\phi_s)p_d(\Omega)\overline{h}_{ed2} - \kappa_{d'} p_{d'}(\Omega)\overline{h}_{ed'1} \leq 0$.

From the Standard Gumbel assumption, we have that

$$\overline{V}^h_{j+1}(\mathbf{x}', \Omega') = \overline{\gamma} + \log\left(\sum_{d_k=1}^{D} \exp(V^h_{j+1}(d_k; \mathbf{x}', \Omega'))\right).$$

Probability of Receiving a Housing Bequest At the end of each period, agents may receive a housing bequest with probability $\pi^b_{ed_0j}$ after having made their consumption, housing, and migration choices. This probability can vary by education group, birthplace, and age. In case they choose not to be homeowners, and if they decide to live in the birthplace (natives), then, if they receive a housing bequest, they become homeowners. The idea is that the property is inherited in the birthplace, and only natives who are not already homeowners can go live there. In all other cases (homeowners or migrants), when agents receive a bequest, they are forced to sell the inherited house, pay the transaction costs ϕ_s, and split the proceeds among κ_b siblings. Parameter κ_b is the same across agents. Hence, the presence of housing bequests may increase the incentives to delay purchasing a house and to remain in, or move to, the birthplace. Notice that the inherited house value depends on prices in the birthplace. Appendix C.1 shows how value functions change in the presence of housing bequests ($\pi^b_{ed_0j} > 0$).

Terminal Value Finally, agents receive a one-time utility from their homeownership status and accumulated liquid and housing wealth when they exit the model after age J.[22] In particular, at age $j = J$, agents' expected value function does not include $\overline{V}^n_{j+1}$ or $\overline{V}^h_{j+1}$. Instead, it includes the terminal value function:

$$V_{J+1}(\mathbf{x}', \Omega') = \omega_1 \frac{(a' + p_{d'}(\Omega)\overline{h}_{ed'2}\mathbb{1}\{h' = 2\})^{1-\gamma}}{1-\gamma} + \omega_2 \mathbb{1}\{h' = 2\},$$

where ω_1 denotes terminal utility from wealth and ω_2 denotes terminal utility from homeownership.

3.3 Markets and Aggregate Shocks

The labor and housing markets are competitive and clear in equilibrium in each location d and period. There is a representative firm in each location that employs college and non-college workers to produce the consumption good. Additionally, a representative construction firm operates in each location and produces housing units, together with a representative real-estate firm that rents out housing units.

The production function of the firm operating in location d is given by

$$Y_d = X_d(\zeta_d L^\rho_{Nd} + (1 - \zeta_d) L^\rho_{Ed})^{1/\rho},$$

[22] The inclusion of this terminal value condition at age J, which corresponds to 64 years old in the data, accounts for the importance of wealth and homeownership during retirement and for bequest motives. This is needed to match the observed (slow) dissaving behavior of older agents (Oswald 2019).

where X_d is the local overall productivity, ζ_d is the local skill-specific productivity, ρ is the CES production function parameter, and L_{ed} is total efficiency units input of education group e in location d. The first-order conditions determine wages w_{ed} of college and non-college workers:

$$w_{ed} = X_d(\zeta_d \mathbb{1}_N + (1-\zeta_d)\mathbb{1}_E)(\zeta_d L_{Nd}^{\rho} + (1-\zeta_d)L_{Ed}^{\rho})^{\frac{1-\rho}{\rho}} L_{ed}^{\rho-1}, \tag{1}$$

where $\mathbb{1}_E$ and $\mathbb{1}_N$ are indicator functions for the college and non-college first-order conditions, respectively.

Additionally, the inverse housing supply function is given by

$$p_d = k_d H_d^{\psi}, \tag{2}$$

where H_d denotes the housing supply stock. This function is obtained from the first-order condition of a representative construction firm that operates with a convex cost technology, as detailed in Appendix C.3. The convexity of the cost function is a reduced form way to capture the scarcity of buildable land and possible inputs and regulation constraints.

Key parameters are the intercept k_d, which captures construction costs that vary across locations, and the inverse housing supply elasticity ψ, which reflects the responsiveness of housing prices to changes in the housing stock. A higher ψ means that prices react more to changes in the housing supply, for instance due to the construction restrictions coming from land unavailability or regulations. Finally, the local price-to-rent ratio, κ_d, is micro-founded in Appendix C.3 as the zero-profit conditions of real-estate firms operating in each d with location-specific monitoring costs.

Aggregate Shocks Three exogenous primitives of the CES technology and inverse housing supply function are subject to aggregate shocks in each location d. These include housing construction costs $\boldsymbol{k} = (k_1, ..., k_D)$, overall productivity $\boldsymbol{X} = (X_1, ..., X_D)$, and skill-specific productivity $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_D)$. Therefore, there exist $3 \times D$ aggregate shocks in total. As already mentioned, aggregate shocks are grouped into $\boldsymbol{\mathcal{Z}} = (\boldsymbol{k}, \boldsymbol{X}, \boldsymbol{\zeta})$ and follow the exogenous Markovian process $\boldsymbol{\mathcal{Z}'} \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}})$.

3.4 Government

The government generates revenue by taxing labor income and provides means-tested transfers, unemployment benefits, and other public goods $\overline{G}$ in each period. Taxes and transfers are functions $T(\cdot)$ and $G^g(\cdot)$ of gross labor income $\tilde{y}(\Omega)$. I assume that $\overline{G}$ does not affect individuals' utility. In counterfactual exercises, the government additionally implements housing policies with associated expenditure $G^p(\Omega)$. Since the Government balances the budget, it follows that expenditure on the public goods $\overline{G}$ also depends on the aggregate states Ω.

4 Estimation

This section presents the estimation procedure. Some of the model parameters are estimated with external data sources, whereas the rest are internally calibrated to ensure that a selection of simulated moments align with their data counterparts. These targeted moments include homeownership, coresidence, and migration rates over the lifecycle.

Moreover, I present a strategy to solve the dynamic spatial equilibrium model in the presence of aggregate shocks. In particular, I use a low-dimensional factor model structure to model agents' forecast rules about future local endogenous prices and wages. The forecast rule accurately predicts both benchmark and counterfactual equilibrium prices.

4.1 Model Inputs

This section describes the inputs which are estimated externally, using the sample restrictions described in Section 2.1. People live for 40 years ($J = 40$), from 25 to 64 years old, in one of 12 locations ($D = 12$). The number of locations is kept relatively low for computational reasons.[23]

The model's exogenous inputs include the parameters affecting labor income (unemployment and income risk, taxes and transfers) and the probability of receiving housing bequests. Some additional inputs, such as the exogenous interest rates and the down-payment requirement, are also directly measured from the data. A few remaining parameters, including the inverse housing supply elasticity and transaction costs, are taken from the literature.

4.1.1 Probability of Being Unemployed, Income Process, Taxes and Transfers

As is standard in the literature, a linear regression framework is used to estimate the transition probabilities between employment statuses, while the income process parameters are estimated with a GMM procedure. Moreover, the tax and transfer functions that I assume are also common choices in the literature.

Employment Status In each period, agents can either be unemployed ($l = 1$) or employed ($l = 2$). The Markovian process for the employment status, $l' \sim \Psi_{edj}(l)$, varies by education, location and age, and is estimated by running the linear probability regression models:

$$\overbrace{\mathbb{1}_{l_{edj+1}=1}}^{\text{Unemployed next period}} = \alpha_t + \alpha_e + \alpha_d + \alpha_1 j + \alpha_2 \log(j)$$
$$\underbrace{\mathbb{1}_{l_{edj+1}=2}}_{\text{Employed next period}} = \beta_t + \beta_e + \beta_d + \beta_1 j + \beta_2 \log(j).$$

The regressions are separately estimated for employed and unemployed individuals by exploiting the yearly panel dimension of the EU-SILC individual-level dataset, which includes

[23] Doubling the number of locations would quadruple the state space: the model would require solving at 1.3 billion points, up from the current 325 million, thus substantially increasing the computational burden.

both coresidents and non-coresidents, and sum to one by construction. Education and location fixed-effects, $\alpha_e, \beta_e, \alpha_d$, and β_d, together with age coefficients $\alpha_1, \beta_1, \alpha_2$, and β_2, are used to estimate the Markovian transition probabilities, whereas year fixed-effects α_t and β_t are included to control for aggregate shocks.

The lifecycle transition probabilities from employment to unemployment status across education groups and locations are plotted in Figure 3a. The east urban location is emphasized for reference, while the semi-transparent lines represent other Spanish locations. The probability of being unemployed varies largely across education groups and locations. The unemployment risk is lower for the college-educated and in urban locations within the same NUTS-1 region. There is also an important lifecycle dimension, as unemployment risk tends to be higher at younger ages and decreases over time.

The opposite transition, from unemployed to employed, is plotted in Appendix Figure E11. Re-employment probabilities decrease sharply at older ages, so that the unemployment state becomes more persistent over the lifecycle. Similarly to the probability of becoming unemployed, there is substantial variation across locations and educational groups.

(a) Employed $\rightarrow$ Unemployed
(East urban location emphasized)

(b) Gross income age profiles Υ_{edj}
(East urban location emphasized)

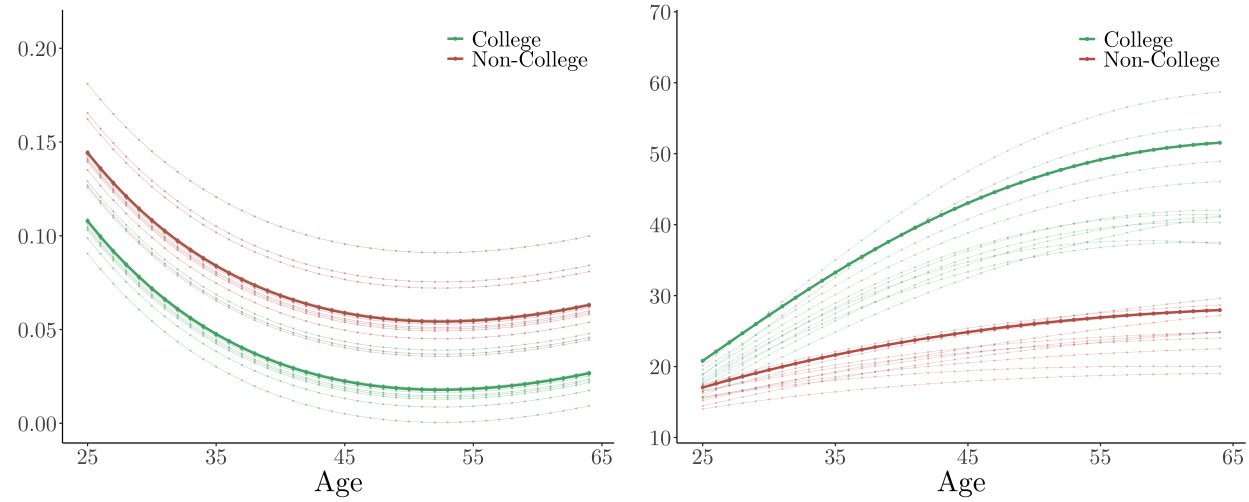


Figure 3: The figures plot age polynomial functions for different locations and education groups (East urban location is emphasized). Data: EU-SILC 2004-2019 (panel **a**), MCVL 2005-2019 (panel **b**).

Labor Income While unemployed people receive constant benefits b, employed workers are paid for each efficiency unit of labor that they supply exogenously,

$$\varpi_{edj} = \exp(\theta_e + z_{ej} + \Upsilon_{edj}).$$

The labor endowment is a function of fixed productivity θ_e, deterministic age-profile Υ_{edj}, and persistent shock z_{ej}. I assume that θ_e is drawn from an education-specific normal distribution and that z_{ej} follows a first-order autoregressive process with persistence parameter ϱ_e. The gross labor income paid to employed workers depends on wage per skill $w_{ed}(\Omega)$ and the labor endowment ϖ_{edj},

$$\tilde{y}_{edj}(\Omega) = [w_{ed}(\Omega)]^{\vartheta_{ed}} \varpi_{edj}.$$

Endogenous wages $w_{ed}(\Omega)$ vary in response to aggregate shocks to primitives, and affect income through the loading coefficient ϑ_{ed}.[24] Therefore, gross log labor income is given by:

$$\ln \tilde{y}_{edj} = \vartheta_{ed} \ln w_{ed}(\Omega) + \theta_e + \Upsilon_{edj} + z_{ej}, \tag{3}$$

$$z_{ej} = \varrho_e z_{ej-1} + \upsilon_{ej-1}, \tag{4}$$

where

$$\theta_e \sim N(0, \sigma^2_{\theta e}), \quad \upsilon_e \sim N(0, \sigma^2_{\upsilon e}), \quad \upsilon_{e0} = 0, \quad z_{e0} = 0.$$

Income Process In order to estimate the income process of equations (3) and (4), I first residualize gross annual labor income from the observed quantities in equation (3). To do so, I assume that location- and education-specific age profiles are given by

$$\Upsilon_{edj} = \theta_{1ed} j + \theta_{2ed} \log(j)$$

and, using the administrative MCVL data, estimate regression

$$\ln \tilde{y}_{edjt} = \vartheta_{ed} \ln w_{edt} + \theta_{1ed} j + \theta_{2ed} \log(j) + \beta' X_t + u_{ej} \tag{5}$$

where t denotes year. The additional controls X_t, not present in equation (3), include year and gender fixed-effects. Local wages w_{edt} are measured as the average gross hourly wages by education, location, and year. These observed time series, consistently, are perfectly matched in the benchmark as equilibrium wages (see Section 4.2.3).

The estimated gross income age profiles Υ_{edj} are plotted in Figure 3b. There are large variations in lifecycle income profiles across location and education groups, with urban locations within NUTS-1 regions and college workers experiencing a substantial premium over rural locations and the non-college educated.

Given regression (5) estimates, residualized labor income is given by

$$\hat{u}_{ej} = \ln \tilde{y}_{edjt} - (\hat{\vartheta}_{ed} \ln w_{edt} + \hat{\theta}_{1ed} j + \hat{\theta}_{2ed} \log(j) + \hat{\beta}' X_t). \tag{6}$$

The income process parameters include the persistence parameters ϱ_N and ϱ_E, the standard deviations of the persistent shocks' innovations σ_{υ_N} and σ_{υ_E}, and the standard deviations of the fixed-effects σ_{θ_N} and σ_{θ_E}. Following Storesletten, Telmer and Yaron (2004), these parameters are estimated by GMM using as population moments the cross-sectional variance of the residualized log income u_{ej} by age. As can be seen in Appendix Figures E12a and E12b, the increasing patterns of earnings inequality over the lifecycle for both college and non-college workers are well captured by the model. More details on the estimation procedure are given in Appendix E.2.

[24] Although local labor markets are perfectly competitive, ϑ_{ed} can be seen as a reduced form way to capture imperfect wage pass-through of labor productivity, which may vary by education and location. For example, there is evidence that labor market power varies across cities and workers' education group (see Hirsch et al. 2022 for Germany and Luccioletti 2022 for Spain).

Taxes Given mean gross income $\overline{y}$, the average tax rate at gross income level $\tilde{y}_{edj}(\Omega)$ is given by:

$$T_t(\tilde{y}_{edj}(\Omega)) = \max\left\{0, 1 - \varsigma_0 \left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}}\right)^{-\varsigma_1}\right\}.$$

This formulation of the income tax function is standard in the literature (Bénabou 2002, Guner, Lopez-Daneri and Ventura 2016). The parameter ς_0 determines the average tax level, while ς_1 determines the progressivity.

Transfers Following Guner, Kaya and Sánchez-Marcos (2023), means-tested transfers are assumed to change linearly with multiples of mean labor income according to the function:

$$G^g(\tilde{y}_{edj}(\Omega)) = \max\left\{0, g_1 - g_2\left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}}\right)\right\}.$$

Parameters g_1 and g_2 are estimated with a linear regression using EU-SILC data. The transfers decline as workers' earnings increase and become zero at around 2.5 times the mean labor income.

4.1.2 Probability of Receiving Housing Bequests

In order to estimate the probability of receiving a housing bequest next period $\pi^b_{ed_0j}$, I turn to the EFF data. This survey asks if respondents have ever inherited one or multiple properties in their lifetime. From this, as detailed in Appendix E.3, we can easily calculate yearly probabilities that align with the cumulative age distribution computed from the data.

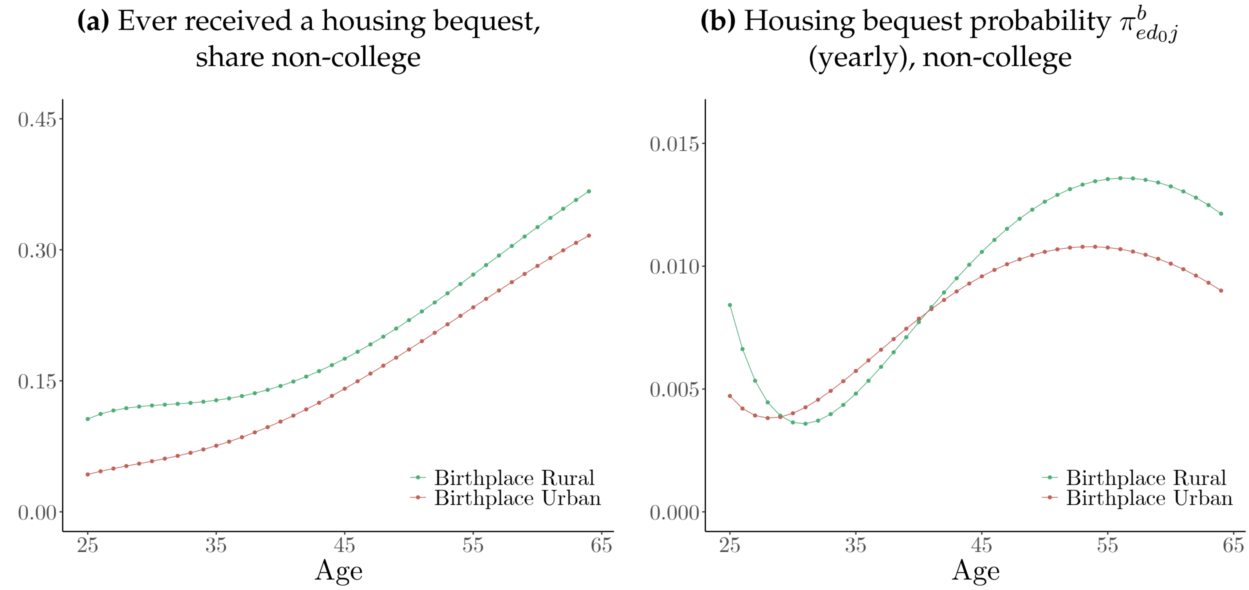


Figure 4: The figures plot the share of people without a college degree who have ever inherited a house during their lifetime (panel a), computed age by age, and the annual probability of receiving a housing bequest (panel b). The equivalent figures for college educated workers are plotted in Appendix Figure E14. Data: EFF 2008-2020.

Our objective is to obtain an individual-level probability of receiving a housing bequest in the next period, both for those who are and aren't currently coresiding. However, only household-level survey information from individuals not residing with their parents is available. Nonetheless, given that individuals currently living with their parents cannot have received a housing bequest from them in the preceding period, the bequest probability can be easily adjusted.[25] Further details are found in Appendix E.3.

Figure 4b plots $\pi^b_{ed_0j}$ for those without college education, while Appendix Figure E14b shows the same probability for college graduates. Panel (a) of Figure 4 depicts the EFF cumulative distribution data, and panel (b) plots the estimated yearly probability. Notice that the yearly likelihood of inheriting a property tends to either increase throughout the lifecycle (for those without college education) or to peak roughly at age 45 (for college graduates). Additionally, this probability tends to be higher for the college-educated and for people born in rural locations.

4.1.3 Exogenous Parameters

The full set of parameters that are set exogenously is listed in Table 1. These include some parameters which are taken from the literature, namely the coefficient of relative risk aversion γ, the CES production function parameter ρ (Acemoglu and Autor 2011), the inverse housing supply elasticity in Spain ψ (Caldera and Johansson 2013), the transaction costs of buying and selling a house in Spain, ϕ_b and ϕ_s (Kaas et al. 2021), and the ς_0, ς_1 parameters for the Spanish tax function (García-Miralles, Guner and Ramos 2019).[26]

Table 1: Parameters set exogenously.

	Parameter	Value	Source
Relative risk aversion	γ	1.5	Standard
CES production function parameter	ρ	0.406	Acemoglu and Autor (2011), U.S.
Inverse housing supply elasticity	ψ	2.21	Caldera and Johansson (2013)
Transaction costs	(ϕ_b, ϕ_s)	(0.067, 0.067)	Kaas et al. (2021)
Down-payment requirement	χ	0.192	Average 2002-2017
Real interest and mortgage rates	(r, r^h)	(0.014, 0.033)	Average 2002-2017
Housing sizes (square meters)	$(\overline{h}_{de1}, \overline{h}_{de2})$	See text	EPF 2016-2019
Number siblings receiving bequests	κ^b	2.5	Census 1991
Prob. receiving housing bequests	$\pi^b_{ed_0j}$	See text	EFF 2008-2020
Prob. changing employment status	$\Psi_{edj}(l)$	See text	EU-SILC 2004-2019
Persistence income process	(ϱ_N, ϱ_E)	(1.00, 1.00)	MCVL 2005-2019
Standard deviation persistent shock	$(\sigma_{\upsilon_N}, \sigma_{\upsilon_E})$	(0.051, 0.056)	MCVL 2005-2019
Standard deviation fixed-effect	$(\sigma_{\theta_N}, \sigma_{\theta_E})$	(0.514, 0.563)	MCVL 2005-2019
Mean gross labor income (thousands)	$\overline{y}$	24.88	MCVL 2005-2019
Tax function	$(\varsigma_0, \varsigma_1)$	(0.898, 0.148)	García-Miralles, Guner and Ramos (2019)
Transfer function	(g_0, g_1)	(0.0227, -0.009)	EU-SILC 2004-2019
Unemployment benefits (thousands)	b	3.3	EU-SILC 2004-2019

Some parameters are computed from the data as simple averages. These include the down-payment requirement χ, the real interest and mortgage rates, r and r^h, housing sizes

[25] Agents in the model can inherit a house multiple times over their lifetimes, which might be viewed as receiving a significant inheritance. Allowing individuals to only inherit only once, although realistic, would require adding a state variable. Moreover, despite the substantial increase in computational costs, results remain almost identical whether or not a state variable for housing bequests is included (see Appendix E.3).

[26] Kaas et al. (2021) report overall transaction costs of 13.5% for Spain. I assume that these costs are equally split among home-buyers and home-sellers.

$\overline{h}_{deh}$, the average number of siblings in the household κ^b, the mean gross labor income $\overline{y}$, and the amount of unemployment benefits b. Finally, some parameters are estimated in the data following the procedures detailed in Sections 4.1.1 and 4.1.2. This is the case for the income process parameters, ϱ_N, ϱ_E, σ_{υ_N}, σ_{υ_E}, σ_{θ_N}, and σ_{θ_E}, the probability implied by $\Psi_{edj}(l)$ of changing employment status, the g_0, g_1 parameters for the transfer function, and the probability of receiving a housing bequest $\pi^b_{ed_0j}$. Parameters' values are either reported in Table 1 or, in the case of housing sizes $(\overline{h}_{de1}, \overline{h}_{de2})$, probabilities $\pi^b_{ed_0j}$ and transitions $\Psi_{edj}(l)$, can be found in Table E4, Figures 3a and 4b, and Appendix Figures E11 and E14.

The initial conditions include agents' fixed types (birthplace, education, migration type, and fixed productivity), initial choices (assets, location, and housing status) and employment status. The initial distributions can vary based on location, education, and native status, but are the same across time. Further details on the estimation of these distributions can be found in Appendix E.4.4.

4.2 Estimating the Equilibrium With Aggregate Shocks

This section outlines the strategy to compute the benchmark and counterfactual equilibria. The key challenge, i.e. how to form expectations over endogenous local prices and wages, is solved with a low-rank forecast rule which accurately predicts the observed time series, that are perfectly matched in the benchmark equilibrium. In counterfactual exercises, the forecast rule is updated to line up with the new prices and wages.

4.2.1 Equilibrium and Aggregate Shocks

Wages and prices in each location, $\mathbf{q} = \{\mathbf{p}, \mathbf{w_N}, \mathbf{w_E}\}$, depend on the equilibrium allocation of agents across states μ and on the location-specific shocks to primitives $\boldsymbol{\mathcal{Z}}$, i.e., $\mathbf{q}(\Omega)$. Two problems arise when trying to infer future prices and wages $\mathbf{q}(\Omega')$ from $\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}})$ and $\mu' \sim \Gamma_\mu(\Omega)$. First, $\Gamma_{\mathcal{Z}}$ requires forming expectations over 36 different processes, which is not computationally feasible.[27] Second, knowing Γ_μ would require agents to keep track of the equilibrium law of motion of μ, an infinitely-dimensional object.

To solve this problem, let's first take a step back. Suppose that we know the equilibrium housing demand $H^D_d(\Omega)$ and labor supply $L^S_{ed}(\Omega)$ by education (efficiency units) in each location. Then, there exists a known equilibrium function $Q(\cdot)$ that maps $\{H^D_d(\Omega), L^S_{Nd}(\Omega), L^S_{Ed}(\Omega)\}$ and $\boldsymbol{\mathcal{Z}}$ to equilibrium prices $\mathbf{q}(\Omega)$:

$$\mathbf{q}(\Omega) = Q(\boldsymbol{\mathcal{Z}}, H^D_d(\Omega), L^S_{Nd}(\Omega), L^S_{Ed}(\Omega)).$$

This function is obtained by plugging housing demand and labor supply into the local inverse housing supply and labor demand functions (1) and (2), i.e.:

[27] Agents would need to integrate their expectations over 36 dimensions. Even by only allowing 2 grid points for each discretized Markovian process, the state space would be multiplied by a factor of 2^{36}, i.e. by almost 70 billion points.

$$p_d(\Omega) = k_d[H_d^D(\Omega)]^{\psi}, \tag{7}$$

$$w_{Nd}(\Omega) = X_d \left\{\zeta_d[L_{Nd}^S(\Omega)]^{\rho} + (1-\zeta_d)[L_{Ed}^S(\Omega)]^{\rho}\right\}^{\frac{1-\rho}{\rho}} \zeta_d[L_{Nd}^S(\Omega)]^{\rho-1}, \tag{8}$$

$$w_{Ed}(\Omega) = X_d \left\{\zeta_d[L_{Nd}^S(\Omega)]^{\rho} + (1-\zeta_d)[L_{Ed}^S(\Omega)]^{\rho}\right\}^{\frac{1-\rho}{\rho}} (1-\zeta_d)[L_{Ed}^S(\Omega)]^{\rho-1}. \tag{9}$$

Assume now that agents form correct expectations over $\mathbf{q}(\Omega')$ and that the model parameters are known. Then, the model can be solved by initially imposing prices and wages $\mathbf{q}(\Omega)$ from the data. In particular, let $\mathbf{q}(\Omega) = \mathbf{q_t}$, where $\mathbf{q_t}$ denotes the yearly time series of prices and wages observed in the data in each location, for years $t \in \{2010, ..., 2019\}$. Then, in each t, one can simulate equilibrium quantities $\{H_{dt}^D(\Omega_t), L_{Ndt}^S(\Omega_t), L_{Edt}^S(\Omega_t)\}$. Finally, given $\{H_{dt}^D(\Omega_t), L_{Ndt}^S(\Omega_t), L_{Edt}^S(\Omega_t)\}$ and and $Q(\cdot)$, we can find the aggregate shocks $\boldsymbol{\mathcal{Z}_t}$ that generate observed prices and wages $\mathbf{q_t}$ in equilibrium. For instance, given the simulated housing demand in urban Madrid, the data price, and the known parameter ψ, it is straightforward to invert (7) to obtain the housing construction cost shock k_{dt} that delivers the observed price in that year. More generally, the procedure to recover $\{k_{dt}, X_{dt}, \zeta_{dt}\}$ in each location is described in Appendices C.2 and C.3.

In other words, as it is usually the case with quantitative spatial models, prices and wages in the data can be perfectly matched in the benchmark equilibrium by inverting the model's primitives. However, solving and simulating the model relies on agents forming correct expectations over $\mathbf{q}_{t+1}$. How can this be the case? We need a forecast rule that is both accurate and low-dimensional. Such a rule is defined in the next section.

4.2.2 Forecast Rule

The forecast rule for observed prices and wages $\mathbf{q}_t$ is estimated using a low-rank factor model of rank two (Bai 2009), so that

$$\mathbf{q}_t \simeq \boldsymbol{\lambda_1} f_{1t} + \boldsymbol{\lambda_2} f_{2t}, \tag{10}$$

where $\boldsymbol{\lambda_k} = (\lambda_{k1}^p, \ldots, \lambda_{k12}^p, \lambda_{k1}^{w_N}, \ldots, \lambda_{k12}^{w_N}, \lambda_{k1}^{w_E}, \ldots, \lambda_{k12}^{w_E})$, for $k = 1, 2$, are the *fixed* loading parameters for each process and location, and (f_{1t}, f_{2t}) are two aggregate (country-level) *time-varying* and orthogonal factors. In particular, the local time series for housing prices and wages by education can be approximated as combinations of just 2 common underlying factors and process- and location-specific parameters, as:

$$\begin{aligned} p_{dt} &\simeq \lambda_{1d}^p f_{1t} + \lambda_{2d}^p f_{2t}, \\ w_{Ndt} &\simeq \lambda_{1d}^{w_N} f_{1t} + \lambda_{2d}^{w_N} f_{2t}, \\ w_{Edt} &\simeq \lambda_{1d}^{w_E} f_{1t} + \lambda_{2d}^{w_E} f_{2t}, \end{aligned}$$

so that the dimensionality of the problem, which is infinite for $\mathbf{q}(\Omega)$, is reduced to 2.

Furthermore, assume that the aggregate factors follow two mutually-independent AR(1) processes,

$$f_{1t+1} = \varrho_1 f_{1t} + \upsilon_1, \tag{11}$$

$$f_{2t+1} = \varrho_2 f_{2t} + \upsilon_2. \tag{12}$$

$$\upsilon_1 \sim N(0, \sigma_{f_1}^2), \;\; \upsilon_2 \sim N(0, \sigma_{f_2}^2).$$

Hence, each period economy can move from any realization of (f_{1t},f_{2t}) to a new one. Agents know how factors (f_{1t},f_{2t}) evolve, and can map each realization of (f_{1t},f_{2t}) into prices and wages in each location using (10). In particular, if agents know parameters ($\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$, ϱ_1, ϱ_2, σ_{f_1}, σ_{f_2}) and observe current factors (f_{1t},f_{2t}), then they can form expectations about future prices and wages in each location following the rule:

$$\mathbf{q}_{t+1} = \boldsymbol{\lambda_1} f_{1t+1} + \boldsymbol{\lambda_2} f_{2t+1} + \boldsymbol{\upsilon_{t+1}}, \tag{13}$$

where $\boldsymbol{\upsilon_q}$ is the approximation error in equation (10).

In Section 4.2.3, I describe the procedure to solve the benchmark equilibrium using the forecast rule (13) and the factors' laws of motion (11) and (12). In Section 4.2.4, I show how to solve for the counterfactual equilibrium where housing policies are introduced. In Appendix D.3, I draw a connection between the low-rank forecast rule and forecast rules in the Krusell and Smith (1998) tradition, as well as highlighting the limitations of traditional Krusell and Smith-type strategies in the present high-dimensional spatial setting. In Appendix D.4 I justify the use of the low-rank forecast rule (13) by performing a series of accuracy tests, both in the benchmark and in the counterfactual equilibria. Finally, in Appendix C.5 I provide a formal definition of the equilibrium.

4.2.3 Estimating the Benchmark Equilibrium

In the benchmark equilibrium, I can choose $\boldsymbol{\mathcal{Z}_t}$ to perfectly match the time series of prices and wages in the data, with $t \in \{2010, ..., 2019\}$. This is done by inverting equilibrium equations (7), (8), and (9), given the model's parameters and the benchmark equilibrium quantities $\{H^D_{dt}(\Omega_t), L^S_{Ndt}(\Omega_t), L^S_{Edt}(\Omega_t)\}$, which are simulated by imposing the equilibrium prices and wages observed in the data. As a result, for each realization of (f_{1t}, f_{2t}), I calculate a set $\boldsymbol{\mathcal{Z}}_t$ so that the model economy matches observed prices and wages in equilibrum.

Moreover, since I know that the $\mathbf{q}_t$ vector of prices and wages observed in the data are equilibrium objects in the benchmark, I can directly estimate equation

$$\mathbf{q}_t = \boldsymbol{\lambda_1} f_{1t} + \boldsymbol{\lambda_2} f_{2t}$$

outside of the model. I perform the estimation using the interactive fixed-effects procedure described in Bai (2009)[28]. As can be seen in Appendix Figures D8, D9, and D10, the prediction

[28]The estimation procedure of the factors and the loading parameters proposed in Bai (2009) involves principal component analysis. The objective of the procedure is to find the matrix of rank 2 that bests approximates the unrestricted 36×10 matrix of fixed effects that would perfectly capture the evolution of local prices and

of the factor model aligns very closely to the prices and wages matched from the data. This is not surprising, since this class of factor models are designed to predict the evolution of the data as best as possible, subject to the low-rank restriction. Further details on the estimates can be found in Appendix D.5.

Finally, given the estimated time series of (f_{1t}, f_{2t}) factors, I can estimate their AR(1) process externally using equations (11) and (12).[29] I assume that each factor follows a three-points Markov chain, which I estimate as discrete approximation of the AR(1) processes.[30] Therefore, all the right-hand side variables and the laws of motion for factors in the forecast rule (13) can be readily estimated from the data.

In Appendix D.1, I present an algorithm to solve the benchmark equilibrium. Knowledge of current factors is enough to form accurate predictions about future prices and wages, and agents do not need to keep track of the evolution of μ_t and $\mathcal{Z}_t$. The estimation of the forecast rule in the benchmark equilibrium does not require simulating the model economy multiple times until convergence in the agents' guesses is achieved. Instead, the rule is estimated only once outside of the model, thanks to the fact that observed prices and wages can be perfectly matched in equilibrium. This hugely simplifies estimation, since, given the high-dimensionality of the spatial model, solving efficiently for the benchmark economy is necessary to be able to calibrate the model's parameters.

This computational advantage distinguishes my approach from canonical strategies based on Krusell and Smith (1998), which typically require multiple simulations to estimate the forecast rules in the benchmark equilibrium. More details on the connection between the low-rank forecast rule and rules in the Krusell and Smith (1998) tradition are given in Appendix D.3. Appendix D.4 justifies the use of the low-rank forecast rule by performing a series of accuracy tests when predicting benchmark equilibrium prices and wages and by testing for the optimal number of factors.

4.2.4 Estimating Counterfactuals

In the counterfactual exercises where housing policies are introduced, a new stochastic steady-state equilibrium is solved. Equilibrium prices and wages are different from the benchmark, even if the process for the exogenous aggregate shocks remains the same. This occurs because policies affect individual decisions, thus modifying the distribution of agents μ_t across individual states (e.g. the share of people renting in a specific urban location). Equilibrium prices and wages are updated using equations (7), (8), and (9), where shocks $\mathcal{Z}_t$ are taken as given and are set equal to their benchmark values.

Since equilibrium prices and wages have changed, the forecast rule (13) estimated from

wages over time.

[29]For benchmark equilibrium prices and wages, I obtain estimates $\varrho_1 = 0.922$, $\varrho_2 = 0.753$ (standard errors 0.055 and 0.094, respectively), and $\sigma_{f_1} = 0.188$, $\sigma_{f_2} = 0.454$.

[30]When solving the model, agents form expectations on future factor realizations based on current realizations and Markov chain probabilities. Moreover, Markov chains are linearly interpolated in the simulations, which results in different probabilities attached to specific values of (f_{1t+1}, f_{2t+1}) depending on the exact (f_{1t}, f_{2t}) realizations.

the data is no longer valid and needs to be updated. Loading parameters ($\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$), factors (f_{1t}, f_{2t}), and the parameters governing their laws of motions (11) and (12) are allowed to change in the counterfactual. At each iteration of the model, I re-estimate the agents' forecast rule using the new set of counterfactual prices and wages, and I repeat the process until convergence in the loading parameters and factors. In this way, the agents' forecast rule is consistent with the new equilibrium prices and wages, under all the different realizations of the aggregate shocks $\boldsymbol{\mathcal{Z}_t}$. Agents know how the updated factors (f_{1t},f_{2t}) evolve, and can map each realization of (f_{1t},f_{2t}) into prices and wages in each location using the new forecast rule.

In Appendix D.2, I present an algorithm to solve the counterfactual equilibrium. Differently from the benchmark, estimating the forecast rule in the counterfactual equilibrium requires simulating the model economy multiple times until convergence in the agents' guesses is achieved. This is related to canonical strategies based on Krusell and Smith (1998). The procedure, however, is not computationally taxing, as the other model parameters have already been estimated and are kept fixed at their benchmark values, so that the counterfactual equilibrium can be solved using a single outer loop. Appendix D.4 justifies the use of the low-rank forecast rule by performing a series of accuracy tests when predicting counterfactual equilibrium prices and wages.

4.3 Calibration

This section describes the calibration strategy. Given the exogenous inputs and the agents' forecast rule, I internally calibrate the remaining parameters to match key moments in the data – including coresidence, homeownership, and migration rates over the lifecycle.

4.3.1 Calibrated Parameters

The remaining model parameters, listed in Table 2, are estimated with the simulated method of moments.[31] Some of the calibrated parameters affect the utility function: the discount factor β, the utility from being a homeowner η_1 and living in the birthplace η_2, and the terminal utility from holding wealth and a house, ω_1 and ω_2. The share of stayer types π_τ, the migration and coresidence cost δ, ξ parameters, and amenities A_{d_j} are also calibrated.

The calibrated parameters indicate that agents gain direct utility from homeownership ($\eta_1 = 0.216$) and from residing in their birthplace ($\eta_2 = 0.007$), beyond the economic benefits these choices provide. When translated into economic terms, an agent consuming the average consumption level would be willing to give up 53.9% of their consumption (10,260 euros) to obtain the utility benefits of homeownership, and 3.2% of their consumption (610 euros) to enjoy the utility of living in their birthplace. In this setting where other advantages of being native, namely the option to coreside and easier access to housing bequests, are explicitly

[31] In particular, I use an exponential natural evolution strategy. This is a numerical optimization algorithms that makes updates based on the natural gradient instead of the plain gradient, and belongs to the family of derivative-free algorithms known as black box optimizers. The Julia package `BlackBoxOptim` is used for estimation.

Table 2: Calibrated parameters.

	Parameter	Estimate
Utility		
Discount factor	β	0.990
Utility homeowner	η_1	0.216
Utility home-bias	η_2	0.007
Terminal value wealth	ω_1	0.001
Terminal value homeowner	ω_2	0.000
Other		
Share stayers	π_τ	0.800

	Parameter	Estimate
Migration cost function		
Intercept	δ_0	1.876
Age	δ_1	0.007
Log age	δ_2	0.745
College	δ_e	-0.056
Coresidence cost function		
Intercept	ξ_0	0.148
Age	ξ_1	0.001

modelled, pure preferences coming from home-bias do not play a quantitatively large role.[32]

Agents in the model also experience disutility when migrating or coresiding with parents. A newborn non-college person consuming the average consumption level would be willing to give up 96.2% of their consumption (18,330 euros) not to pay the one-off migration utility cost. Moreover, they would sacrifice 43% of their consumption (8,200 euros) not to experience the disutility from coresiding. The estimated discount factor is close to standard values in the literature ($\beta = 0.99$).

A large share of the population (80%) is estimated to be of the "stayer" type, i.e. with prohibitively high migration costs. Koşar, Ransom and Van der Klaauw (2022) use the NY Fed's Survey of Consumer Expectations to elicit respondents' migration probabilities for a set of hypothetical scenarios. They find very large non-pecuniary moving costs for 52% of the population ("rooted") and infinitely large moving costs for 12% of the population ("never-movers"). These estimates may serve as a lower bound for Spain, given that the Spanish internal migration rate is lower than in the U.S.[33]

4.3.2 Targeted Moments

The internally calibrated parameters are estimated to match key moments in the data. The simulated moments and their data counterparts are listed in Table 3. They include the share of homeowners (overall and among coresidents), the share of coresidents, the migration rate among the college-educated, the share of people not migrating, and the median wealth to income ratio at age 50. They also include statistics that capture the lifecycle evolution of housing status and migration. Specifically, the targeted moments are homeownership, coresidence, and migration rates during the initial 10 ages in the model (25-34) as well as the last 10 ages (55-64). The migration flows across locations are also targeted, and are reported in Appendix Table E3.

The moments of Table 3 are simple averages or ratios. Their estimation in the data and model is straightforward, with the exception of the share of individuals who never migrate. The challenge arises due to the short 4-year panel dimension of the EU-SILC, which makes the

[32]This is different from Zerecero (2021), where all the benefits from being natives are loaded in the home-bias preference parameter.

[33]The annual migration rate is around 3% across U.S Metropolitan Statistical Areas and 2% across States (Molloy, Smith and Wozniak 2011), whereas it is 0.8% across Spanish locations.

Table 3: Targeted moments. Data: Census, EU-SILC, MCVL, EFF.

	Data	Model
Lifecycle Homeownership Rates		
Ages 25-34	0.481	0.459
Ages 55-64	0.880	0.919
Lifecycle Coresidence Rates		
Ages 25-34	0.400	0.416
Ages 55-64	0.060	0.060
Lifecycle Migration Rates		
Ages 25-34	0.015	0.015
Ages 55-64	0.005	0.005

	Data	Model
Housing Tenure		
Share homeowners, not coresidents	0.900	0.919
Share homeowners	0.735	0.731
Share coresidents	0.184	0.205
Internal Migration		
Migration rate, college-educated	0.010	0.010
Share never migrating	0.894	0.889
Wealth		
Median wealth-income ratio, age 50	7.002	7.549

data unsuitable for estimating the proportion of individuals who never relocate throughout their lifecycle. To address this issue, I also use information from the longer MCVL panel. A detailed explanation of this inference procedure can be found in Appendix E.4.2.

As can be seen in Table 3, the model captures well the lifecycle evolution in housing tenure and migration rates. As in the data, homeownership increases over the lifecycle, whereas coresidence and migration rates decrease. The cross-sectional data moments, related to housing tenure choices, migration, and wealth, are also well captured in the model. Moreover, the model replicates the migration flows to different locations observed in the data (Table E3).

Some moments relate more strongly to some parameters than others. Migration and coresidence rates are connected to the corresponding migration and coresidence cost function parameters. In particular, intercepts (δ_0, ξ_0) are related to the average migration and coresidence rates, whereas age coefficients (δ_1, δ_2, ξ_1) are related to their evolution over the lifecycle. Moreover, the migration college coefficient δ_e is identified by the migration rate among the college-educated. The homeownership rates over the lifecycle, instead, are closely linked to the period and terminal utility from homeownership parameters (η_1, ω_1). Finally, the share of stayers π_τ is identified by the share of people never migrating, and the median wealth to income ratio is closely related to the discount factor β.

5 Validation

I validate the model by showing that it replicates a set of untargeted moments. First, I run a policy within the model that mirrors a place-based homeownership subsidy that was recently introduced in Spain. As a result of the increase in homeownership following the policy, the model yields a migration elasticity that is similar to the one estimated in the data. Second, I find that the model reproduces the lifecycle homeownership gap between natives and migrants and the income premium enjoyed by migrants over the lifecycle. As a final non-targeted moment, the model reproduces the observed negative relationship between local homeownership rates and wealth inequality across locations.

5.1 Place-Based Subsidy for Young Homebuyers

In the model, homeownership reduces internal migration through the existence of monetary transaction costs associated with selling a house. To validate this key channel in the model, I take advantage of the quasi-experimental nature of a policy that subsidized homeownership for young individuals residing in small municipalities. I run the same policy in the model and, as a result of the increase in homeownership following the subsidy, I find a negative migration elasticity that is close to the one estimated in the data. Notably, transaction costs – the key parameter driving homeowners' lower mobility in the model – are externally set and taken from the literature (Kaas et al. 2021), which makes the elasticity a fully untargeted moment.

The policy, introduced in early 2018, gives a subsidy to first-time home buyers in cities with less than five thousand inhabitants consisting of 10,800 euros, or up to 20% of the house price if the amount is lower. To qualify, the buyer has to be under 35 years old, earn a gross annual income of less than 19,400 euros, and use the purchased house as their primary residence. Furthermore, the house's price cannot exceed 100,000 euros.[34]

After purchasing the home with this subsidy, the buyer has the freedom to sell it under specific conditions: relocating for job reasons, acquiring another home in the same or in a different location, or if five years have passed since the subsidy was granted. Should one relocate without meeting these conditions before the five-year mark, a proportional repayment of the subsidy is required. For instance, moving after two years without qualifying for one of the exceptions requires returning 5,400 euros of the subsidy.

Policy in The Data To analyze the policy's impact in the data, it is crucial to identify the small municipalities where the policy was introduced. The MCVL does not distinguish between cities with population below 40,000. Instead, I use the Residential Variation Statistics (*Estadística de Variaciones Residenciales*, or EVR) and the Continuous Register Statistics (*Estadística del Padrón Continuo*, or EVR) datasets. These combined datasets provide data on the universe of Spanish movers and stayers, respectively, and include an indicator for residents in cities with population lower than 10,000. However, they lack information on homeownership, which is needed to analyse the first stage impact of the policy, and on income, which could be used to restrict the treatment to the low-income recipients.

To address the first limitation, I estimate the first stage using the Household Budget Survey (*Encuesta de Presupuestos Familiares*, or EPF). This dataset not only includes an indicator for those living in cities with a population of less than 10,000, but also crucially gives information on homeownership status. To account for the second limitation, the treatment does not condition on any income data. Treated individuals are residents of cities with fewer than 10,000 inhabitants who are younger than 35, not all of which are eligible for the policy. For instance, some of them might live in cities with populations between 5,000 and 10,000 or earn more than 19,400 euros, rendering them ineligible. Therefore, the analysis should be

[34]The policy, originally introduced to increase emancipation among young people and to reduce rural depopulation, has been renewed in 2021 with slight modifications, including a higher municipality size threshold of 10,000 inhabitants. For data availability reasons and to isolate the effect of the policy from the impact of the COVID-19 pandemic, the analysis focuses on the 2016-2019 period, i.e. two years before and after the first introduction of the policy.

understood as an intention-to-treat exercise. The datasets cover the years 2016 to 2019, encompassing two years before and after the implementation of the policy in early 2018.

The event studies estimated in the data take the form:

$$y_{it} = \alpha_{\text{small1}} + \alpha_t \times \alpha_{\text{small2}} + \alpha_r + \alpha_{rt} + \alpha X_{it} + \epsilon_{it}, \tag{14}$$

where y_{it} is the outcome, which is either the homeownership status (first stage) or the individual migration event (reduced form), α_{small} is an indicator function for the set of treated cities (municipalities with less than 10,000 inhabitants), and α_t, α_r, and α_{rt} are year, region, and region-year fixed-effects, respectively. In the baseline specification, additional controls X_{it} include gender, age and age squared. Migration occurs when the following year's location of residence is different from the current one. The sample is restricted to individuals younger than 35, the age eligibility threshold for the subsidy, and to individuals living in municipalities with less than 20,000 inhabitants. This ensures a suitable control group: residents aged below 35 from slightly larger municipalities than the population threshold required for the subsidy, specifically those with populations between 10,000 and 20,000.

The two event studies for the first stage and the reduced form, centered around the 2018 policy year, are depicted in Figures 5a and 5b. Consistent with the model's predictions, I observe that the policy increased homeownership and decreased out-migration. As can be seen in columns (1) and (2) of Appendix Table F5, the homeownership rate among the treated increased by 0.115 and the annual migration rate decreased by 0.002 on average. These estimates come from the first stage and reduced form difference-in-differences version of regression (14). More details on this diff-in-diff specification can be found in Appendix F.1 and equation (40).

The lack of significant pre-trends in the event studies (Figures 5a and 5b) is in line with the conditional exogeneity assumption of the treatment. As further exogeneity checks, I run two placebo event studies where the sample is restricted to people aged 37-40, who are just above the age eligibility threshold and could not have received the subsidy in any of the two post-treatment years of the event study. The treatment and control groups are otherwise defined as in the baseline regressions, i.e. people living in municipalities below or above 10,000 inhabitants (but less than 20,000 inhabitants). In line with the exclusion restrictions, the placebo treatment does not have a significant effect on the outcomes (Appendix Figures F15a and F15b).

Moreover, I estimate the migration regression by restricting the sample to people *born* in municipalities with less than 20,000 inhabitants, and defining α_{small} in regression (14) as an indicator function for people *born* in municipalities with less than 10,000 inhabitants. The birthplace treatment, arguably more likely to be exogenous than a treatment based on current residence, produces results that are similar to the baseline regression (see Appendix Figure F16).[35] In all specifications, COVID-19 years are excluded to isolate the impact of the policy

[35]The birthplace treatment specification can only be estimated for the reduced form, and not for the first stage, for data availability reasons.

from the disruptive effect on homeownership and migration decisions caused by the pandemic. Event-study specifications that include the COVID-19 years until 2021 can be found in Appendix Figures F15c and F15d.

I combine the first stage and reduced form estimates to obtain a migration elasticity with respect to changes in homeownership. As reported in column (2) of Table 4, new homeowners decrease their annual migration by around 1.83 percentage points as a result of the policy, a substantial effect given the average migration rate of 0.008. This is in line with estimates obtained from running a separate regression with household fixed-effects and the same control variables using the EU-SILC panel, a different dataset encompassing years 2004-2019 and all locations (column 1 of Table 4). The elasticity estimated with the panel is very similar to the one obtained from the policy experiment (-0.0186 instead of -0.0183).

I argue that the estimated reduction in migration rates following the increase in homeownership due to the policy reflects the causal effect of homeownership on internal migration. Beyond the exogeneity issues previously discussed and addressed, another potential concern challenges this interpretation. Specifically, the policy's influence on reduced migration might be direct, that is, it may not be entirely mediated by the policy's effect on homeownership. In such a scenario, the observed increase in homeownership may partially arise from individuals who opt to remain as a result of the policy, which also enables them to purchase a house. This could primarily be driven by the policy's repayment rules, which require that a homebuyer who relocates within five years for non-professional reasons and without acquiring another house must make a proportional repayment of the subsidy. Such rules may thus deter new home-buyers from moving for reasons which are not directly related with their status as homeowners (e.g., transaction costs of selling).

To address this concern, I turn to the model. By simulating a version of the policy that does not have these (mild) repayment requirements, I estimate an elasticity that is virtually identical to the one I obtain when simulating the actual policy (-0.021). Moreover, by additionally simulating the policy in a counterfactual equilibrium without transaction costs of selling the house ϕ_s, the key friction that limits homeowners' mobility, I find no reduced form effects of the policy on migration rates. Therefore, I conclude that a reverse causality effect of migration on homeownership is not driving the results, and that estimates from the quasi-experiment reflect the causal effects of homeownership on migration rates.

Policy in The Model The place-based subsidy to young low-income homebuyers living in small cities is simulated in the model. To do so, I need to add an additional state variable that takes on three values: not eligible for the policy, eligible and not recipient, and recipient. This distinction is needed because people who already received the subsidy cannot ask for it again, since the policy only applies to first-time homebuyers. Moreover, in the version of the policy that reproduces the repayment rules, recipients of the policy may be obliged to pay part of the subsidy back in case they move, while non-recipients are not.

The eligibility criteria reflect the income and age thresholds from the 2018 policy. This place-based policy is implemented exclusively in locations Northwest rural, Center rural, and

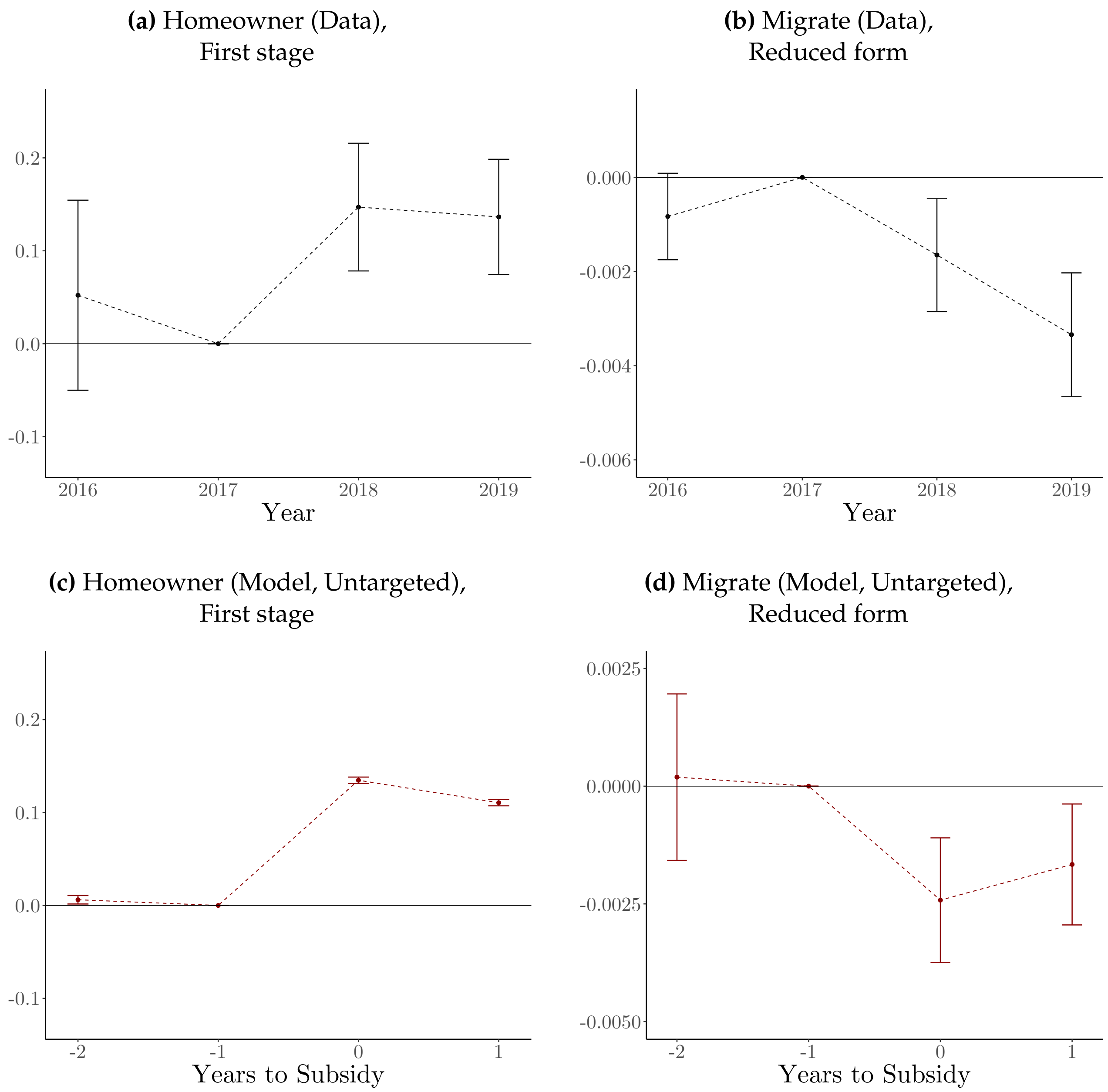


Figure 5: Treated in panels **(a)** and **(b)**: People aged less than 35 living in small cities (<10k inhabitants). Control group: same age living in slightly larger cities (10k-20k). Treatment year: 2018. Migration across regions (NUTS-1) and rural-urban areas (rural <10k, rural >10k or urban >40k). Treated in panels **(c)** and **(d)**: People who received the subsidy in the counterfactual model with the policy. Control group: same people as the treated, but in the baseline model without the policy. Other included controls and fixed-effects: gender (in the data), age, age squared, region, and region-year. Clustered (locations) standard errors. Data: EPF, EPC, EVR 2016-2019.

South rural, the rural locations with the lowest mean housing prices. Within these locations, average housing prices fluctuate between 100,000 and 150,000 euros, aligning more closely with the policy's 100k eligibility threshold than other rural areas. Yet, not every resident in these three locations meeting the age and income requirements is automatically eligible. When modeling the policy, only a random 8.75% share of them is considered eligible, aligning with the proportion of smaller municipalities (those with populations under 5,000) within these rural locations. Finally, the one-off subsidy for recipient home-buyers consists of 10,800 euros, an amount that, in line with the policy requirements, is always higher than 20% of the local house prices.

The 2018 policy is not included in the baseline, which is estimated using price and wage data from 2010 to 2019. Rather, it is simulated in a counterfactual exercise. Since the focus of the diff-in-diff and event study analyses is on the short-term effect of the policy in the two subsequent years after it is put in place, I do not allow model prices and wages to adjust in general equilibrium. More details on the policy implementation in the model are given in Appendix F.1.

The treatment group comprises individuals receiving the subsidy in this counterfactual scenario, while the control group consists of these same individuals in the baseline model where the policy is absent. In order words, using the structure of the model, causal identification is achieved by directly comparing *potential outcomes*. I estimate the event study within the model using the same regression specification as its data counterpart, equation (14). However, since I am directly measuring the effect on the treated rather than the intention-to-treat, I use individual fixed-effects α_{i1} and α_{i2} instead of the small municipalities indicator functions α_{small1} and α_{small2}. Moreover, I don't control for gender (absent in the model) and I center the event studies around the year the individual receives the subsidy, either 2018 or 2019. As in the data, the sample is restricted to the years 2016-2019.

The evolution of the event studies for both the first stage (Figure F17a) and the reduced form (Figure F17b) is qualitatively similar to their data counterparts (Figures 5a and 5b). There are no pronounced pre-trends before the subsidy, and, mirroring the effect in the data, homeownership increases and migration decreases as a result of the policy. The corresponding difference-in-differences average effects are reported in columns (3) and (4) of Appendix Table F5. The magnitude of the effect on homeownership and migration in the model is, by construction, higher than in the data. This is because, in the model, I am measuring the impact on the treated rather than the intention-to-treat.

I can perform a back-of-the-envelope adjustment of the model estimates by multiplying them by the take-up rate in the model (among the eligible and non-eligible individuals aged less than 35 living in small cities), to obtain some approximate intention-to-treat estimates that can be more easily compared with the *absolute* changes in the data. Since the take-up rate in the model is 13.97%, back-of-the-envelope first stage and reduced form model estimates are 0.6697 $\times$ 0.1397 = 0.0936 and -0.0141 $\times$ 0.1397 = -0.0020, respectively. These estimates are close to their (untargeted) data counterpart (0.1151 and -0.0021), as can be seen in Table F5. A similar adjustment to the event study estimates (Figures 5c and 5d) delivers results which are very close to their data counterparts (Figures 5a and 5b).

Importantly, the relative effect on migration with respect to homeownership is also similar. When combining the diff-in-diff first stage and reduced form results, I obtain a migration elasticity with respect to changes in homeownership of -0.0211, which is close to the ones estimated in the data. This can be seen by comparing the model elasticity of column (3) in Table 4 with the data panel elasticity of columns (1) or the data policy elasticity of column (2). Notably, transaction costs – the key parameter driving homeowners' lower mobility in the model – are externally set and taken from the literature (Kaas et al. 2021), which makes the elasticity a fully untargeted moment.

Table 4: Estimated migration elasticity with respect to changes in homeownership.

	Migrate		
	Data Panel	Data Policy	Model Policy
	(1)	(2)	(3)
Homeowner	-0.0186*	-0.0183**	-0.0211***
	(0.0089)	(0.0099)	(0.0054)
Household FE	✓		
Individual FE			✓
Year-Region FE	✓	✓	✓
Region FE	✓	✓	✓
Observations	73,570	4,535,448	17,363

Note: Other included controls: age, age squared, and, in (1) and (2), gender. Clustered (location) std. errors in (1) and (3), bootstrap std. errors in (2), *p<0.1, **p<0.05, ***p<0.01. Data: EPF, EPC, EVR 2016-2019, EU-SILC 2004-2019.

As a final counterfactual experiment, I simulate the policy in an economy without transaction costs of selling the house ($\phi_s = 0$). As can be seen in the event studies of Figure F18, the policy increases homeownership but has no significant reduced form effects on migration rates. Therefore, I conclude that the estimates from the quasi-experiment reflect the causal effects of homeownership on migration rates. The reduction in migration coming from changes in homeownership appears to be mainly driven by the transaction costs of selling the house ϕ_s.

5.2 Natives Are More Likely to be Homeowners Despite Earning Less

An additional set of untargeted moments predicted by the model is illustrated in Figure 6. These figures delineate the lifecycle profiles of homeownership rates and income, distinguishing between natives and migrants. The model accurately predicts that natives have a higher propensity for homeownership compared to migrants (Figure 6a). The homeownership gap is higher at younger ages and shrinks as agents get older, although it remains persistent at around 10 percentage points.

Interestingly, natives are more likely to be homeowners even though they earn less over the lifecycle. This feature of the data is also matched in the model as an untargeted moment (Figure 6b). The higher income observed among migrants can be partially attributed to a larger proportion of college-educated individuals within this group. Additionally, this income advantage may serve as a compensating differential offsetting the inherent benefits enjoyed by natives, such as easiest access to bequests, the option for coresidence, and the home-bias utility. By delving into the mechanism, I conclude that the option to coreside with parents is the key element behind the gap in homeownership and income gap.

Mechanism To uncover the determinants of the homeownership and income gap between

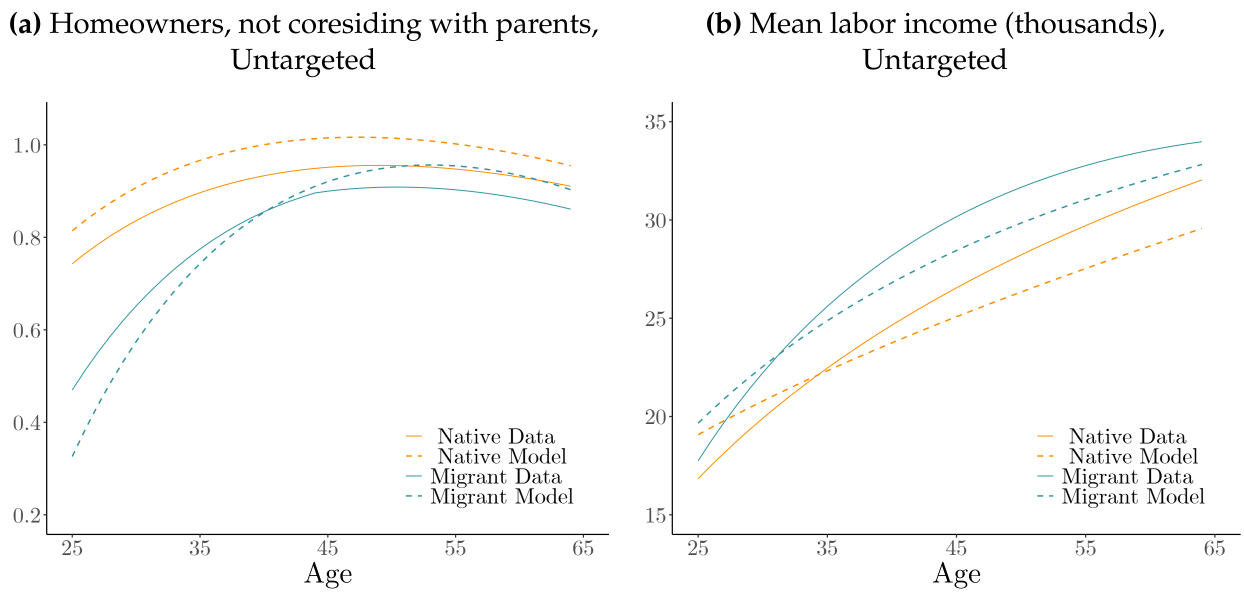


Figure 6: The figures plot the untargeted lifecycle profiles of homeownership rate and income, separately for natives and migrants. Data: Census 2011 (panel **a**), MCVL 2005-2019 (panel **b**).

natives and internal migrants, I explore the effects of coresidence, housing bequests, and the home-bias exogenous utility parameter. By iteratively shutting down one of these three channels while maintaining the other two, I conclude the option for natives to coreside is the key driver of the gaps. If I keep the housing bequests and home-bias preference while eliminating the coresidence option, I find that both the homeownership and income gaps between natives and migrants vanish (Figure F19).[36]

Coresidence has a positive impact on savings and on (housing) wealth accumulation. This is different from Kaplan (2012), where the option to live with parents leads to overall *lower* saving rates. In his framework, this happens because coresidence reduces the precautionary savings motive by providing an insurance mechanism against negative income shocks. While this mechanism is also incorporated in my model, it is quantitatively less important than the direct effect coming from the absence of housing costs during coresidence. Moreover, differently from Kaplan (2012), my model incorporates homeownership paired with a downpayment constraint. Coresidence offers agents a way to overcome this credit friction. By saving while they are living with their parents, they can obtain sufficient funds for the downpayment, which leads to subsequent housing wealth accumulation.[37]

5.3 High Homeownership Locations Have Lower Wealth Inequality

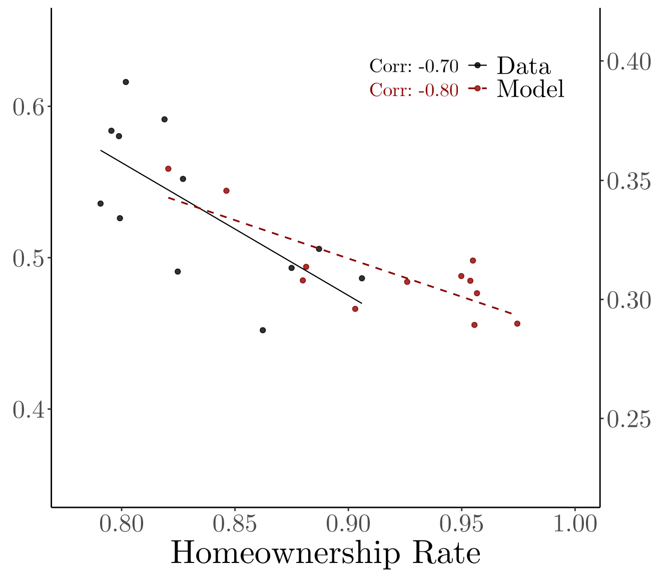
The model also predicts the untargeted inverse relationship observed between the local homeownership rate and wealth inequality across locations. This relationship is shown in Figure

[36] Housing bequests do not significantly drive this gap because when migrants inherit property that they must sell due to their non-residence in their birthplace, they can still use the proceeds to purchase a new house if they choose. Furthermore, the exogenous home-bias preferences play virtually no role because they are estimated to be small (refer to Section 4.3.1).

[37] Additionally, the analysis in Kaplan (2012) focuses on low-income young agents, who experience particularly high income risk. In this population subgroup, the precautionary saving motive is likely strongest and the access to homeownership is low (see also the related discussion in Rosenzweig and Zhang 2019).

7, where net wealth inequality is measured using the Gini coefficient. It should be noted that, although the model closely mirrors the observed untargeted negative correlation (-0.8 in the model compared to -0.7 in the data), it underestimates the level of wealth inequality by approximately 40%. This discrepancy arises largely because the model omits much of the wealth heterogeneity among homeowners, given that housing prices in the model are constant within locations and education groups. Supporting this, a Gini decomposition exercise reveals that 66% of wealth inequality is driven by variation in wealth among homeowners, which in Spain is largely explained by differences in housing values.[38]

Figure 7: Gini of net wealth,
Untargeted



Note: The figure plots the untargeted relationship across locations between the local Gini of net wealth and the local homeownership rate among not coresidents. The left vertical axis refers to the data, whereas the right axis refers to the model. Data: EFF 2008-2020.

Mechanism A negative relationship between homeownership and wealth inequality has been observed across countries (Kaas, Kocharkov and Preugschat 2019, Kindermann and Kohls 2018). My analysis uncovers that a similar relationship also holds across locations within a single country. In particular, the -0.7 correlation between local homeownership rates and Gini of net wealth that I observe across Spanish locations (Figure 7) is comparable to the -0.89 correlation that Kaas, Kocharkov and Preugschat (2019) find when comparing large European countries (Appendix Figure F20). This literature emphasizes that widespread access to homeownership lifts the wealth of the wealth poor relatively more, and tends to decrease wealth inequality. In Appendix F.3, I exploit the EFF panel dimension and find evidence that this mechanism is also at play across Spanish individuals and locations.

[38]The Gini decomposition exercise follows Mookherjee and Shorrocks (1982) and is performed using EFF 2005-2020 data. The R package `dineq` is used for the estimation. The main sources of variations for the Gini coefficient are wealth inequality within homeowners (66%) and between homeowners and renters (27%). The remaining variation comes from within renters inequality (1.5%) and the residual term (5.5%).

Using the unconditional quantile regression framework developed by Firpo, Fortin and Lemieux (2009), I find that homeownership increases net wealth along the entire distribution and, in relative terms, it does so especially among the wealth poor (Appendix Figure F21). As detailed in Appendix F.3, the regression accounts for household fixed-effects and other observables (including age, income, and education level), so that results are not simply driven by the selection of wealthier households into homeownership. Rather, the effect is identified from the variation in homeownership status occurring within households, keeping all other observables fixed. Buying the main residence is found to increase net wealth by a factor of 6 in the first decile and to more than double it in the second decile. As a comparison, the increase in net wealth for people in the top decile of the distribution is of just 4%. The estimated positive effect of homeownership on wealth accumulation reveals that Spanish households are not willing to fully substitute housing wealth with other types of financial wealth. This may be happening because, as in the model, people derive some direct utility from holding wealth in the form of housing, or because they value homeownership's insurance against rental price volatility.

Moreover, using the Recentered Influence Function (RIF) for the Gini coefficient (Firpo, Fortin and Lemieux 2009), a statistical tool that allows to compute the effect of individual observations on aggregate statistics, I show that wealth inequality tends to reduce when the share of wealth accounting to the bottom 20% increases and that of the top 20% reduces. Appendix Figure F22 reports the results, whereas Appendix F.3 provides details on the estimation procedure. This result, combined with the unconditional quantile effects of homeownership on net wealth, explains why higher homeownership rate is associated with lower wealth inequality: homeownership allows poorer households to accumulate relatively more wealth, which makes the wealth distribution more equal.

Accordingly, by estimating a regression that uses the RIF of the Gini of net wealth as outcome (Firpo, Fortin and Lemieux 2009) and controls for household fixed-effects and other observables, I find that homeownership tends to decrease wealth inequality (Appendix Table F6). Moreover, by interacting the homeownership treatment with the initial level of households' wealth, I find that this result is driven by households that start in the bottom 20% of the wealth distribution. In particular, the estimated negative effect of homeownership on the Gini of net wealth is highest in the bottom 10%, is still negative and significant for household with initial wealth between the first and second decile, and loses significance for richer households (Appendix Table F7).

A similar mechanism is at play in the model. Locations in the model with the highest homeownership rate have on average a 2 percentage points higher share of local net wealth accounting to the bottom 20% of the distribution, and around a 2 pp. lower share accounting to the top 20% (Appendix Figure F23). Therefore, simularly to the data, the higher aggregate homeownership rate reduces wealth inequality to the extent that it contributes to the relative increase in net wealth in the bottom 20% of the distribution.

6 Should Governments Subsidize Homeownership?

Next, I use the model to analyze the welfare implications of housing policies. Mortgage interest deductions are found to increase welfare, whereas rent subsidies for young low-income individuals decrease it. Targeting policies to specific locations reduces the benefits of mortgage deductions but mitigates the welfare costs of rent subsidies. Policies that reduce the transaction costs of buying lead to small welfare gains, while reducing the transaction costs of selling barely affects welfare. Since homeownership allows to insure against rental price volatility, the presence of aggregate shocks in the economy amplifies the welfare effects of policies affecting the homeownership rate.

6.1 Mortgage Interest Deductions

I study the welfare implications of mortgage interest tax deductions, by running in the model a policy of this kind that was in place in Spain until 2013. The policy allowed homeowners to annually deduct 1,300 euros from their labor income taxes while they were repaying their mortgage. I simulate three different versions of the policy. First, I simulate a version mirroring the actual policy in Spain, which applies across all locations. Then, I only target residents in the three highest-price urban locations. Finally, I simulate a version where only the three lowest-price rural locations residents are targeted.[39] I find that the untargeted version of the policy increases welfare by 1.64% in consumption-equivalent terms, whereas targeting high-price urban locations increases it by 0.74%. On the other hand, a policy that only targets low-price rural residents barely affects welfare.

To evaluate the effect of mortgage interest deductions, I compare the benchmark equilibrium with counterfactual equilibria where different policies are implemented. In the counterfactuals, I allow housing prices, wages, and income to adjust in response to policies. Housing prices and wages respond to changes in the spatial allocation of labor that, in turn, affect local housing demand and labor supply. The level of taxes ς_0 is also modified, keeping the tax progressivity parameter ς_1 fixed, to cover the additional revenue required by the policy. This takes into account changes in individuals' taxable income due to migration and general equilibrium effects. Comparisons with the counterfactual equilibrium are only made after prices, wages, and taxes have already converged, ignoring the transitional dynamics between the stochastic steady-states.

Figures 8 and 9 plot the equilibrium outcomes of policies as colored dots, which are compared with the central black dot representing the benchmark equilibrium. In general, policies that increase homeownership tend to decrease internal migration, due to the transaction costs associated with selling the house (Figure 8a). Policy-induced increases in homeownership are also typically linked to lower levels of wealth inequality (Figure 8b), since homeownership tends to lift net wealth of individuals at the bottom of the distribution relatively more than

[39]The three high-price urban locations belong to the Northeast, East, and Madrid regions. The low-price rural locations belong to the Northwest, Center, and South regions.

the rest (see Section 5.3). Finally, policies that increase the urban population share are linked to lower spatial income dispersion (Figure 9a) and higher spatial price dispersion (Figure 9b) across locations. This happens because the influx of new migrants to high-price, high-wage urban locations tends to further increase urban prices and to reduce urban wages, due to the shifts in local housing demand and labor supply.

(a) Homeownership and Migration **(b)** Homeownership and Wealth Inequality

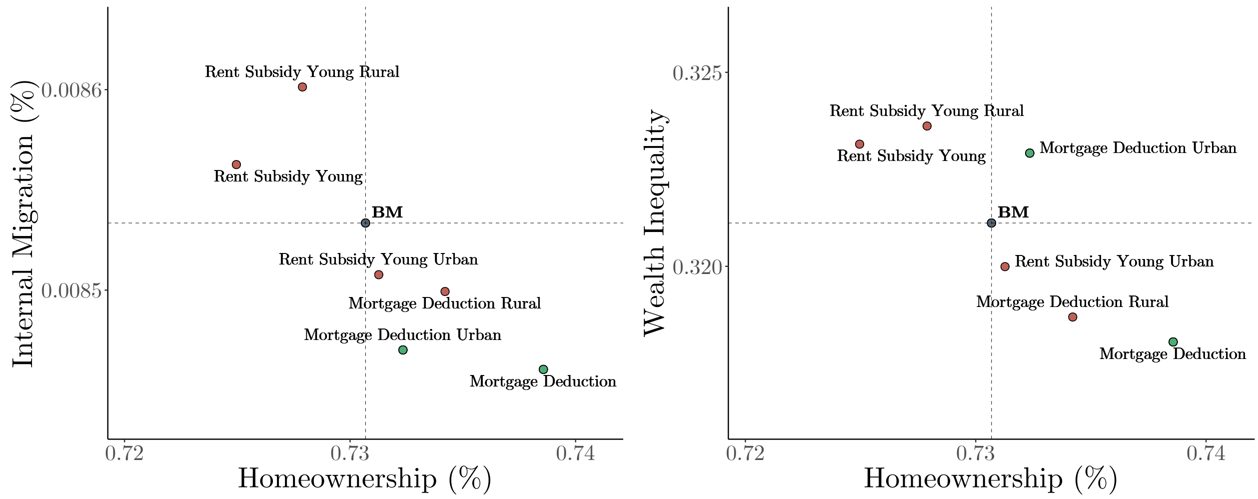


Figure 8: The figures plot the equilibrium outcomes of policies. Green dots correspond to welfare-improving policies, whereas red dots are used for policies that decrease welfare. The black *BM* dot represents the benchmark economy. Wealth inequality is measured as the Gini coefficient of net wealth.

Welfare is measured as the percentage change in consumption that the average newborn agent would require, or give up, in order to be indifferent between the counterfactual and the benchmark equilibria. More details on how welfare is computed are given in Appendix G.1. As shown in Table 5, untargeted mortgage interests deduction policies increase welfare by 1.64% and are supported by a majority (83.1%), meaning that the ex-ante lifetime utility of most newborn agents is higher in the counterfactual than in the benchmark.

Mortgage interest deductions increase homeownership by 0.2 percentage points among non-coresidents and by 0.8 pp. at the individual level (the share of coresidents declines from 20.5% to 19.8%). As a result of the increase in the homeownership rate, both internal migration and wealth inequality decline (by -0.86% and -0.95%, respectively). The lower urban share following the introduction of the policy also translates into lower urban prices and spatial price dispersion. However, it leaves spatial income dispersion across location (variance of log mean income by location) roughly unchanged. Overall, the average newborn is better after mortgage deductions are introduced, despite the 2 percentage points increase in the average tax rate needed to finance the policy.

In the counterfactual exercise in which mortgage interest deductions are only targeted to the three highest-price urban locations, welfare increases by 0.74%. This policy, however, is not supported by a majority, as only 37.9% of agents are better off after its introduction. Despite their pro-home buying nature, targeted mortgage interest deductions increase the homeownership rate only marginally at the individual level (0.2 pp.) and leave it roughly constant among non-coresidents. This happens because the policy leads to an increase in

Table 5: Welfare effects and share supporting housing policies.

	With Aggregate Shocks		Without Aggregate Shocks	
	Welfare Effect	Share Support	Welfare Effect	Share Support
Mortgage Interest Deduction				
All locations	0.0164	0.831	0.0027	0.684
High-price urban	0.0074	0.379	0.0034	0.418
Low-price rural	0.0009	0.324	-0.0078	0.337
Rent Subsidy to Young Low-Income				
All locations	-0.0134	0.149	-0.0064	0.195
High-price urban	-0.0055	0.086	-0.0030	0.120
Low-price rural	-0.0009	0.093	-0.0040	0.115

both urban prices and the urban population share. Despite the introduction of the mortgage interest deductions, many agents living in the targeted urban areas are unable to become homeowners due to higher housing prices. This increase in prices stems from the pressure on housing demand coming from the influx of migrants and the higher share of local residents who stop coresiding. Since the homeownership rate remains roughly constant and housing prices become more spatially dispersed, wealth inequality increases as a result of the policy.

Finally, a mortgage interest deduction policy that only targets low-price rural residents barely affects welfare (Table 5). Similarly to its untargeted version, this policy increases homeownership and reduces internal migration (Figure 8a). However, it also increases spatial income dispersion by 0.4% (Figure 9a), as the population in urban productive cities declines (-0.5%). The resulting net effect of the policy on welfare is barely positive (0.1%), and the support among newborn agents is relatively low (32.4%).

The tax increases needed to finance the policy offset most (64%) of the partial equilibrium welfare benefits. When additionally allowing prices and wages to adjust in equilibrium, welfare gains further decrease for both the untargeted policy and the one targeted to urban locations. The losses are especially high in the targeted policy and for low-income agents, due to the increase in housing prices. Agents in this income group, indeed, are less likely to be homeowners and are more exposed to price fluctuations. Conversely, wage and price general equilibrium effects tend to benefit low-income agents if the policy is targeted to rural location. However, these effects, which are driven by the decrease in housing prices, do not fully offset the partial equilibrium welfare losses.

Finally, a comparison of the welfare effects of the policy in a version of the model that is re-calibrated without aggregate shocks reveals that the existence of aggregate uncertainty makes mortgage deduction policies more valuable to agents (see Table 5). This is due to the pro-homeownership nature of the policy and to the fact that homeownership provides insurance against rental price volatility, which is valued by risk-averse agents. Absent aggregate uncertainty, welfare gains are around six times lower for the untargeted version of the policy and two times lower for the version that only targets high-price urban locations. Moreover, a mortgage interest deduction policy targeted at low-price rural locations would lead to welfare losses (-0.78%) in a model without aggregate shocks, whereas it leaves welfare mostly unaffected in the presence of aggregate uncertainty.

Notice that it is not ex-ante clear whether homeownership increases or decreases agents' insurance against adverse shocks. Indeed, homeowners are less likely to self-insure against negative income shocks by migrating to less affected locations, due to the transaction costs of selling their house. However, the counterfactual exercises reveal that the net insurance value of homeownership is positive.

(a) Urban Share and Income Dispersion **(b)** Urban Share and Price Dispersion

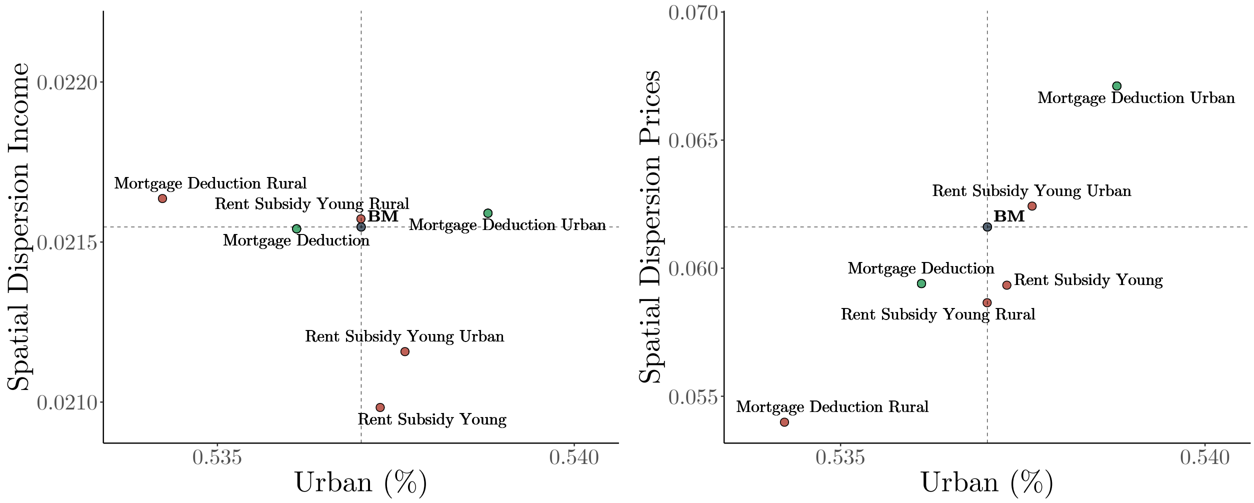


Figure 9: The figures plot the equilibrium outcomes of policies. Green dots correspond to welfare-improving policies, whereas red dots are used for policies that decrease welfare. The black *BM* dot represents the benchmark economy. Spatial dispersion is measured using variance of the log of the variable.

6.2 Subsidy to Young Low-Income Renters

I also study the welfare implications of a rent subsidy policy introduced in 2018 for young, low-income individuals. The eligibility criteria include a maximum gross annual income threshold of 19,500 euros and an age limit of 35 years.[40] The subsidy provides an annual benefit of 3,000 euros, a significant sum for the low-income recipients of the policy given that it amounts to around 30% of the average rent. As in the mortgage interest deductions counterfactual, I allow prices, wages, and taxes to adjust in general equilibrium in response to the subsidy, and study versions of the policy which are both untargeted and targeted by location.

As shown in Figure 8a, the untargeted policy decreases homeownership (-1.7%) and increases internal migration by 0.3% relative to the benchmark. It also attracts workers to urban locations (0.1%), thus decreasing the spatial dispersion of income by -2.6% (Figure 9a). Despite the redistributive nature of the policy, rent subsidies to low-income agents are found to increase wealth inequality by 0.6% (Figure 8b). This occurs because lower-income agents are more likely to rent to receive the subsidy, which reduces their incentives to accumulate housing wealth through homeownership.

[40]The maximum permissible annual rent is set by Spanish law at 7,200 euros, or 10,800 euros in the most expensive ares. Rents in all model locations are lower than those thresholds.

Table 5 indicates that the rent subsidy diminishes welfare by 1.34% and is supported by only 14.9% of agents. Due to restrictiveness of the eligibility criteria, the benefits of rent subsidies are concentrated among few individuals. On the other hand, all employed workers need to bear the tax increase required finance the policy, although small (0.4 percentage points higher average tax rate). The take-up of the policy is low because few people find it optimal to rent, even with the subsidy: the majority of agents in the model are natives, who are better off coresiding, rather than renting, until they have enough funds to pay the down-payment and buy a house.

Finally, Table 5 reveals that targeting rent subsidies to high-price urban or low-price rural locations marginally mitigates the negative welfare effect of the policy. Support for the targeted policies, however, is even lower than for the untargeted rent subsidies, as only around 9% of newborn agents are better off with them than without.

6.3 Transaction Costs

I also simulate policies that target both the transaction costs of buying (ϕ_b) and selling (ϕ_s). Unlike mortgage interest deductions and rent subsidies, these policies have not been implemented in Spain. Reducing the transaction cost of buying has the same effect as subsidizing the mortgage down-payment, whereas reducing the transaction cost of selling increases migration among homeowners. Therefore, ϕ_b and ϕ_s are two key frictions that limit homeownership and internal migration decisions. As a result, policies that target transaction costs can have a substantial influence on the model outcomes and welfare.

I reduce the transaction costs of buying from 0.067 to 0.032, taking away the fraction of ϕ_b coming from real estate transfer taxes (RETT).[41] As usual, the revenue needed to finance the policy is covered by increasing the average income tax rate.

Following the introduction of the policy, there is a large increase in the share of individuals who buy (4 percentage points). Nonetheless, internal migration increases by 16%, mainly due to a substantial decrease in coresidence (–2.5 pp., or -12%). Subsidizing the down-payment makes coresiding less attractive and increases migration, as less people need to stay in their birthplace in order to live with their parents and overcome the financial friction in the mortgage market.

Nonetheless, the welfare gains from this policy are small (0.3%). Although the partial equilibrium gains are substantial (2.5%), most of these benefits disappear as taxes adjust (1.4%) and due to the large increase in house prices following the rise in the homeownership rate.[42]

Finally, I simulate a policy that reduces the transaction costs of selling by removing RETT.

[41] RETT are assumed to be 7% in Spain (Kaas et al. 2021), and are split between sellers and buyers.

[42] I have also simulated a larger policy that uses the same revenue that is needed to finance the untargeted version of mortgage interest deductions (increase in average taxes by 2 percentage points). Such a policy reduces the housing value share that is devoted to the down-payment and transaction costs from 26.7% to 17.7%. The policy has large partial equilibrium gains (5.2%). However, most of these gains disappear after taxes are increased, while keeping prices and wages fixed (2.2%). When allowing for the full general equilibrium effects, house prices increase by a whopping 7.4%, and the policy is found to *decrease* welfare by 1.4%.

As a result, ϕ_s decreases from 0.067 to 0.032. Following the introduction of the policy, the share of individuals who become homeowners increases by 0.6 percentage points, coresidence decreases (–0.4 pp.), and internal migration increases by 8%. Nonetheless, the policy leads to virtually zero general equilibrium gains. The benefits are mainly concentrated among homeowners selling their houses, while taxes increase for everybody.

7 Conclusion

Should governments promote homeownership? Although housing policies may overcome financial frictions and, by increasing homeownership, insure against rent volatility, they can also reduce internal migration to productive locations. To address this question, I build a spatial equilibrium model populated by finitely-lived agents making dynamic coresidence, homeownership, migration, and saving decisions.

Homeownership provides utility and insurance against aggregate rental price risk, but reduces migration as homeowners are forced to sell and pay the associated transaction costs before moving. The relationship between homeownership and migration, however, is not one-sided, as migration decisions themselves can impact future homeownership prospects. Non-migrant workers can live with their parents for a time, and coresidence can allow them to save and buy a house earlier than migrants. Additionally, by remaining in or returning to their birthplace, they have easier access to housing bequests.

I develop a new strategy to solve dynamic spatial models with aggregate uncertainty by modelling agents' expectations about local endogenous prices and wages with lower-rank factors. The model is calibrated for Spain, a country with high homeownership and low internal migration despite significant spatial differences in income and unemployment risk. I use quasi-experimental evidence from recent place-based policies that subsidized homeownership to validate the model.

Mortgage interest deductions increase welfare by 1.64%, have majority support, and reduce wealth inequality. However, they decrease internal migration and the share of people living in productive locations. Conversely, rent subsidies to young low-income individuals decrease welfare by 1.34%, despite increasing internal migration and the urban population. Targeting policies to specific locations reduces the benefits of mortgage deductions but mitigates the welfare costs of rent subsidies. Policies that reduce the transaction costs of buying lead to small welfare gains, while reducing the transaction costs of selling does not affect welfare. Finally, the presence of aggregate shocks increases the welfare benefits of pro-homeownership policies.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4, 1043–1171. Elsevier. 24

Artuç, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American economic review*, 100(3): 1008–1045. 4

Bada, Oualid, and Dominik Liebl. 2014. "The R Package phtt: Panel Data Analysis with Heterogeneous Time Trends." *Journal of Statistical Software*, 59(6): 1–34. 78

Bai, Jushan. 2009. "Panel Data Models With Interactive Fixed Effects." *Econometrica*, 77(4): 1229–1279. 1, 4, 26, 27, 78

Bai, Jushan, and Serena Ng. 2002. "Determining the number of factors in approximate factor models." *Econometrica*, 70(1): 191–221. 78

Bilal, Adrien. 2023. "Solving Heterogeneous Agent Models With the Master Equation." NBER Working Paper 31103. 4, 5

Bilal, Adrien, and Esteban Rossi-Hansberg. 2021. "Location as an Asset." *Econometrica*, 89(5): 2459–2495. 5

Bilal, Adrien, and Esteban Rossi-Hansberg. 2023. "Anticipating Climate Change Across the United States." NBER Working Paper 31323. 4

Budí-Ors, Tomás, and Josep Pijoan-Mas. 2022. "Macroeconomic Development, Rural Exodus, and Uneven Industrialization." Working Paper. 58

Bénabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70(2): 481–517. 23

Caldera, Aida, and Åsa Johansson. 2013. "The Price Responsiveness of Housing Supply in OECD Countries." *Journal of Housing Economics*, 22(3): 231–249. 24

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica*, 87(3): 741–835. 4

Cardaliaguet, Pierre, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. 2019. *The Master Equation and the Convergence Problem in Mean Field Games.* Princeton University Press. 4

Carroll, Christopher D. 2006. "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems." *Economics letters*, 91(3): 312–320. 82

Davis, Morris A, and Stijn Van Nieuwerburgh. 2015. "Housing, Finance, and the Macroeconomy." In *Handbook of Regional and Urban Economics*. Vol. 5, 753–811. Elsevier. 4

De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." *The Review of Economic Studies*, 84(1): 106–142. 5, 53, 54, 56, 63

Den Haan, Wouter J. 2010. "Assessing the Accuracy of the Aggregate Law of Motion in Models With Heterogeneous Agents." *Journal of Economic Dynamics and Control*, 34(1): 79–99. 77

Desmet, Klaus, and Esteban Rossi-Hansberg. 2014. "Spatial Development." *American Economic Review*, 104(4): 1211–43. 4

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg. 2018. "The Geography of Development." *Journal of Political Economy*, 126(3): 903–983. 4

Díaz, Antonia, Álvaro Jáñez, and Felix Wellschmied. 2023. "Geographic Mobility Over the Life-Cycle." Working Paper. 5

DiPasquale, Denise, and Edward L Glaeser. 1999. "Incentives and Social Capital: Are Homeowners Better Citizens?" *Journal of Urban Economics*, 45(2): 354–384. 1

Disney, Richard, John Gathergood, Stephen Machin, and Matteo Sandi. 2023. "Does Homeownership Reduce Crime? A Radical Housing Reform from the UK." *Economic Journal*, Forthcoming. 1

Eeckhout, Jan, and Nezih Guner. 2017. "Optimal Spatial Taxation: Are Big Cities Too Small?" Working Paper. 4

Engelhardt, Gary V. 1996. "Consumption, Down Payments, and Liquidity Constraints." *Journal of Money, Credit and Banking*, 28(2): 255–271. 1

Esteve, Albert, and David S Reher. 2021. "Rising Global Levels of Intergenerational Coresidence Among Young Adults." *Population and Development Review*, 47(3): 691–717. 5

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. 2019. "State Taxes and Spatial Misallocation." *The Review of Economic Studies*, 86(1): 333–376. 4

Fernández-Villaverde, Jesús, Samuel Hurtado, and Galo Nuño. 2023. "Financial Frictions and the Wealth Distribution." *Econometrica*, 91(3): 869–901. 4

Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica*, 77(3): 953–973. 40, 96, 97, 98

Floetotto, Max, Michael Kirker, and Johannes Stroebel. 2016. "Government Intervention in the Housing Market: Who Wins, Who Loses?" *Journal of Monetary Economics*, 80: 106–123. 4

Forte-Campos, Victor, Enrique Moral-Benito, and Javier Quintana. 2021. "A Cost of Living Index for Spanish Cities." *Banco de España Article*, 29: 21. 14

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the US Declined?" *Journal of Urban Economics*, 102: 76–90. 4

García-Miralles, Esteban, Nezih Guner, and Roberto Ramos. 2019. "The Spanish Personal Income Tax: Facts and Parametric Estimates." *SERIEs*, 10: 439–477. 24

Giannone, Elisa, Qi Li, Nuno Paixao, and Xinle Pang. 2023. "Unpacking Moving." Unpublished. 4

Greaney, Brian. 2023. "Homeownership and the Distributional Effects of Uneven Regional Growth." SSRN 4406024. 4

Grevenbrock, Nils, Alexander Ludwig, and Nawid Siassi. 2023. "Homeownership rates, housing policies, and co-residence decisions." *Macroeconomic Dynamics*, 1–24. 5

Guner, Nezih, Ezgi Kaya, and Virginia Sánchez-Marcos. 2023. "Labor Market Institutions and Fertility." Working Paper. 23

Guner, Nezih, Martin Lopez-Daneri, and Gustavo Ventura. 2016. "Heterogeneity and Government Revenues: Higher Taxes at the Top?" *Journal of Monetary Economics*, 80: 69–85. 23

Hirsch, Boris, Elke J Jahn, Alan Manning, and Michael Oberfichtner. 2022. "The Urban Wage Premium in Imperfect Labor Markets." *Journal of Human Resources*, 57(S): S111–S136. 22

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics*, 11(2): 1–39. 4

Kaas, Leo, Georgi Kocharkov, and Edgar Preugschat. 2019. "Wealth Inequality and Homeownership in Europe." *Annals of Economics and Statistics*, 136: 27–54. 1, 5, 39, 95, 96, 97

Kaas, Leo, Georgi Kocharkov, Edgar Preugschat, and Nawid Siassi. 2021. "Low Homeownership in Germany—A Quantitative Exploration." *Journal of the European Economic Association*, 19(1): 128–164. 4, 14, 24, 32, 37, 45

Kaplan, Greg. 2012. "Moving Back Home: Insurance Against Labor Market Risk." *Journal of Political Economy*, 120(3): 446–512. 5, 14, 38

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2020. "The Housing Boom and Bust: Model Meets Evidence." *Journal of Political Economy*, 128(9): 3285–3345. 4, 76

Kennan, John, and James R Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica*, 79(1): 211–251. 5, 14

Kindermann, Fabian, and Sebastian Kohls. 2018. "Rental Markets and Wealth Inequality in the Euro-Area." Unpublished. 5, 39

Kleinman, Benny, Ernest Liu, and Stephen J Redding. 2023. "Dynamic Spatial General Equilibrium." *Econometrica*, 91(2): 385–424. 4

Koşar, Gizem, Tyler Ransom, and Wilbert Van der Klaauw. 2022. "Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities." *Journal of Econometrics*, 231(1): 123–147. 30

Krusell, Per, and Anthony A Smith, Jr. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106(5): 867–896. 1, 4, 27, 28, 29, 76, 77

Luccioletti, Claudio. 2022. "Labor Market Power Across Cities." Working Paper. 22

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*. ed. by P. Zarembka. Academic Press. 16

Molloy, Raven, Christopher L Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic perspectives*, 25(3): 173–196. 30

Mookherjee, Dilip, and Anthony Shorrocks. 1982. "A Decomposition Analysis of The Trend in UK Income Inequality." *The Economic Journal*, 92(368): 886–902. 39

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics*, 5(1): 65–103. 14

OECD. 2021. *Brick by Brick: Building Better Housing Policies.* 1

Oswald, Florian. 2019. "The Effect of Homeownership on the Option Value of Regional Migration." *Quantitative Economics*, 10(4): 1453–1493. 4, 14, 18

Paz-Pardo, Gonzalo. 2023. "Homeownership and Portfolio Choice Over the Generations." *American Economic Journal: Macroeconomics*, Forthcoming. 1, 5, 14

Piazzesi, Monika, and Martin Schneider. 2016. "Housing and Macroeconomics." *Handbook of Macroeconomics*, 2: 1547–1640. 4

Rosenzweig, Mark, and Junsen Zhang. 2019. "Housing Prices, Inter-Generational Co-Residence, and "Excess" Savings by the Young: Evidence Using Chinese Data." NBER Working Paper 26209. 5, 6, 11, 15, 38, 60

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 999–1033. 16

Sinai, Todd, and Nicholas S Souleles. 2005. "Owner-Occupied Housing as a Hedge Against Rent Risk." *The Quarterly Journal of Economics*, 120(2): 763–789. 4

Sommer, Kamila, and Paul Sullivan. 2018. "Implications of Us Tax Policy for House Prices, Rents, and Homeownership." *American Economic Review*, 108(2): 241–274. 4

Srinivas, Vishnu. 2019. "Explaining the Increase in Young Adults Living With Parents." *Journal of Economic Issues*, 53(4): 1017–1028. 5

Storesletten, Kjetil, Christopher I Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing Over the Life Cycle." *Journal of Monetary Economics*, 51(3): 609–633. 22, 83

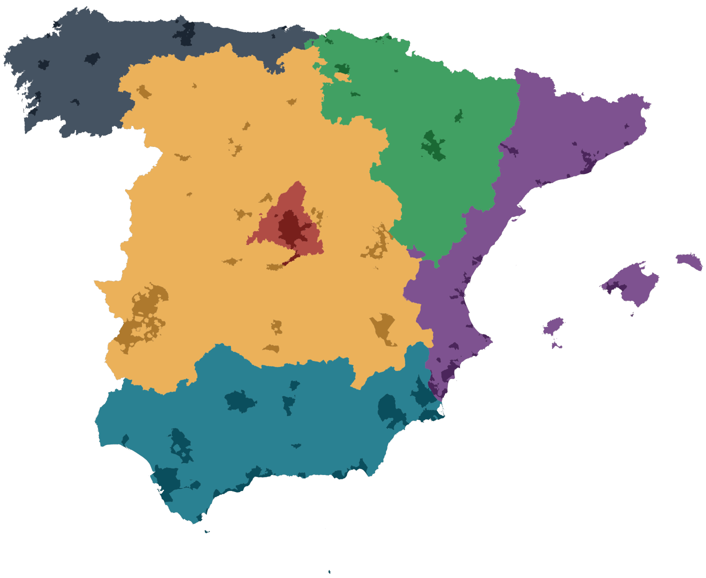
Zerecero, Miguel. 2021. "The Birthplace Premium." Unpublished. 5, 14, 30

Appendix

This Appendix is organized as follows. Section A contains additional tables and figures referenced in the text. Section B provides details about the data. Section C presents further details on the model. Section D contains additional information on the equilibrium. Section E provides details on the estimation strategy. Section F contains details on the validation. Finally, Section G provides additional information on the policy counterfactuals.

A Additional Figures and Tables

Figure A1: Locations: Combination of Regions and Urban-Rural Municipalities. [Back]



Note: This map shows the locations used in the data and model. Locations are defined as combinations of NUTS-1 regions in peninsular Spain and groups of municipalities classified as either urban or rural. The regions are represented by distinct colors, with darker shades used to indicate urban municipalities within each region. Peninsular Spain comprises six distinct NUTS-1 regions, resulting in a total of 12 locations when combined with urban and rural areas. The Canary Islands, which belong to the Atlantic Ocean northwest of Africa, and Ceuta and Melilla, Spanish enclaves lying in mainland Africa, are excluded. Urban municipalities are defined by the Spanish Ministry of Transport and Mobility (*áreas urbanas*) and have at least 40,000 residents.

Definition of Locations Locations in the data and model are combinations of NUTS-1 regions and urban and rural areas within each region. The NUTS classification is a hierarchical system developed by the European Union to divide its territory for statistical purposes. NUTS-1 regions refer to major socio-economic regions, with an average population size between 3 and 7 millions. There are six distinct NUTS-1 regions in peninsular Spain. I use official definitions of urban areas constructed by Spain's Ministry of Housing in 2008. Urban areas group municipalities linked by commuting and employment patterns. In order to identify urban areas in the MCVL data, which does not distinguish between cities with population below 40,000, I follow De la Roca and Puga (2017) and exclude from urban areas the municipalities with a population lower than 40k. The resulting 168 urban municipalities cover 80% of the urban population identified by the Spain's Ministry of Housing and 55% of the Spanish population. The remaining 7,968 municipalities are classified as rural.

Figure A1 shows the locations used in the data and model. The six NUTS-1 regions are represented by distinct colors, with darker shades used to indicate urban municipalities within each region. Locations group together different urban areas and rural areas within the same regions, and are thus 12 in total. The same urban location may encompass multiple non-adjacent urban areas. While it would be ideal to treat each urban area as a separate location, doing so would be computationally too costly, as the high degree of heterogeneity in the model translates into a large state space. For a discussion of the computational challenges associated with increasing the number of locations, see footnote 23. Locations within NUTS-1 areas can thus be interpreted as "representative" rural and urban locations in those regions.[43]

While it's essential to maintain a limited number of locations, differentiating between rural and urban areas is crucial for two main reasons. First, there is a marked difference in housing and labor markets in urban versus rural areas within the same region, leading to large differences in wages and prices (Figures D8, D9, and D10). Ignoring this distinction in the model, for instance by treating each region as a single location, would omit key sources of variations in the data.[44] Second, the place-based policy used to validate the model is targeted to small rural locations. Therefore, the inclusion of rural locations is necessary to accurately simulate the policy in the model.

Lifecycle Housing Tenure and Migration Figure A2 plots the lifecycle evolution of housing tenure choices (panel a), i.e. homeownership, coresidence, and renting, and of internal migration rates (panel b). The age fixed-effects predicting internal migration (panel b) are estimated using the MCVL for precision. Nonetheless, the convex and decreasing lifecycle pattern is similar when estimated with the EU-SILC.

Migration Regressions Table A1 reports the estimated coefficients of a regression that takes the form:

$$\text{Migrate}_{it} = \alpha_i + \alpha_r + \alpha_{rt} + \tau \text{Homeowner}_{it} + \alpha X_{it} + \epsilon_{it}$$

The outcome variable is the migration event, which occurs when the following year's location of residence is different from the current one. The specification includes region (NUTS-1) and region-year fixed effects α_r and α_{rt} (column 1 of Table A1), and, in column (2), additionally includes household fixed-effects α_i. Additional controls X_{it} include age, age squared, income level, and indicators for college-educated, married, parent, employed, and gender. The regressions are estimated with the EU-SILC 2004-2019 panel.

[43]Notice for example that, although Alicante and Barcelona belong to the same urban location, the model's wages and housing prices for locations are determined as population weighted averages in the data (perfectly matched in the benchmark equilibrium). Since Barcelona has a population that is more than 5 times higher than Alicante's, its relative weight in determining the location's final wages and price is substantially higher.

[44]The amount of variation in wages captured by the 12 locations alone is roughly equal to the amount captured by the 49 provinces in peninsular Spain (5%), which indicates the importance of differentiating between rural vs. urban areas.

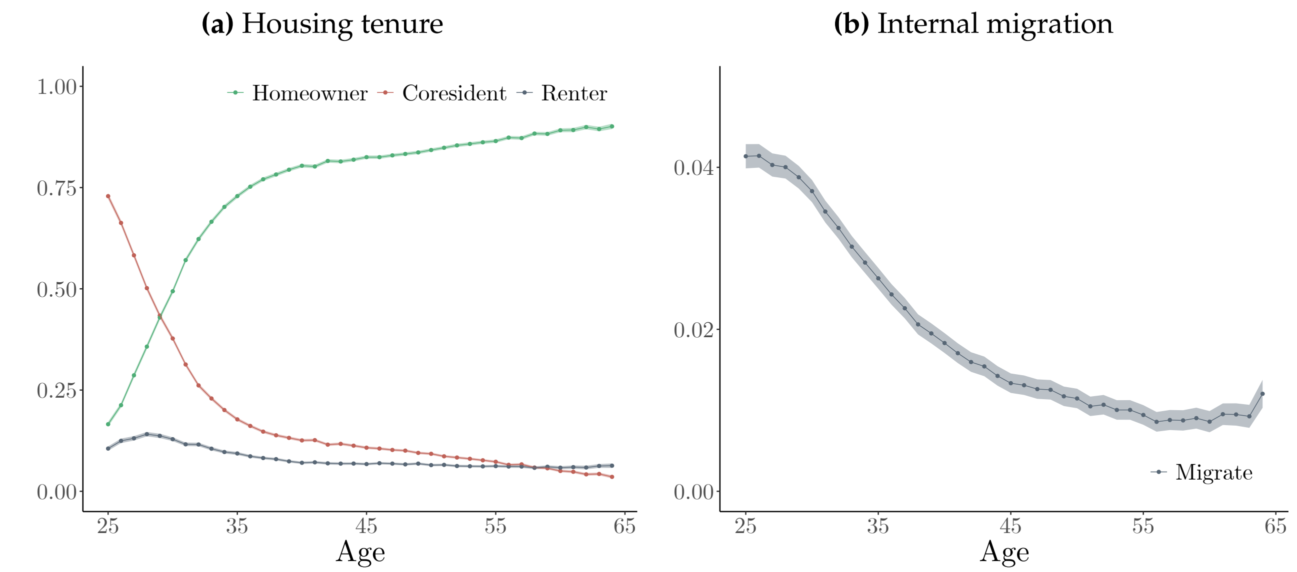


Figure A2: The figures plot age fixed-effects with 95% confidence intervals (heteroskedasticity-robust standard errors). Year fixed-effects are also included in panel **b**. Data: Census 2011 (panel **a**), MCVL 2005-2019 (panel **b**). [Back]

In the specification of column (1), I find that homeowners are less likely to migrate than renters after controlling for age and other observables. It should be noted that the only significant predictors of migration behavior include homeownership status, age, and college education. Each of these characteristics is specifically modeled as migration cost shifters within the theoretical framework, as detailed in Section 3.2.

When I additionally include household fixed-effects (column 2), the coefficients associated with homeownership status and age keep their significance. The effect of college education becomes insignificant, likely due to the limited within-household variation in educational attainment that can be exploited in a 4-year panel. Looking at the magnitude of the effects, homeownership is the single most important predictor of migration rates: the probability of moving reduces by around 1.92 percentage points for household heads who become homeowners.

Migration Event Studies and Income In Figure A3 I plot three event studies centered around migration events. Specifically, I plot the τ_j coefficients in a a series of regressions that take the form:

$$\log y_{it} = \alpha_t + \sum_{\substack{j=-4 \\ j\neq -1}}^{4} \tau_j \mathbb{1}\left\{(t-t_0)=j\right\} + \tau_{-5}\mathbb{1}\left\{(t-t_0)\leq -5\right\} + \tau_5 \mathbb{1}\left\{(t-t_0)\geq 5\right\} + \alpha X_{it} + \epsilon_{it},$$

where the outcome y_{it} is annual labor income, α_t are year fixed-effects, τ_j are the event study fixed effects relative to the migration event occurring at t_0, and τ_{-5} and τ_5 are the average effects before and after five years from the event, which are included to avoid collinearity between the year fixed-effects and the τ_j yearly coefficients. Finally, X_{it} contains additional controls, which include indicators for gender, college-educated, sector (3-digits NACE), permanent contract, part-time contract, public employee, and occupational skills (five groups from low-skilled to very high-skilled, as in De la Roca and Puga (2017)). The regressions are

Table A1: Migration occurs when the following year's location of residence is different from the current one. Region fixed-effects refer to NUTS-1 regions. Standard errors are clustered at the location level, *p<0.1, **p<0.05, ***p<0.01. Data: EU-SILC 2004-2019. [Back]

	Migrate	
	(1)	(2)
Homeowner	-0.0045**	-0.0192**
	(0.0019)	(0.0089)
Age	-0.0012**	-0.0049*
	(0.0005)	(0.0023)
Age^2	0.0000*	0.0000*
	(0.0000)	(0.0000)
College	0.0024*	-0.0008
	(0.0013)	(0.0069)
Married	-0.0007	-0.0048
	(0.0012)	(0.0076)
Parent	-0.0023	-0.0039
	(0.0016)	(0.0065)
Employed	0.0006	-0.0003
	(0.0010)	(0.0031)
Income	0.0003	-0.0006
	(0.0002)	(0.0004)
Male	-0.0012	-0.0049
	(0.0020)	(0.0077)
Household FE		✓
Region FE	✓	✓
Year-Region FE	✓	✓
Observations	72,593	72,593
R^2	0.0527	0.3882
Migration Rate	0.0082	0.0082

estimated with MCVL 2005-2019 data.

The event studies are estimated using different sample restrictions and migration events. In the regression plotted in panel A3a, the sample is restricted to migration events from rural locations, and an interaction is added to differentiate between moves to urban vs. rural locations. The same is done for the regression of panel A3b, with the difference that the sample is restricted to migration events that originate from urban locations. Finally, panel A3c plots the event study coefficient estimated from the full sample, but adding an interaction with a college identifier to the τ_j coefficients.

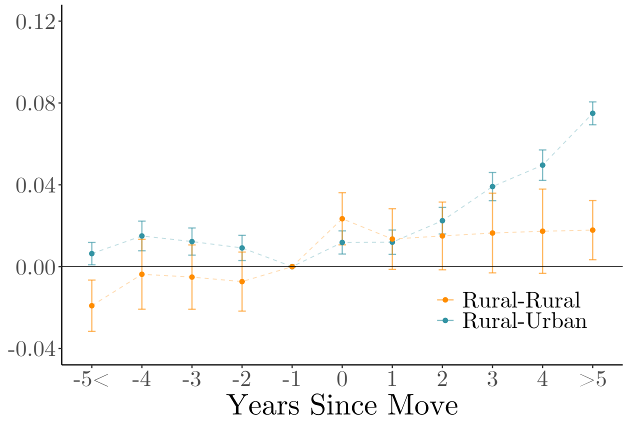
The event studies reveal that internal migrants experience persistent income gains after moving. This is especially true when moving to urban locations (panels A3a and A3b) and for the college-educated workers (panel A3c).

Income and Coresidence Regressions Table A2 reports the estimated coefficients of a regression that takes the form:

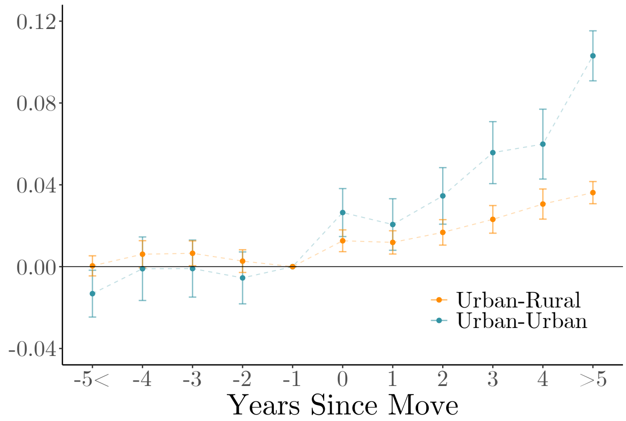
$$y_{it} = \alpha_i + \alpha_d + \alpha_t + \alpha X_{it} + \epsilon_{it}.$$

The outcome y_{it} is either log labor income (columns 1 and 2) or a coresident dummy (columns 3 and 4). The regression includes location and year fixed effects α_d and α_t, and, in columns

(a) **Log income, migration event from urban locations**



(b) **Log income, migration event from rural locations**



(c) **Log income, migration event by education group**

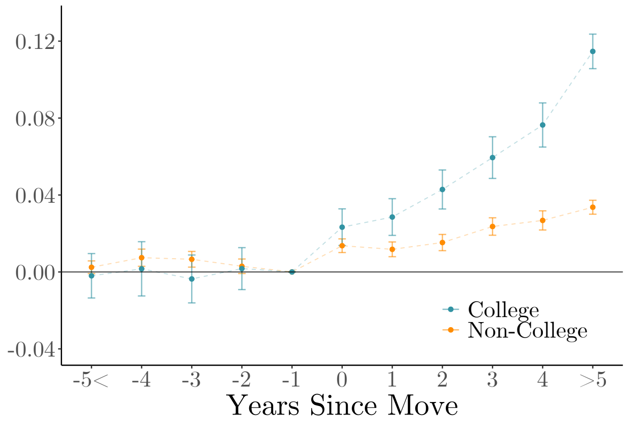


Figure A3: The figures plot the fixed-effects corresponding to the years from the migration event, with 95% confidence intervals (heteroskedasticity-robust standard errors). Other included fixed-effects are: year, gender, college-educated, sector (3-digits NACE), permanent contract, part-time contract, public employee, occupational skills (five groups from low-skilled to very high-skilled, as in De la Roca and Puga (2017)). Data: MCVL 2005-2019. [Back]

(2) and (4), additionally includes individual fixed-effects α_i. Controls X_{it} include age, age squared, college, and indicators for natives (in columns 1 and 2) or log labor income levels (in columns 3 and 4). The regressions are estimated with the MCVL 2005-2019 data.

The first two columns of Table A2 reveal that natives, all else equal, gain less than internal migrants. This earnings gap might arise partly from compensating differentials – for instance, natives might have a home-bias, making them more willing to forego some income to remain in their birthplace. Another contributing factor could be selection, which may play a role if migrants, on average, have higher unobservable ability. Even after accounting for individual fixed-effects (as shown in column 2), the wage premium for migrants remains, although reduced to roughly half, 1.7% as opposed to 3.2%. This suggests that selection on fixed unobservables alone does not explain the entirety of the gap, and that compensating differentials are likely to also play a role.

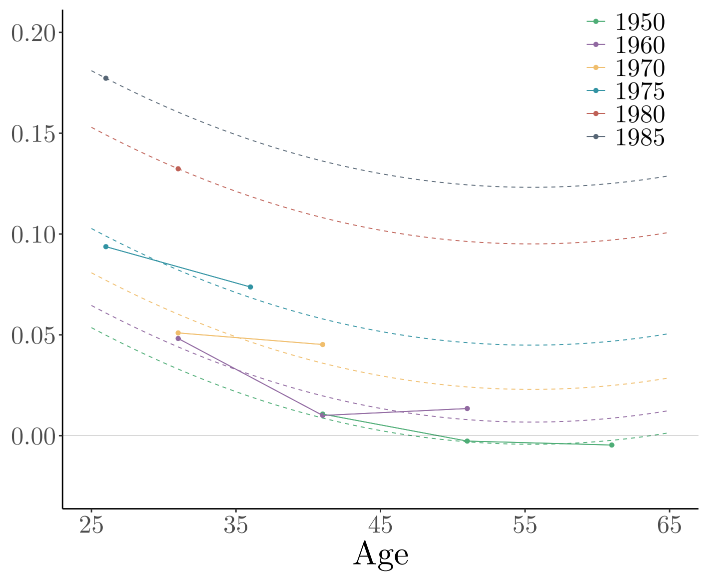
Column (3) of Table A2 shows that coresidents are, all else equal, negatively selected in

Table A2: Standard errors are heteroskedasticity-robust, *p<0.1, **p<0.05, ***p<0.01. Data: MCVL 2005-2019. [Back]

	Log Labor Income		Coresident	
	(1)	(2)	(3)	(4)
Age	0.0282***	-0.1865***	-0.0397***	-0.0143***
	(0.0001)	(0.0021)	(0.0000)	(0.0007)
Age^2	-0.0004***	-0.0004***	0.0007***	0.0007***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
College	0.4552***	0.1080***	-0.0087***	-0.1415***
	(0.0009)	(0.0015)	(0.0006)	(0.0011)
Native	-0.0314***	-0.0169***		
	(0.0007)	(0.0017)		
Log Labor Income			-0.0258***	-0.0042***
			(0.0003)	(0.0003)
Individual FE		✓		✓
Location FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
R^2	0.1250	0.7154	0.1476	0.7117
Observations	5,253,926	5,253,926	5,253,926	5,253,926

terms of labor income. Interestingly, even after accounting for individual fixed-effects, the association between income and coresidence remains negative and significant (column 4). This suggests that the option to live with parents can be used as insurance against negative income shocks occuring during the lifecycle.

Figure A4: Natives-migrants homeownership gap by cohort



Note: The dashed line in the figure depicts, for different cohorts, age polynomial functions estimated in a regression where the dependent variable is the partialled-out homeownership gap by age between natives and migrants. The fixed-effects used to partial-out homeownership are: location, gender, college-educated, married, parent, employed (reference group: male, non-college, single, not parent, employed). Data: Census 1991, 2001, 2011. [Back]

Natives-Migrants Homeownership Gap by Cohort When computing the homeownership gap between natives and migrants (Figure 2a), accounting for cohort effects is crucial. Over the past 60 years, Spain has moved from a dictatorship to a democratic system and completed the structural transformation process. These changes have had profound impacts on both the price-to-income ratio and internal migration rates (Budí-Ors and Pijoan-Mas 2022).

To adjust for cohort effects, I first separately estimate with 1991, 2001, and 2011 Census data three regressions of the form:

$$\text{Homeowner}_{itq} = \alpha_{jnq} + \beta X_{itq} + \epsilon_{itq}.$$

The dependent variable is a homeownership dummy, α_{jnq} is an age-native fixed-effects for cohort (i.e. birth year) q, and X_{itq} contains indicators for location, gender, college-educated, married, parent, and employed.

I then obtain the partialled-out homeownership rates by age and native status, $\widehat{\alpha}_{j0q}$ (migrants at age j) and $\widehat{\alpha}_{j1q}$ (natives at age j), for three different cohorts at each age. Then, after defining the partialled-out homeownership gap between natives and migrants by age and cohort, $\text{Gap}_{jq} = \widehat{\alpha}_{j1q} - \widehat{\alpha}_{j0q}$, I estimate the regression

$$\text{Gap}_{jq} = \alpha_{q0} + \alpha_1 \text{Age}_j + \alpha_2 \text{Age}_j^2 + \epsilon_{jq} \tag{15}$$

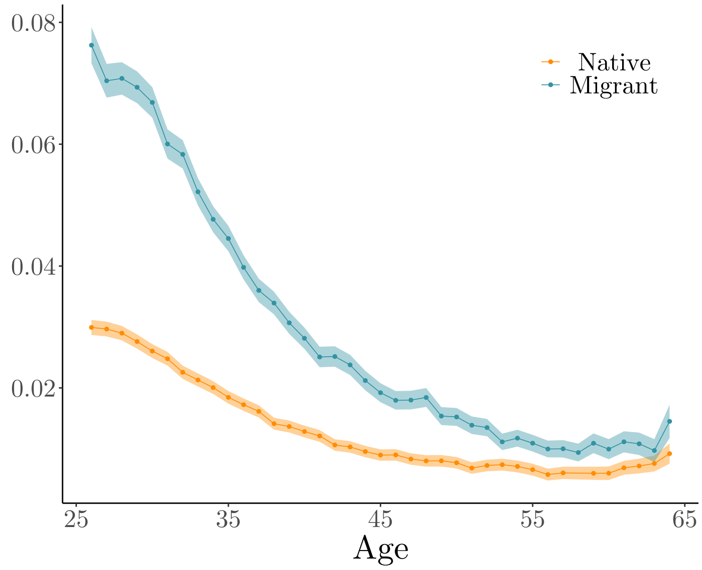
by fitting the homeownership gaps with a polynomial in age and a cohort-specific intercept. The dashed lines in Figure A4 represent the fitted polynomial functions for different cohorts, whereas the dots linked by solid lines depict Gap_{jq}.

As can be seen in Figure A4, the homeownership gap between natives and migrants has been increasing with every new cohort. A naive cross-sectional analysis that ignored cohort effects would conclude that the gap is at around 20 percentage points at age 25 and disappears by age 64. However, when we account for cohort effects, it becomes clear that this gap persists throughout the lifecycle for the most recent cohorts.

The baseline Figure 2a uses Census 2011 data. I don't adjust for cohort effects for ages ranging from 25 to 35; instead, I rely on cross-sectional data to determine the homeownership gap between migrants and natives. It's likely that cohort effects during the initial 10-year age period aren't as pronounced. Conversely, the large changes in the coresidence rates observed among young people, which evolve very differently between natives and migrants before age 35 (see 2c), have a large effect on the homeownership gap due to the selected type of people that stop coresiding each year. The cohort analysis of regression (15) is unable to account for these trends, since the age polynomial function is constant across the lifecycle, which is why I choose not to use those estimates for ages 25-35. However, starting at age 36, I do adjust for cohort effects, applying the homeownership gap derived from regression (15) for the 1975 cohort (who turned 36 in 2011).

Migration Rates Among Natives and Migrants Figure A5 plots age fixed-effects of a regression where the dependent variable is the migration event. Other included fixed-effects are location, gender, college-educated, married, parent, and employed. Migrants are more likely to change residence again in the future, especially at younger ages.

Figure A5: Share migrating



Note: The figure plots age fixed-effects of a regression where the dependent variable is the migration event. Other included fixed-effects are: location, gender, college-educated, married, parent, employed (reference group: male, non-college, single, not parent, employed). Confidence intervals at 95% level with heteroskedasticity-robust standard errors. Data: MCVL 2005-2019. [Back]

Housing Expenditure Among Natives and Migrants Figure A6 plots age fixed-effects of two regressions where the dependent variable is the imputed housing expenditure for natives and migrants. The regressions, that also control for location, gender, college education, and employment status fixed-effects, are estimated using the Household Budget Survey (*Encuesta de Presupuestos Familiares*, or EPF). The limitation of the EPF is that it only contains household-level expenditure data, which cannot be readily attributed to household members. As a result, we must make assumptions about how much coresidents contribute to their parents' housing expenses.

In the left panel (Figure A6a), I make the assumption that coresidents do not contribute to the overall household consumption. Conversely, in the right panel (Figure A6b), I make the opposite extreme assumption: coresidents cover their entire OECD equivalence share of expenses. This share ranges between 11.1% to 16.7%, depending on the number of siblings. The expenses taken into account cover housing, maintenance, and utilities costs for the main residence.

The EPF does not include birthplace information, and thus does not allow to differentiate between natives and migrants directly. To navigate this, I estimate overall expenditure by age and assume that the consumption behavior of natives and migrants is observationally equivalent *within* the group of people who are coresiding and *within* those who are not coresiding. Then, I impose the lifecycle coresidence shares by native status, computed using the

2011 Census, and obtain the plots of Figure A6.

Regardless of the assumptions made in panels A6a and A6b, a substantial expenditure gap exists between natives and migrants. Depending on how much coresidents pay for household expenditure, natives under the age of 35 save, on average, between 2,300 and 3,000 euros more in annual housing costs than their migrant counterparts. This higher saving rate can be attributed to the economies of scale associated with coresidence and to the fact that natives are more likely to coreside than migrants.

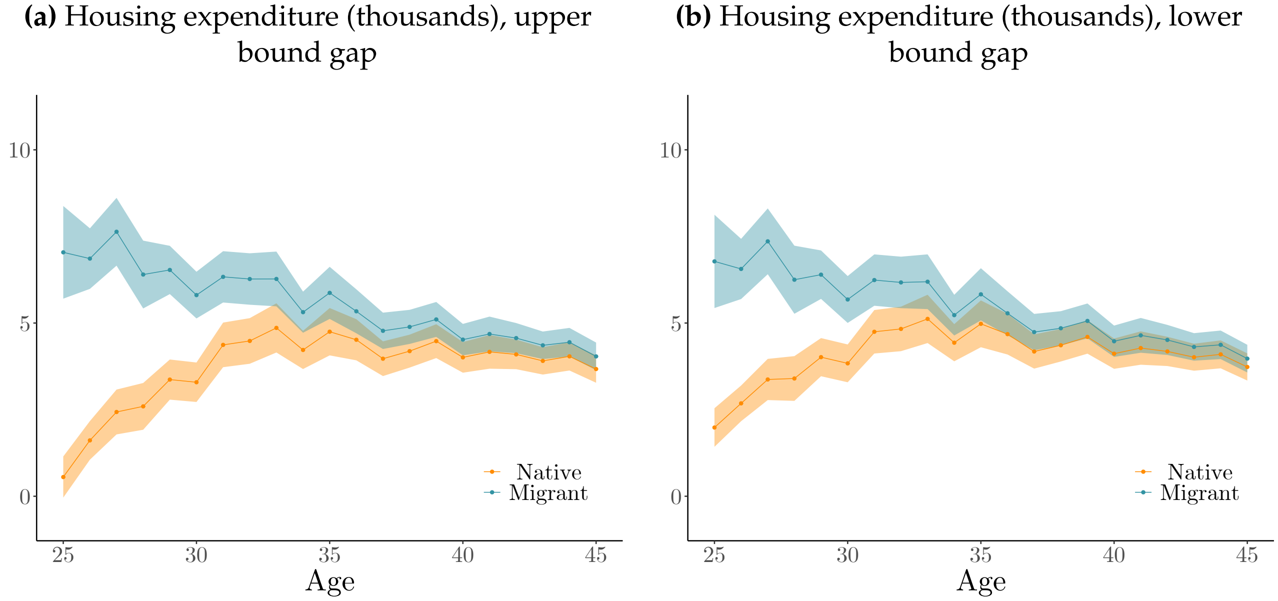


Figure A6: Age fixed-effects are plotted. Other included fixed-effects: location, gender, college, employed. Assumptions: 1. Native coresidents pay either no share of household consumption (left panel) or 11.1-16.7% of it, based on number of siblings and OECD equivalence scale (right panel); 2. When not living with parents, natives and migrants' expenditure is assumed to be observationally equivalent. Categories: housing, maintanance, utilities (no second houses). Confidence intervals at 95% level with heteroskedasticity-robust standard errors. Data: EPF 2016-2019. [Back]

Rosenzweig and Zhang (2019) use unique individual-level expenditure data for China and find that parents' saving rates reduce by 12 percentage points when their children coreside with them. This is consistent with the assumption of panel A6a that children pay no share of housing costs when coresiding, since parents expenditure goes down by an amount that is roughly equal to the children's OECD equivalence share. Accordingly, agents in the model pay no housing rent when living with parents. Nonetheless, it's important to highlight that the results from Rosenzweig and Zhang (2019) are not precisely estimated. This leaves open the possibility that children might be contributing to some part of the household expenditure, as assumed in panel A6b.

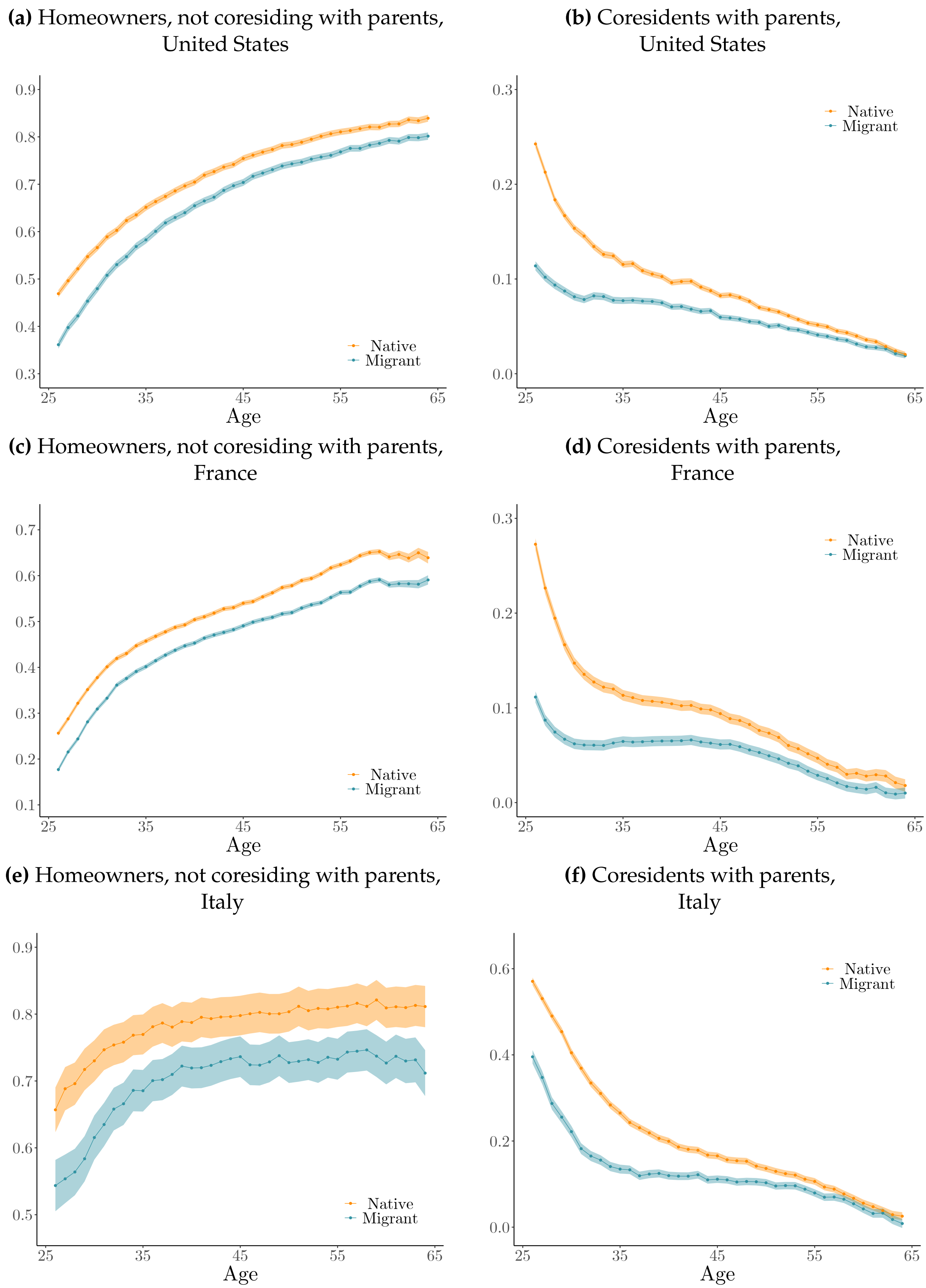


Figure A7: The figures plot age fixed-effects. Other included fixed-effects are: location, gender, college-educated, married, parent, employed, and, in the U.S. sample, race (reference group: male, non-college, single, not parent, employed, white). Location: MSA (United States), NUTS-3/*Département* (France), NUTS-2/*Regione* (Italy); Native: living in State of birth (United States), living in NUTS-3 of birth (France, Italy). Confidence intervals at 95% level with heteroskedasticity-robust standard errors are plotted. Data: ACS 5-years 2007-2011, 2011 Census (France, Italy). [Back]

International Comparison: Natives-Migrants Homeownership and Coresidence Gaps The homeownership gap between observationally equivalent natives and internal migrants is also observed in France, Italy, and the United States, as shown in Figures A7a, A7c, and A7e. The choice of these countries is driven by the availability of ACS and Census data on coresidence, homeownership, and both current and birthplace locations within OECD countries, together with data on the observables used as controls: indicators for location, gender, college-educated, married, parent, employed, and, in the U.S. sample, race.

Current and birthplace locations in each country are defined by using the smallest available geographical units that can be assimilated to cities. The chosen definitions for current locations are Metropolitan Statistical Areas in the United States, NUTS-3 regions (*Départements*) in France and NUTS-2 regions (*Regioni*) in Italy. Furthermore, natives are defined as people living in the state of birth (in the United States), or in the NUTS-3 region of birth (in France or Italy).

Figures A7b, A7d, and A7f illustrate that in each of these countries, natives have a higher tendency to coreside with parents compared to migrants. This pattern, also observed in the Spanish context (Figure 2c), suggests that coresidence can be a key factor influencing the homeownership gap between natives and migrants. The unavailability of data on the share of people living in inherited houses does not allow me to compare across countries the importance of housing bequests in generating the homeownership gap between natives and migrants (Figure 2b).

B Data Appendix

[Back]

Several cross-sectional and panel datasets are used in the analysis, encompassing Census, survey, administrative, and online sources. All samples are restricted to individuals aged between 25 and 64 who are Spanish-born citizens and currently active, i.e. employed or unemployed. Descriptions of these data sources are provided below.

MCVL The continuous Work History Sample (*Muestra Continua de Vidas Laborales*, or MCVL) is a 4% non-stratified sample of individuals affiliated to the Spanish social security. This panel records job changes and contractual modifications within the same firm. Information on wages is provided for the entire working life of the sampled individuals, when available. I focus on 2005–2019, the period in which job spells are matched with tax record data that provide uncensored earnings, and compute yearly full-time equivalent wages using the available information on working hours. For a small number of cases, the computed wages are much higher than workers' contributions to social security. To prevent these outliers from affecting the results, I remove the observations corresponding to the top 1% of the wage distribution. Labor income is deflated using the 2014 Consumer Price Index.

Furthermore, the MCVL provides information on workers' gender and age, which are contained in social security records. The sample is also matched with Spain's Continuous Register Statistics (*Estadística del Padrón Continuo*), so that individual characteristics such as province of birth and residence, educational attainment, and an indicator for those coresiding with parents can be recovered.[45] Data on municipalities of birth and residence are also provided when these municipalities have population over 40,000. This information is used in combinations with the province data (mappable to NUTS-1 regions) to retrieve birthplace and current location, as explained in Appendix A.

Employers assign workers to different social security contribution groups that are highly related to the level of education required to perform the job. Following De la Roca and Puga (2017), I organize these groups into five skill categories: very high-skilled, high-skilled, medium-high-skilled, medium-low-skilled, and low-skilled occupations. For example, the upper contribution group, which includes very high-skilled occupations, is reserved for jobs that require an engineering or bachelor's degree and for top managerial positions. Finally, the NACE 3-digit sector of the establishment and its location are also reported. The final dataset consists of 5.25 million observations.

EU-SILC The European Union Statistics on Income and Living Conditions (EU-SILC) is a survey that spans the years 2004-2019 and contains 74,000 Spanish households. It is a household-level panel with a 4-year rotating structure. It includes information on homeownership, coresidence, household-level income, transfers, unemployment benefits, and household members' employment status. It also includes information on NUTS-1 regions of residence and on the size of the municipality of residence (indicator for municipalities with pop-

[45]The Continuous Register Statistics contains information on the household composition (date of birth and gender of each individual living in the household). I count a person as coresiding with parents if they are living with someone that is at least 18 years older than them.

ulation larger than 50,000). By approximating the 40,000 cutoff with the 50,000 one, urban and rural locations of residence can be defined accordingly.

The household-level annual migration rate matched in the model comes from the EU-SILC rather than the MCVL. A number of reasons motivate this choice. First, the EU-SILC provides representative information on employed and unemployed household heads, while the MCVL only contains information on the unemployed receiving unemployment benefits. Second, the migration behavior of individuals in the EU-SILC can be compared with changes in homeownership, a piece of information that is missing in the MCVL. The elasticity between migration and homeownership is an important moment that is compared to the model counterpart. Lastly, while the EU-SILC regularly surveys the municipality of residence, this data is reliant on individuals' self-reporting in the MCVL, in line with Spanish bureaucratic procedures (*Empadronamiento*). This can introduce time lags in updates and potentially bias the migration information coming from the administrative data.

EFF The Survey of Household Finances (*Encuesta Financiera de las Familias*, or EFF) is a household-level panel that covers the years 2005-2020 and comprises 15,000 households. The Bank of Spain conducts this survey triennally. However, since each wave contains data from two successive years, only one year out of three contains no data. The dataset covers all the important variables that are relevant for the analysis: these include location (current and birthplace), homeownership, coresidence, and income (only at the household-level).

The EFF is also the only dataset with wealth information. The net wealth information used for the analysis comprises all household wealth, including financial assets, real estate, businesses shares, private pension plans, and other real valuables minus total debt. Finally, the data contains information on housing bequests received at some point in life (not necessarily during the sample period), which is needed to estimate the probability to inherit a dwelling in the model.

The Bank of Spain, as the EFF data provider, has authorized remote access to restricted geographic information (current and birthplace) in compliance with privacy guidelines. The 2005-2006 wave lacks data on birthplace. Whenever this piece of data is needed for the analysis, I use the 2008-2020 version of the panel. Urban and rural locations are classified within NUTS-1 regions following the 40,000 population threshold rule for municipalities in urban areas.

Census The 2011 Census of Population and Housing includes 1.3 million observations. It contains cross-sectional information on location (province of birth, municipality of residence), homeownership (whether the current dwelling has been paid, inherited, or is still mortgaged), and coresidence. The municipalities of residence can be directly mapped to rural and urban locations. The 1991 and 2001 Censuses contain similar data and can be used to account for cohort effects.

EVR, EPC The combined Residential Variation Statistics (*Estadística de Variaciones Residenciales*, or EVR) and the Continuous Register Statistics (*Estadística del Padrón Continuo*, or EVR) provide data on the *universe* of Spanish movers and stayers, respectively. They include an indicator for residents in cities with population lower than 10,000 and between 10,000 and 20,000 - among other population thresholds. The granualarity and precision of these data sources allows me to study the migration effect of place-based policy introduced in small municipalities.

EPF The Household Budget Survey (*Encuesta de Presupuestos Familiares*, or EPF) provides expenditure data and additional variables, including homeownership status, on a yearly sample of approximately 24,000 households. It also has indicators for residents in small municipalities, which enables me to examine the impact of place-based policy on homeownership. Average housing sizes are also measured with this data source.

Idealista *Idealista* is the leading real estate web portal in Spain. Its database covers the quasi-universe of dwellings that have been listed on the internet. Their website offers publicly available reports with housing prices and rents times series for both NUTS-3 and NUTS-2 regions, as well as for a selection of municipalities, encompassing all urban municipalities used in the analysis.[46] I use this data source focusing on the period between 2010 and 2019, to avoid missing data in some locations before 2010.

Price and rent series for urban locations are computed using population-weighted averages of time series at the urban municipality level. For rural locations series, a two-step approach is used. First, I compute prices and rents at the *rural* NUTS-2 levels as the residual series that would yield the *aggregate* observed NUTS-2 series, when combined in a population weighted-average with the observed *urban* NUTS-2 prices and rents. Then, I aggregate the rural NUTS-2 series at the rural location level using population weighted averages.

[46]The data can be accessed at https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/.

C Model Appendix

[Back]

C.1 Individual Problem

[Back]

Dynamic Problem: Coresidents and Renters Let $V_j^n(d'; \mathbf{x}, \Omega)$ denote the value function of agents who enters the period as non-homeowners, conditional on choosing location d'. These agents choose between coresiding, renting, and buying a house by solving

$$V_j^n(d'; \mathbf{x}, \Omega) = \max\{V_j^{c,n}(d'; \mathbf{x}, \Omega), V_j^{r,n}(d'; \mathbf{x}, \Omega), V_j^{o,n}(d'; \mathbf{x}, \Omega)\}. \tag{16}$$

Agents who choose to coreside in d' solve

$$V_j^{c,n}(d'; \mathbf{x}, \Omega) = \max_{c>0} \left\{u_j(c, h' = 0, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi,l,\boldsymbol{\mathcal{Z}}} \left[\overline{V}_{j+1}^n(\mathbf{x}', \Omega')\right]\right\}, \tag{17}$$

$$\begin{aligned} &\text{subject to} \\ &a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c) \geq 0, \\ &\varpi' \sim \Gamma_{edj}(\varpi), \quad l' \sim \Psi_{edj}(l), \\ &\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \quad \mu' \sim \Gamma_\mu(\Omega), \end{aligned}$$

where $\overline{V}_{j+1}^n(\mathbf{x}', \Omega')$ is the value function for agents who start next period as non-homeowners taking into account expected migration choices, which I describe below.

Those who choose to rent in d' solve

$$V_j^{r,n}(d'; \mathbf{x}, \Omega) = \max_{c>0} \left\{u_j(c, h' = 1, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi,l,\boldsymbol{\mathcal{Z}}} \left[\overline{V}_{j+1}^n(\mathbf{x}', \Omega')\right]\right\}, \tag{18}$$

$$\begin{aligned} &\text{subject to} \\ &a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c - \underbrace{\kappa_{d'} p_{d'}(\Omega) \overline{h}_{ed'1}}_{\text{Rent}}) \geq 0, \\ &\varpi' \sim \Gamma_{edj}(\varpi), \quad l' \sim \Psi_{edj}(l), \\ &\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \quad \mu' \sim \Gamma_\mu(\Omega). \end{aligned}$$

Finally, agents who choose to buy in d' solve

$$V_j^{o,n}(d'; \mathbf{x}, \Omega) = \max_{c>0} \left\{u_j(c, h' = 2, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi,l,\boldsymbol{\mathcal{Z}}} \left[\overline{V}_{j+1}^h(\mathbf{x}', \Omega')\right]\right\}, \tag{19}$$

$$\begin{aligned} &\text{subject to} \\ &a' = (1 + r\mathbb{1}_{a\geq 0} + r^h \mathbb{1}_{a<0})(a + y(w_{ed}(\Omega), \varpi, l) - c - (1 + \underbrace{\phi_b}_{\text{Transaction cost of buying}}) p_{d'}(\Omega) \overline{h}_{ed'2}), \\ &a' \geq -(1-\chi) p_{d'}(\Omega) \overline{h}_{ed'2} \ \text{ if } j<J, \\ &a' \geq 0 \ \text{ if } j=J, \\ &\varpi' \sim \Gamma_{edj}(\varpi), \quad l' \sim \Psi_{edj}(l), \\ &\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \quad \mu' \sim \Gamma_\mu(\Omega), \end{aligned}$$

where $\overline{V}^h_{j+1}(\mathbf{x}', \Omega')$ is the value function for agents who start next period as homeowners taking into account expected migration choices, which I describe below.

Given $V^n_j(d'; \mathbf{x}, \Omega)$, the migration decision of agents entering the period as non-homeowners is determined by

$$V^n_j(\mathbf{x}, \Omega, \varepsilon) = \max_{d' \in D} \left\{ V^n_j(d'; \mathbf{x}, \Omega) + \varepsilon_{d'} \right\}, \quad \varepsilon \overset{\text{iid}}{\sim} \text{Standard Gumbel}. \tag{20}$$

Given the iid Gumbel assumption on the vector of preference shocks $\varepsilon = (\varepsilon_1, ..., \varepsilon_D)$ for location d' at age j, there exists a simple expression for $\overline{V}^n_{j+1}(\mathbf{x}', \Omega')$, i.e.,

$$\begin{aligned} \overline{V}^n_{j+1}(\mathbf{x}', \Omega') &= \mathbb{E}_\epsilon[V^n_{j+1}(\mathbf{x}', \Omega', \varepsilon)] \\ &= \overline{\gamma} + \log\left(\sum_{d_k=1}^{D} \exp(V^n_{j+1}(d_k; \mathbf{x}', \Omega')) \right), \end{aligned}$$

where $\overline{\gamma}$ is the Euler-Mascheroni constant.

Dynamic Problem: Homeowners Let $V^h_j(d'; \mathbf{x}, \Omega)$ denote the value function of agents who enter the period as homeowners, conditional on choosing location d'. These agents choose between selling the house, by either coresiding or renting, and staying homeowners, either in the current or in a different location. This is done by solving

$$V^h_j(d'; \mathbf{x}, \Omega) = \max\{V^{c,h}_j(d'; \mathbf{x}, \Omega), V^{r,h}_j(d'; \mathbf{x}, \Omega), V^{o,h}_j(d'; \mathbf{x}, \Omega)\}. \tag{21}$$

Homeowners who want to migrate, always need to sell their house first.

Agents who choose to coreside in d' solve

$$\begin{aligned} V^{c,h}_j(d'; \mathbf{x}, \Omega) = \max_{c>0} & \left\{ u_j(c, h' = 0, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{\mathcal{Z}}} \left[\overline{V}^n_{j+1}(\mathbf{x}', \Omega') \right] \right\}, \\ & \text{subject to} \\ & a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c + (1 - \underbrace{\phi_s}_{\substack{\text{Transaction cost} \\ \text{of selling}}}) p_d(\Omega) \overline{h}_{ed2}) \geq 0, \\ & \varpi' \sim \Gamma_{edj}(\varpi), \;\; l' \sim \Psi_{edj}(l), \\ & \boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \;\; \mu' \sim \Gamma_\mu(\Omega). \end{aligned} \tag{22}$$

Notice that only natives are able to coreside, i.e. it must be that $d' = d_0$.

Those who choose to rent in d' solve

$$\begin{aligned} V^{r,h}_j(d'; \mathbf{x}, \Omega) = \max_{c>0} & \left\{ u_j(c, h' = 1, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{\mathcal{Z}}} \left[\overline{V}^n_{j+1}(\mathbf{x}', \Omega') \right] \right\}, \\ & \text{subject to} \\ & a' = (1+r)(a + y(w_{ed}(\Omega), \varpi, l) - c + (1 - \phi_s) p_d(\Omega) \overline{h}_{ed2} - \kappa_d p_{d'}(\Omega) \overline{h}_{ed'1}) \geq 0, \\ & \varpi' \sim \Gamma_{edj}(\varpi), \;\; l' \sim \Psi_{edj}(l), \\ & \boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}), \;\; \mu' \sim \Gamma_\mu(\Omega). \end{aligned} \tag{23}$$

Finally, homeowners who choose to keep their house in $d' = d$ (non-migrants) solve

$$V_j^{o,h}(d';\mathbf{x},\Omega) = \max_{c>0}\left\{u_j(c,h'=2,d',\mathbf{x}) + \beta\mathbb{E}_{\varpi,l,\boldsymbol{\mathcal{Z}}}\left[\overline{V}_{j+1}^h(\mathbf{x}',\Omega')\right]\right\}, \tag{24}$$

$$\begin{aligned}
&\text{subject to}\\
&a' = (1 + r\mathbb{1}_{a\geq 0} + r^h\mathbb{1}_{a<0})(a + y(w_{ed}(\Omega),\varpi,l) - c),\\
&a' \geq a\left(\frac{(J-1)-j}{J-j}\right)\mathbb{1}_{a<0} \ \text{ if } j<J,\\
&a' \geq 0 \ \text{ if } j=J,\\
&\varpi' \sim \Gamma_{edj}(\varpi),\ \ l' \sim \Psi_{edj}(l),\\
&\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}),\ \ \mu' \sim \Gamma_\mu(\Omega),
\end{aligned}$$

i.e., don't have housing costs but need to repay their mortgage following a linear schedule. Instead, homeowners who choose to migrate and to buy a house in d' solve

$$V_j^{o,h}(d';\mathbf{x},\Omega) = \max_{c>0}\left\{u_j(c,h'=2,d',\mathbf{x}) + \beta\mathbb{E}_{\varpi,l,\boldsymbol{\mathcal{Z}}}\left[\overline{V}_{j+1}^h(\mathbf{x}',\Omega')\right]\right\}, \tag{25}$$

$$\begin{aligned}
&\text{subject to}\\
&a' = (1 + r\mathbb{1}_{a\geq 0} + r^h\mathbb{1}_{a<0})(a + y(w_{ed}(\Omega),\varpi,l) - c + (1-\phi_s)p_d(\Omega)\overline{h}_{ed2} - (1+\phi_b)p_{d'}(\Omega)\overline{h}_{ed'2}),\\
&a' \geq -(1-\chi)p_{d'}(\Omega)\overline{h}_{ed'2} \ \text{ if } j<J,\\
&a' \geq 0 \ \text{ if } j=J,\\
&\varpi' \sim \Gamma_{edj}(\varpi),\ \ l' \sim \Psi_{edj}(l),\\
&\boldsymbol{\mathcal{Z}}' \sim \Gamma_{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}),\ \ \mu' \sim \Gamma_\mu(\Omega).
\end{aligned}$$

Given $V_j^h(d';\mathbf{x},\Omega)$, the migration decision of agents entering the period as homeowners is determined by

$$V_j^h(\mathbf{x},\Omega,\varepsilon) = \max_{d'\in D}\left\{V_j^h(d';\mathbf{x},\Omega) + \varepsilon_{d'}\right\},\quad \varepsilon \overset{\text{iid}}{\sim} \text{Standard Gumbel}. \tag{26}$$

Again, from the Standard Gumbel assumption, we have that

$$\begin{aligned}
\overline{V}_{j+1}^h(\mathbf{x}',\Omega') &= \mathbb{E}_\epsilon[V_{j+1}^h(\mathbf{x},\Omega,\varepsilon)]\\
&= \overline{\gamma} + \log\left(\sum_{d_k=1}^{D}\exp(V_{j+1}^h(d_k;\mathbf{x}',\Omega'))\right).
\end{aligned}$$

The resulting final decision rules include: housing tenure decisions to coreside, rent, and buy, i.e., $h_j^c(\mathbf{x},\Omega) \in \{0,1\}$, $h_j^r(\mathbf{x},\Omega) \in \{0,1\}$, and $h_j^o(\mathbf{x},\Omega) \in \{0,1\}$, migration decisions $d_j(\mathbf{x},\Omega) \in \{1,...,12\}$, and consumption choices $c_j(\mathbf{x},\Omega)$.

Probability of Receiving a Housing Bequest Agents receive housing bequests with probability $\pi_{ed_0j}^b$ at the end of the period, after having made their consumption, housing, and migration choices. In case they choose not to be homeowners, and if they decide to live in the birthplace (natives), then, if they receive a housing bequest, they become homeowners. The idea is that the property is inherited in the birthplace, and only natives who are not already homeowners can go live there. In all other cases (homeowners or migrants), when agents receive a bequest, they are forced to sell the inherited house, pay the transaction costs ϕ_s, and split the proceeds among κ_b siblings. Parameter κ_b is the same across agents. Notice that the inherited house value depends on prices in the birthplace.

let $\mathbf{x}'_\mathbf{b}$ be the resulting state vector whenever a housing bequest is received at age j. Then we have

$$\begin{cases} h'_b = 2 & \text{if } h' \neq 2 \text{ and } d' = d_{j_0} \\ a'_b = a' & \\ h'_b = h' & \\ a'_b = a' + \dfrac{p_{d_0}(\Omega)\overline{h}_{ed_0 2}}{\kappa_b}(1-\phi_s) & \text{otherwise} \end{cases}$$

Although budget constraints are unchanged, the function maximized by agents who decide to live outside their birthplace or to buy a house has a different expected value function. Consider the case of an agent who decides to buy (or keep) their house for the next period. Then, the function is given by:

$$\max_{c>0} \left\{ u_j(c, h' = 2, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{z}} \left[(1 - \pi^b_{ed_0 j}) \overline{V}^h_{j+1}(\mathbf{x}', \Omega') + \pi^b_{ed_0 j} \overline{V}^h_{j+1}(\mathbf{x}'_\mathbf{b}, \Omega') \right] \right\}.$$

Similarly, a non-homeowner agent that decides to live outside their birthplace and not to buy solves:

$$\max_{c>0} \left\{ u_j(c, h', d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{z}} \left[(1 - \pi^b_{ed_0 j}) \overline{V}^n_{j+1}(\mathbf{x}', \Omega') + \pi^b_{ed_0 j} \overline{V}^n_{j+1}(\mathbf{x}'_\mathbf{b}, \Omega') \right] \right\}.$$

The function of an agent who can become homeowner by inheriting, say an individual who decides to coreside and to stay in the birthplace, is instead given by:

$$\max_{c>0} \left\{ u_j(c, h' = 0, d', \mathbf{x}) + \beta \mathbb{E}_{\varpi, l, \boldsymbol{z}} \left[(1 - \pi^b_{ed_0 j}) \overline{V}^n_{j+1}(\mathbf{x}', \Omega') + \pi^b_{ed_0 j} \overline{V}^h_{j+1}(\mathbf{x}'_\mathbf{b}, \Omega') \right] \right\}.$$

Notice that the expected value function now contains both $\overline{V}^n_{j+1}$ and $\overline{V}^h_{j+1}$.

Terminal Value Finally, agents receive a one-time utility from their homeownership status and accumulated liquid and housing wealth when they exit the model after age J. In particular, at age $j = J$, agents' expected value function does not include $\overline{V}^n_{j+1}$ or $\overline{V}^h_{j+1}$. Instead, it includes the terminal value function

$$V_{J+1}(\mathbf{x}', \Omega') = \omega_1 \frac{(a' + p_{d'}(\Omega)\overline{h}_{ed'2} \mathbb{1}\{h' = 2\})^{1-\gamma}}{1-\gamma} + \omega_2 \mathbb{1}\{h' = 2\},$$

as in

$$\max_{c>0} \left\{ u_J(c, h', d', \mathbf{x}) + \beta \mathbb{E}_{\boldsymbol{z}} \left[V_{J+1}(\mathbf{x}', \Omega') \right] \right\}.$$

C.2 Labor Markets

[Back]

Firms Output in location d produced by the representative firm is given by

$$Y_d = X_d(\zeta_d L_{Nd}^{\rho} + (1-\zeta_d)L_{Ed}^{\rho})^{\frac{1}{\rho}},$$

where X_d is the overall labor productivity, ζ_d is the skill-specific productivity, and L_{ed} is total efficiency units input of education group e in location d. Local labor markets are competitive. The first-order conditions determine wages w_{ed} of college and non-college labor

$$w_{ed} = X_d(\zeta_d \mathbb{1}_N + (1-\zeta_d)\mathbb{1}_E)(\zeta_d L_{Nd}^{\rho} + (1-\zeta_d)L_{Ed}^{\rho})^{\frac{1-\rho}{\rho}} L_{ed}^{\rho-1}, \tag{27}$$

where $\mathbb{1}_E$ and $\mathbb{1}_N$ are indicator functions for the college and non-college first-order conditions, respectively. The firms' output consists of the freely tradable consumption good, which acts as the numeraire.

Individuals The supply of efficiency-units of labor, given the distribution of agents μ_{edj} across education, location, and age states, employment shares $\widehat{l}_{edj} = \mathbb{E}[\mathbb{1}\{l_{edj} = 2\}]$, age income profiles Υ_{edj}, and distribution function $\varphi_{edj}(\theta_e, z_{ej})$, is given by

$$L_{ed}^S = \sum_{j=1}^{J} \mu_{edj}\widehat{l}_{edj} \int \exp(\theta_e + z_{ej} + \Upsilon_{edj})\varphi_{edj}(\theta_e, z_{ej})\, d\theta_e\, dz_{ej}. \tag{28}$$

Estimation of the Aggregate Shocks Let t denote years. Aggregate shocks hit location d every year, and are thus indexed by t, i.e. (ζ_{dt}, X_{dt}). The distribution of agents across individual states in period t is given by μ_t. In the benchmark equilibrium, ζ_{dt} and X_{dt} can be inverted so that the observed time series of wages in the data are matched in equilibrium, i.e., $w_{Ndt} = w_{Nd}(\Omega)$ and $w_{Edt} = w_{Ed}(\Omega)$. Divide the first-order conditions and manipulate to obtain:

$$\zeta_{dt} = \frac{\frac{w_{Ndt}}{w_{Edt}}\left(\frac{L_{Ndt}}{L_{Edt}}\right)^{1-\rho}}{1 + \frac{w_{Ndt}}{w_{Edt}}\left(\frac{L_{Ndt}}{L_{Edt}}\right)^{1-\rho}}.$$

Given the substitution parameter ρ and substituting the supply of labor into the firms' first-order conditions,

$$L_{edt} = L_{ed}^S(\Omega_t), \tag{29}$$

derive ζ_{dt} by perfectly matching wages within locations and in all years (i.e. under all realizations of the aggregate shocks, as described in Section 4.2.3). Finally, use any of the two location-specific first-order conditions to derive X_{dt} in all years and locations.

C.3 Housing Markets

[Back]

Firms The representative construction firm operating in location d sells houses at price p_d (per square meters) and has a convex technology cost of production

$$C(H_d) = k_d \frac{H_d^{\psi+1}}{\psi + 1},$$

where H_d is the total housing stock supplied, measured in square meters, and $k_d > 0$, $\psi > 0$.[47] The convexity of the cost function is a reduced form way to capture the scarcity of buildable land and possible inputs and regulation constraints, while intercept k_d captures construction costs that may vary across location and periods.

Profit maximization in competitive housing markets leads to the first-order condition

$$p_d = k_d H_d^{\psi}, \tag{30}$$

which is the inverse housing supply function. The inverse housing supply elasticity ψ reflects the responsiveness of housing prices to changes in the housing stock.

Finally, there is a representative real-estate firm operating in each location that rents out housing units at price ι_d. Assuming that the outside-option investment has returns r and that operating costs of real-estate firms (e.g. monitoring costs, depreciation of the housing unit) are a fixed, location-dependent proportion $\bar{\iota}_d$ of the housing price, the zero-profit conditions is given by

$$\iota_d - \bar{\iota}_d p_d = r p_d.$$

Therefore, denoting $\kappa_d = \bar{\iota}_d + r$, housing rents (per square meter) are given by

$$\iota_d = \kappa_d p_d, \tag{31}$$

where κ_d is the price-to-rent ratio.

Individuals Renters and homeowners demand housing sizes of $\overline{h}_{1ed}$ and $\overline{h}_{2ed}$ square meters, respectively. These quantities are fixed choices but are allowed to vary by education group and location.

Given the distribution of agents μ_{edj} across education, location, and age states, renter shares $\widehat{h}_{1edj} = \mathbb{E}[\mathbb{1}\{h_{edj} = 1\}]$, and homeowner shares $\widehat{h}_{2edj} = \mathbb{E}[\mathbb{1}\{h_{edj} = 2\}]$, housing demand is given by

$$H_d^D = \sum_{j=1}^{J} \sum_{e=1}^{2} \mu_{edj} (\widehat{h}_{1edj} \overline{h}_{1ed} + \widehat{h}_{2edj} \overline{h}_{2ed}). \tag{32}$$

Estimation of the Aggregate Shocks Let t denote years. Aggregate shocks hit location d every year, and are thus indexed by t, i.e. k_{dt}. The distribution of agents across individual states in period t is given by μ_t. In the benchmark equilibrium, k_{dt} can be inverted so that the

[47] The housing stock depreciates fully each period, so that the flow of construction investment in every period equals the housing stock H_d. In an exercise that analyses the transitions between steady states, the depreciation rate of the housing stock would determine the length of the transitions.

observed time series of prices in the data are matched in equilibrium, i.e., $p_{dt} = p_d(\Omega)$.

Given the inverse housing supply elasticity ψ and benchmark equilibrium quantities

$$H_{dt} = H_d^D(\Omega_t), \tag{33}$$

derive k_{dt} by perfectly matching prices (per square meter) under all realizations of the aggregate shocks (Section 4.2.3). Finally, κ_d is estimated by computing the ratio between average housing prices and rents by location.

C.4 Government [Back]

Budget The government budget balances in each period in which aggregate shocks hit the economy. Therefore, expenditure on policies, transfers and unemployment benefits, and other public goods (denoted by $G^p(\Omega)$, $G^g(\Omega)$ and $\overline{G}(\Omega)$ respectively) equals revenues from income taxes ($T(\Omega)$). Expenditure on public goods $\overline{G}(\Omega)$ does not affect individuals' utility.

Given the distributions of transfer recipients and taxpayers by gross labor income and given the shares of workers that receive transfers and pay taxes, $\varphi^g_{edj}(\tilde{y}_{edj}(\Omega))$, $\varphi^\varsigma_{edj}(\tilde{y}_{edj}(\Omega))$, $\widehat{l}^g_{edj}(\Omega)$ and $\widehat{l}^\varsigma_{edj}(\Omega)$ respectively, and given mean gross income levels $\overline{y}$ and the overall share of unemployed $\bar{l}_1(\Omega)$, the government budget constraint is given by

$$T(\Omega) = \overline{G}(\Omega) + G^g(\Omega) + G^p(\Omega), \tag{34}$$

where

$$G^g(\Omega) = \int \widehat{l}^g_{edj}(\Omega)\left(g_1 - g_2\left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}}\right)\right)\varphi^g_{edj}(\tilde{y}_{edj}(\Omega))\, d\tilde{y}_{edj}(\Omega) + \bar{l}_1(\Omega)b,$$

$$T(\Omega) = \int \tilde{y}_{edj}(\Omega)\widehat{l}^\varsigma_{edj}(\Omega)\left(1 - \varsigma_0\left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}}\right)^{-\varsigma_1}\right)\varphi^\varsigma_{edj}(\tilde{y}_{edj}(\Omega))\, d\tilde{y}_{edj}(\Omega).$$

Estimation Given benchmark tax revenues and transfers, $\overline{G}(\Omega)$ is chosen to balance the budget when $G^g(\Omega) = 0$ (benchmark equilibrium under all realizations of the aggregate shocks, as described in Section 4.2.3). In counterfactual exercises, only $G^p(\Omega)$, $G^g(\Omega)$ and $T(\Omega)$ change, and the level of labor income taxes ς_0 is adjusted to balance the budget. Given the benchmark tax progressivity, this corresponds to a proportional change in taxes along the income distribution.

C.5 Equilibrium [Back]

Let individual states be given by x and aggregate states by $\Omega = (\boldsymbol{\mathcal{Z}}, \mu)$, where μ is the distribution of agents across x.

A *recursive competitive equilibrium* consists of value functions $\{V_j^n(\mathbf{x}, \Omega, \varepsilon), V_j^n(d'; \mathbf{x}, \Omega),$ $V_j^{c,n}(d'; \mathbf{x}, \Omega), V_j^{r,n}(d'; \mathbf{x}, \Omega), V_j^{o,n}(d'; \mathbf{x}, \Omega), V_j^h(\mathbf{x}, \Omega, \varepsilon), V_j^h(d'; \mathbf{x}, \Omega), V_j^{c,h}(d'; \mathbf{x}, \Omega), V_j^{r,h}(d'; \mathbf{x}, \Omega),$ $V_j^{o,h}(d'; \mathbf{x}, \Omega)\}$, decision rules $\{h_j^c(\mathbf{x}, \Omega), h_j^r(\mathbf{x}, \Omega), h_j^o(\mathbf{x}, \Omega), d_j(\mathbf{x}, \Omega), c_j(\mathbf{x}, \Omega)\}$, local housing prices

$\mathbf{p}(\Omega) = \{p_1(\Omega), ..., p_{12}(\Omega)\}$, local housing rents $\boldsymbol{\iota}(\Omega) = \{\iota_1(\Omega), ..., \iota_{12}(\Omega)\}$, local wages by education $\mathbf{w_e}(\Omega) = \{w_{e1}(\Omega), ..., w_{e12}(\Omega)\}$, labor income tax level $\varsigma_0(\Omega)$, and aggregate law of motion Γ_μ such that:

1. Individual optimize by solving problems (16)–(26), with associated value functions $\{V_j^n, V_j^{c,n}, V_j^{r,n}, V_j^{o,n}, V_j^h, V_j^{c,h}, V_j^{r,h}, V_j^{o,h}\}$ and decision rules $\{h_j^c, h_j^r, h_j^o, d_j, c_j\}$.

2. Housing prices $\mathbf{p}(\Omega)$ clear each local housing market, i.e., for all d

$$p_d(\Omega) = k_d[H_d^D(\Omega)]^\psi,$$

where housing demand is given by

$$H_d^D(\Omega) = \sum_{j=1}^{J}\sum_{e=1}^{2} \mu_{edj}(\Omega)\left[\widehat{h}_{1edj}(\Omega)\overline{h}_{1ed} + \widehat{h}_{2edj}(\Omega)\overline{h}_{2ed}\right],$$

with population shares $\mu_{edj}(\Omega)$, renter shares $\widehat{h}_{1edj}(\Omega)$, and homeowner shares $\widehat{h}_{2edj}(\Omega)$.

3. Housing rents $\boldsymbol{\iota}(\Omega)$ are given by

$$\iota_d(\Omega) = \kappa_d p_d(\Omega).$$

4. Wages by education $\mathbf{w_e}(\Omega)$ clear each local labor market, i.e., for all d

$$w_{Nd}(\Omega) = X_d\left\{\zeta_d[L_{Nd}^S(\Omega)]^\rho + (1-\zeta_d)[L_{Ed}^S(\Omega)]^\rho\right\}^{\frac{1-\rho}{\rho}}\zeta_d[L_{Nd}^S(\Omega)]^{\rho-1},$$

$$w_{Ed}(\Omega) = X_d\left\{\zeta_d[L_{Nd}^S(\Omega)]^\rho + (1-\zeta_d)[L_{Ed}^S(\Omega)]^\rho\right\}^{\frac{1-\rho}{\rho}}(1-\zeta_d)[L_{Ed}^S(\Omega)]^{\rho-1},$$

where labor supply is given by

$$L_{ed}^S(\Omega) = \sum_{j=1}^{J}\mu_{edj}(\Omega)\widehat{l}_{edj}(\Omega)\int \exp(\theta_e + z_{ej} + \Upsilon_{edj})\varphi_{edj}(\theta_e, z_{ej};\Omega)\,d\theta_e\,dz_{ej},$$

with population shares $\mu_{edj}(\Omega)$, employment shares $\widehat{l}_{edj}(\Omega)$, and distribution function $\varphi_{edj}(\theta_e, z_{ej};\Omega)$.

5. The level of taxes $\varsigma_0(\Omega)$ balances the Government budget, i.e.

$$T(\Omega) = \overline{G}(\Omega) + G^g(\Omega) + G^p(\Omega),$$

where

$$G^g(\Omega) = \int \widehat{l}^g_{edj}(\Omega) \left(g_1 - g_2 \left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}} \right) \right) \varphi^g_{edj}(\tilde{y}_{edj}(\Omega)) \, d\tilde{y}_{edj}(\Omega) + \bar{l}_1(\Omega) b,$$

$$T(\Omega) = \int \tilde{y}_{edj}(\Omega) \widehat{l}^\varsigma_{edj}(\Omega) \left(1 - \varsigma_0(\Omega) \left(\frac{\tilde{y}_{edj}(\Omega)}{\overline{y}} \right)^{-\varsigma_1} \right) \varphi^\varsigma_{edj}(\tilde{y}_{edj}(\Omega)) \, d\tilde{y}_{edj}(\Omega),$$

$\overline{G}(\Omega)$ denotes expenditure on public goods, and $G^p(\Omega)$ denotes expenditure on policies. Population shares receiving transfers and paying taxes, $\widehat{l}^g_{edj}(\Omega)$ and $\widehat{l}^\varsigma_{edj}(\Omega)$, and distribution functions $\varphi^g_{edj}(\tilde{y}_{edj}(\Omega))$ and $\varphi^\varsigma_{edj}(\tilde{y}_{edj}(\Omega))$ are implied by Ω.

6. The equilibrium law of motion Γ_μ is induced by the exogenous processes for idiosyncratic and aggregate shocks, as well as the decision rules. Therefore, Γ_μ is consistent with individual choices.

D Estimation Strategy With Aggregate Shocks [Back]

D.1 Solution Algorithm for the Benchmark Equilibrium [Back]

Below, I present an algorithm to compute the benchmark equilibrium:

1. Estimate the external parameters, including the exogenous laws of motions Ψ_{edj} and Φ_{edj}, the benchmark forecast rule (13), and the factors' laws of motion (11) and (12).

2. Guess the internally calibrated parameters.

3. Given the model's parameters, solve and simulate the model by:

   i. Imposing prices and wages $\mathbf{q_t}$ observed in the data

   ii. Using the estimated forecast rule for $\mathbf{q_{t+1}}$

4. Given the simulated equilibrium housing demand and labor supply $\{H_d^D(\Omega_t)$, $L_{Nd}^S(\Omega_t)$, $L_{Ed}^S(\Omega_t)\}$, use equations (7), (8), and (9) to find parameters $\boldsymbol{\mathcal{Z}_t}$ that are consistent with the observed prices and wages.[48]

5. If the fit of the targeted moments is good, stop. Otherwise, update the guesses for the internally calibrated parameters and repeat steps 3, 4, and 5 until the data moments are fitted well.

D.2 Solution Algorithm for the Counterfactual Equilibrium [Back]

Below, I present an algorithm to compute the counterfactual equilibrium:

1. Start with the benchmark parameters, which include parameters $\boldsymbol{\mathcal{Z}_t}$. These parameters are kept fixed in the counterfactual.

2. Given the model's parameters, solve and simulate the model *with the counterfactual* by initially imposing prices, wages, forecast rule, and factors' laws of motion as in the benchmark.

3. Given counterfactual equilibrium housing demand and labor supply $\{\widehat{H}_d^D(\Omega_t)$, $\widehat{L}_{Nd}^S(\Omega_t)$, $\widehat{L}_{Ed}^S(\Omega_t)\}$ and the benchmark parameters $\boldsymbol{\mathcal{Z}_t}$, use equations (7), (8), and (9) to update local prices and wages, $\widehat{\mathbf{q}}_\mathbf{t}$. A dampening parameter $k \in [0, 1]$ can be used for the update.

4. Given the updated counterfactual prices and wages $\widehat{\mathbf{q}}_\mathbf{t}$, estimate the new benchmark forecast rule (13) and the factors' laws of motion (11) and (12). New factors $(\widehat{f}_{1t}, \widehat{f}_{2t})$, and parameters $(\widehat{\boldsymbol{\lambda}}_\mathbf{1}, \widehat{\boldsymbol{\lambda}}_\mathbf{2}, \widehat{\varrho}_1, \widehat{\varrho}_2, \widehat{\sigma}_{f_1}, \widehat{\sigma}_{f_2})$ are obtained.

[48]The price-to-rent ratio κ_d of equation (31), is estimated by computing the ratio between average housing prices and rents by location, as explained in Appendix C.3.

5. Solve and simulate the model *with the counterfactual* by imposing the *updated* prices, wages, forecast rule, and factors' laws of motion.

6. Repeat steps 3., 4., and 5. until the new $(\widehat{f}_{1t}, \widehat{f}_{2t}, \widehat{\boldsymbol{\lambda}}_{\mathbf{1}}, \widehat{\boldsymbol{\lambda}}_{\mathbf{2}})$ guesses are close to the guesses in the previous iteration.

D.3 Connection With Krusell and Smith (1998) [Back]

In this section, I show how my low-rank forecast rule relates to typical forecast rules in the Krusell and Smith (1998) tradition. Let's consider a Krusell and Smith-type linear forecast rule that uses two moments from the distribution of current prices and wages $\mathbf{q}(\Omega)$, $m_1(\mathbf{q}(\Omega))$ and $m_2(\mathbf{q}(\Omega))$:

$$\mathbf{q}(\Omega') = a_1(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')m_1(\mathbf{q}(\Omega)) + a_2(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')m_2(\mathbf{q}(\Omega)), \tag{35}$$

where $a_1(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')$ and $a_2(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')$ are the time-varying coefficients mapping moments to predictions, which depend on the realizations of the exogenous aggregate shocks $\boldsymbol{\mathcal{Z}}$ and on their known laws of motion.[49]

In my model, aggregate shocks $\boldsymbol{\mathcal{Z}}$ are known functions of factors f_1 and f_2. My forecast rule (13) is equal to the one in equation (35) if $a_1(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')m_1(\mathbf{q}(\Omega)) = \boldsymbol{\lambda}_{\mathbf{1}} f_1$, and $a_2(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')m_2(\mathbf{q}(\Omega)) = \boldsymbol{\lambda}_{\mathbf{2}} f_2$, where expectations over future f_1' and f_2' are based on factors' laws of motions (11) and (12).

The forecast rule for the benchmark equilibrium is estimated only once, outside of the model, thanks to the fact that observed prices and wages can be perfectly matched in equilibrium. This is a key computational advantage relative to Krusell and Smith-type forecast rule, since, given the high-dimensionality of the spatial model, solving efficiently for the benchmark economy is necessary to be able to calibrate the model's parameters. The forecast rule is re-estimated in the benchmark equilibrium. This procedure, however, is not computationally taxing, as the other model parameters have already been estimated and are kept fixed at their benchmark values, so that the counterfactual equilibrium can be solved using a single outer loop

My strategy has the added advantage of greatly reducing the problem's dimensionality. In particular, the number of forecast rules (35) in a Krusell and Smith-type strategy would be prohibitively high. Even by only allowing 2 grid points for each discretized Markovian process, the unrestricted process for $\boldsymbol{\mathcal{Z}}$ would imply that $a_1(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')$ and $a_2(\boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{Z}}')$ have length 2^{36}, i.e. almost 70 billions.

D.4 Accuracy of the Forecast Rule [Back]

[49] For example, this equation is analogous to equation (14) in Kaplan, Mitman and Violante (2020). In the paper, they use a strategy based on Krusell and Smith (1998) to estimate a dynamic model with homeownership and aggregate price flactuations.

I have assumed that the correct forecast rule for equilibrium prices $\mathbf{q}$ is a factor model of rank two, i.e.,

$$\mathbf{q} \simeq \boldsymbol{\lambda_1} f_1 + \boldsymbol{\lambda_2} f_2.$$

However, the correct model may also be

$$\mathbf{q} \simeq \boldsymbol{\lambda_1} f_1 + \boldsymbol{\lambda_2} f_2 + \boldsymbol{\lambda_3} f_1 f_2, \tag{36}$$

or

$$\mathbf{q} \simeq \boldsymbol{\lambda_1} f_1 + \boldsymbol{\lambda_2} f_2 + \boldsymbol{\lambda_3} f_1 f_2 + \boldsymbol{\lambda_4} f_1^2 + \boldsymbol{\lambda_5} f_2^2,$$

or a different function with higher-order interactions between the factors. I show that the prediction properties of the forecast rule (13) do not improve if I use the mildest possible relaxation (36) as an alternative specification:

$$\mathbf{q}' = \boldsymbol{\lambda_1} f_1' + \boldsymbol{\lambda_2} f_2' + \boldsymbol{\lambda_3} f_1' f_2' + \boldsymbol{\upsilon_q}. \tag{37}$$

To do so, I compute the prediction errors and R^2 of the forecast rules (13) and (37), both in the benchmark and in the counterfactual equilibria.

In the benchmark, both forecast rules exhibit high R^2 values, although the relaxed forecast rule (37) has a marginally higher R^2 of 0.996 compared to 0.991 in (13). The R^2, however, is not a sufficient statistic to measure the accuracy of forecast rules and can sometimes be misleading (Den Haan 2010). When examining the mean and median forecasting errors, the baseline forecast rule (13) performs marginally worse in terms of the former but better in the latter, with errors of 4.8% (mean) and 2.3% (median). In contrast, the errors for (37) are 4.3% and 3.3%, respectively. Therefore, the overall prediction properties are not unambiguously improved when relaxing the forecast rule.

As pointed out by Den Haan (2010), the accuracy of Krusell and Smith (1998)-type forecast rules over longer horizons may be poor, even if 1-year ahead prediction properties are good. Therefore, I repeat the same exercise for 25-year ahead prediction errors and R^2. To simplify the exercise, I check the accuracy of the forecast rule based on *average* guesses about future prices and wages.

In particular, using the laws of motion (11) and (12), I impose $f_1' = \varrho_1 f_1$ and $f_2' = \varrho_2 f_2$ for 1-year ahead average predictions, and $f_1' = \varrho_1^{25} f_1$ and $f_2' = \varrho_2^{25} f_2$ for 25-year ahead average predictions. The prediction properties of the baseline forecast rule do not worsen significantly over longer time horizon. The mean and median prediction errors go from 5.7% and 4.4% to 6.6% and 5.2%, respectively, whereas the R^2 increases when predicting 25-year ahead data, from 0.991 to 0.993.[50]

[50] Prediction errors are higher in this exercise because I am restricting agents to only use ϱ_1 and ϱ_2 to predict future $\mathbf{q}$, instead of taking into account the full exogenous laws of motion for f_1 and f_2 as when solving for the equilibrium. Therefore, a large part of these deviations comes from the underlying volatility of the exogenous stochastic factor processes, rather than from errors in the baseline forecast rule.

Additionally, the exercise is repeated in the counterfactual equilibrium where housing policies are introduced.[51] The forecasting properties of the baseline rule (13) and relaxed rule (37) in the counterfactual remain very similar to the benchmark equilibrium: the R^2 are 0.992 and 0.996, the mean prediction errors are 4% and 4%, and the median prediction errors are 1.8% and 2.9%, respectively. Again, the overall prediction properties are not unambiguously improved when relaxing the forecast rule. Therefore, I use the baseline forecast rule (13) to solve for the benchmark and counterfactual equilibria.

Selecting the Number of Factors Finally, I test for the correct number of factors, denoted by $k \in \{1, ..., K\}$, using the strategy proposed in Bai and Ng (2002). This method introduces a penalty for overfitting in the minimization problem, which increases linearly with k and counterbalances the fit improvement due to the inclusion of additional common factors.

The optimal number of factors is estimated to be either two or three, depending on the specific criteria.[52] In practice, the fit of the data is very similar whether two or three factors are used. However, the computational burden of a model with a three factors' forecast rule is much higher.[53] Therefore, I opt to maintain the baseline specification involving only two factors (f_{1t}, f_{2t}) for the forecast rule.

D.5 Fit of the Factor Model [Back]

Using subscript t for convenience, prices and wages can be written as function of the two aggregate (country-level) factors f_{1t} and f_{2t}, which are orthogonal and time-varying, and the set of fixed loading parameters $\boldsymbol{\lambda_d}$:

$$\begin{aligned} p_{dt} &= \lambda^{p}_{1d} \times f_{1t} + \lambda^{p}_{2d} \times f_{2t}, \\ w_{edt} &= \lambda^{w_e}_{1d} \times f_{1t} + \lambda^{w_e}_{2d} \times f_{2t}. \end{aligned}$$

The factors and loading parameters are estimated outside of the model using the interactive fixed effects strategy developed in Bai (2009).[54] Data cover the period 2010-2019, and come from Idealista (housing prices) and the MCVL (hourly wages).

As can be seen in Figures D8, D9, and D10, the factor model fits the data very well. The prediction captures in a parsimonious way the evolution of local prices and wages between 2010 and 2019. Aggregate price and wage shocks are correlated between themselves, across locations, and over time. These realistic features of the data are well captured by the factor model structure.

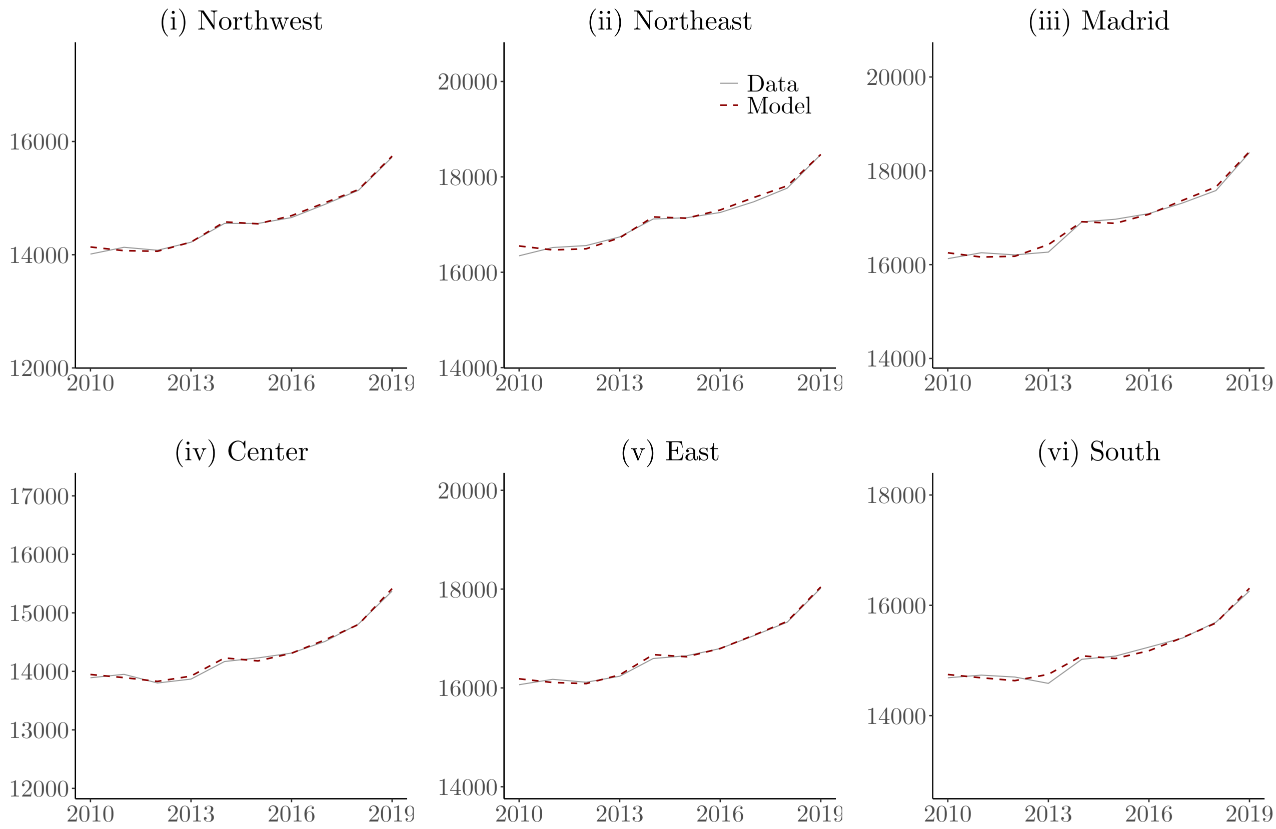
[51] I report here the results following the introduction of the untargeted mortgage interest deduction policy. Results are comparable when using other housing policies.

[52] I use the three criteria that satisfy the requirements of Theorem 2 in Bai and Ng (2002), denoted in the paper by PC_{p1}, PC_{p2}, and PC_{p3}. The criteria, although asymptotically equivalent, have different properties in finite samples. The optimal number of factors is 3 according to test PC_{p1} and PC_{p2}, whereas it is 2 according to PC_{p3}.

[53] Since I discretize factors using 3 grid points, using three factors instead of two requires solving the model at 975 million points instead of the current 325 million.

[54] Estimation is carried out with the R package `phtt` (Bada and Liebl 2014).

(a) Non-college wages, urban locations



(b) Non-college wages, rural locations

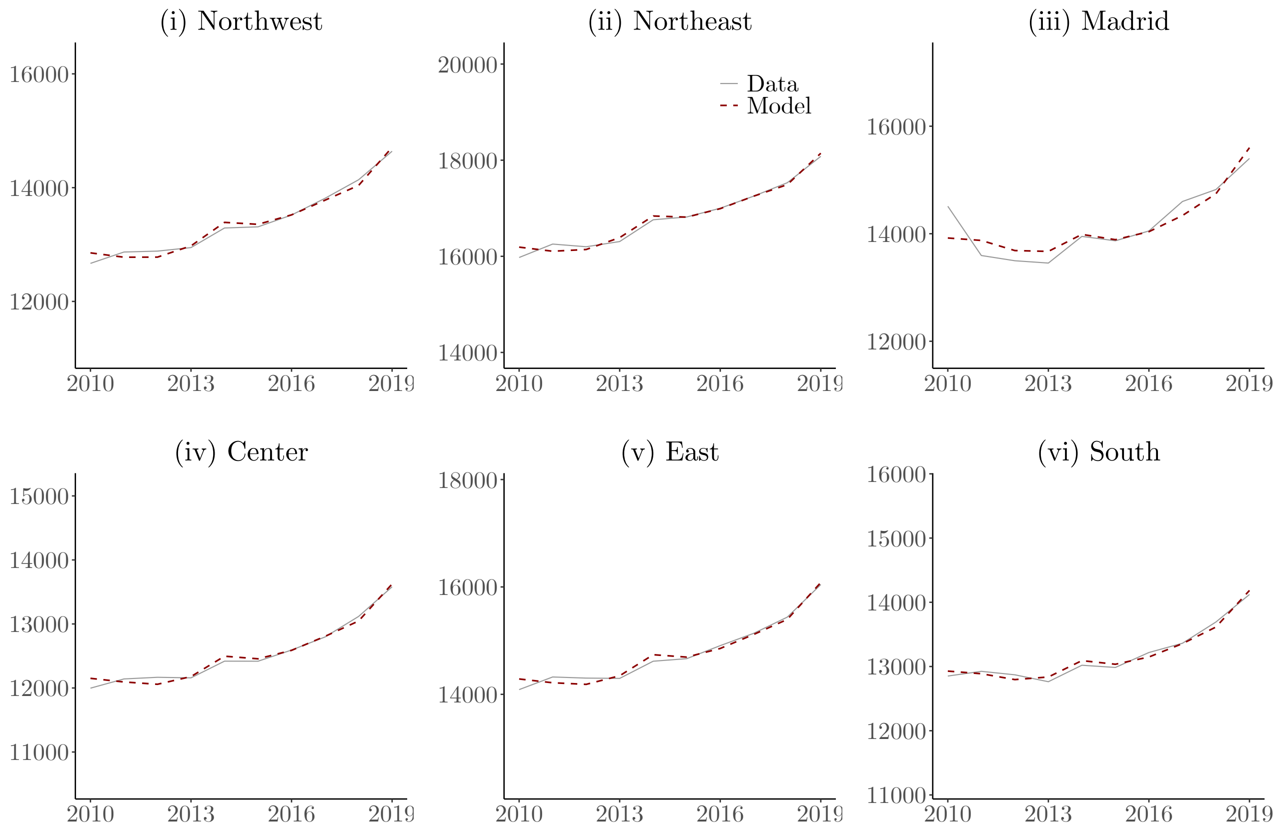
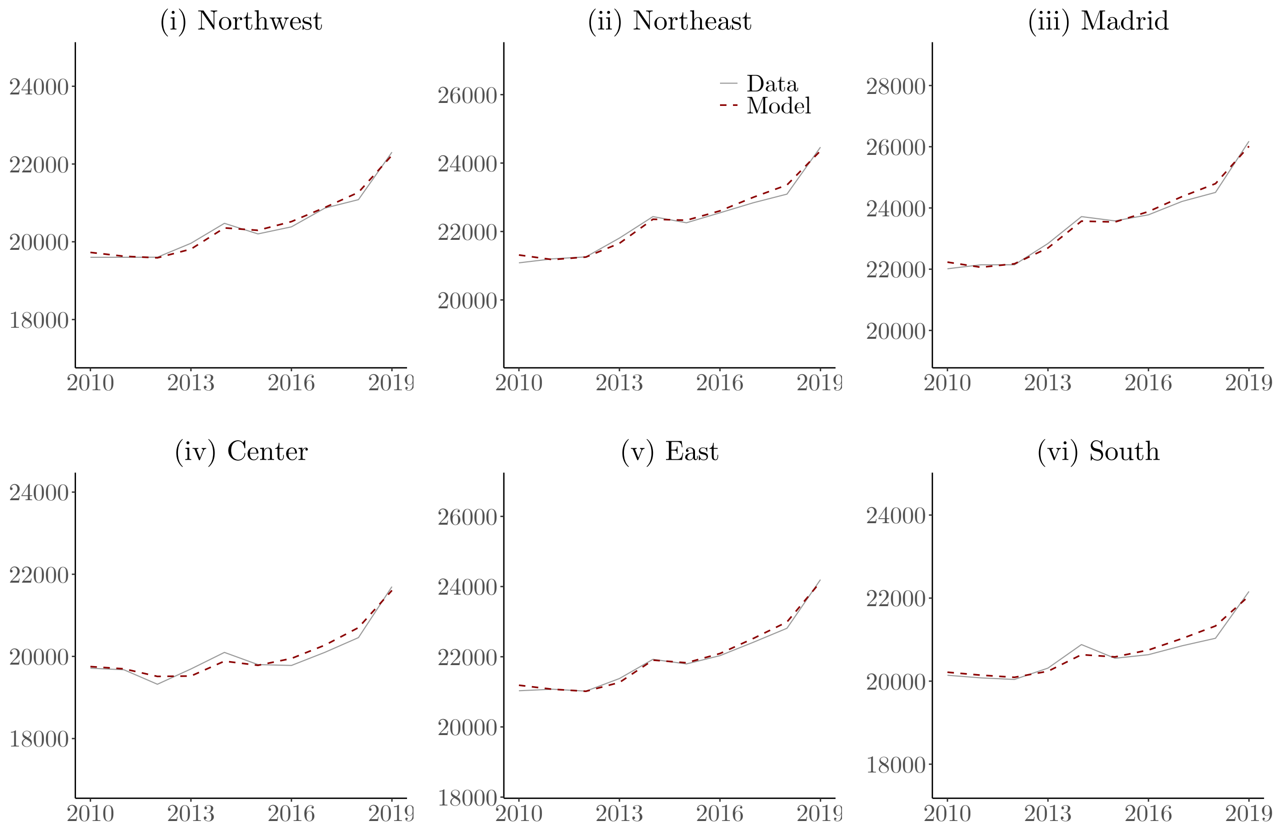


Figure D8: The figures plot mean non-college annual gross hourly wages (in thousands). Data: MCVL. [Back]

(a) College wages, urban locations



(b) College wages, rural locations

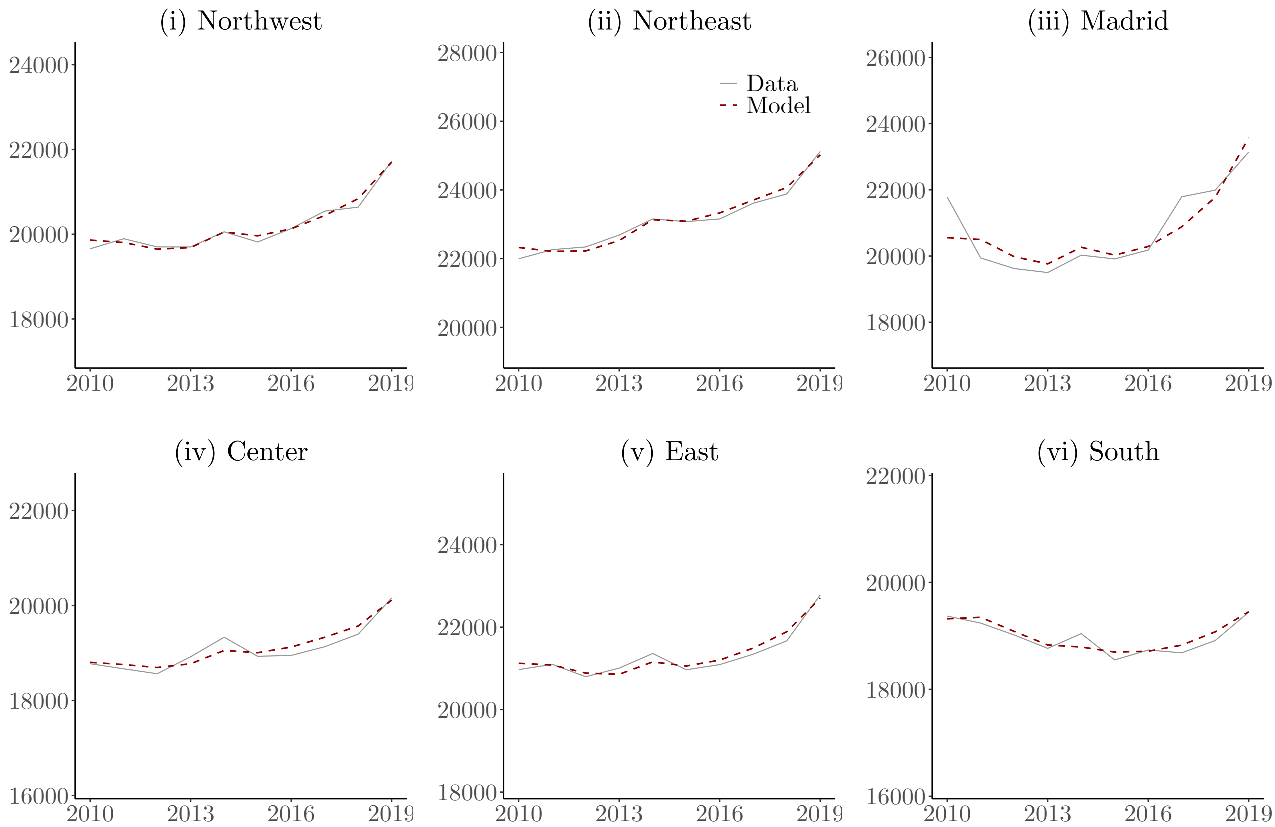


Figure D9: The figures plot mean college annual gross hourly wages (in thousands). Data: MCVL. [Back]

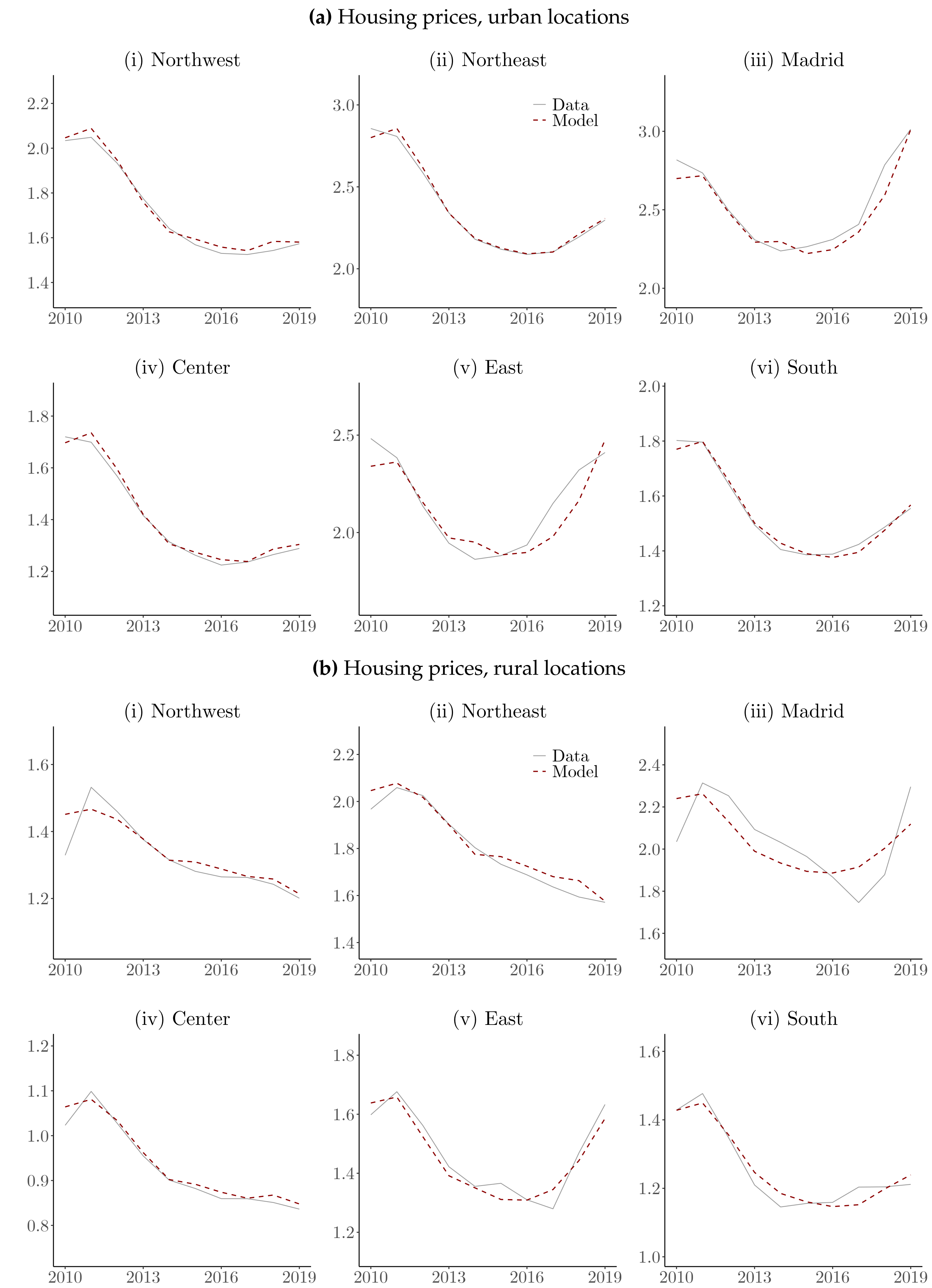


Figure D10: The figures plot mean housing prices per square meter (in thousands). Data: Idealista. [Back]

The estimated factor f_{1t} increases, whereas factor f_{2t} decreases over the 2010-2019 period. Since loading parameter $\lambda_{1d}^{w_e}$ is estimated to be positive for all locations and education types and $\lambda_{2d}^{w_e}$ is estimated to be close to zero, predicted wages increase over time across all location and education groups (Figures D8 and D9).

The evolution of housing prices, however, is more heterogenous across locations. Loading parameters λ_{2d}^{p} are positive everywhere, which, together with the decreasing evolution of the f_{2t} factor, tends to push prices down over the 2010-2019 period. Meanwhile, λ_{1d}^{p} is positive in high-productivity areas like Madrid and the East Regions, but negative in less productive areas like the Northwest and Central regions. In places where λ_{1d}^{p} is positive, the f_{1t} factor drives housing prices upwards. Consequently, by 2019, prices in these areas return to their 2010 levels, as shown in Figure D10.

D.6 Further Remarks

[Back]

The large degree of heterogeneity in terms of agents' types and aggregate states makes the model computationally costly to solve. Moreover, the discrete choice nature of the decision problem and the existence of income thresholds for the amount of taxes paid and transfers or subsidies received – whose eligibility depends on local wages and hence on the migration choice – introduce non-linearities in the value function. Therefore, the endogenous grid method developed by Carroll (2006) cannot be directly applied to speed up the estimation strategy. Hence, the model is solved using value function iteration.

For efficient model estimation, it's crucial to minimize the number of state variables. As detailed in Appendix E.3, there is no need to incorporate a housing bequest indicator. Moreover, my formulation of the mortgage repayment schedule only requires current age j and current assets a_j (see Section 3.2). Even without including additional state variables, the model yields realistic repayment patterns, with cross-sectional mortgage debt declining smoothly over the lifecycle. Additionally, in large parts of the value function iteration procedure, tracking the birthplace d_0 is not necessary. By knowing the agents' current location (state) and future location (decision), I can determine their status as natives or migrants in the subsequent period, which is enough to compute current utility.

E Estimation of Model Inputs

[Back]

E.1 Probability of Being Unemployed

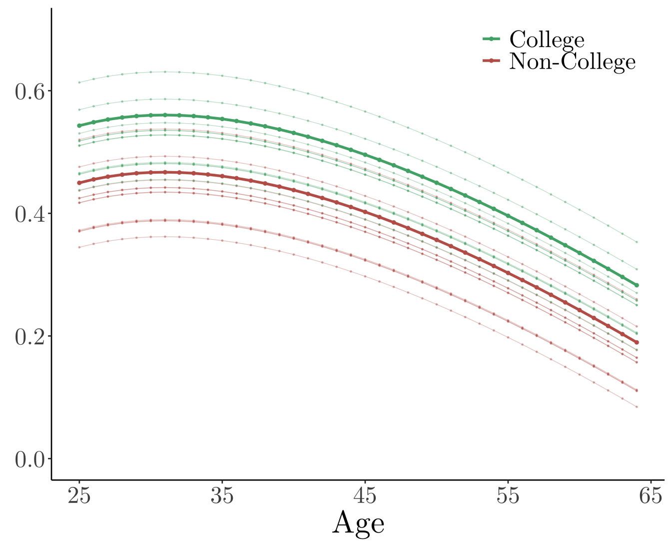
[Back]

The probability of transitioning from employment to unemployment by education and location is plotted in Figure E11. These probabilities are estimated using the EU-SILC 2004-2019 panel following the procedure described in Section 4.1.1, i.e. by running the linear probability regression models:

$$\overbrace{\mathbb{1}_{l_{edj+1}=1}}^{\text{Unemployed next period}} = \alpha_t + \alpha_e + \alpha_d + \alpha_1 j + \alpha_2 \log(j)$$
$$\underbrace{\mathbb{1}_{l_{edj+1}=2}}_{\text{Employed next period}} = \beta_t + \beta_e + \beta_d + \beta_1 j + \beta_2 \log(j).$$

Re-employment probabilities decrease sharply at older ages, so that the unemployment state becomes more persistent over the lifecycle. Similarly to the probability of becoming unemployed (Figure 3a), there is substantial variation across locations and educational groups. Re-employment probabilities are higher for the college-educated and for urban residents.

Figure E11: Employed → Unemployed
(East urban location emphasized)



Note: The figure plots age polynomial functions for different locations and education groups (East urban location is emphasized). Data: EU-SILC 2004-2019. [Back]

E.2 Income Process

[Back]

The parameters of the income process are estimated using the generalized method of moments. Following a standard choice in the literature, I use the cross-sectional variance of log income by age as moment conditions (Storesletten, Telmer and Yaron 2004).

From equation (3), residualized log income is given by

$$\begin{aligned}\tilde{u}_{ej} &= \ln \tilde{y}_{edjt} - \Upsilon_{edj} - \vartheta_{ed} \ln w_{edt} \\ &= \theta_e + z_{ej}\end{aligned}$$

Given education type e, the variance of $\tilde{u}_{ej}$ for $j \in \{1, ..., J\}$ is

$$Var(\tilde{u}_{ej}) = \sigma^2_{\theta e} + \sigma^2_{ve} \sum_{j'=0}^{j-1} \varrho_e^{2j'}.$$

Accordingly, the J moment conditions for education e are given by:

$$\begin{aligned} g_1(\tilde{\boldsymbol{u}}_{\boldsymbol{e}}, \sigma_{\theta e}, \sigma_{ve}, \varrho_e) &= \tilde{u}^2_{e1} - \sigma^2_{\theta e} + \sigma^2_{ve}, \\ g_2(\tilde{\boldsymbol{u}}_{\boldsymbol{e}}, \sigma_{\theta e}, \sigma_{ve}, \varrho_e) &= \tilde{u}^2_{e2} - \sigma^2_{\theta e} + \sigma^2_{ve}(1 + \varrho_e^2), \\ &\vdots \\ g_J(\tilde{\boldsymbol{u}}_{\boldsymbol{e}}, \sigma_{\theta e}, \sigma_{ve}, \varrho_e) &= \tilde{u}^2_{eJ} - \sigma^2_{\theta e} + \sigma^2_{ve} \sum_{j'=0}^{J-1} \varrho_e^{2j'}, \end{aligned}$$

with $\mathbb{E}\left[g(\tilde{\boldsymbol{u}}_{\boldsymbol{e}}, \sigma_{\theta e}, \sigma_{ve}, \varrho_e)\right] = 0$.

The GMM uses these moments conditions and the variance of log residual income coming from the data. The grey solid lines of Figures E12c and E12d depict these empirical variances for the non-college and college educated, respectively. Residualized income in the data is given by equation (6), and can be computed by regressing out observables from regression (5), which is estimated with the MCVL. As described in Appendix B, I remove outlier observations corresponding to the top 1% of the wage distribution. I also exclude workers whose annual income is lower than the level of unemployment benefits (3,300 in the benchmark economy). Such low income levels can be observed in the data when workers are only employed for some months of the year and are unemployed (or out of the labor force) in the rest of the year. This status is treated as unemployment in the model, and, accordingly, it does not affect the estimated income process of employed workers.

The variance of log income tends to increase with age for both college and non-college workers. However, especially for non-college workers (Figure E12c), the lifecycle pattern becomes non-monotonic from age 55 through 64. This is likely due to the pre-retirement choices of workers, which happen at different points in life for different individuals. Given this idyosincratic behavior of the last 10 ages in the model, I estimate the income process by restricting the sample to workers aged 25-54 ($J = 30$), both for the non-college and for the college educated.

The resulting lifecycle variance profile predicted by the GMM model is plotted with red dashed lines in Figures E12a and E12b (non-college and college workers), as well as in Figures E12c and E12d. The model fits the data well, both for the 25-54 age groups and for the full 25-64 group. For comparison, in Figure E13 I plot with red dashed lines the predicted lifecycle income variances that I obtain by estimating the GMM with the unrestricted 25-64

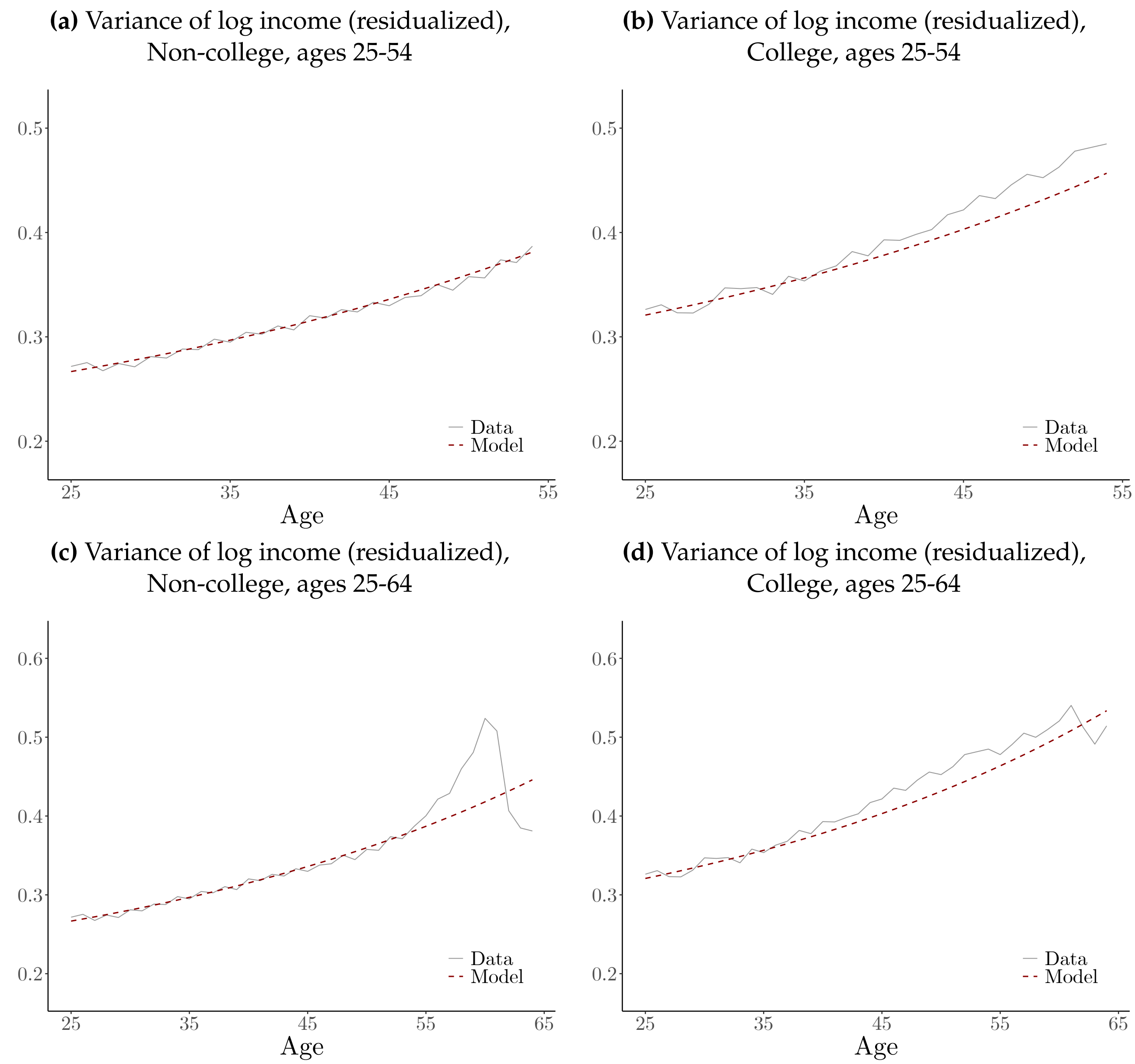


Figure E12: The figures plot the variance of log residualized income over the lifecycle for non-college and college educated workers and for different age groups (25-54 and 25-64). The solid grey line depicts the data, whereas the red dashed line represents the lifecycle variance profiles predicted by the GMM model used to estimate the income process. The model is estimated by restricting the sample to workers aged between 25 and 54. Data: MCVL 2005-2019. [Back]

sample. Due to the non-monotonic behavior observed in the data at ages 55-64, the model overpredicts earnings uncertainty at ages 35-54 for non-college workers. Overall, the estimation that uses the restricted 25-54 version of the sample appears to fit better the lifecycle evolution of the variance of log income.

E.3 Probability of Receiving Housing Bequests [Back]

The share of individuals who have ever inherited a house at different ages throughout their lives is plotted in panel (a) of Figure 4 for those without a college degree, and in Figure E14 for those with a college education. This data, sourced from the EFF 2008-2020 and grouped by birthplace (rural vs. urban), needs to be transformed into an annual probability,

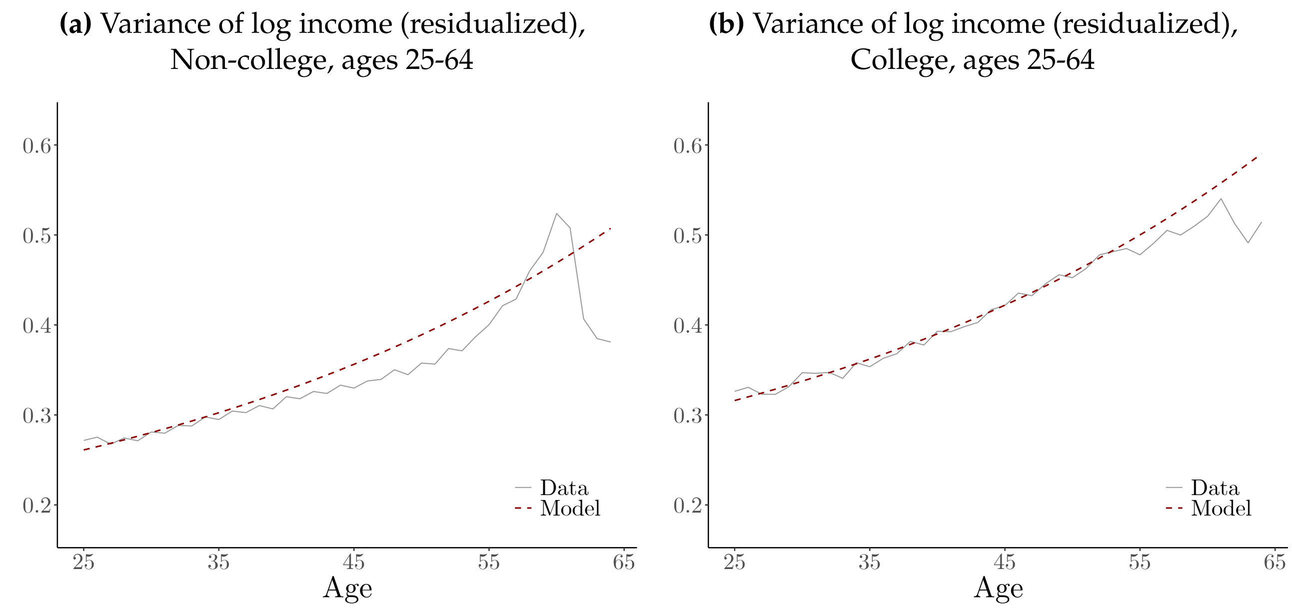


Figure E13: The figures plot the variance of log residualized income over the lifecycle for non-college and college educated workers aged 25-64. The solid grey line depicts the data, whereas the red dashed line represents the lifecycle variance profiles predicted by the GMM model used to estimate the income process. The model is estimated by restricting the sample to workers aged between 25 and 64. Data: MCVL 2005-2019. [Back]

$\pi^b_{ed_0 j}$, of receiving a housing bequest. This probability varies by types and serves as input for the model.

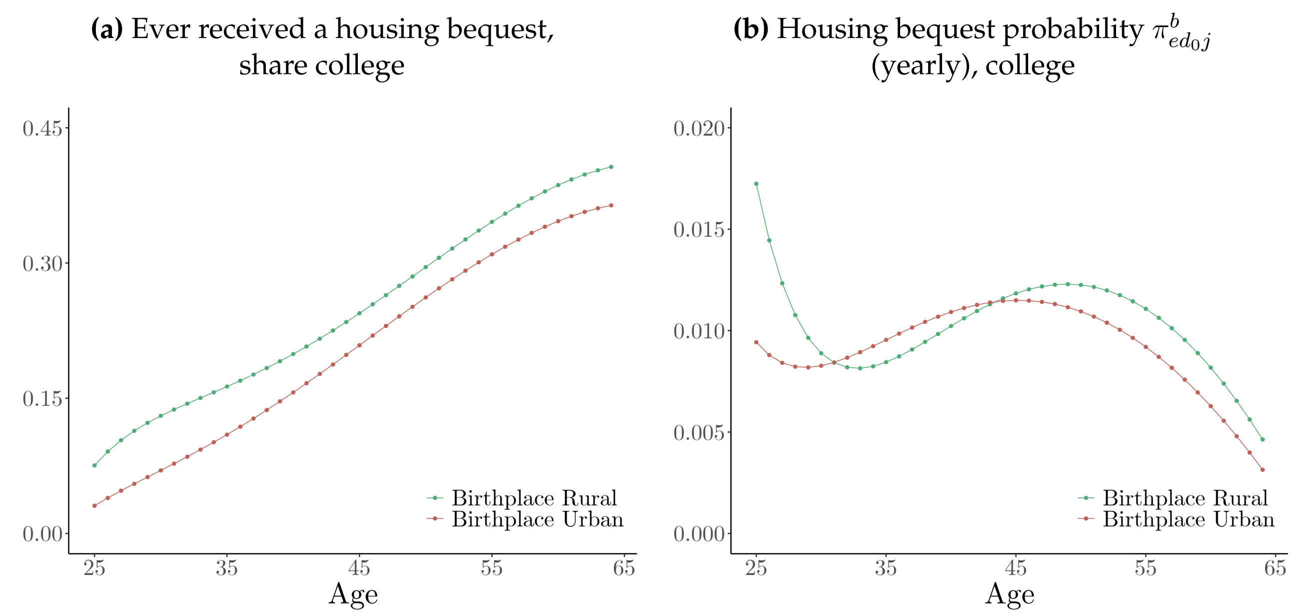


Figure E14: The figures plot the share of college-educated individuals who have ever inherited a house during their lifetime (panel a), computed age by age, and the annual probability of receiving a housing bequest (panel b). The equivalent figures for workers without a college decree are plotted in Appendix Figure 4. Data: EFF 2008-2020. [Back]

This conversion would be straightforward if one were to add an additional state variable, indicating that the agent has received a housing bequest in the past. In that case, calling $\pi^{b,\text{cum}}_{ed_0 j}$

the share of people who have ever inherited a house at age j, the probability of receiving a housing bequest would be

$$\widehat{\pi}^{b}_{ed_0 j} = \pi^{b,\text{cum}}_{ed_0 j+1} - \pi^{b,\text{cum}}_{ed_0 j} \tag{38}$$

for those who have not received a housing bequest in the past, and $\widehat{\pi}^{b}_{ed_0 j} = 0$ for those who have already received a housing bequest.

Introducing an extra state variable to the benchmark is computationally taxing. Specifically, it would require solving the model at 650 million points instead of the current 325 million. Below, I outline an approach to compute the annual probability of receiving a housing bequest, ensuring its consistency with the lifecycle profile depicted in Figure E14a, without adding an extra state variable. This method only works by allowing agents the possibility of inheriting a house multiple times over their lives, which might be viewed as receiving a significant inheritance. While comparing both options (either including or not including a state variable for housing bequests), I find that the results remain almost identical.

To adjust equation (38) to determine the annual probability of inheriting a house in a model without bequest state variables, I use:

$$\pi^{b}_{ed_0 j} = \frac{\pi^{b,\text{cum}}_{ed_0 j+1} - \pi^{b,\text{cum}}_{ed_0 j}}{1 - \pi^{b,\text{cum}}_{ed_0 j}}. \tag{39}$$

It's easy to see that this probability mirrors the cumulative patterns shown in Figure E14a when all agents have the opportunity to inherit a house multiple times. The rise in the cumulative share of people receiving a bequest by age, indeed, is driven by the fraction $(1 - \pi^{b,\text{cum}}_{ed_0 j})$ that has yet to receive one, i.e.

$$0 \times \frac{\pi^{b,\text{cum}}_{ed_0 j+1} - \pi^{b,\text{cum}}_{ed_0 j}}{1 - \pi^{b,\text{cum}}_{ed_0 j}} \times \pi^{b,\text{cum}}_{ed_0 j} + 1 \times \frac{\pi^{b,\text{cum}}_{ed_0 j+1} - \pi^{b,\text{cum}}_{ed_0 j}}{1 - \pi^{b,\text{cum}}_{ed_0 j}} \times (1 - \pi^{b,\text{cum}}_{ed_0 j}),$$

which aligns with the cumulative rise $\pi^{b,\text{cum}}_{ed_0 j+1} - \pi^{b,\text{cum}}_{ed_0 j}$ observed in the data.

The share $\pi^{b,\text{cum}}_{ed_0 j}$ is estimated from the EFF data. In this dataset, the heads of surveyed households report if they have ever inherited one or more properties. Being a household-level survey, the probability must be adjusted for coresidents. Specifically, in the EFF data we can only estimate the probability of receiving a housing bequest conditional on not coresiding, denoted by $\pi^{b,h_{j+1}>0}_{ed_0 j+1}$. Yet, our goal is the unconditional probability, $\pi^{b}_{ed_0 j+1}$. However, note that:

$$\begin{aligned}
\pi^{b}_{ed_0 j+1} &= \mathbb{E}(\mathbb{1}\{h_{j+1} = 0\}) \times \pi^{b,h_{j+1}=0}_{ed_0 j+1} + \mathbb{E}(\mathbb{1}\{h_{j+1} > 0\}) \times \pi^{b,h_{j+1}>0}_{ed_0 j+1} \\
&= \mathbb{E}(\mathbb{1}\{h_{j+1} > 0\}) \times \pi^{b,h_{j+1}>0}_{ed_0 j+1}
\end{aligned}$$

since $\pi^{b,h_{j+1}=0}_{ed_0 j+1} = 0$, i.e. people who are currently coresiding with their parents cannot have received a housing bequest from them in the previous period. $\mathbb{E}(\mathbb{1}\{h_{j+1} > 0\})$ is measured in the 2011 Census using the share of people coresiding by age. Thus, converting the conditional housing inheritance probability to the unconditional one is straightforward.

As a concluding step, the final probability of receiving a housing bequest is achieved by evenly distributing the housing inheritance probability for newborns, $\pi^b_{ed_0 1}$, throughout the lifecycle. This strategy prevents abrupt shifts in the lifecycle probability patterns due to starting conditions. The resulting lifecycle probability $\pi^b_{ed_0 j}$ is depicted in Figure 4b for people without a college education and in Figure E14b for the college-educated.

E.4 Other Inputs and Moments [Back]

E.4.1 Migration Shares by Destination [Back]

Migration shares by destination are computed using the MCVL. This administrative dataset is preferred over the EU-SILC due to its greater number of observations. Migration events to certain Spanish locations are relatively infrequent, and using a dataset with millions of observations reduces sampling error.

Table E3: Amenity estimates and targeted migration shares by destination. [Back]

	Parameter	Estimate
Amenities		
Northwest Rural	A_{Nwr}	-0.124
Northwest Urban	A_{Nwu}	-0.121
Northeast Rural	A_{Ner}	-0.172
Northeast Urban	A_{Neu}	-0.322
Madrid Rural	A_{Mr}	-0.029
Madrid Urban	A_{Mu}	0.009
Center Rural	A_{Cr}	-0.070
Center Urban	A_{Cu}	-0.152
East Rural	A_{Er}	0.000
East Urban	A_{Eu}	-0.007
South Rural	A_{Sr}	-0.046
South Urban	A_{Su}	-0.029

	Data	Model
Migration shares		
Northwest Rural	0.043	0.043
Northwest Urban	0.040	0.056
Northeast Rural	0.030	0.029
Northeast Urban	0.024	0.018
Madrid Rural	0.057	0.058
Madrid Urban	0.157	0.150
Center Rural	0.106	0.114
Center Urban	0.061	0.060
East Rural	0.174	0.158
East Urban	0.166	0.151
South Rural	0.059	0.061
South Urban	0.085	0.100

E.4.2 Share of Never-Movers [Back]

This section outlines the procedure used to infer the share of never-movers from EU-SILC and MCVL data. The procedure combines information on the annual migration rate from the EU-SILC with the lifecycle migration patterns inferred from the longer MCVL panel.

Let $\widehat{\text{moved}}_0, \widehat{\text{moved}}_1, ..., \widehat{\text{moved}}_{\overline{k}}$ be the share of people that have migrated a total of $k = 0, 1, ..., \overline{k}$ times, respectively, over the period of K years observed in the data. By construction, in the context of yearly migration events, $\overline{k} \leq K$. Since one extra year is needed to compute migration episodes, the EU-SILC (a 4-years rotating panel) has $K = 3$ whereas the MCVL (a 14-years panel) has $K = 13$.

The ideal dataset for computing $\widehat{\text{moved}}_0$, the share of people that never moved over the lifecycle, would have $K = J = 40$. I leverage the extended panel dimension of the MCVL panel to approximate this ideal dataset. Define the ratio between the number of individuals migrating more than once ($k > 1$) and those migrating just once as

$$\overline{m}_k = \frac{\widehat{\text{moved}}_k}{\widehat{\text{moved}}_1}.$$

Then, the annual migration rate can be computed as

$$\begin{aligned} \text{Annual Migration Rate} &= \frac{\widehat{\text{moved}}_1 + 2 \times \widehat{\text{moved}}_2 + ... + \overline{k} \times \widehat{\text{moved}}_{\overline{k}}}{K} \\ &= \widehat{\text{moved}}_1 \left(\frac{1 + 2 \times \overline{m}_2 + ... + \overline{k} \times \overline{m}_{\overline{k}}}{K} \right) \end{aligned}$$

The share of people moving only once, $\widehat{\text{moved}}_1$, that is consistent with the EU-SILC annual migration rate of 0.0082 and the MCVL lifecycle migration quantities $\overline{m}_2, ..., \overline{m}_{\overline{k}}$, and K is then given by

$$\widehat{\text{moved}}_1 = K \times \frac{\text{Annual Migration Rate}}{(1 + 2 \times \overline{m}_2 + ... + \overline{k} \times \overline{m}_{\overline{k}})} = 0.0569.$$

Finally, the share of never-movers can be computed as

$$\widehat{\text{moved}}_0 = 1 - \widehat{\text{moved}}_1 (1 + 2 \times \overline{m}_2 + ... + \overline{k} \times \overline{m}_{\overline{k}}) = 0.8938.$$

E.4.3 Housing Sizes

[Back]

Fixed housing sizes by housing tenure, education, and location (urban vs. rural) are estimated using simple conditional averages from EPF 2016-2019. As can be seen in Table E4, house sizes tend to be larger for homeonwers, the college-educated, and rural residents.

Table E4: Housing sizes (in square meters) by tenure, education, and location. Data: EPF 2016-2019. [Back]

	Homeowner		Renter	
	College	Non-College	College	Non-College
Rural	129	116	92	84
Urban	106	93	83	76

E.4.4 Initial Conditions

[Back]

Agents are born with four fixed types: birthplace location d_0, education level e, migration type τ, and individual-level fixed productivity θ_e, allowed to vary by education. Birthplace and education types are drawn from two empirical categorical distributions estimated with Census 2011 data. Migration types are drawn from a Bernoully distribution with calibrated parameter π_τ, the probability of drawing a "stayer" type. Finally, θ_e is drawn from from two education-specific normal distributions with standard deviations σ_{θ_N} and σ_{θ_E}, estimated with the income process using MCVL 2005-2019 data (Appendix E.2).

In the first period of the model, individuals also draw initial assets a, location d and housing status h. Initial assets follow a log normal distribution estimated using non-housing net wealth data from the EFF 2005-2020. The categorical distributions for locations and housing status are instead estimated with the 2011 Census. The initial housing tenure shares (coresidents, renters, and homeonwers) are separately estimated by native status, as natives are more likely to coreside than migrants. Agents are born with some mortgage debt in case they enter the model as homeowners. Debt is computed using EFF 2005-2020 data, separately by education level. Initial mortgage debt is a fixed share of housing wealth.

A fraction of migrants effectively behave as natives: they derive home-bias utility from living in the initial location and, as long as they stay in that location, can coreside with parents.[55] The probability of belonging to this group of migrants is given by the ratio of the share of migrant coresidents at age 25 over the share of native coresidents at age 25, computed separately within education groups and locations. Finally, initial employment status by education and location is drawn from a categorical distribution estimated using EU-SILC 2004-2019 data.

[55]This is needed to match the fact that some migrants coreside in the data (see Figure 2c). An alternative strategy would be to classify as natives all migrants that were coresiding at 25 (the first age in the model), since these people are likely able to also coreside in the future by staying in the same city (e.g., because their parents migrated with them before they turned 25). Yet, due to the constraints in available data, this method cannot be adopted. The census data is cross-sectional, lacking information about, for example, whether a 45-year-old migrant who is currently a non-coresident had previously coresided in the same city at the age of 25.

F Validation [Back]

F.1 Place-Based Subsidy for Young Homebuyers [Back]

Policy in The Data As a robustness exercise, I estimate two placebo event analyses on individuals aged 37-40, just beyond the age eligibility for the subsidy. Treatment and control definitions remain consistent with the main regressions. As expected, these placebo treatments don't significantly impact the outcomes (see Figures F15a and F15b).

Additionally, I restrict the sample to only include people *born* in municipalities with populations under 20,000, and treat people who were born in locations with less than 10,000 inhabitants. Results are similar to the baseline event studies (Figure F16). Event-study versions including COVID-19 years up to 2021 are in Figures F15c and F15d.

The difference-in-differences regression takes the form:

$$y_{it} = \alpha_{\text{small}} + \tau \text{Treated}_{it} + \alpha_r + \alpha_{rt} + \alpha X_{it} + \epsilon_{it}, \tag{40}$$

where y_{it} is the outcome, which is either the homeownership status (first stage) or the individual migration event (reduced form), α_{small} is an indicator function for the set of treated cities (municipalities with less than 10,000 inhabitants), τ is the coefficient associated to the treatment, an indicator equal to one for years 2018-2019 and small municipality residents, and zero otherwise, whereas α_r and α_{rt} are region and region-year fixed-effects, respectively. In the baseline specification, additional controls X_{it} include gender, age and age squared. The sample restriction is the same as in the event studies (regression (14)): individuals younger than 35 living in cities with less than 20,000 inhabitants, so that the control group is composed of people living in cities with size just above the policy population threshold. The data covers the years 2016 to 2019, encompassing two years before and after the implementation of the policy in early 2018.

Policy in The Model After buying a house with the subsidy, the policy allows the homebuyer to resell the property under certain circumstances: if they are relocating for work, if they are buying a new house in the same or a different area, or once five years have passed since receiving the subsidy. Conversely, if the individual chooses to migrate before this five-year period without satisfying one of these conditions, they are obligated to pay back a proportionate amount of the subsidy. For example, migrating after just two years without meeting any of the exceptions would require a 5,400 euro repayment of the subsidy.

There repayment rules and exceptions are reproduced when simulating the policy in the model. First, I add an additional state variable that takes on three values: not eligible for the policy, eligible and not recipient, and recipient. This differentiation is required as individuals who have already received the subsidy stop being eligible. Additionally, while subsidy recipients might need to reimburse a part of the subsidy if they migrate, non-recipients don't face this constraint.

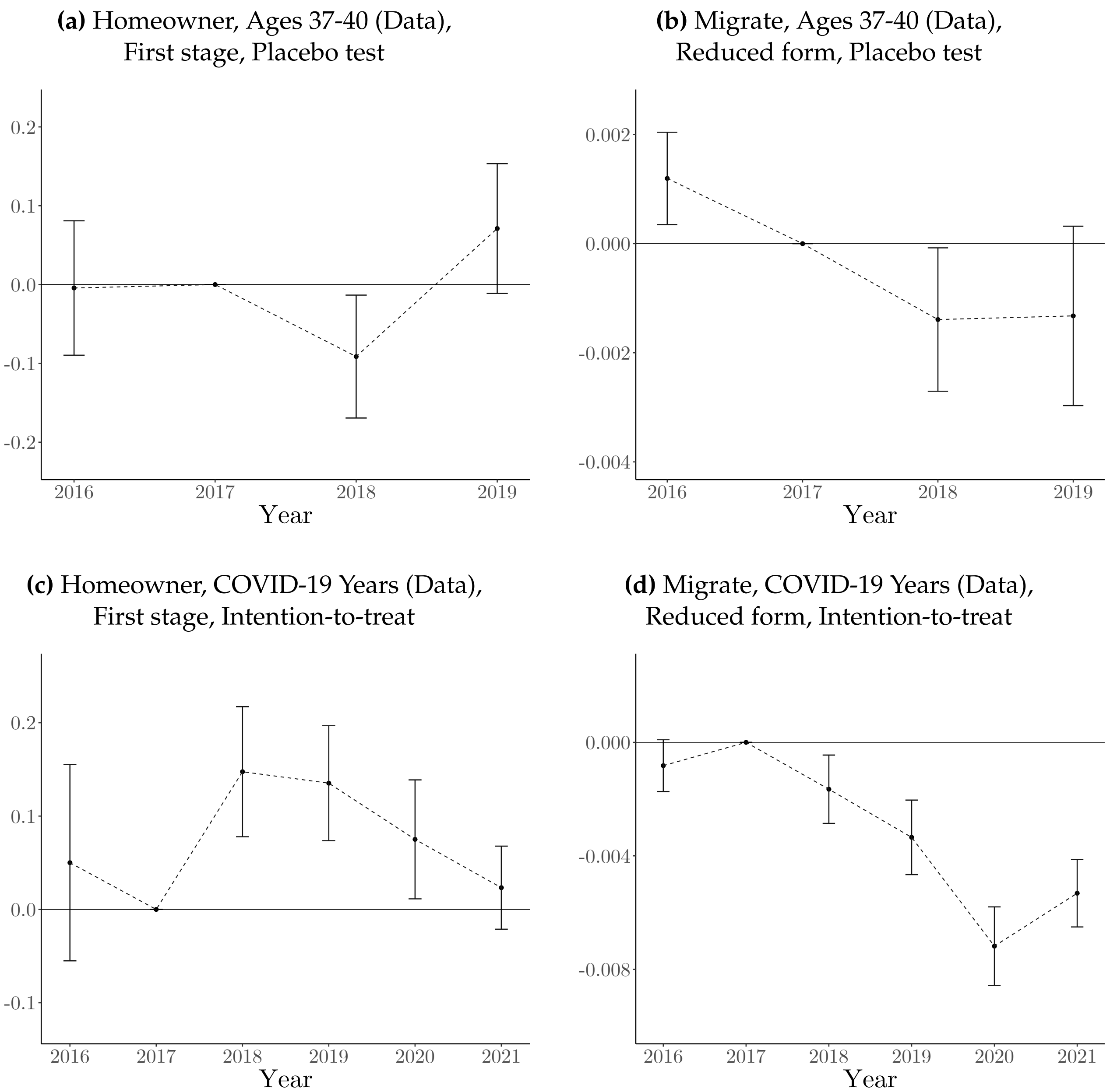


Figure F15: Treated: People living in small cities (<10k inhabitants), aged between 37 and 40 (panels **(a)** and **(b)**) or less than 35 (panels **(c)** and **(d)**). Control group: same age living in slightly larger cities (10k-20k). Treatment year: 2018. Migration across regions (NUTS-1) and rural-urban areas (rural <10k, rural >10k or urban >40k). Other included controls and fixed-effects: gender (in the data), age, age squared, region, and region-year. Clustered (locations) standard errors. Data: EPF, EPC, EVR 2016-2019. [Back]

Keeping track of the years since the policy was received, to ensure the five-year threshold is met, would be computationally too costly. However, I can use the mortgage repayment schedule, the agents' age, and the current level of assets to infer the years since the property's purchase. In particular, I compute the level of negative assets realized in equilibrium for people with a mortgage, averaged by location, education, age, and years since the house was bought. These averages define the thresholds above which I classify individuals with the corresponding types (location, education, and age) as having been homeowners for a certain amount of years.

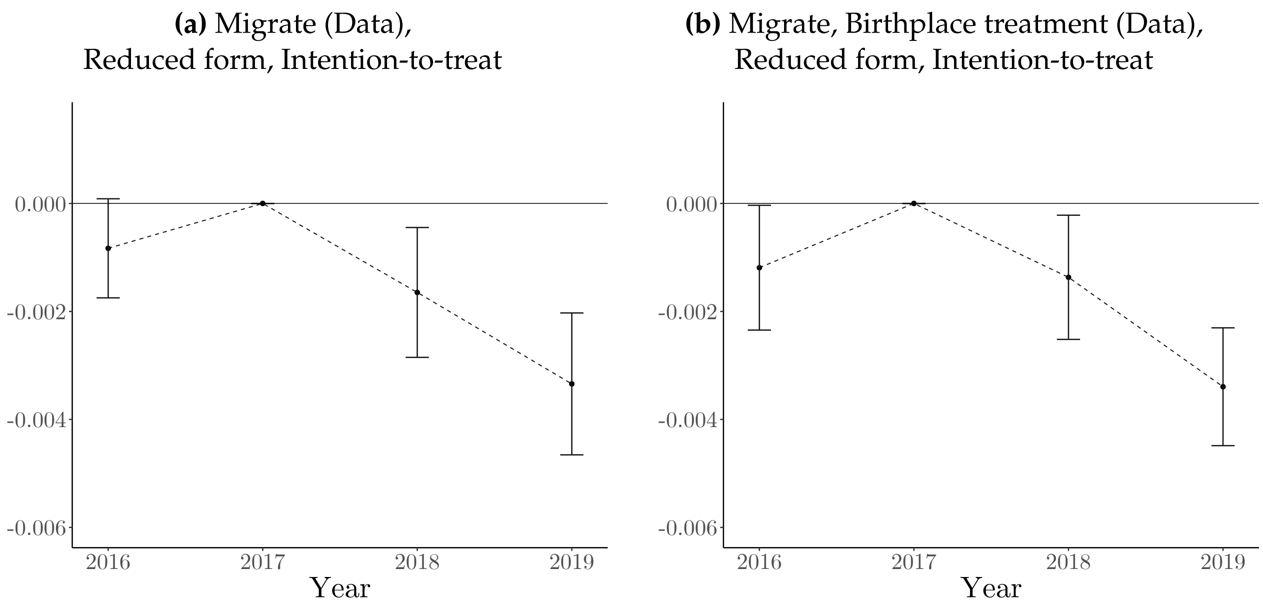


Figure F16: Treated: People aged less than 35 living (panel **(a)**) in small cities (<10k inhabitants) or born (panel **(b)**) in small cities. Control group: same age living, or born, in slightly larger cities (10k-20k). Treatment year: 2018. Migration across regions (NUTS-1) and rural-urban areas (rural <10k, rural >10k or urban >40k). Other included controls and fixed-effects: gender (in the data), age, age squared, region, and region-year. Clustered (locations) standard errors. Data: EPF, EPC, EVR 2016-2019. [Back]

Table F5: Regressions (1) and (2) also include a control for gender. Standard errors are clustered at the location level, **p*<0.1, ***p*<0.05, ****p*<0.01. Data: EPF, EPC, EVR 2016-2019. [Back]

	Data, Treated Locations:		Model, Treated Individuals:	
	Homeowner	Migrate	Homeowner	Migrate
	(1)	(2)	(3)	(4)
Treated	0.1151**	-0.0021***	0.6697***	-0.0141***
	(0.0502)	(0.0006)	(0.0156)	(0.0034)
Age	0.0150	0.0206***	0.2736***	-0.0004
	(0.0645)	(0.0018)	(0.0193)	(0.0055)
Age^2	0.0003	-0.0004***	-0.0154***	-0.0002
	(0.0011)	(0.0000)	(0.0011)	(0.0003)
Treated Locations FE	✓	✓		
Individual FE			✓	✓
Year-Region FE	✓	✓	✓	✓
Region FE	✓	✓	✓	✓
R^2	0.09850	0.00314	0.68957	0.36216
Observations	2,056	4,533,392	17,363	17,363

Recipients of the policy who move before the five-year mark and don't buy a house in the new location, are forced to make a proportional repayment of the policy. In the model, I cannot distinguish between work and non-work related moves, whereas the policy allows recipients to migrate if the reason is job-related. Non-work related moves among homeowners aged 25-40 that do not buy a house elsewhere in the following year are 71% of the total in the U.S. (CPS 2009-2019).[56] Thus, in the model, when repayments arise, they are adjusted down by 29%. Finally, I simulate the subsidy exclusively during the years it was implemented in

[56] No similar survey data is available for Spain.

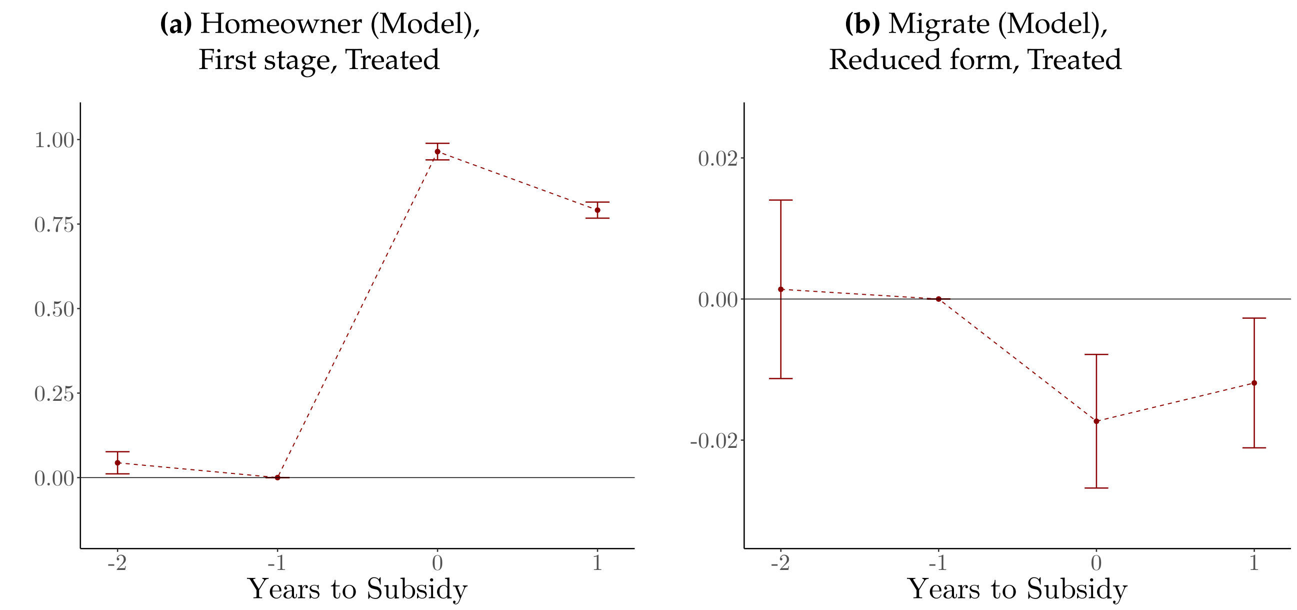


Figure F17: Treated: People who received the subsidy in the counterfactual model with the policy. Control group: same people as the treated, but in the baseline model without the policy. Included controls and fixed-effects: age, age squared, region, and region-year. Clustered (locations) standard errors. [Back]

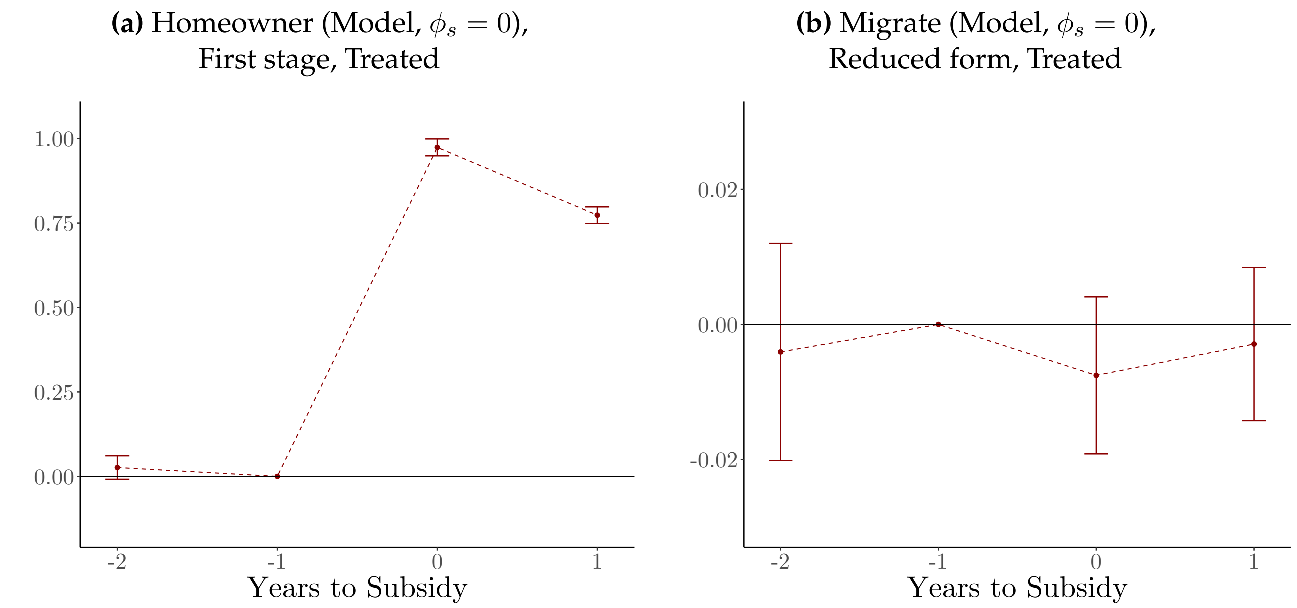


Figure F18: Treated: People who received the subsidy in the counterfactual model with the policy and $\phi_s = 0$. Control group: same people as the treated, but in the baseline model without the policy. Included controls and fixed-effects: age, age squared, region, and region-year. Clustered (locations) standard errors. [Back]

Spain (2018-2021). This is done to estimate the short-term effects of the policy and mimic the event study design estimated in the data.

Figure F18 plots the simulated effects of the policy in a counterfactual exercise without transaction costs of selling houses ($\phi_s = 0$). The policy increases homeownership without significantly influencing migration rates in the reduced form. Therefore, I conclude that transaction costs ϕ_s are the key driver of reduced migration among homeowners.

F.2 Coresidence Explains Higher Homeownership Among Natives [Back]

Figure F19 plots the untargeted lifecycle profiles of homeownership rate and income, separately for natives and migrants, in the counterfactual economy where the is no coresidence option. When I take away this option, I find that both the homeownership and income gaps between natives and migrants vanish.

Coresidence offers natives a way to overcome the mortgage down-payment friction. By saving while they are living with their parents, they can obtain sufficient funds for the down-payment, which leads to subsequent higher homeownership rates. The higher income observed among migrants largely serves as a compensating differential offsetting this benefit enjoyed by natives.

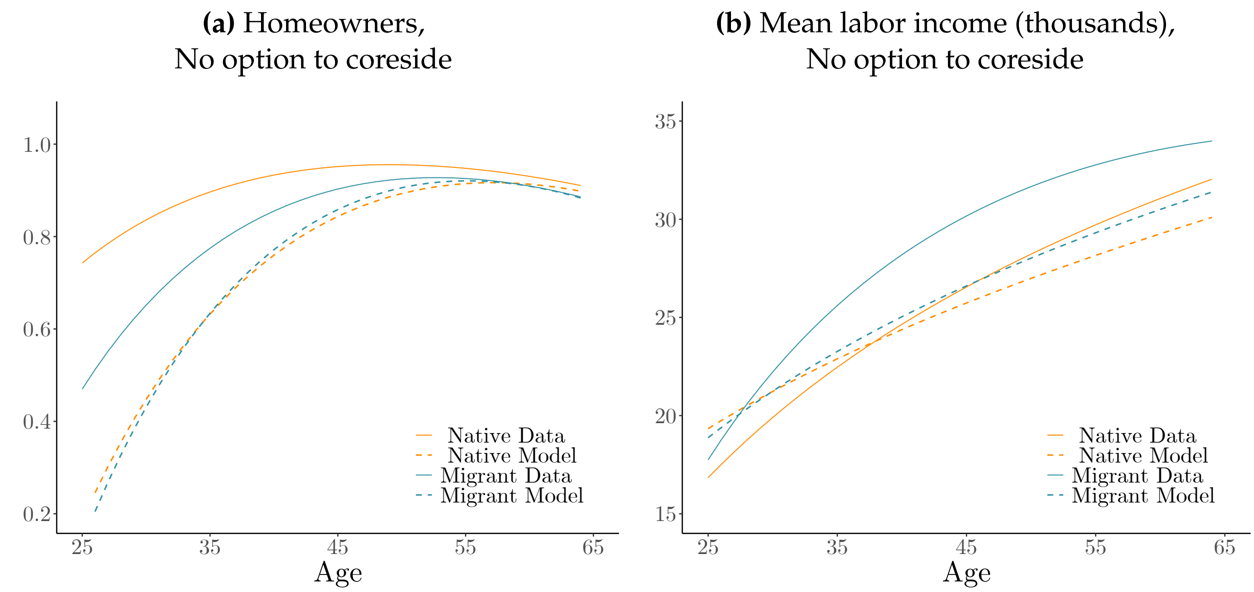
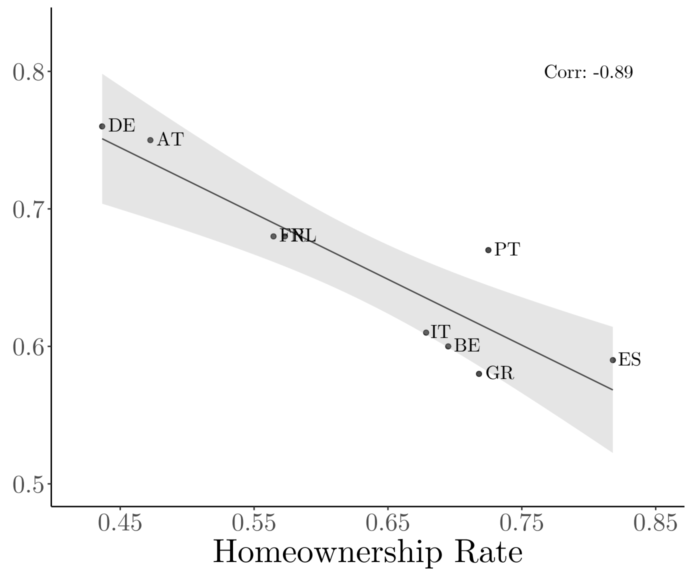


Figure F19: The figures plot the lifecycle profiles of homeownership rate and income, separately for natives and migrants, in the counterfactual economy where the is no coresidence option. Data: Census 2011 (panel **a**), MCVL 2005-2019 (panel **b**).

F.3 Why Does Homeownership Reduce Wealth Inequality? [Back]

Figure F20 plots the relationship between the local Gini of net wealth and the local homeownership rate across large European countries. The source is Figure 1 in Kaas, Kocharkov and Preugschat (2019). They use data coming from the Household Finance and Consumption Survey (HFCS), which is the European equivalent of the EFF data I use to compute the same relationship across Spanish locations (Figure 7).

Figure F20: Gini of net wealth



Note: The figure plots the relationship between the local Gini of net wealth and the local homeownership rate among not coresidents across large European countries. Source: Kaas, Kocharkov and Preugschat (2019), using HFCS 2013-2016 data. [Back]

Figure F21 plots the unconditional quantile effects of homeownership on net wealth along the wealth distribution (Firpo, Fortin and Lemieux 2009). Figure F21a plots the estimates for the entire distribution, whereas Figure F21b focuses on the top 50%. The estimated unconditional quantile regression takes the form:

$$\text{RIF}\left\{y_{it}, \text{Quantile}_k(F_{y_{it}})\right\} = \alpha_i + \alpha_t + \tau \text{Homeowner}_{it} + \alpha \text{X}_{it} + \epsilon_{it}$$

for quantiles $k \in \{10, 20, ..., 90\}$. The outcome variable RIF $\{y_{it}, \text{Quantile}_k(F_{y_{it}})\}$ measures the influence of the individual observation indexed by i and t on the unconditional quantile k of the net wealth distribution. We are interested in parameter τ, the effect of homeownership on the recentered influence function. The regression includes household and year fixed-effects, α_i and α_t, and individual controls X_{it}: household income and number of members, household head's age and age squared, and indicators for self-employed, college-educated, married, and parent. The data used to estimate the regression comes from EFF 2005-2020.

As in Kaas, Kocharkov and Preugschat (2019), unconditional quantile effects are reported in terms of semi-elasticities, by dividing them by each corresponding quantile value. Household fixed-effects are included. As can be seen in Figure F21, homeownership is found to increase net wealth along the entire distribution and, in relative terms, it does so especially among the wealth poor. Becoming homeowners increases the net wealth by a factor of 6 in the first decile and more than doubles it in the second decile. The increase in net wealth for people in the top decile of the distribution is comparatively much lower, amounting to about 4%.

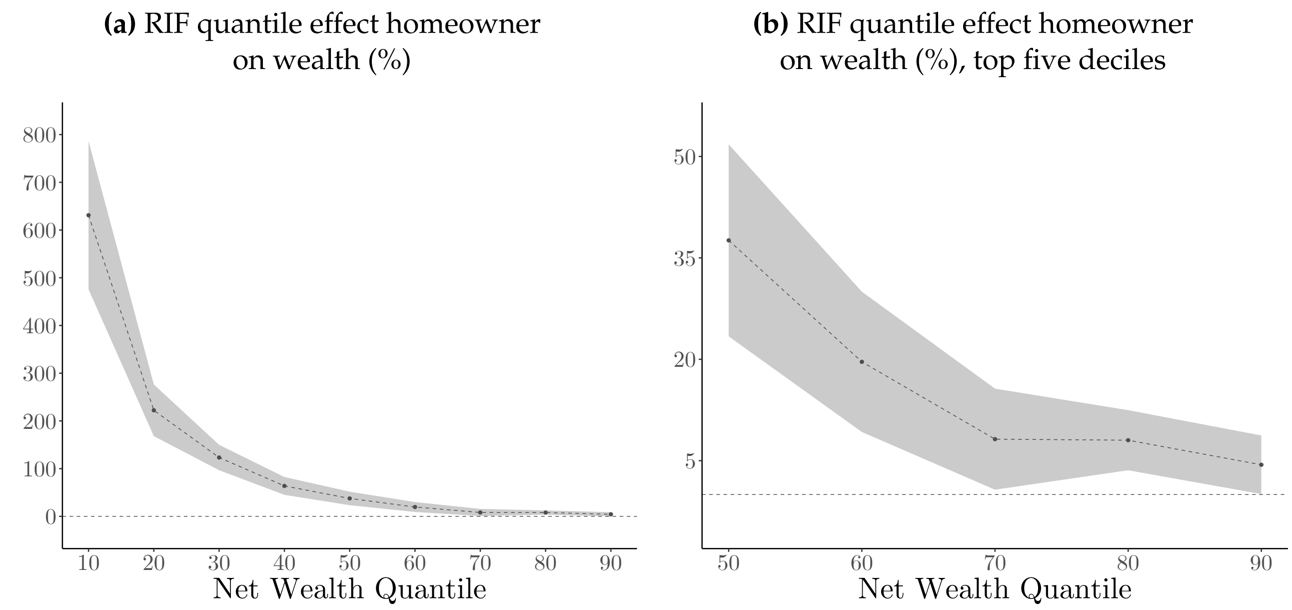


Figure F21: The figures plot the unconditional quantile effects of homeownership on net wealth along the wealth distribution (Firpo, Fortin and Lemieux 2009). Quantile effects are reported in terms of semi-elasticities, by dividing them by each corresponding quantile value (Kaas, Kocharkov and Preugschat 2019). Household and year fixed-effects are included. The other included controls are household income and number of members, household head's age and age squared, and indicators for self-employed, college-educated, married, and parent. Data: EFF 2005-2020. [Back]

Figure F22 plots the Recentered Influence Function (RIF) of the net wealth Gini coefficient by quantile (Firpo, Fortin and Lemieux 2009), computed using EFF 2005-2020 data. The recentered influence function measures the predicted effect of each individual observation on the overall Gini of net wealth. Positive values in the bottom and top two deciles of the wealth ditribution reveal that increasing the mass of households in the bottom or top 20% increases inequality. Conversely, the negative RIF values from quantile 20 to 80 tell us that moving households towards the middle 60% of the distribution reduces the net wealth Gini coefficient.

To estimate the effect of homeownership on wealth inequality, I run the following regression:

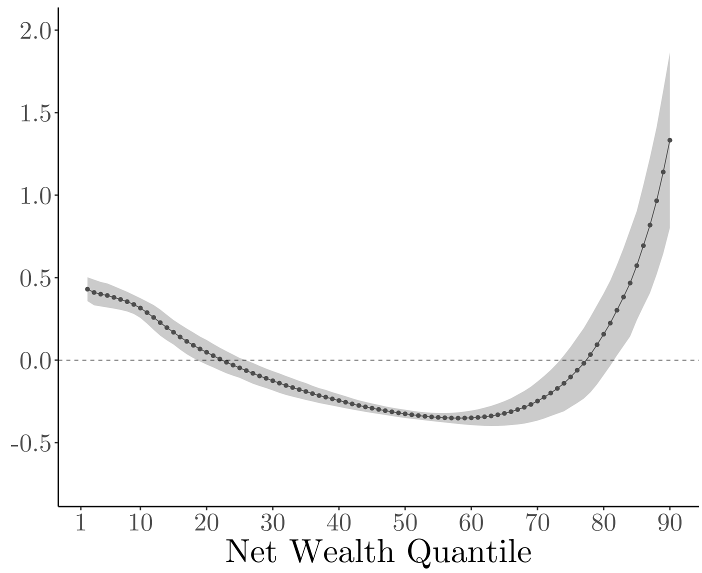
$$\text{RIF}\left\{y_{it}, \text{Gini}(F_{y_{it}})\right\} = \alpha_i + \alpha_t + \tau \text{Homeowner}_{it} + \alpha X_{it} + \epsilon_{it}, \tag{41}$$

where $\text{RIF}\left\{y_{it}, \text{Gini}(F_{y_{it}})\right\}$ measures the influence of the individual observation indexed by i and t on the overall Gini of net wealth, α_i and α_t are household and year fixed-effects, and controls X_{it} include household income and number of members, household head's age and age squared, and indicators for self-employed, college-educated, married, and parent. The data used to estimate the regression comes from EFF 2005-2020.

Results are reported in Table F6. Homeownership is estimated to reduce wealth inequality, both when household fixed-effects are not included (column 1) and when they are included (column 2). In the baseline specification with household fixed-effects, a 10 percentage points increase in the homeownership rate is estimated to reduce the Gini of net wealth by around -0.019 points.

In column (2) of Table F7, I report results for a regression that is analogous to (41), with

Figure F22: Recentered Influence Function of Gini net wealth



Note: The figure plots the Recentered Influence Function of the net wealth Gini coefficient by quantile (Firpo, Fortin and Lemieux 2009). Data: EFF 2005-2020. [Back]

Table F6: Year fixed-effects are always included, whereas household fixed-effects are only included in column (2). Other included controls are: household income level, indicators for college-educated and married, and other demographic variables (age, age squared, number of household members, indicators for parent and self-employed). Heteroskedasticity-robust standard errors, *p<0.1, **p<0.05, ***p<0.01. Data: EFF 2005-2020. [Back]

	RIF of Gini net wealth	
	(1)	(2)
Homeowner	-0.4273***	-0.1938***
	(0.0600)	(0.0521)
Income	0.0165**	0.0151
	(0.0081)	(0.0114)
College	-0.4530*	0.0020
	(0.2569)	(0.0616)
Married	-0.1340*	0.0100
	(0.0730)	(0.0431)
Demographic Variables	✓	✓
Household FE		✓
Year FE	✓	✓
R^2	0.11	0.77
Observations	20,783	20,783

the only difference that the homeownership treatment is interacted with a categorical variable measuring the level of households' wealth in the first available panel period: bottom 10%, between 10% and 20%, or above the bottom 20% of the wealth distribution. This specification reveals that the average negative effect of homeownership on the Gini of net wealth, reported in column (1) for comparison, is fully driven by the bottom of the distribution. In particular, the estimated impact is highest in the bottom 10%, is still negative and significant for household with initial wealth between the first and second decile, and loses significance for richer households.

Table F7: Year and household fixed-effects are included. Other included controls are: household income level, indicators for college-educated and married, and other demographic variables (age, age squared, number of household members, indicators for parent and self-employed). Heteroskedasticity-robust standard errors, *p<0.1, **p<0.05, ***p<0.01. Data: EFF 2005-2020. [Back]

	RIF of Gini net wealth	
	(1)	(2)
Homeowner	-0.1938***	
	(0.0521)	
Homeowner, Wealth$_{t=1}$ < p10		-0.3978***
		(0.1129)
Homeowner, Wealth$_{t=1}$ $\in$ [p10, p20]		-0.2179***
		(0.0625)
Homeowner, Wealth$_{t=1}$ > p20		-0.0184
		(0.1231)
Demographic Variables	✓	✓
Household FE	✓	✓
Year FE	✓	✓
R^2	0.77	0.77
Observations	20,783	20,783

Figure F23 plot the share of local net wealth accounting to the top and bottom 20% by location, as simulated in the benchmark model. The horizzontal axis measures the simulated local homeownership rate. Locations in the model with the highest homeownership rate have on average a 2 pp. higher share of local net wealth accounting to the bottom 20% of the distribution, and around a 2 pp. lower share accounting to the top 20%.

(a) Local share of net wealth in bottom 20% by location (Model)

(b) Local share of net wealth in top 20% by location (Model)

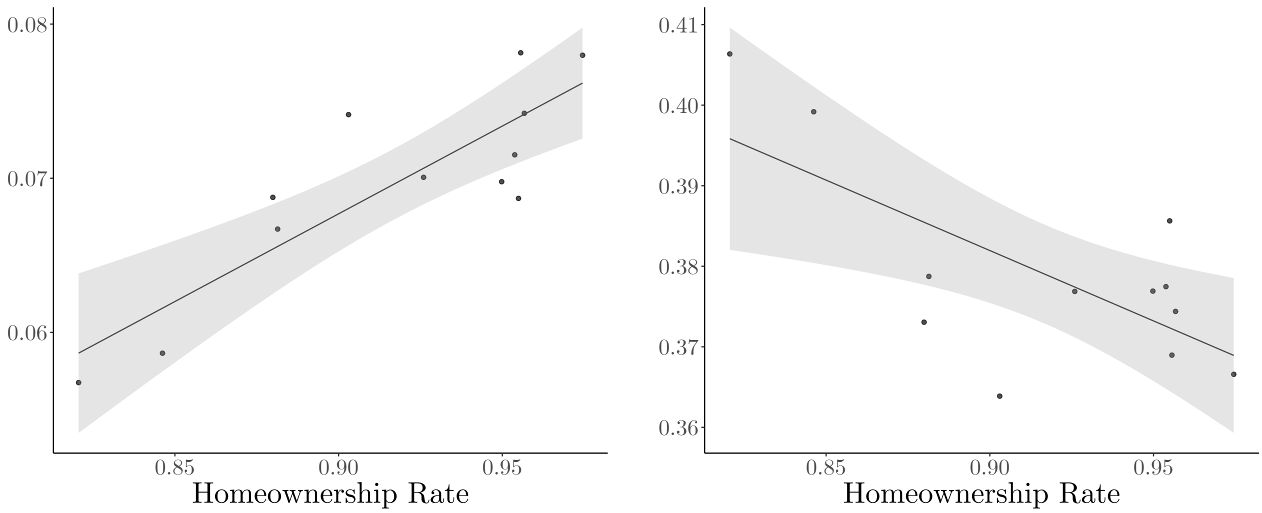


Figure F23: The figures plot the share of local net wealth accounting to the top and bottom 20% by location. This is a simulated outcome from the benchmark model. The horizzontal axis measures the simulated local homeownership rate. [Back]

G Policy Counterfactuals

[Back]

G.1 Welfare Measure

[Back]

Let lifetime utility of newborns ($j = 1$) in the benchmark and in the policy counterfactual with consumption tax Δc be denoted by V and $\hat{V}(\Delta c)$, respectively. The consumption-equivalent tax Δc is defined such that

$$
\begin{aligned}
&V - \hat{V}(\Delta c) = 0 \\
&\qquad\qquad V = \frac{1}{N}\sum_{i=1}^{N}\left\{u_1(c_i^*, h_i^*, d_i^*, \mathbf{x_i}) + \beta\mathbb{E}_{\varpi,l,\boldsymbol{z}}\left[\overline{V}_2^{f(h_i^*)}(\mathbf{x}', \Omega')\right]\right\} \\
&\qquad\qquad \hat{V}(\Delta c) = \frac{1}{N}\sum_{i=1}^{N}\left\{u_1((\Delta c)\hat{c}_i, \hat{h}_i, \hat{d}_i, \mathbf{x_i}) + \beta\mathbb{E}_{\varpi,l,\boldsymbol{z}}\left[\overline{V}_2^{f(\hat{h}_i)}(\mathbf{x}', \Omega')\right]\right\}
\end{aligned}
$$

where (c_i^*, h_i^*, d_i^*) and $(\hat{c}_i, \hat{h}_i, \hat{d}_i)$ denote optimal choices in the benchmark and counterfactual equilibria, $f(\cdot)$ takes value h (homeowner) for $h_i^* = \hat{h}_i = 2$ and value n (non-homeowner) otherwise, and N is the number of simulated individuals.

The consumption tax parameter Δc adjusts all agents' consumption uniformly either upwards or downwards. A value of $\Delta c > 1$ suggests that, following the policy implementation, agents require a higher level of consumption to be indifferent with the benchmark. Conversely, $\Delta c < 1$ indicates that agents would give up part of their consumption to keep the policy.

My measure of welfare is given $1 - \Delta c$. In other words, welfare measures the percentage change in consumption that the average newborn agent would require, or give up, in order to be indifferent between the counterfactual and the benchmark. I don't analyze welfare along the transition path between the stochastic steady-states. In particular, I compare newborn agents who are either born in the benchmark equilibrium or in the counterfactual equilibrium after prices, wages, and taxes have already converged.